



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Desarrollo de una estrategia innovadora en el diseño de *starters* para la industria láctea

Rodrigo Achigar

Manuscrito presentado como requisito para la obtención del título
de Doctor en Biotecnología

Montevideo, abril de 2018

Tutor de tesis: Dra. María Julia Pianzzola

Co-tutor: Dr. Sylvain Moineau

Agradecimientos

En primer lugar, quiero agradecer a mi tutora María Julia, por haberme abierto las puertas hace ya casi diez años. Por haberme dado la confianza y tener la paciencia para acompañarme hasta acá.

También agradecerle a mi co-tutor Sylvain Moineau y a todo su equipo en la Université Laval. En especial a Denise Tremblay y a Simón Labrie por su invaluable ayuda.

A mis compañeros de la vieja guardia del 12 con quienes compartimos tantos años, Cesar, Diego, las Marianas, Vicki, MIL, Vale, Gabi, MIS, Gianna y especialmente a Analía.

También agradecerles a mis estudiantes, que de una forma u otra se han visto salpicados por esta tesis, pero en especial a Maru, Vale y Felipe que han estado en la primera línea de batalla de este proyecto.

A todo el Laboratorio de Biotecnología de la Universidad ORT y el CBI+I, por haberme apoyado en este proyecto a todos los niveles.

A la empresa Christian Hansen A/S y todo el equipo de BactoPhi, especialmente al Dr. Alfonso H. Magadán por hacer posible gran parte de este proyecto, y por ser un maestro y amigo invaluable.

A la Agencia Nacional de Investigación e Innovación, Biomás S.A. y Diagramma S.A. por apoyar y financiar este proyecto.

A mis amigos, porque todavía no entienden que es bien lo que hago y no tienen voluntad de entenderlo, lo que los hace los amigos perfectos.

A Mamá y Papá por ser un apoyo incondicional en todo, y permitir hasta hoy que yo pueda ser lo que quiero ser; y a mi hermana, la contadora que más sabe de biología de este país.

A los miembros del tribunal por dedicar su tiempo a corregir y mejorar esta tesis.

Por último, agradecerles a ellas, los amores de mi vida. Por compartir la carga en este camino como solo ellas han sabido hacer. Por y para ellas ha sido todo este esfuerzo.

Para Momi y Cande.

Resumen

Un proceso biotecnológico que depende del uso de bacterias para producir un producto puede verse alterado por los fagos. Problemas de este tipo han sido reportados extensamente en la industria láctea. La consiguiente lisis de un gran número de células sensibles retrasará o incluso detendrá el proceso de fermentación de la leche, lo que traerá como consecuencia productos de baja calidad y caída de productividad en la planta. En el peor de los casos, la leche inoculada deberá descartarse perdiéndose grandes volúmenes de materia prima. Durante décadas, la industria láctea ha tenido que lidiar con esta problemática, dependiendo de una serie de medidas de control, que van desde el rediseño de fábricas, saneamientos mejorados, cambios en los procesos y rotación de starters. Una aproximación más efectiva, es la utilización de cepas que sean naturalmente resistentes. Cuando se aísla un nuevo fago, las empresas proveedoras pueden intervenir reemplazando las cepas sensibles por otras que no lo sean, y reintroducir el producto en el mercado. La escasa disponibilidad de cepas para sustituir en las rotaciones, así como el acceso a colecciones de fagos hacen que esta estrategia sea poco sostenible en el tiempo, aún así, ha mostrado ser la más exitosa hasta el momento.

El diseño de starters presenta una oportunidad de desarrollo tecnológico especialmente para una industria como la uruguaya que debe competir en calidad para ganar mercados de exportación, y que además es totalmente dependiente de la importación para abastecerse de insumos que no necesariamente han sido desarrollados según las necesidades puntuales de la industria local. En este trabajo nos proponemos el desarrollo de una metodología de obtención de cepas de *Streptococcus thermophilus* mejoradas para su uso en la industria láctea nacional, con especial énfasis en la búsqueda de soluciones a las infecciones fágicas en planta que constituyen el principal problema de productividad.

Índice

Agradecimientos	3
Resumen	5
índice	6
Lista de Figuras	10
Lista de Tablas	15
Guía de lectura	17
Capítulo 1: Introducción	18
Fermentaciones microbianas	18
Bacterias ácido-lácticas	20
Generalidades de las bacterias ácido-lácticas	20
Evolución de las BAL	21
Adaptación al ambiente lácteo	23
Metabolismo	25
Metabolismo de nitrógeno	27
Cultivos <i>starters</i>	30
<i>Streptococcus thermophilus</i>	34
Taxonomía	35
Atributos tecnológicos y funcionales de <i>S. thermophilus</i>	38
Metabolismo de azúcares	39
Metabolismo de nitrógeno	42
Bacteriófagos	43
Caudovirales	46
Fagos y las fermentaciones lácticas	49
Fuentes de contaminación	50

Estrategias de control	52
Mecanismos de defensa	54
Inmunidad innata: sistemas de restricción- modificación (R-M)	54
Inmunidad adaptativa: el sistema CRISPR-Cas	55
Funcionamiento	56
Sistema CRISPR-Cas de tipo IIA	59
Industria láctea nacional	61
Objetivos	64
Objetivos específicos	64
Antecedentes y estrategia	65
Capítulo 2: Caracterización de cepas de <i>Streptococcus thermophilus</i>	66
Resumen	66
Materiales y métodos	67
Cepas bacterianas	67
Bacteriófagos	67
Detección del gen PrtS	67
Actividad proteolítica en placa	67
Actividad Ureasa	68
Fermentaciones en leche	68
Sensibilidad a bacteriófagos	68
Electroforesis en campo pulsado (PFGE)	68
Amplificación aleatoria de ADN polimórfico (RAPD)	69
Evaluación del contenido de los loci CRISPR	69
Extracción de proteínas bacterianas	69
Extracción de ADN viral	70

Digestión con extractos celulares	70
Digestión de ADN viral con Dpnl	70
Resultados	72
Atributos tecnológicos	72
Actividad proteolítica	73
Actividad ureasa	75
Curvas de fermentación	77
Sensibilidad a bacteriófagos	81
Caracterización molecular	84
Electroforesis en campo pulsado (PFGE)	86
Amplificación aleatoria de ADN polimórfico (RAPD)	87
Contenido de los loci CRISPR	89
Secuencias espaciadoras	90
Inmunidad innata: sistemas de restricción-modificación (R-M)	101
Capítulo 3: Análisis de genomas virales	105
Resumen	105
Antecedentes	105
Materiales y métodos	106
Microscopía electrónica	106
Secuenciación de genomas virales y análisis	106
Resultados	107
Morfología	107
Características genómicas	108
Capítulo 4: Generación de mutantes fago-resistentes	136
Resumen	136
Materiales y métodos	137

Alineamientos	137
Mapeo de protoespaciadores	137
Cálculo del CAI	137
Selección de mutantes fago-resistentes	137
Cálculo de la eficiencia de plaqueo (E.O.P.)	138
Resultados	139
Protoespaciadores	139
Mapeo de protoespaciadores	139
Búsqueda de sesgos en la adaptación	149
Generación de mutantes resistentes	154
Diversidad de protoespaciadores	160
Capítulo 5: Diseño de cepas resistentes	168
Resumen	169
Antecedentes	168
Estrategia	169
Materiales y métodos	171
Cepas bacterianas y plásmidos	171
Transformación de <i>Lactococcus lactis</i> por electroporación	171
Transformación de <i>Streptococcus thermophilus</i> por electroporación	172
Transformación de <i>Lactobacillus spp.</i> por electroporación	172
Transformación de <i>E. coli.</i> por electroporación	173
Ensamblaje de construcciones en pGh9	174
Modificación del fenotipo gal+ en <i>Lactobacillus</i> mediante recombineering	175
Resultados	176
Diseño del locus CRISPR donante	176
Sistema de modificación CRISPR1-Cas9	179

Integración y silenciamiento por recombineering	183
Resistencia adquirida	189
Conclusiones finales	191
Bibliografía	194

Lista de Figuras

Figura 1.1. Representación esquemática del sistema proteolítico en bacterias ácido-lácticas	29
Figura 1.2. Observación al microscopio (100x) de una tinción de Gram de <i>Streptococcus thermophilus</i>	35
Figura 1.3. Morfología de los viriones de las tres familias que componen el orden de los <i>Caudovirales</i> (Kurtboke 2012)	46
Figura 1.4. Estructura, espaciadores y PAMs de los diferentes tipos de sistema CRISPR	57
Figura 1.5. Esquema del proceso de adquisición de nuevos espaciadores mediado por Cas1 y Cas2	58
Figura 1.6. Diferentes tipos y mecanismos de síntesis de crARNs	59
Figura 1.7. Estructura y esquema de funcionamiento del sistema CRISPR tipo IIA	61
Figura 1.8. Evolución de las toneladas de queso exportadas por Uruguay	62
Figura 1.9. Variación anual del precio de la tonelada de queso exportado por Uruguay	62
Figura 1.10. Distribución del mercado exportador años 2013, 2015 y 2017	63
Figura 2.1: Electroforesis en gel de agarosa de los productos de amplificación para el gen <i>pvtS</i> .	73

Figura 2.2. Ensayo semi-cuantitativo de actividad proteolítica.	74
Figura 2.3. Ensayo semi-cuantitativo de actividad proteolítica para la cepa Uy12.	75
Figura 2.4. Ensayo a microescala de actividad ureasa.	76
Figura 2.5. Matriz esquemática del fenotipo mostrado por las cepas de la colección	76
Figura 2.6. Curvas de fermentación para las 50 cepas de la colección.	77
Figura 2.7. Curvas de fermentación para las cuatro cepas evaluadas en su grado de proteólisis.	78
Figura 2.8. Gráfico de q_{pH} para las cepas evaluadas en su capacidad proteolítica.	79
Figura 2.9: Curvas de fermentación de cepas con actividad ureasa positiva.	80
Figura 2.10: Curvas de fermentación de cepas con actividad ureasa negativa	80
Figura 2.11. Dendrograma obtenido a partir de la matriz de sensibilidad resistencia.	85
Figura 2.12. Electroforesis en campo pulsado para muestras de ADN digerida con la enzima <i>SmaI</i> .	86
Figura 2.13. Patrones de RAPD obtenidos para algunas cepas de la colección utilizando el primer OPI-02	87
Figura 2.14. Dendrograma obtenido con los datos de RAPD de las cepas de la colección.	88
Figura 2.15. Alineamiento y agrupamiento en <i>clústeres</i> de <i>loci</i> CRISPR1 de las cepas de la colección.	94
Figura 2.16. Alineamiento y agrupamiento en <i>clústeres</i> de <i>loci</i> CRISPR3 de las cepas de la colección.	95
Figura 2.17. Esquema del procedimiento utilizado en la búsqueda sistemas de defensa tipo R-M en las cepas de la colección	101

Figura 2.18. Mapa del genoma del fago 73.	102
Figura 2.19. Mapa del genoma del fago 73 ubicando los 26 sitios de corte para la enzima DpnI	103
Figura 2.20. Digestión de ADN genómico del fago 73 con la enzima DpnI.	104
Figura 3.1. Micrografías de los fagos	107
Figura 3.2. A: organización del módulo de replicación tipo 7201. B: organización del módulo de replicación tipo Sfi21	109
Figura 3.3. Representación gráfica y comparación de los genomas de fagos de la colección.	111
Figura 4.1. Alineamiento de los genomas tipo cos de la colección.	140
Figura 4.2. Representación gráfica de los protoespaciadores CRISPR1 de los diez genomas de fagos tipo cos de la colección.	142
Figura 4.3. Representación gráfica de los protoespaciadores CRISPR3 de los diez genomas de fagos tipo cos de la colección.	143
Figura 4.4. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 53.	144
Figura 4.5. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 73	145
Figura 4.6. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 94.	145
Figura 4.7. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 107.	146
Figura 4.8. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 115.	146

Figura 4.9. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 123.	147
Figura 4.10. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 128.	147
Figura 4.11. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 140.	148
Figura 4.12. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago 142.	148
Figura 4.13. Frecuencia de protoespaciadores CRISPR3 (izquierda) y CRISPR1 (derecha) para el fago ST.	149
Figura 4.14. Valores de CAI para cada gen del fago 53.	151
Figura 4.15. Tipos de estructuras formadas durante la replicación del genoma de fagos tipo Cos.	152
Figura 4.16. Esquema circular del genoma del fago 53	153
Figura 4.17. Esquema del <i>locus</i> CRISPR3 de la cepa Uy01.	156
Figura 4.18. Alineamiento entre la secuencia del espaciador cuatro y el extremo 3' de la secuencia líder del <i>locus</i> CRISPR3.	156
Figura 4.19. Ubicación de los espaciadores adquiridos por las BIMs de Uy01 en el genoma del fago.	157
Figura 4.20. Frecuencia de los espaciadores adquiridos por las BIMs de Uy01	157
Figura 4.21. Curvas de pangenomas para <i>Bacillus anthracis</i> y <i>Streptococcus agalacticae</i>	162
Figura 4.22. Curva construida con los protoespaciadores CRISPR1.	165
Figura 4.23. Curva construida con los protoespaciadores CRISPR3.	165

Figura 4.24. Curva construida con los protoespaciadores CRISPR1 de los fagos cos de la colección más DT1.	166
Figura 4.25. Curva construida con los protoespaciadores CRISPR3 de los fagos cos de la colección más DT1.	166
Figura 5.1. Estructura del <i>locus</i> CRISPR1 diseñado en este trabajo.	178
Figura 5.2. <i>Locus</i> CRISPR1 y genes <i>Cas</i> asociados de <i>Streptococcus thermophilus</i>	180
Figura 5.3. Mapa del vector pGh9Target.	182
Figura 5.4. Mapa del vector pGh9Donor	184
Figura 5.5. evento de integración esperado.	185
Figura 5.6. Esquema de funcionamiento del sistema Lambda Red.	187
Figura 5.7. Secuencia del promotor utilizado.	188
Figura 5.8. Expresión de la proteína reportera mCherry dirigida por el promotor seleccionado.	188

Lista de Tablas

Tabla 2.1: matriz de sensibilidad resistencia.	83
Tabla 2.2: <i>Loci</i> CRISPR amplificados en cada cepa.	91
Tabla 2.3: <i>Clústeres</i> en los que se agruparon los <i>loci</i> CRISPR de cada cepa.	96
Tabla 2.4: espaciadores de mayor frecuencia en la colección de cepas	97
Tabla 2.5: Resultado del BLAST realizado a los espaciadores de alta frecuencia que no presentaron protoespaciadores homólogos en los genomas de fagos uruguayos.	98
Tabla 2.6: Incongruencias en la matriz sensibilidad resistencia	100
Tabla 3.1. Dimensiones de las partículas virales	107
Tabla 3.2. Características generales de los fagos de la colección	108
Tabla 3.3. Características generales del fago 53	112
Tabla 3.4. Características generales del fago 73	113
Tabla 3.5. Características generales del fago 93	116
Tabla 3.6. Características generales del fago 94	118
Tabla 3.7. Características generales del fago 107	120
Tabla 3.8. Características generales del fago 115	123
Tabla 3.9. Características generales del fago 123	125
Tabla 3.10. Características generales del fago 128	127
Tabla 3.11. Características generales del fago 140	129
Tabla 3.12. Características generales del fago 142	131
Tabla 3.13. Características generales del fago ST	133

Tabla 4.1. Espaciadores obtenidos en el ensayo de generación de resistentes de Uy01 versus el fago 53	158
Tabla 4.2. Panviroma calculado con los protoespaciadores tipo CRISPR1	163
Tabla 4.3. Panviroma calculado con los protoespaciadores tipo CRISPR3	163
Tabla 5.1. Primers utilizados para el ensamblaje de los distintos plásmidos generados en este trabajo	174
Tabla 5.2. Eficiencia de transformación para los diferentes vectores (ufc/ μ g)	181
Tabla 5.3. Eficiencia de transformación para los diferentes vectores (ufc/ μ g)	183
Tabla 5.4. oligos diseñados para noquear los genes <i>repA</i> y <i>emR</i> .	189

Guía de lectura

Este manuscrito se encuentra dividido en 5 capítulos independientes. El primero es una introducción general donde se plantea el marco teórico, la problemática a abordar, los objetivos, antecedentes y estrategia de trabajo. Los restantes cuatro capítulos contienen resultados, discusión, conclusiones parciales y materiales y métodos. A continuación del quinto capítulo, se encuentra un apartado final con las conclusiones generales y perspectivas. Por último, al final del manuscrito se encuentra la lista de referencias consultadas en este trabajo

1

Introducción

Fermentaciones microbianas

La elaboración de alimentos fermentados es una de las tecnologías más tempranamente desarrolladas por el hombre en su historia. Hace ocho mil años, nuestros antepasados prepararon las primeras cervezas y panes. Sin tener conciencia, los habitantes de la Media Luna Fértil (hoy Irak) fueron los primeros en utilizar organismos vivos en los procesos de fermentación. La existencia de microorganismos no se conoció hasta las observaciones realizadas por Anton van Leeuwenhoek en 1676, quién reportó la presencia de lo que hoy conocemos como protozoarios y algas en muestras de agua del lago Delft. Sin embargo, ninguno de estos agentes estaba firmemente implicado en la producción de alimentos o su deterioro y no fue hasta la aparición de los trabajos pioneros de Needham, Spallanzani en la disputa por la generación espontánea de microorganismos y hasta que Pasteur y Bassi de Lodi presentaran la teoría germinal de las enfermedades infecciosas en los siglos XVIII y XIX que se conoció que la cerveza, el pan, el queso, el vino y la mayoría de los alimentos fermentados provienen de procesos sumamente antiguos y que, en algunos casos, estos no han cambiado mucho en hasta el día de hoy.

Históricamente, el uso de microorganismos en estos procesos de fermentación ocurrió en forma accidental. Sin embargo, se descubrió que la adición de una parte del producto de un proceso previo al nuevo lote podría servir para "iniciar" el proceso. A esta práctica se la conoce como "*back slopping*" (se traduce al español literalmente como "retroceso", aunque la forma correcta de nombrarla es "re-inoculación").

A diferencia de los productos biotecnológicos modernos, en su mayor parte, consideramos a los alimentos fermentados como productos básicos de alto volumen y bajo valor. Sin embargo, para productos como la cerveza, ahora existe una enorme comprensión científica

de la ciencia que sustenta el producto, una ciencia que, no obstante, está atemperada por las presiones de la tradición.

Los componentes clave de un sistema de fermentación son el microorganismo y su materia prima. Para algunos productos, como el vino y la cerveza, existe una modificación radical de las propiedades de la materia prima, haciéndolos más palatables. Por ejemplo, mientras que para el caso de la cerveza, los extractos de granos de la pre-fermentación son de sabor desagradable en otros casos, como en el jugo de uva, es más aceptable. Para otros productos, el organismo es menos central, aunque sigue siendo importante. Es el caso del pan, aunque no todos los estilos involucran a la levadura en su producción. Para productos como el queso, el resultado final es bastante distinto a las materias primas. Como norma general, el producto final (alimento fermentado) es en realidad el medio de crecimiento “gastado”, y enriquecido con los desechos metabólicos de los microorganismos fermentadores (ácidos, alcoholes, etc.).

Los organismos empleados en la fermentación de alimentos son muchos y diversos. Los principales actores son las bacterias del ácido láctico en los productos lácteos, y la levadura en la producción de bebidas alcohólicas y pan. Las bacterias del ácido láctico también pueden desempeñar un papel positivo en la producción de ciertos tipos de vinos y cervezas, pero igualmente son los principales organismos de descomposición de dichos productos (Anon 2005).

El objetivo original y primario de la fermentación en la producción de alimentos, era obtener un efecto preservador sobre los mismos, pero con el desarrollo de nuevas tecnologías de preservación efectivas, éste ha dejado de ser el principal objetivo. Muchos de estos alimentos son elaborados por su sabor, textura y aroma únicos, los cuales son apreciados por los consumidores. Aun así, incluso en estas situaciones, las condiciones generadas por la fermentación del sustrato son esenciales para la preservación y seguridad microbiológica del producto. La preservación de alimentos mediada por fermentación implica la oxidación de carbohidratos y derivados, para generar productos finales los cuales son generalmente ácidos, alcoholes y dióxido de carbono. Estos productos ayudan a controlar el crecimiento de microorganismos que puedan deteriorar el alimento, y, como la oxidación es parcial, el alimento retiene suficiente energía como para seguir siendo nutritivo para el consumidor. Estrictamente hablando, una fermentación es el consumo anaerobio de carbohidratos para obtener energía por parte de los microorganismos. Sin embargo, este término comúnmente

incluye todos los procesos biológicos, tanto aerobios como anaerobios, llevados a cabo por el microorganismo durante la elaboración del alimento (Achigar 2014).

Las fermentaciones para la elaboración de alimentos se han practicado durante milenios, resultando en variedad de productos que van desde los derivados de la carne y las plantas hasta los derivados de la leche. En cada caso, el proceso de fermentación implica la oxidación de carbohidratos para generar una gama de productos que son principalmente ácidos orgánicos, alcohol y dióxido de carbono.

Dichos productos tienen un efecto conservante al limitar el crecimiento de la flora degradadora y/o patógena en el alimento. También se pueden producir otros productos deseables, que afectan la calidad del alimento, incluidos los sabores (diacetilo y acetaldehído), así como compuestos que pueden tener implicaciones positivas para la salud (vitaminas, antioxidantes, péptidos bioactivos, etc.). Quitando las fermentaciones alcohólicas que involucran levaduras como los principales agentes de transformación, los microorganismos principales en la elaboración de productos fermentados son las bacterias ácido-lácticas (BAL).

Bacterias ácido-lácticas

Generalidades de las bacterias ácido-lácticas

Las bacterias ácido-lácticas son microorganismos Gram positivos, anaerobios facultativos, catalasa-negativos, no móviles, no formadores de esporas, cocos o bacilos con un contenido GC inferior al 55%. Según la clasificación actual, pertenecen al *phylum Firmicutes*, clase *Bacilli*, orden *Lactobacillales*. Las diferentes familias incluyen a los *Aerococcaceae*, *Carnobacteriaceae*, *Enterococcaceae*, *Lactobacillaceae*, *Leuconostocaceae* y *Streptococcaceae*.

A nivel filogenético, las BAL se clasifican de acuerdo a la secuencia del ARNr 16S. Las BAL ancestrales, habrían sido organismos tipo bacilo habitantes del suelo, que subsecuentemente perdieron genes y su fenotipo asociado, a la vez que se adaptaban a nuevos nichos ecológicos nutricionalmente ricos.

Los miembros más comunes del grupo que se utilizan para usos alimentarios incluyen los lactococos para la fabricación de queso, *Streptococcus thermophilus* para la fabricación de queso y yogur y varios miembros del género *Lactobacillus* para una variedad de fermentaciones lácteas, cárnicas y vegetales.

La producción de alimentos fermentados se basa en el uso de cultivos iniciadores. Un cultivo iniciador o *starter* puede definirse como una preparación microbiana de elevada concentración celular de al menos un microorganismo, la cual es agregada a una materia prima para producir un alimento acelerando y dirigiendo su proceso de fermentación.

Tradicionalmente, la fermentación de productos lácteos se daba por bacterias ácido-lácticas presentes de forma natural en la leche. Aunque existen doce géneros de bacterias del ácido láctico, hoy en día son utilizados principalmente cuatro géneros en la producción de cultivos *starters*: *Lactococcus*, *Leuconostoc*, *Streptococcus* y *Lactobacillus*. La inclusión de miembros del género *Enterococcus* en este grupo ha sido controvertida, ya que estos son organismos fecales y se reconocen como patógenos oportunistas (Franz et al. 1999). Sin embargo, frecuentemente se encuentran representantes de este género en cultivos iniciadores indefinidos y se reconoce su aporte en los procesos maduración y al desarrollo de sabor en el queso (Dudley & Steele 2005; Crow et al. 2001; McSweeney et al. 1993; Broome et al. 1990).

Debido a que este tipo de producción de alimentos fermentados se daba por el desarrollo de esta microflora presente de forma natural en la materia prima, la calidad del producto final dependía de la carga y diversidad microbiana en la misma.

Evolución de las BAL

Las BAL pertenecen al *phylum Firmicutes*, clase *Bacilli*. Se cree que el ancestro común para esta clase tuvo un potencial de codificación de aproximadamente 3.000 genes, mientras que, en el presente, las BAL generalmente tienen genomas pequeños de hasta 3.5 megabases (Mb) y un potencial de codificación promedio de 2.000 genes. Las especies con los genomas más pequeños son las mejor adaptadas a vivir en entornos nutricionalmente ricos, mientras que las que tienen hábitats más versátiles tienen los genomas más grandes. La evolución hacia genomas pequeños generalmente es el resultado de la pérdida de genes y transferencia horizontal (Makarova & Koonin 2007), con un papel menor desempeñado por la duplicación de genes.

El análisis de la evolución entre los miembros de las bacterias ácido-lácticas muestra que el camino desde el ancestro común de todos los *Bacilli* hasta llegar a los *Lactobacillales* está marcado por la pérdida de aproximadamente 1.000 genes. Esta pérdida involucra el conjunto completo de genes implicados en la esporulación, genes codificantes de citocromos tipo hemo-cobre, algunos cofactores y genes codificantes de catalasa. Algunas especies (por ejemplo, *Lactobacillus. delbrueckii* subespecie *bulgaricus* y *Streptococcus thermophilus*) muestran una mayor prominencia aún de pérdida de genes y un alto número de pseudogenes, posiblemente relacionada con la especificidad del nicho ecológico que ocupan (Makarova & Koonin 2007).

Durante la evolución posterior de cada linaje de *Lactobacillales*, se produjeron pérdidas de genes adicionales y diferenciales, lo que ha dado como resultado un ordenamiento evolutivo representado por una filogenia con cinco ramas principales: las rama *Lactococcus-Streptococcus* y *Enterococcus-Tetragenococcus* claramente separadas en un lado, la rama *Leuconostoc-Oenococcus-Weissella-Fructobacillus* por otro lado, la rama *Pediococcus*-grupo *L. casei* y por último, la más distante, la rama *L. delbrueckii*, la cual integra a las especies del grupo *L. acidophilus*.

Las bacterias ácido-lácticas también han adquirido variedad de genes por eventos de duplicación y transferencia horizontal (HGT). Los datos disponibles sugieren que la adaptación a entornos ricos en nutrientes fue la principal fuerza impulsora de estas adquisiciones de genes. Entre los 86 genes adquiridos que se infiere han dado forma al ancestro común de los *Lactobacillales*, solo dos no tienen ortólogos fuera de *Lactobacillales*. (Makarova et al. 2006).

Un ejemplo de adquisición de genes parálogos es proporcionado por la presencia de dos genes codificantes para la enzima glucolítica enolasa en lactobacilos. El análisis filogenético muestra que una de las enolasas en *Lactobacillales* es la versión ancestral presente en prácticamente todos los *Firmicutes*, mientras que la última aparentemente ha sido adquirida por algún ancestro de las BAL, posiblemente por transferencia horizontal de una *Actinobacteria*.

Un proceso adicional de adquisición de genes se ilustra mediante la familia de enzimas involucradas en la biosíntesis de peptidoglicano. El gen que codifica la UDP-N-acetilmuramil tripéptido sintasa (*murE*) está presente en la mayoría de los *Bacilli*; sin embargo, en los

lactobacilos (a excepción de *L. plantarum*), un parálogo de esta enzima, posiblemente derivado de una *Actinobacteria*, ha reemplazado al gen ancestral (Makarova & Koonin 2007).

Los genes que proporcionan resistencia a los antibióticos en otras bacterias se han encontrado en lactobacilos sensibles a los antibióticos, donde se asocian con fenotipos no relacionados con la resistencia a los medicamentos. Un ejemplo es un parálogo de la lisil-tRNA sintetasa de clase II, generalmente implicado en la resistencia a antibióticos tipo oxacilina, que en los lactobacilos se fusiona con un dominio asociado a la membrana y está probablemente implicado en la biosíntesis de la pared celular (Lahtinen et al. 2011).

El papel de la actividad del fago/profago en la ganancia de genes, se ilustra mediante la observación de que 11 genes presentes en ambos *L. johnsonii* y *L. lactis* subsp. *cremoris* se localizan cerca del sitio IS de un profago, lo que sugiere eventos de transducción en la historia evolutiva de estas especies.

Todas las secuencias genómicas de BAL están caracterizadas por una gran presencia de secuencias de inserción (IS), lo que indica el papel central en la evolución de este grupo. Un buen ejemplo de plasticidad mediada por IS es el de *L. helveticus*, el cual se ha diversificado desde su ancestro habitante de la flora intestinal no solamente por la pérdida selectiva de genes sino también por la adquisición de más de 200 secuencias de inserción, encontrándose en su genoma 16 agrupamientos de marcos abiertos de lectura (ORFs) flanqueados por secuencias de inserción (Lahtinen et al. 2011).

Adaptación al ambiente lácteo

Las BAL se encuentran asociadas a la producción de prácticamente todos los productos lácteos fermentados. Sin embargo, muchas de las especies utilizadas en estos procesos están estrechamente relacionadas filogenéticamente con microorganismos incapaces de crecer en leche. Por tanto, se puede afirmar que la capacidad de metabolizar lactosa y degradar caseína son las principales adaptaciones al ambiente lácteo.

Los primeros estudios publicados sobre genética molecular de BAL, mostraron que los genes vinculados al metabolismo de lactosa en cepas mesófilas se encuentran codificados en plásmidos, mientras que para especies termófilas al menos parte de la maquinaria caseinolítica se encuentra codificada en plásmidos, mientras que los genes asociados al metabolismo de la lactosa se encuentran en el cromosoma (Lahtinen et al. 2011).

La comparación entre las cepas *L. helveticus* DPC4571 (cepa utilizada como starter industrial) y la cepa *L. acidophilus* NCFM (de origen intestinal) mostró que comparten un 98,4% de identidad en sus genes 16S ribosomales. Esto sugiere un estrecho relacionamiento a pesar de ocupar nichos ecológicos tan diferentes. El análisis de sus genomas completos mostró que el 75% de los genes de DCP4571 poseen ortólogos en el genoma de NCFM (O'Sullivan et al. 2009). Las diferencias se encuentran en 123 genes exclusivos en DCP4571, y 503 genes exclusivos en NCFM. Entre estos 626 genes nicho-específicos, aquellos vinculados a proteólisis y resistencia a bacteriófagos resultaron ser específicos para el ambiente lácteo. La especificidad de los genes relacionados a la proteólisis puede parecer obvia ya que para crecer en leche es necesario degradar caseína en alguna medida, pero en lo que tiene que ver con resistencia a bacteriófagos la correlación no parece tan evidente. La explicación a este fenómeno puede ser la presión selectiva a la que se someten las cepas comerciales para su uso a nivel industrial (Lahtinen et al. 2011).

Dos de las adaptaciones al ambiente lácteo más importantes han sido observadas en especies de *Lactobacillus casei* (Cai et al. 2009) y *Lactococcus lactis* (Siezen et al. 2010). Mediante análisis de MLST, se demostró que *Lactobacillus casei* se separó en dos grupos principales hace aproximadamente 1,5 millones de años. Una nueva separación en diferentes cepas ocurrió hace aproximadamente 10.000 años. Al comparar los genomas completos de siete aislamientos de diferentes partes del mundo, se observó que cinco de estos han perdido casi un quinto de los genes presentes en la cepa de referencia ATCC 334. Los genes perdidos corresponden a proteínas hipotéticas, genes de fagos, metabolismo de carbohidratos y transporte de aminoácidos. Particularmente, se observó la pérdida de una región de 129 kb, la cual codifica el 15% del total de genes relacionados con la utilización de carbohidratos (Lahtinen et al. 2011).

En el caso particular de *Lactococcus lactis*, una comparación entre cepas aisladas de ambientes lácteos y de ambientes no lácteos ha mostrado la gran adaptación sufrida por las primeras. Al comparar los genomas de las cepas KF147 aislada de material vegetal, y la cepa láctea IL1403, se observan grandes similitudes a nivel nucleotídico y estructural. Sin embargo IL1403 ha perdido los genes involucrados en la degradación de la pared celular de células vegetales y de los sistemas de internalización y utilización de estos productos de degradación por lo que no es capaz de crecer en matrices vegetales (Siezen et al. 2010).

Metabolismo

El objetivo de los cultivos *starter* en los productos lácteos fermentados es propiciar los cambios químicos, sensoriales y nutricionales típicamente asociados con ese alimento. Probablemente la característica más importante de las BAL es su capacidad de fermentar los carbohidratos y producir ácido láctico como producto final primario o como uno de los productos finales. Es importante tener en cuenta que las bacterias del ácido láctico pueden cambiar su metabolismo en respuesta a diferentes condiciones, lo que resulta en diferentes productos finales.

Una de las características más importantes de los cultivos lácteos *starters* es su capacidad de metabolizar la lactosa. La lactosa es el carbohidrato principal en la leche y se encuentra a concentraciones de 45 a 50 g/L. Es un disacárido formado por un residuo de glucosa y uno de galactosa unidos por un enlace $\beta 1 \rightarrow 4$. Su capacidad para fermentar lactosa es el criterio de selección más importante para los cultivos iniciadores.

Las bacterias ácido-lácticas se clasifican en dos categorías principales: organismos homofermentadores y heterofermentadores. Esta clasificación está basada en las vías de metabolización de la lactosa y los productos finales obtenidos de la fermentación. Los cultivos iniciadores homofermentativos incluyen *Lactococcus*, *Streptococcus* y algunos lactobacilos. Estos microorganismos metabolizan azúcares utilizando la vía glucolítica (o Embden-Meyerhof-Parnas (EMP)), donde una mol de glucosa se fermenta dando lugar a dos moles de ácido láctico y se producen dos ATP. Los cultivos iniciadores heterofermentadores, que incluyen *Leuconostoc* y lactobacilos, fermentan un mol de glucosa a un mol de ácido láctico, un mol de CO₂ y un mol de etanol o un mol de ácido acético utilizando la ruta de fosfoacetolasa que produce un ATP. La fermentación de azúcares por la vía de la fosfoacetolasa produce solo la mitad de la energía que la ruta glucolítica (Marth & Steele 2001; Vasiljevic & Shah 2008). Las diferencias en los productos finales metabólicos tienen un impacto significativo en el producto final.

En la vía homoláctica, la lactosa puede ingresar a la célula como azúcar libre o como fosforilada. El mecanismo de transporte de lactosa en la célula está determinado en gran medida por la vía de hidrólisis del azúcar internalizado. Por lo tanto, la lactosa puede ser transportada a la célula por dos mecanismos diferentes. El más eficiente es el sistema fosfotransferasa fosfoenolpiruvato dependiente de lactosa (PEP-PTS), donde la lactosa se fosforila durante su transporte a través de la membrana. Una vez dentro de la célula, la

lactosa es hidrolizada por la fosfo- β -galactosidasa, produciéndose una molécula de glucosa y una de galactosa-6-fosfato. La glucosa resultante se metaboliza mediante glucólisis (vía EMP) en ácido láctico. La galactosa-6-fosfato es convertida en tagatosa a través de la ruta de la tagatosa, y luego convertida en triosas que finalmente entran en la ruta glucolítica para producir ácido láctico. El otro mecanismo de transporte igualmente común es la internalización de la lactosa sin modificar, mediado por permeasas. Una vez dentro de la célula, la β -galactosidasa hidroliza la lactosa. La glucosa que se produce de nuevo se metaboliza mediante glucólisis, mientras que la galactosa se convierte en glucosa a través de la vía Leloir y mediante glucólisis en ácido láctico. En todas estas rutas metabólicas, el piruvato es el intermediario clave cuya conversión da como resultado la producción de cuatro moles de ácido láctico en condiciones normales. Es importante destacar que algunas bacterias *starters* como *S. thermophilus*, *Lb. delbrueckii* subsp. *bulgaricus*, y algunos *Lb. delbrueckii* subsp. *lactis* fermenta solo la glucosa y transporta la fracción de galactosa fuera de la célula en cantidades proporcionales a la lactosa internalizada, mediante un sistema antiporte lactosa-galactosa. Por lo tanto, el sistema de permeasas que participa en el transporte de lactosa a la célula también excreta galactosa fuera de la célula (Vasiljevic & Shah 2008).

En la vía EMP, con glucosa como sustrato, dos moléculas de ATP son utilizadas para convertir glucosa en fructosa-1,6-bifosfato. La hidrólisis de esta molécula genera dos compuestos de 3 carbonos, los cuales terminan convirtiéndose en fosfoenolpiruvato (PEP) generando ATP y NADH. El PEP puede utilizarse en el sistema de transporte PEP-PTS (mencionado anteriormente) o convertirse en piruvato generando ATP en el proceso. El piruvato es convertido a lactato por la enzima lactato deshidrogenasa, reciclándose poder reductor. El tipo de lactato producido por el microorganismo (L(+)-, D(-)- o DL-lactato) está dado por el tipo de la lactato deshidrogenasa que posea (L, D o una mezcla). La reacción global implica la producción de dos moléculas de ATP y dos de lactato a partir de una molécula de glucosa. El ácido láctico es excretado al ambiente (Lahtinen et al. 2011).

En la fermentación heteroláctica, al igual que en la homoláctica, la lactosa es transportada al interior de la célula a través de un sistema de permeasas. La lactosa es hidrolizada en glucosa y galactosa por la β -galactosidasa, convirtiéndose esta última en glucosa a través de la vía Leloir. Las bacterias con metabolismo heteroláctico, consumen hexosas y producen una mezcla de ácido láctico, CO₂ y acetato o etanol. Estos organismos (miembros del género *Leuconostoc* y *Lactobacillus*), carecen de la enzima fructosa bifosfato aldolasa,

por lo tanto, no pueden llevar a cabo la vía EMP completamente, pero poseen glucosa 6-P deshidrogenasa y xilulosa 5-P fosfocetolasa, lo que les permite metabolizar las hexosas por la vía fosfogluconato-fosfocetolasa para obtener energía. Esta ruta, tiene una primera fase oxidativa, en la cual a partir de glucosa se obtiene un compuesto de 5 carbonos (ribulosa 5-fosfato) y CO₂. En la fase no oxidativa, la ribulosa 5-fosfato es convertida en xilulosa 5-fosfato, la cual es hidrolizada a gliceraldehído 3-fosfato y acetil fosfato. El gliceraldehído 3-fosfato es metabolizado a lactato siguiendo desde ese punto la vía EMP, mientras que el acetil fosfato puede oxidarse a acetato o reducirse a etanol, dependiendo del potencial óxido-reducción (O-R) del medio. Los productos finales son excretados al medio.

El balance global de la fermentación heteroláctica es un mol de ATP por mol de hexosa (se producen dos moles de ATP por vía del gliceraldehído 3-fosfato a lactato, y se gasta uno en la fosforilación de la glucosa). Sin embargo, es posible obtener un ATP más, si el acetil fosfato es oxidado a acetato, obteniéndose un balance global de dos moles de ATP por mol de glucosa consumido (Lahtinen et al. 2011).

Metabolismo de nitrógeno

Las BAL se caracterizan por ser nutricionalmente exigentes. Requieren de sustancias promotoras de crecimiento, como aminoácidos y vitaminas, ya que tienen capacidades limitadas para sintetizarlas ellos mismos (Mierau et al. 1996). Cada cepa posee requerimientos específicos de aminoácidos diferentes. Como ejemplo, para el caso de los lactococos los requerimientos incluyen glutamato, metionina, valina, leucina, isoleucina e histidina que son necesarios. Otras cepas pueden requerir fenilalanina, tirosina, lisina y alanina. La leche es un medio de crecimiento adecuado para los cultivos *starters*; sin embargo carece de los niveles necesarios de péptidos y aminoácidos libres para soportar el correcto desarrollo de las bacterias ácido lácticas (Vasiljevic et al. 2005). Por lo tanto, las bacterias del ácido láctico han desarrollado sistemas proteolíticos complejos que incluye proteinasas que descomponen las caseínas en péptidos, peptidasas que degradan los péptidos, y sistemas de transporte que translocan aminoácidos y péptidos a través de la membrana citoplasmática hacia el citoplasma (Kunji et al. 1996). Estos sistemas les proporcionan a las BAL la capacidad de utilizar caseína como fuente adicional de aminoácidos y nitrógeno para el crecimiento (Smid et al. 1991).

La actividad proteolítica de las BAL tienen un papel fundamental en las propiedades del producto lácteo final (Broadbent & Steele 2007). Los péptidos y aminoácidos producidos por estos organismos como consecuencia de la proteólisis de caseína, sirven como precursores para la formación de los compuestos responsables de sabor y aroma (Ott, J. E. Germond, et al. 2000). El sistema proteolítico en lactococos se ha estudiado ampliamente (Broadbent & Steele 2007) y actualmente se considera el modelo estándar (Lahtinen et al. 2011). La actividad caseinolítica es llevada a cabo por una serín proteasa de la familia de las subtilisinas llamada PrtP, anclada a la pared celular (Figura 1.1). Esta enzima degrada caseína obteniéndose péptidos de tamaño variable. Los péptidos largos (4 a 18 aminoácidos) son transportados por un sistema de transporte de oligopéptidos (Opp), mientras que los di y tripéptidos son transportados por otro sistema (D) Estos sistemas son del tipo ABC . Dentro de las células, los péptidos son procesados por peptidasas intracelulares hasta obtener aminoácidos libres (Lahtinen et al. 2011).

Muchas BAL producen NH_3 a partir de aminoácido arginina a través de la ruta de la arginina desaminasa, vía que puede servir como fuente de energía para algunos cultivos iniciadores (Laht et al. 2002). La producción de amonio en los procesos de fermentación puede convertirse en un problema operativo, ya que éste neutraliza la producción de ácido y por tanto afecta el rendimiento de la producción. En esta vía, la arginina se convierte en citrulina por acción de la arginina desaminasa con liberación de NH_3 . La citrulina se fosforila a carbamoil-fosfato por medio de la enzima carbamoiltransferasa. El carbamoil-fosfato es luego hidrolizado a CO_2 y NH_3 con liberación de un mol de ATP por la acción de una carbamato kinasa.

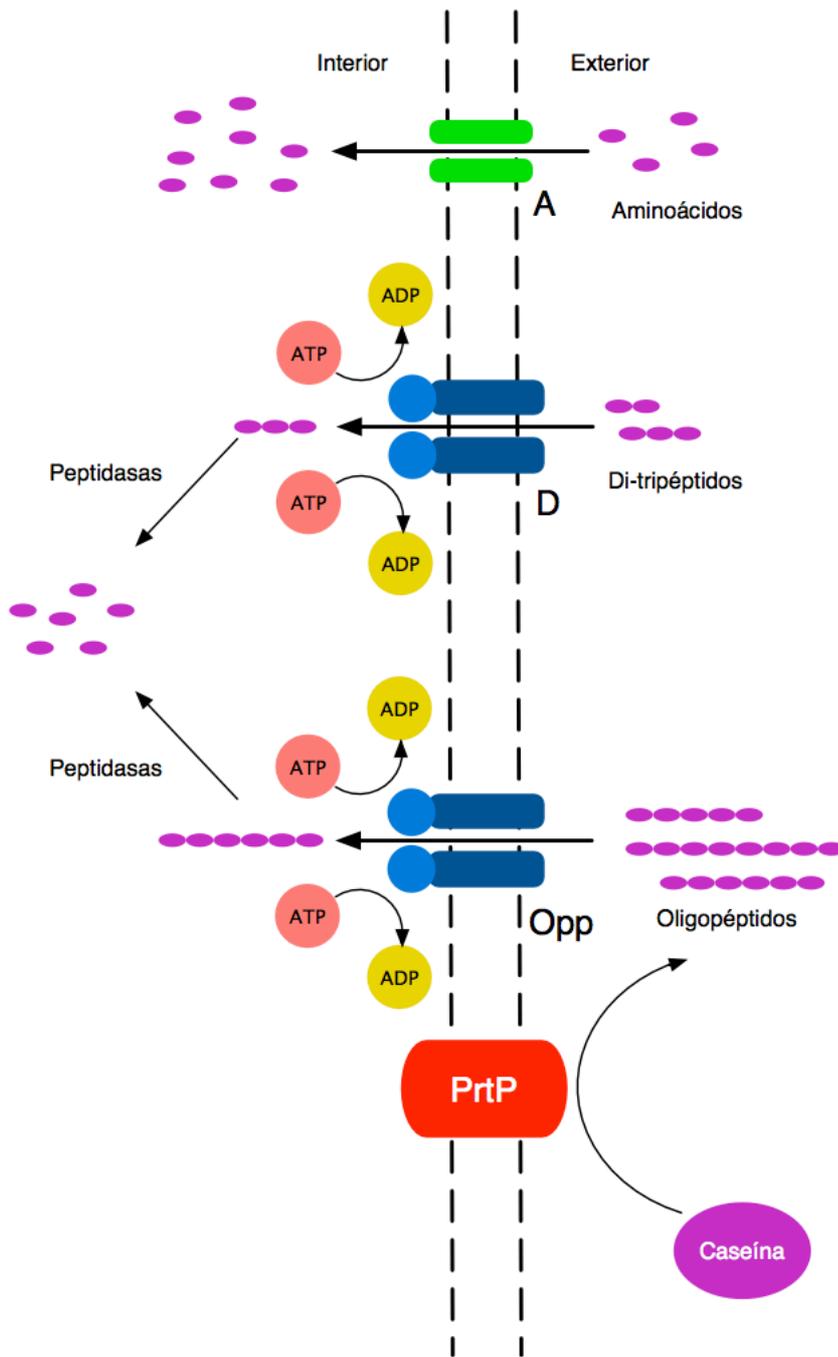


Figura 1.1. Representación esquemática del sistema proteolítico en bacterias ácido-lácticas

La tasa de acidificación es un rasgo dependiente de cada cepa, que puede estar influenciado por varios factores, como el metabolismo lactosa-galactosa, el sistema proteolítico y la actividad ureolítica. La actividad de ureasa es un rasgo fenotípico ampliamente distribuido en cepas de *S. thermophilus* (Mora et al. 2004) y solo se han descrito unas pocas cepas ureolíticas lentas o negativas (Mora et al. 2002). En algunos casos la actividad ureolítica de las cepas de *S. thermophilus* puede ser responsable de una baja velocidad de acidificación de las mismas (Martin et al. 1997). Específicamente, la ureasa (urea amidohidrolasa, EC 3.5.1.5) cataliza la hidrólisis de urea (presente en la leche

a una concentración de aproximadamente 0.2-0.4 g/l) para producir amoníaco y carbamato que se descompone espontáneamente para producir una segunda molécula de amoníaco y ácido carbónico. Estas reacciones conducen a una disminución más lenta del pH, amortiguando la producción de ácido láctico durante el proceso de fermentación (Mora et al. 2004). Los altos niveles de urea en la leche son causados por multiplicidad de factores, siendo hoy en día la principal causa la alimentación con raciones altas en proteínas con las que se alimenta el ganado lechero.

Algunas BAL son capaces de secretar polisacáridos extracelulares (EPS). Estos han sido objeto de estudio debido a su potencial como prebióticos, especialmente los heteropolisacáridos que contienen gluco- y fructo-oligosacáridos (GOS y FOS respectivamente). Aunque los EPS proporcionan un carácter viscoso a la matriz donde los microorganismos están creciendo, el papel fisiológico exacto de estas moléculas no está claro. Los EPS contribuyen a la mejora de la reología, textura y sensación en la boca de determinados productos lácteos como el yogurt, además de funcionar como agentes espesantes (Welman & Maddox 2003). Se ha reportado que los cultivos productores de EPS mejoran las propiedades funcionales del queso mozzarella bajo en grasa y parcialmente desnatado (Petersen et al. 2000); y las propiedades viscoelásticas del queso Cheddar reducido en grasa (Hassan et al. 2005).

Cultivos *starters*

Como se mencionó anteriormente, en el inicio de la elaboración de alimentos fermentados el proceso recaía sobre los procesos que ocurrían de forma natural (fermentación espontánea) en la materia prima. La fermentación espontánea se optimizó a través de la re inoculación o *back slopping*, es decir, la inoculación de la materia prima con una pequeña cantidad de una fermentación exitosa previamente realizada. Por lo tanto, el *back slopping* resultó en la selección de las cepas mejor adaptadas al proceso. Esta práctica utiliza un cultivo *starter* seleccionado de forma de acortar el proceso de fermentación y reducir el riesgo de falla de la misma. En la actualidad es una práctica corriente en la elaboración de quesos artesanales, en la producción de chucrut y masa fermentada, y particularmente en la elaboración de productos para los cuales la ecología microbiana y el papel preciso de la población microbiana no son bien conocidos.

Hoy en día, la producción de alimentos mediante fermentación espontánea y *back slopping* representa un método de preservación barato y confiable en los países menos desarrollados, mientras que en los países occidentales la producción de alimentos fermentados se ha convertido en procesos de gran escala y representa una rama importante de la industria alimentaria .

La adición directa de cultivos iniciadores seleccionados a las materias primas ha sido un gran avance en el procesamiento de alimentos fermentados, ya que resulta en un alto grado de control sobre el proceso de fermentación y la estandarización del producto final. Las cepas con las características fisiológicas y metabólicas apropiadas se aíslan de hábitats naturales o de productos fermentados con éxito.

Un cultivo iniciador o *starter* es una preparación microbiana que se agrega intencionalmente para lograr cambios deseables bajo condiciones controladas de fermentación. Estas preparaciones pueden incluir bacterias del ácido láctico, bacterias propiónicas, así como otras bacterias, levaduras y mohos que maduran en la superficie. La mayoría de los starters utilizados en la industria láctea están compuestos por BALs. El laboratorio de Chr. Hansen en 1878 fue el primero en introducir cultivos iniciales lácticos comerciales para la fabricación de queso y manteca en América del Norte. Inicialmente, estos eran cultivos líquidos en botellas de vidrio que requerían algunos repiques antes de su uso para la fabricación de alimentos. Éstos eran producidos en Dinamarca y enviados a Estados Unidos junto con coagulantes lácteos, distribuidos desde la ciudad de Nueva York. En 1880 Las operaciones de Chr. Hansen se mudaron a Little Falls, NY. En ese momento, Little Falls era el centro de la producción de queso en los EE. UU (Ozer & Akdemir-Evrendilek 2014).

Inicialmente, las cepas utilizadas en los cultivos iniciadores fueron recolectadas a lo largo de los años en fábricas que tuvieran éxito y uniformidad en la elaboración de queso a partir de leche cruda. Los primeros cultivos *starter* comerciales consistían en una mezcla de lactococos y leuconostoc (mejoradores del sabor). Finalmente, los cultivos fueron purificados; se cultivaron en leche desnatada pasteurizada y se mantuvieron como cultivos estables mixtos de cepas múltiples (Cogan et al. 2007). Los primeros *starters* liofilizados estuvieron disponibles a mediados de la década de 1950. En 1975, el *Marschall Dairy Laboratory* introdujo los primeros cultivos iniciadores congelados y altamente concentrados para la utilización directa en la preparación de queso, crema agria, etc. (Ozer & Akdemir-Evrendilek 2014).

Uno de los grandes desafíos para el desarrollo de cultivos iniciadores económicamente viables, fue crear dosis de alta concentración. Inicialmente, se usaron sistemas de control de pH externos y posteriormente, internos para hacer crecer aún más los cultivos iniciadores a un número mayor. Un sistema de control externo consiste en la regulación del pH mediante neutralización del ácido láctico producido por los microorganismos mediante la adición de hidróxido de amonio. De esta forma se evita la inhibición por ácido de las bacterias y permite que el cultivo alcance muy altas concentraciones. Este método se utilizó para producir cultivos a granel a principios de la década de 1960 y aún es una práctica utilizada. Los sistemas de control de pH interno logran el mismo efecto con el uso de sales tampón incorporadas en el medio. La estabilidad de los cultivos y la consecuente alta actividad de los mismos se alcanzaron mediante la aplicación de tecnologías de criopreservación, principalmente por nitrógeno líquido (Høier et al. 2010). Todos los tipos de cultivos iniciadores se produjeron mediante estos procedimientos descritos. Hoy en día, todas las empresas que venden cultivos *starters* a la industria láctea utilizan estos procesos básicos de elaboración o algunas modificaciones según el organismo.

Los cultivos lácteos iniciadores se agrupan comúnmente en cultivos mesófilos (temperatura óptima de crecimiento de aproximadamente 30°C) y cultivos termófilos (temperatura óptima de aproximadamente 42-45°C). Además, los cultivos *starter* se clasifican como cultivos definidos o cultivos mixtos (complejos). Los cultivos definidos son elaborados a partir de cepas puras bien caracterizadas y cuidadosamente seleccionadas en base a varias características importantes, tales como resistencia a fagos, producción de ácido, utilización de citrato, producción de aromas y sabores, etc. Los cultivos lácteos mixtos son formulados a menudo con un número indefinido de cepas de una misma especie, o incluso de diferentes especies. Por último, puede definirse una tercera categoría de cultivos indefinidos, que, como su nombre indica, son cultivos mixtos cuya composición no es realmente conocida (Fox et al. 2000).

Los cultivos lácteos mesófilos incluyen principalmente *Lactococcus* y *Leuconostoc*, y también en menor grado de importancia *Pediococcus*. Las principales especies de *Lactococcus* son *Lactococcus lactis* subsp. *cremoris* y *Lactococcus lactis* subsp. *lactis*. Estas dos subespecies se pueden diferenciar entre sí por la capacidad de *L. lactis* subsp. *lactis* de crecer a 40°C en leche con 4% de NaCl y 0,1% azul de metileno, a pH 9,2, y por la formación de NH₃ a partir de arginina (Stiles & Holzapfel 1997). Además, *L. lactis* subsp. *lactis* posee una enzima glutamato descarboxilasa, mientras que *L. lactis* subsp. *cremoris*

no, permitiendo la producción de ácido γ -aminobutírico a partir de glutamato por este organismo. *Lactococcus lactis* biovar. *diacetylactis* tiene la capacidad de metabolizar citrato para producir diacetilo. En el género *Leuconostoc*, dos especies son importantes en la fermentación láctica. *Leuconostoc mesenteroides* subsp. *cremoris* y *Leuconostoc lactis*, ambos heterofermentadores. Las especies del género *Leuconostoc*. También tienen la capacidad de metabolizan el citrato para producir CO₂, diacetilo y acetato, compuestos importantes en la formación de los ojos (en quesos Gouda y Edam) y sabor (por ejemplo, en el requesón). Los organismos que tienen la capacidad de metabolizar el citrato se consideran productores de aroma. En el pasado, los cultivos mesófilos mixtos se han clasificado como L, D, LD u O. Esta clasificación es basada en si los cultivos contenían *Leuconostoc spp* (L), *L. lactis* biovar. *diacetylactis* (D), ambos (LD), o sin productores de sabor (O) respectivamente (Vasiljevic & Shah 2008). Los cultivos mixtos mesófilos típicamente se componen por un 90% de organismos productores de ácido y 10% de organismos productores de aroma.

Como se mencionó anteriormente, los cultivos termófilos tienen una temperatura óptima de crecimiento de aproximadamente 42-45°C. Los cultivos termófilos más importantes utilizados en la elaboración de productos lácteos incluyen *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus delbrueckii* subsp. *lactis* y *Lactobacillus helveticus*. Estos microorganismos pueden inocularse de forma combinada (como en el caso del yogurt) o de forma individual (como por ejemplo en la producción de algunos quesos). La relación simbiótica entre *S. thermophilus* y *Lb. delbrueckii* subsp. *bulgaricus* está muy bien establecida. Estos dos organismos se usan en la producción de yogurt. *Lb. delbrueckii* subsp. *bulgaricus* produce además de ácido láctico a partir de lactosa, pequeños péptidos y aminoácidos como resultado la proteólisis de la caseína, los cuales estimulan el crecimiento de *S. thermophilus*. Por su parte, *S. thermophilus* produce pequeñas cantidades de ácido fórmico, el cual estimula el crecimiento de *Lb. delbrueckii* subsp. *bulgaricus* (Vasiljevic & Shah 2008). Esta relación simbiótica proporciona una mayor tasa de crecimiento y, por ende, una mayor producción de ácido.

Los cultivos adjuntos se usan típicamente en los quesos para proporcionar un desarrollo de sabor adicional. Estos cultivos pueden aumentar la intensidad y cambiar el equilibrio del sabor del queso. Un ejemplo de esto es la adición de organismos termófilos los cuales son capaces de sobrevivir a las temperaturas de cocción de la cuajada y participar más tarde en el desarrollo del sabor. Sin embargo, estos organismos no crecen ni contribuyen a la

producción de ácido o sabor a temperaturas inferiores a 25°C. Microorganismos mesófilos tales como *Lactobacillus casei* y *Lactobacillus paracasei* también se han usado tradicionalmente en la mejora del sabor (por ejemplo, en el queso Cheddar). Hoy en día, la falta de sabor en el queso se da como consecuencia de la alta calidad de la leche cruda (bajos recuentos microbianos), la pasteurización y la mejora de la higiene en las plantas queseras.

Otro grupo importante de microorganismos en la elaboración de quesos son las bacterias ácido-lácticas no starters o NSLABs (*Non Starter Lactic Acid Bacteria*). Durante la fabricación, las bacterias lácticas starters, son inoculadas en grandes cantidades y los recuentos alcanzan valores de 10⁹/g de cuajada, pero luego mueren rápidamente. Mientras la población starter declina, otro importante grupo microbiano (las NSLABs) prolifera en el queso. Las NSLABs son bacterias contaminantes, que pueden provenir de la propia leche utilizada en la elaboración (pequeño número de sobrevivientes a la pasteurización que proliferan post fermentación) o del ambiente de la planta. Al ser un grupo adventicio (y por tanto no controlado) tienen un impacto sobre el producto final que no se puede generalizar (es decir, que no puede definirse como “bueno” o “malo”). Entre los impactos que pueden producir estos microorganismos, se encuentran fermentaciones tardías y proteólisis de la cuajada, procesos que afectan los perfiles de sabor, aroma y textura de los quesos. Las especies estudiadas que producen un mayor impacto son : *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus casei*, *Lactobacillus brevis* y *Lactobacillus pentosus*.

Más recientemente, se han agregado microorganismos probióticos como cultivos adyuvantes en la elaboración de los productos lácteos fermentados tradicionales, incluidos el yogur y los quesos fermentados de leche. *Bifidobacterium spp.* y *Lactobacillus acidophilus* se han utilizado como cultivos probióticos adjuntos debido a la acumulación de datos sobre sus beneficios para la salud (Ozer & Akdemir-Evrendilek 2014).

Streptococcus thermophilus

Posiblemente, *Streptococcus thermophilus* (Figura 1.2)sea una de las principales bacterias consumidas por el hombre. Se cree que *S. thermophilus* ha sido utilizada en la elaboración de alimentos desde que los animales fueran domesticados y el acceso a la leche algo común, hace aproximadamente 10.000 años (Floch et al. 2016). Hoy en día, es considerada

la segunda especie más importante de BAL industrial después de *Lactococcus lactis*, con un valor de mercado de alrededor de 40 mil millones de dólares (Iyer et al. 2010).



Figura 1.2. Observación al microscopio (100x) de una tinción de Gram de *Streptococcus thermophilus*

Taxonomía

La dificultad de la ubicación taxonómica de *S. thermophilus* es reflejo de su reciente emergencia desde un ancestro común con otros estreptococos. Originalmente *S. thermophilus* fue considerada una especie en sí misma por Orla-Jensen en 1919 (Schleifer et al. 1991), pero esto ha ido cambiando a lo largo de los años. *S. thermophilus* y *Streptococcus salivarius* están altamente relacionados por lo que durante mucho tiempo hubo dudas sobre si los dos deberían considerarse una misma especie o no; incluso algunos investigadores propusieron que *S. thermophilus* debería ser una subespecie de *S. salivarius*. En 1984, los experimentos de hibridación de ADN-ADN colocaron a *S. thermophilus* bajo el paraguas de *S. salivarius*, pasando a ser nombrada *S. salivarius* subsp. *thermophilus* (Farrow & Collins 1984). Schleifer y colaboradores llevaron a cabo más experimentos de hibridación ADN-ADN bajo condiciones estrictas varios años después, y determinaron que estos microorganismos debían volver a adquirir el *status* de especies diferentes (Schleifer et al. 1991).

Entre los estreptococos, *S. thermophilus* se encuentra dentro del grupo *viridians*, informalmente considerado como el grupo de los comensales alfa-hemolíticos. Este grupo también incluye otras dos especies, *S. salivarius* y *Streptococcus vestibularis*. Estas dos especies comensales de la cavidad oral y el intestino humano, mientras que el reservorio

de *S. thermophilus* no ha sido identificado fuera de la leche. *S. salivarius* también posee un historial de uso en la elaboración alimentos en varias partes del mundo, con registros que documentan su presencia en cultivos *starters* naturales utilizados en fermentaciones tradicionales en Europa y África. Sin embargo, *S. salivarius* no se encuentra tan bien adaptada al ambiente lácteo como *S. thermophilus* (Floch et al. 2016).

Al menos media docena de genomas completos de *S. thermophilus* han sido secuenciados en su totalidad, y del mismo modo otras 47 cepas han sido estudiadas utilizando un enfoque genómico comparativo (Rasmussen et al. 2008; Hols et al. 2005; Bolotin et al. 2004). Esta información ha aportado evidencia sustancial que respalda el rol significativo de la transferencia horizontal de genes en la adaptación de *S. thermophilus* al ambiente lácteo. Más aún, estos eventos siguen produciéndose por lo que puede decirse que *S. thermophilus* aún sigue adaptándose a su entorno.

Mientras que *S. thermophilus* todavía parece estar adquiriendo nuevo material, solo el 58% de los genes identificados forma el genoma *core* (Rasmussen et al. 2008). Esto es probablemente, como se mencionó anteriormente, reflejo de su aparición relativamente reciente como una nueva especie desde su ancestro comensal del grupo *salivarius* (Rasmussen et al. 2008). La fuerza principal que ha guiado y guía este cambio, es la necesidad de adaptación a la vida en un medio nutricionalmente limitado según los requerimientos bacterianos. La fuente principal de nutrientes para crecimiento en leche son carbohidratos fermentables, que en gran medida se encuentran en forma de lactosa. Una de las características más interesantes notadas en el análisis inicial de la secuencia del genoma fue que *S. thermophilus* tiene aproximadamente 1500 genes (80%) con ortólogos a genes encontrados en otras especies de estreptococos, lo que indica que *S. thermophilus* y sus parientes patógenos y comensales todavía comparten una parte sustancial de su fisiología y metabolismo general (Bolotin et al. 2004). Los genes no esenciales relacionados típicamente con la virulencia en el género *Streptococcus* se encuentran inactivos o directamente no se detectaron en *S. thermophilus* (Hols et al. 2005). Aproximadamente el 10% de los ORFs en *S. thermophilus* son pseudogenes, sus funciones originales son innecesarias para el crecimiento en la leche o se han adaptado. Un ejemplo de adaptación, son los genes necesarios para la síntesis de exopolisacáridos (EPS). Éstos, están relacionados con los genes que participan en la síntesis de la cápsula (un factor de patogenicidad) en *Streptococcus pneumoniae* y *Streptococcus agalactiae* (Bolotin et al. 2004).

Además de la disminución y la pérdida de genes, la transferencia horizontal de genes ha contribuido en gran medida a la configuración del genoma de *S. thermophilus* (Bolotin et al. 2004). Muchos genes adquiridos por *S. thermophilus* parecen provenir de otras especies de bacterias ácido lácticas, como *L. lactis* y *Lactobacillus delbrueckii*, y por lo tanto contribuyen a su adaptación al medio ambiente de la leche (Floch et al. 2016).

En los genomas de las cepas de referencia LMG18311 y CNRZ1066, existen 450 secuencias de inserción, algunas con contenido GC anómalo (es decir con un porcentaje diferente al del resto del promedio del genoma) y asociadas con genes relevantes para la adaptación al ambiente lácteo. Aproximadamente el 75% de las secuencias de inserción están asociadas con el cambio en el orden de genes de *S. thermophilus* con respecto al genoma de *S. pyogenes*, lo que sugiere que estas secuencias desempeñan un papel importante en la conformación del genoma. Un ejemplo particularmente interesante de transferencia horizontal es una región de 17 kb encontrada dentro de un gen *pepD* truncado, presente en ambas cepas de *S. thermophilus*. Podría considerarse como un punto caliente de transferencia lateral de genes, ya que contiene tres de las seis copias de la secuencia de inserción 1191 presentes en la cepa LMG18311 y constituye un mosaico de fragmentos de ADN con más del 90% de identidad con ADN de *Lactobacillus bulgaricus*, y dos subespecies de *L. lactis* (*lactis* y *cremoris*), todas bacterias cuyo nicho es precisamente la leche. Curiosamente, la región flanqueante izquierda se encuentra conservada en dos especies de estreptococos (*Streptococcus equii* y *S. mutans*), y, de forma similar, la región flanqueante derecha se encuentra conservada en *S. equii*. Esta conservación apoya la hipótesis de que las inserciones tuvieron lugar en el genoma de *S. thermophilus* y no en un estreptococo ancestral. El fragmento ADN de *L. bulgaricus* (3.6 kb) presente en este mosaico, incluye una copia única del gen *metC*, el cual permite la biosíntesis de metionina, un aminoácido muy raro en la leche. El alto nivel de identidad (95%) de las secuencias *metC* en ambas especies, revela un evento de transferencia muy reciente. Estas dos especies se encuentran filogenéticamente muy distanciadas, pero son utilizadas en conjunto para la elaboración de yogurt. Esto afirma el concepto de que la proximidad ecológica es más importante que la proximidad filogenética a la hora de que se den eventos de HGT (Bolotin et al. 2004).

Los estudios de genómica comparativa muestran que la aparición de *Streptococcus thermophilus* como especie se ha dado principalmente por eventos de pérdida de funciones a pesar del importante rol de la transferencia horizontal. Es la primera instancia donde se

observa evolución regresiva en un nicho alimenticio en lugar de un patógeno o en situaciones simbiote-huésped (Cole et al. 2001; Wren 2000). El decaimiento masivo en el número de genes dio como resultado la inactivación y pérdida de la mayoría de los determinantes de virulencia. Esto proporciona un fuerte argumento genómico a favor del estatus “Generalmente reconocido como seguro” (GRAS por sus siglas en inglés) de *Streptococcus thermophilus* (Bolotin et al. 2004). Esta gran cantidad de información genómica refuerza el hecho de que el consumo masivo de esta bacteria por parte de los seres humanos no conlleva ningún riesgo para la salud (Floch et al. 2016).

Atributos tecnológicos y funcionales de *S. thermophilus*

S. thermophilus es de gran importancia para la industria alimentaria ya que como se mencionó anteriormente, se utiliza ampliamente para la fabricación de productos lácteos (Hols et al. 2005). Además del uso tradicional en combinación con *L. delbrueckii* subsp. *bulgaricus* en yogur, *S. thermophilus* se usa para producir diversas variedades de queso, como el queso suizo, el parmesano, el provolone, la mozzarella y asiago (Parente & Cogan 2004). También se ha utilizado recientemente para la producción de Cheddar en combinación con cultivos *starter* mesófilos (Awad et al. 2005).

Una de las principales funciones de *S. thermophilus* en la fermentación de la leche es proporcionar una acidificación rápida. Marino y colaboradores encontraron que *S. thermophilus* es la especie predominante y con una tasa de acidificación mayor entre las BAL a lo largo de la maduración del queso Montasio artesanal. Concluyeron que la velocidad de acidificación es un rasgo metabólico dependiente de cada cepa, y que puede ser influenciado por varios factores, como el metabolismo lactosa-galactosa, el sistema proteolítico y la actividad ureasa (Marino et al. 2003). Además del ácido láctico, también produce bajos niveles de formiato, acetoína, diacetilo, acetaldehído y acetato como productos finales adicionales (Ott, J.-E. Germond, et al. 2000).

Por lo tanto, el papel de *S. thermophilus* en la fermentación de la leche no está relacionado únicamente con la producción de ácido láctico. Este espectro de cualidades tecnológicas representa el grado de diversidad fenotípica existente dentro de la especie (Iyer et al. 2010).

Metabolismo de azúcares

El papel principal de *S. thermophilus* en las fermentaciones de lácteos es la rápida conversión de lactosa en lactato, pero también la producción de otros compuestos que contribuyen al sabor y la textura. Cinco diferentes azúcares pueden ser fermentados por *S. thermophilus*: lactosa, sacarosa, glucosa, galactosa y fructosa (Vaughan et al. 2001; Vaillancourt et al. 2002), siendo los dos últimos azúcares solo fermentados por un número limitado de cepas. Del mismo modo, la diversidad de los productos finales de la fermentación es bastante limitada.

Los estreptococos patógenos exhiben una cierta capacidad para degradar polímeros complejos tales como almidón o glucanos (dextranos). Enzimas específicas, como la dextranasa y la mutanasa, son producidas por estreptococos orales (*S. mutans*, *S. salivarius*) y están implicadas en la remodelación de estructuras de glucano en la placa dental (Hayacibara et al. 2004). Otras enzimas, como la levansucrasa, están directamente implicadas en la síntesis de los fructanos a partir de sacarosa y contribuyen a la formación de biofilms (Rathsam et al. 1993). En los genomas completos de 4 estreptococos patógenos (*S. pneumoniae*, *S. pyogenes*, *S. agalactiae* y *S. mutans*), se identificaron genes que codifican dextranasas, amilasas y pululanases. Sin embargo, la capacidad para degradar, remodelar o sintetizar estos polímeros complejos se reduce severamente en *S. thermophilus*. El genoma contiene tres pseudogenes que codifican los restos de una dextranasa, una pululanasa y una levansucrasa, sin embargo, se identificó un gen completo de amilasa (*amyL*). Esto está en consonancia con el hallazgo de que esta especie muestra una baja actividad amilolítica (Slos et al. 1991).

La sacarosa y la fructosa son los únicos azúcares que son transportados por sistemas fosfotransferasa fosfoenolpiruvato-dependientes (PTS) en *S. thermophilus*, pero con una eficacia menor que la lactosa (Van Den Bogaard et al. 2004). La glucosa es un azúcar no-PTS en *S. thermophilus* y un sustrato pobre para su crecimiento (Poolman 1993). Sorprendentemente, *S. thermophilus* está muy adaptado al crecimiento en lactosa, que es transportada por *LacS*. Una vez dentro de la célula, la b-galactosidasa (*LacZ*) hidroliza la lactosa en glucosa y galactosa. Si bien todas las cepas utilizan el residuo de glucosa como fuente de energía, la galactosa generalmente no se metaboliza y es expulsada al medio externo por *LacS*, que funciona de esta forma como un *antiporter* lactosa/galactosa (Gunnewijk & Poolman 2000). La no utilización de cómo fuente de energía, no es el resultado

de falta de información genética, ya que los genes que codifican la vía Leloir (*galK-TEM*) están presentes en el genoma de numerosas cepas incapaces de metabolizarla (Van Den Bogaard et al. 2004). El análisis de mutantes *Gal+* espontáneos ensayos de complementación genética utilizando el gen *galK* de *S. salivarius*, han establecido e que la baja actividad galactoquinasa es la principal limitante en la utilización de galactosa en *S. thermophilus* (Vaillancourt et al. 2004; Vaughan et al. 2001)

En vista de su importancia para la fermentación en la leche, el metabolismo de la lactosa de *S. thermophilus* ha sido estudiado ampliamente. Los genes implicados en el metabolismo de la galactosa (vía de Leloir) y la lactosa (transporte e hidrólisis) están genéticamente vinculados y forman un grupo de genes (*galKTEMlacSZ*), que está altamente conservado entre diferentes cepas (Ercolini et al. 2005). Asimismo, este *clúster* se encuentra conservado *S. salivarius* (Vaillancourt et al. 2002), lo que refleja la estrecha relación evolutiva entre estas bacterias. Aguas arriba del gen *galK* de *S. thermophilus*, se encuentra un gen orientado divergentemente que codifica un regulador transcripcional llamado *galR*. Se ha demostrado que *galR* regula positivamente la transcripción de los dos operones que componen el *clúster* gal-lac (*galKTEM* y *lacSZ*), mientras que regula negativamente su propia expresión (Vaughan et al. 2001). Por el contrario, el homólogo a *galR* de *S. mutans*, regula negativamente la expresión de los genes *galKTE*, lo que demuestra que diferentes estreptococos emplean modos variables de regulación utilizando elementos genéticos homólogos (Vaughan et al. 2001; Ajdić & Ferretti 1998). Tras la inducción mediada por galactosa o lactosa, los genes *galKTE* de *S. thermophilus* se transcriben a un nivel muy bajo, lo que explica el fenotipo *Gal-* mostrado por muchas cepas de esta especie. Curiosamente, se pudieron obtener fácilmente mutantes espontáneos capaces de fermentar galactosa a partir de la cepa *S. thermophilus* CNRZ302 y se demostró que expresan los genes *galKTE* a un nivel mucho más alto en comparación con el tipo silvestre. El análisis molecular reveló que esto se debía a mutaciones específicas dentro de la región promotora de *galK* (Vaughan et al. 2001), lo que se corresponde con el hallazgo de mutaciones específicas de este mismo promotor en aislados de *S. thermophilus* con fenotipo *Gal+* (Van Den Bogaard et al. 2004).

Además de la regulación positiva por *galR*, el promotor *lacS* de *S. thermophilus* se encuentra regulado por la proteína de control de catabolito A, o CcpA, lo que coincide con la presencia de un elemento sensible a CcpA (*cre*) en el promotor *lacS*. CcpA se ha estudiado en muchas bacterias Gram positivas donde esta proteína es responsable de la represión del catabolismo cuando las células se cultivan en fuentes de carbono PTS, de las cuales la

glucosa es generalmente la preferida. Como se indicó anteriormente, *S. thermophilus* prefiere la lactosa sobre la glucosa y transporta este disacárido a través del transportador no-PTS LacS. La disrupción funcional de *ccpA* en *S. thermophilus* CNRZ302, libera la represión de *lacSZ* cuando las células crecen en lactosa (Van Den Bogaard et al. 2000). Además de regular los niveles de LacS y LacZ, CcpA también regula los niveles de expresión de tres enzimas clave de la ruta glucolítica en *S. thermophilus*. La expresión de los genes codificantes para la lactato deshidrogenasa (*ldh*), fosfo-fructoquinasa (*pfk*) y piruvato quinasa (*pyk*) parecen estar regulada de manera *ccpA*-dependiente cuando las células crecen en lactosa (Van Den Bogaard et al. 2000). También se ha reportado que existe un control mediado por CcpA de las mismas enzimas glucolíticas para otras bacterias, incluidas *S. bovis* y *L. lactis* (Luesink et al. 1998), lo que sugiere que el control global de del flujo glucolítico mediado por catabolito es un mecanismo de regulación conservado en las bacterias del ácido láctico. A nivel fisiológico, cepas mutantes *ccpA*- mostraron una tasa de crecimiento reducida en lactosa, mientras que en el nivel de metabolitos, mostraron un fuerte aumento en la absorción de lactosa, en la hidrólisis intracelular de la misma, una expulsión masiva de galactosa y glucosa, y una reducida tasa de formación de lactato (Van Den Bogaard et al. 2000). A raíz de esto, se ha concluido que el control del promotor *lacS* de CcpA juega un papel clave en el ajuste de la absorción de lactosa, la tasa de hidrólisis y capacidad glucolítica total (Hols et al. 2005).

El control mediado por CcpA está estrechamente relacionado con el estado metabólico y energético de la célula a través de la modulación del estado de fosforilación de la proteína fosfotransportadora HPr. La fosforilación diferencial de HPr en un residuo de histidina (HPr-His-P) y / o serina (HPr-Ser-P) es un determinante clave en el metabolismo del carbono en muchas bacterias, y se ve afectada principalmente por metabolitos intracelulares que reflejan el estado metabólico y energético de la célula, incluido el intermediario glucolítico fructosa bifosfato (FBP) y el ATP (Hols et al. 2005). En *S. thermophilus*, HPr-Ser-P es la forma dominante de HPr durante el crecimiento exponencial en presencia de exceso de lactosa (Gunnewijk & Poolman 2000). En estas condiciones, HPr-Ser-P forma un complejo con CcpA, el cual reprime la expresión de *lacSZ* e induce la expresión de los genes *ldh*, *pyk* y *pfk*, optimizando así el transporte de lactosa y el flujo glucolítico (Gunnewijk & Poolman 2000). En condiciones limitantes de lactosa, se produce una transición del estado de fosforilación de HPr, que conduce a la formación de HPr-His-P como la forma dominante (Gunnewijk & Poolman 2000), lo que conduce a una disminución flujo glucolítico y una

expresión aumentada de *lacSZ* (Van Den Bogaard et al. 2000). HPr tiene una función adicional en *S. thermophilus*: puede transferir su grupo fosforilo al dominio tipo PTS IIA de LacS, estimulando la tasa de intercambio galactosa-lactosa de este transportador (Titgemeyer & Hillen 2002). Por lo tanto, HPr tiene un doble papel en la regulación del metabolismo de lactosa en *S. thermophilus*.

Metabolismo de nitrógeno

Dado que la leche contiene bajas cantidades de aminoácidos libres y péptidos cortos, el crecimiento óptimo de *S. thermophilus* en este medio requiere de la hidrólisis de caseína, la internalización y la hidrólisis de los péptidos resultantes, o la biosíntesis de aminoácidos *de novo* (Iyer et al. 2010). Sin embargo, parece ser menos exigente que muchas otras BAL como *L. lactis* u otros lactobacilos. Análisis de omisión de un único aminoácido revelaron que el número y el tipo de aminoácidos esenciales requeridos para el crecimiento dependen de la cepa (Garault et al. 2000). Mientras que algunas cepas de *S. thermophilus* no mostraron ningún requerimiento especial, otras mostraron auxotrofías para al menos cuatro aminoácidos. Experimentos de omisión múltiple de aminoácidos han demostrado que el agotamiento tanto del glutamato como de la glutamina, o de cisteína y metionina, anula completamente el crecimiento de *S. thermophilus*, lo que indica que las vías de síntesis de éstos no son completamente funcionales (Letort & Juillard 2001).

La relativamente alta conservación de genes funcionales para la biosíntesis de aminoácidos en *S. thermophilus* refleja la importancia de la síntesis para crecer en leche. Esto también se ve reflejado por la baja cantidad de pseudogenes asociados a la biosíntesis de aminoácidos. Solamente los genes *alaD*, *dapE*, *ilvD2* e *yhcE* involucrados en la biosíntesis de alanina, lisina, aminoácidos ramificados y con sulfuro, aparecen como pseudogenes en la cepa LMG 18311, mientras que la cepa CNRZ1066 carece de estos genes además de *aspC3*, que codifica la aspartato aminotransferasa (Hols et al. 2005).

S. thermophilus posee un sistema proteolítico similar al que se encuentra en otros BAL y está compuesto por una proteasa extracelular anclada a células capaz de hidrolizar caseína, un conjunto de sistemas de transporte de aminoácidos y péptidos y un conjunto de peptidasas intracelulares implicadas en la hidrólisis de péptidos derivados de caseína. Aunque *S. thermophilus* está potencialmente equipado con más de 20 enzimas proteolíticas, su sistema proteolítico sigue estando poco caracterizado y su funcionamiento

se infiere en gran medida de sistemas más estudiados como el de *Lactococcus lactis* (Hols et al. 2005).

La proteasa extracelular PrtS de *S. thermophilus* es una serín-proteasa de la familia de las subtilisinas, y se encuentra anclada en la pared celular (Fernandez-Espla et al. 2000). PrtS está presente en solo unas pocas cepas de *S. thermophilus* y su presencia se ha asociado con un crecimiento rápido y altas tasas de acidificación en la leche (Fernandez-Espla et al. 2000). PrtS es esencial para el crecimiento óptimo de *S. thermophilus* cuando se encuentra sola en la leche. Sin embargo, cuando se cultiva conjuntamente con una cepa de *Lactobacillus bulgaricus*, *S. thermophilus* es capaz de crecer óptimamente utilizando péptidos liberados ésta (Courtin et al. 2002). Esta cooperación podría explicar la ausencia de PrtS en numerosas cepas de *S. thermophilus* (Hols et al. 2005).

Bacteriófagos

Los virus son partículas infecciosas extremadamente pequeñas, generalmente de tamaños del orden de nanómetros, que no son visibles en un microscopio óptico y pueden pasar a través de filtros de porcelana fina. Existen en una gran variedad de formas e infectan prácticamente todos los sistemas vivos. Todos los virus tienen un genoma, típicamente compuesto por un solo tipo de ácido nucleico, pudiendo ser una o varias moléculas de ADN o ARN. Este genoma se encuentra rodeado por una capa protectora estable llamada cápside y, algunas veces, por capas adicionales que pueden ser muy complejas y contener carbohidratos, lípidos y proteínas adicionales. Los virus que tienen solo una capa de proteína se denominan virus "desnudos" o virus sin envoltura. Muchos virus tienen una envoltura (virus envueltos) formada a partir de la membrana lipídica de la célula huésped durante la liberación del virus (Kurtboke 2012).

Los virus interactúan con diferentes tipos de células en organismos vivos y producen diferentes tipos de enfermedades. Cada virus infecta un cierto tipo de célula que generalmente se llama célula "huésped". La principal característica de cualquier enfermedad viral es la lisis celular, proceso que involucra la liberación de la partículas virales conllevando la muerte de esta. En organismos multicelulares, si mueren suficientes células, todo el organismo sufrirá problemas. Algunos virus pueden causar infecciones de por vida o crónicas, donde los virus continúan reproduciéndose en el cuerpo a pesar de los mecanismos de defensa del huésped. Otros virus causan infecciones de por vida porque el

virus permanece dentro de su célula huésped en un estado latente (latente) como es el caso del herpes, pero el virus puede reactivarse y producir nuevos ataques de enfermedad en cualquier momento si el sistema de defensa del huésped se debilita (Kurtboke 2012).

Los virus tienen dos fases en su ciclo de vida: fuera de las células y dentro de las células que infectan. Las partículas virales fuera de las células pueden sobrevivir durante un largo tiempo aún en condiciones adversas como entidades inertes llamadas viriones. Fuera de las células no pueden reproducirse porque carecen de la maquinaria necesaria para replicar su propio genoma y sintetizar sus proteínas. Los virus infectan las células del huésped, reconociendo receptores específicos en su superficie. Los receptores virales son moléculas superficiales de las células hospedadoras involucradas en funciones celulares rutinarias. El virus es capaz de unirse a este dado que una porción de un complejo molecular en la superficie viral (típicamente espículas) tiene una forma complementaria a la forma de la parte soluble externa del receptor. Luego de producirse esta unión, el virus debe encontrar la manera de ingresar a la célula. Tanto los virus con envoltura como los sin envoltura usan proteínas presentes en sus superficies para unirse e ingresar a la célula hospedadora mediante endocitosis (Lopez & Arias 2010). Las vesículas endocíticas transportan las partículas virales al área perinuclear de la célula huésped, donde las condiciones para la replicación viral son óptimas. Existe otra forma de infección que no involucra el ingreso total de la partícula viral, sino que el virus inyecta directamente su genoma directamente al citoplasma de las células (a veces acompañado por proteínas específicas), quedando la cápside en el exterior.

Los virus que infectan bacterias son llamados bacteriófagos, existen en todos los entornos naturales y están directamente relacionados con la cantidad de bacterias presentes. Como consecuencia, representan la entidad biológica más abundante en la Tierra, con un estimado de 10^{32} bacteriófagos en el planeta (Wommack & Colwell 2000). La secuenciación de genomas bacterianos ha revelado que los elementos genéticos fágicos representan una importante fuente de diversidad de secuencias y pueden influir potencialmente en la patogenicidad y la evolución de las bacterias (Kurtboke 2012).

Los bacteriófagos fueron descubiertos hace más de un siglo. En 1896, Ernest Hanbury Hankin, un bacteriólogo británico (1865 -1939), informó que “algo” en las aguas de los ríos en la India tenía propiedades antibacterianas inesperadas contra el cólera, y que podía pasar a través de un filtro de porcelana muy fina y mantener esta característica distintiva.

En 1915, el bacteriólogo británico Frederick Twort (1877-1950), superintendente del *Brown Institution of London*, descubrió un pequeño agente que mataba colonias de bacterias en cultivos en crecimiento. Publicó los resultados, pero su trabajo posterior fue interrumpido por el comienzo de la Primera Guerra Mundial y la escasez de fondos. Felix d'Herelle (1873-1949) descubrió de forma independiente en el Instituto Pasteur de Francia en 1917 un agente que mataba bacterias. Observó que cultivos bacterianos de disentería desaparecían con la adición de un filtrado libre de bacterias obtenido de las aguas residuales. También fue d'Herelle quien identificó que estos agentes eran microbianos y no químicos, a pesar de que no fue hasta 1940 qué, con el desarrollo de la microscopía electrónica, se reconoció el carácter viral de estos (Kurtboke 2012). Se dio cuenta de que estos organismos se propagaban a expensas de las bacterias, por lo que los denominó bacteriófagos, " que procede del griego "*phagein*" que significa literalmente comedores de bacterias (Martha & Clokie 2008). También es responsable de reconocer su potencial importancia clínica.

El *International Comitee on Taxonomy of Viruses* (ICTV) actualmente reconoce la existencia de un orden, 13 familias y 31 géneros de fagos. Los viriones pueden tener simetría binaria, cúbica o helicoidal, o ser pleomórficos. La mayoría de los fagos poseen un genoma de ADN de doble hebra (dsADN), pero existen grupos de bacteriófagos con genomas de ADN de cadena simple, ARN doble hebra y ARN de cadena simple. Los fagos con cola y simetría bilateral representan aproximadamente el 96% de los fagos, y se clasifican en el orden *Caudovirales*, que a su vez se subdivide en tres grandes familias: *Siphoviridae*, *Myoviridae* y *Podoviridae* (Figura 1.3). Por el contrario, los fagos cúbicos, filamentosos y pleomórficos comprenden menos de 190 virus solamente (3,6% de los fagos) y se clasifican en 10 familias pequeñas (Calendar & Abedon 2005).

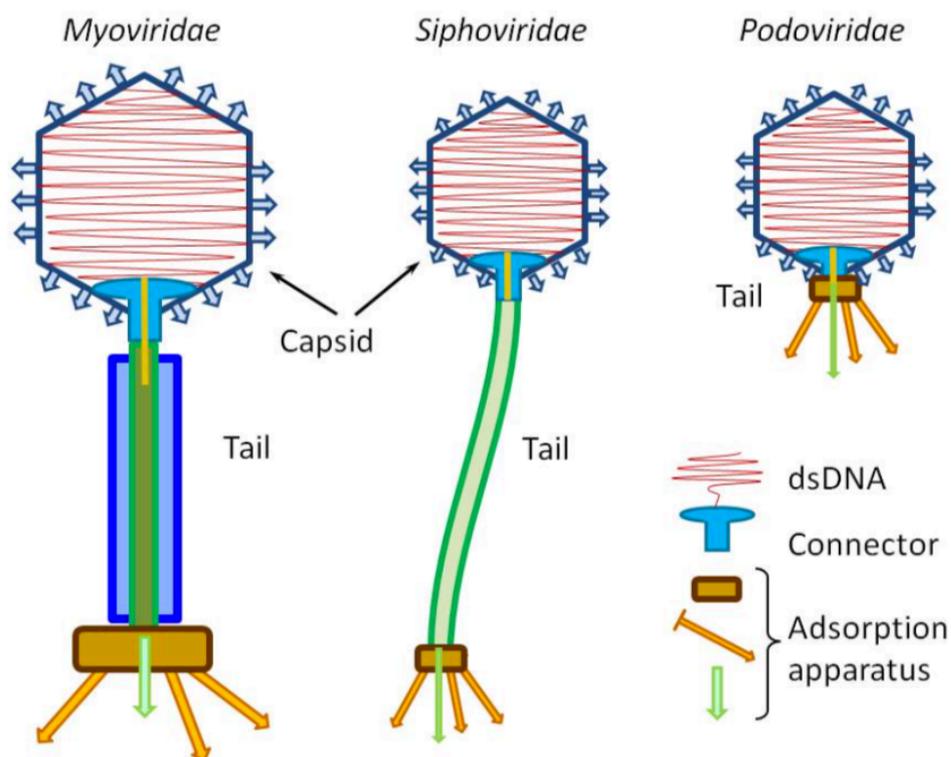


Figura 1.3. Morfología de los viriones de las tres familias que componen el orden de los *Caudovirales* (Kurtboke 2012)

Caudovirales

Los elementos estructurales básicos de estos bacteriófagos son la cápside que protegen el genoma oculto dentro de la misma, y estructuras adicionales que proporcionan la interfaz necesaria con la superficie bacteriana para la liberación del genoma en su interior. Los *Caudovirales* se caracterizan por poseer genomas dsADN y por una organización general común de las partículas virales compuesta por una cápside y una cola. Las diferentes especies de fagos pueden variar tanto en tamaño de longitud como en longitud del genoma. Son probablemente el grupo evolutivamente más dinámico de la biosfera. Estudios de genómica comparativa de varios genomas secuenciados han revelado pistas sobre el mecanismo de evolución de los fagos y han aportado evidencias sobre la importancia de fenómenos de recombinación no homóloga o ilegítima entre diferentes genomas como fuente de generación de diversidad, y eventos de recombinación homóloga como mecanismo de reordenamiento de secuencias (Achigar 2014). Todos los genomas de *Caudovirales* secuenciados hasta el momento poseen una naturaleza modular. Un genoma con estructura en mosaico se puede definir como una secuencia formada por varios fragmentos o módulos genéticos definidos, que pueden estar compuestos por un gen o

grupo de genes, capaces de moverse libremente en la población (Hendrix 2003). Esto genera que la sumatoria de todos los módulos en la población conformen un gran pool de genes que pueden tener muchísimas combinaciones para formar un genoma funcional. Una alta frecuencia en eventos de transferencia horizontal de genes es la causa principal de esta estructura en mosaico. La recombinación ilegítima es un mecanismo de intercambio de material genético entre secuencias que no comparten similitud. Este mecanismo aporta una gran variabilidad, ya que mediante estos nuevos reordenamientos los fagos son capaces de adquirir nuevas características a partir de material genético sin ninguna similitud con el propio (Hatfull 2010). Los eventos de recombinación (ilegítima u homóloga) entre fagos virulentos pueden ocurrir entre fagos que infecten simultáneamente al mismo huésped o con elementos genéticos propios de la bacteria. Esto también se cumple para fagos temperados, los cuales pueden recombinar también con otros profagos presentes en el genoma bacteriano (Achigar 2014).

El ciclo de vida de un bacteriófago puede dividirse en tres etapas: adsorción, infección y liberación (Calendar & Abedon 2005). Cada virión tiene una cápside poliédrica, predominantemente icosaédrica, que protege al genoma. La cápside se compone de muchas copias de una o varias proteínas diferentes y tiene una organización muy estable. La cola y sus estructuras relacionadas son elementos indispensables que aseguran la entrada del ácido nucleico viral en la bacteria huésped durante el proceso de infección. La cola sirve como un transmisor de señal y, posteriormente, como una “tubería” a través de la cual se inyecta el ADN en la célula huésped durante la infección. Las colas pueden ser cortas o largas; estas últimas se dividen en colas contráctiles y no contráctiles. Las colas largas se componen típicamente de muchas copias de varias proteínas dispuestas con simetría helicoidal. Todos los tipos de colas tienen apéndices externos unidos al extremo distante de la cola y a menudo incluyen una placa base con varias fibras y una punta o aguja, que reconocen específicamente los receptores de membrana de la bacteria (Leiman et al. 2010). Una vez que el receptor ha sido encontrado por esta aguja de la cola, (proceso que ocurre mediante múltiples eventos de unión reversible) la placa de base y las fibras de la cola se involucran en el proceso haciendo que el bacteriófago se una de forma irreversible (São-José et al. 2006). Este acoplamiento (unión irreversible) del fago induce la liberación del genoma a través del tubo de cola hacia la célula bacteriana (Kurtboke 2012).

La infección puede dividirse en un período eclipse el cual puede extenderse durante un período variable (por ejemplo durante el ciclo lisogénico, este período puede prolongarse

indefinidamente) y un período de maduración de la progenie (Calendar & Abedon 2005). Luego de inyectado el genoma viral dentro del hospedero comienza el proceso de expresión de las proteínas virales y replicación, con la consecuente detención de la actividad metabólica de la bacteria y degradación del genoma hospedero. La expresión de las proteínas se da en orden, comenzando por los genes tempranos y seguidas por proteínas estructurales y aquellas involucradas en el ensamblaje de los viriones y liberación de la progenie (genes tardíos) (Zinno, 2008). La replicación del genoma comienza luego de que los genes tempranos se han expresado y se da mediante la producción de una única molécula de ADN doble hebra que contiene múltiples copias del genoma viral, el cual luego es procesado para dar múltiples copias individuales. En el caso de los virus lisogénicos, luego de inyectado el genoma viral al interior celular se expresan una serie de proteínas tempranas que inducen la integración del genoma del fago en el cromosoma bacteriano.

El paso siguiente a la replicación del genoma y expresión de las proteínas estructurales, es el empaquetamiento del ADN. Éste está mediado por enzimas llamadas terminasas y existen dos mecanismos diferentes de procesamiento del ADN con múltiples copias del genoma fágico (concatémero) que permite dividir a los bacteriófagos en dos grupos: los fagos tipo *cos* y los fagos tipo *Pac* (Achigar 2014).

Los fagos tipo *cos* poseen secuencias específicas que flanquean cada copia del genoma y son reconocibles por una enzima llamada terminasa. La terminasa reconoce una secuencia *cos* y se une al ADN cortándolo y generando un extremo cohesivo por donde comenzará el empaquetamiento. Luego reconoce la siguiente secuencia *cos* en el concatémero y realiza la misma acción, por lo que una copia completa del genoma viral comprendida entre esos sitios *cos* se empaqueta en cada ciclo (Hernández Magadán 2007).

En el caso de los fagos tipo *Pac*, la terminasa comienza el empaquetamiento reconociendo una secuencia específica llamada sitio *Pac*. Luego, ésta comienza a introducir el concatémero al interior de la procápside hasta que se encuentra llena, produciéndose un corte inespecífico en la secuencia de ADN. Este corte genera un extremo cohesivo por el que comienza la siguiente ronda de empaquetamiento. El ADN ingresa a la procápside hasta llenarse, produciéndose un nuevo corte inespecífico y así sucesivamente. La capacidad de almacenamiento de la cápside es levemente mayor que el tamaño real del genoma del fago, por lo que se generan extremos repetidos en el genoma empaquetado. A este fenómeno se lo conoce como redundancia terminal. Otra consecuencia de este mecanismo de

empaquetamiento es que diferentes viriones de la misma progenie posean secuencias diferentes en los extremos de sus genomas, fenómeno conocido como permutación circular (Hernández Magadán 2007).

Por último, luego que el ADN ha sido empaquetado y los viriones ensamblados, ocurre la liberación. Ésta puede ocurrir por varios mecanismos dependiendo del tipo de fago, incluyendo lisis, extrusión o gemación. La lisis por lo general depende de enzimas codificadas por el propio fago (lisinas y holinas) las cuales se encargarán de degradar tanto la membrana como la pared celular, liberando los viriones al ambiente (Calendar & Abedon 2005).

Todos los fagos de *S. thermophilus* son miembros de la familia *Siphoviridae*, orden *Caudovirales*, caracterizados por cabezas isométricas y colas largas no contráctiles. Las cápsides de los fagos son generalmente de alrededor de 60nm de diámetro, y sus colas oscilan entre 220 y 330 nm. (Quiberoni et al. 2010). Los fagos de *S. thermophilus* poseen algunas características particulares (resistencia térmica, velocidad de replicación) que restringen el desarrollo de estrategias para reducir las infecciones en ambientes industriales . Fagos con estas características pueden causar el fracaso de cultivos iniciadores, incluso si están presentes en bajo número al principio de la fermentación. Además, los fagos pueden permanecer en el aire por largos periodos, Todos estos factores contribuyen al problema de la dispersión de fagos dentro del ambiente industrial (Quiberoni et al. 2010), y son detallados a continuación.

Fagos y las fermentaciones lácticas

Una de las funciones clave de los fagos en la naturaleza, es equilibrar la población bacteriana en todos los entornos compartidos, desafiándolas a evolucionar rápidamente. Los fagos a veces también pueden convertirse en un gran problema para la industria. Un proceso biotecnológico que depende del uso de bacterias para producir un producto puede verse alterado por los fagos. Problemas de este tipo han sido reportados en las industrias de alimentos, productos químicos, farmacéutica, etc. Sin embargo, la industria que más extensamente ha documentado estos fenómenos es la industria láctea (Garneau & Moineau 2011).

La fabricación de queso requiere la inoculación de 10^7 células bacterianas por mL de leche pasteurizada para llevar a cabo una fermentación controlada y obtener productos finales de alta calidad. En el entorno no estéril de la leche cruda o tratada térmicamente, las BAL añadidas entrarán en contacto con los fagos virulentos que se encuentran en la leche. Aunque la concentración inicial de fagos suele ser baja en la leche, una población específica de fagos puede aumentar rápidamente si el cultivo iniciador contiene alguna cepa sensible. La consiguiente lisis de un gran número de células sensibles retrasará o incluso detendrá el proceso de fermentación de la leche, lo que traerá como consecuencia productos de baja calidad y caída de productividad en la planta. En el peor de los casos, la leche inoculada debe descartarse perdiéndose grandes volúmenes de materia prima. Durante décadas, la industria láctica ha tenido que lidiar con esta problemática, dependiendo de una serie de diferentes medidas de control, que van desde el rediseño de fábricas, saneamientos mejorados, cambios en los procesos y rotación de *starters*.

Fuentes de contaminación

Los fagos pueden provenir de diversas fuentes. Es de vital importancia conocer las posibles fuentes de fagos para limitar su entrada dentro de las instalaciones de fabricación, lo que podría ser perjudicial para el proceso de fermentación.

Cualquier materia prima que entre en una instalación de fermentación puede estar contaminada. Por ejemplo, es sabido que la leche cruda, que es un nicho ecológico por excelencia de las BAL, contiene fagos (Kutter & Sulakvelidze 2004). Debido a que la leche se obtiene de diferentes tambos y se almacena mezclada, la diversidad fágica se amplifica en los silos de almacenamiento. Como los fagos se pueden propagar fácilmente en un medio líquido como la leche y, como también pueden difundirse en medios similares a geles, solo se necesitan unas pocas células sensibles para que los títulos virales aumenten rápidamente en un entorno determinado (Müller-Merbach et al. 2007). Ensayos realizados previamente en Uruguay, mostraron que casi un 20% de la leche que entra a planta está contaminada con fagos de *Streptococcus thermophilus*, y casi el 80% de los sueros de fermentaciones problemáticas presentan altos títulos virales (Achigar 2014).

Dependiendo de la frecuencia de los ataques de fagos y el tamaño de las instalaciones, puede ser aconsejable analizar la leche (u otros ingredientes) para detectar la presencia de fagos antes de comenzar el proceso de fermentación para confirmar que la carga de fagos

inicial no representa un riesgo significativo. Varias metodologías han sido puestas a punto para la detección temprana de estos microorganismos; en Uruguay se pusieron a punto metodologías de detección precoz basadas en qPCR que demostraron ser efectivas para identificar ataques de fagos y focos de infección en planta (Achigar 2014).

Si se piensa que los ingredientes representan un riesgo, pueden tratarse para reducir los niveles de fagos o usarse para otros procesos que no se verán afectados por los fagos (elaboración de productos lácteos no fermentados). La industria tiende a reutilizar proteínas y grasa del suero de leche para mejorar el sabor o la textura de un producto final y aumentar sus rendimientos (Garneau & Moineau 2011). En el proceso de generación de concentrados (remoción del agua del suero de leche), puede darse la concentración de partículas virales junto con los componentes secos del suero, generando así un foco de contaminación importante, que puede desencadenar fallas en nuevos procesos. Las alternativas para mitigar esto, es el tratamiento del suero para eliminar los fagos, o utilizar los concentrados en productos diferentes al de origen del suero. Por ejemplo, si el suero de leche se recogió de una fermentación de queso cheddar preparada con cultivos iniciadores mesófilos, los subproductos de este suero de leche podrían utilizarse de forma segura en la fabricación de yogur o en un proceso de queso que requiera cultivos termófilos (Garneau & Moineau 2011). Ensayos llevados a cabo en plantas de Uruguay, mostraron que la práctica de reciclado de materias primas (en particular la grasa) es una de las causantes de fallas graves a nivel industrial (Achigar 2014).

Otro foco de contaminación posible, son los propios cultivos *starters*. Cuando un fago temperado entra en una cepa, puede iniciar el ciclo lítico o integrarse al genoma bacteriano entrando en ciclo lisogénico. Diferentes condiciones a las que están sometidas las bacterias como el calor, las sales, los antimicrobianos, la inanición o los rayos ultravioleta pueden inducir al profago y desencadenar el ciclo lítico (Madera et al. 2009). Por lo tanto, el uso de cepas lisogénicas en un cultivo *starter* puede conducir a la lisis celular durante la fermentación. La inducción también puede ocurrir naturalmente y puede alcanzar una frecuencia de hasta 9% (Garneau & Moineau 2011).

Mientras que la contaminación por fagos de la leche es probablemente la fuente primaria de fagos en planta, las rutas de diseminación de contaminantes pueden ser más difíciles de identificar. Hace algunos años, una serie de investigaciones determinaron el impacto de la presencia de fagos de *Lactococcus* en el aire de las plantas (Verreault et al. 2011; Neve et

al. 2003). Un alto nivel de fagos transportados por el aire en el medio ambiente puede significar que ha ocurrido una propagación reciente o lo que es aún más importante, es un signo de que probablemente estén a punto de desencadenarse fallas de producción (Garneau & Moineau 2011). Estos ensayos mostraron que pueden llegar a haber niveles de hasta 10^3 partículas virales por m^3 de aire y más de 10^3 por cm^2 en las superficies de trabajo. Esto muestra la omnipresencia de los fagos en las plantas de fermentación y confirma las evidencias anecdóticas de que un buen sistema de ventilación en una fábrica de queso es un parámetro crítico para controlar adecuadamente la diseminación del fagos, así como la importancia de las buenas prácticas de fabricación y limpieza, y el entrenamiento del personal para evitar la contaminación de los productos y materias primas.

Estrategias de control

Las estrategias de control para estos problemas son varias y presentan diferentes grados de efectividad. Un buen procedimiento de desinfección es sin duda uno de los factores clave para evitar las contaminaciones microbianas y también es la forma más eficaz de reducir la propagación de fagos dentro de la instalación. Se han probado varios desinfectantes y condiciones en diferentes fagos de BAL. El ácido peracético sería el biocida más eficaz, mientras que el etanol y el isopropanol no suelen serlo. El hipoclorito de sodio tuvo un efecto variable a través de los estudios pero también fue efectivo contra la mayoría de los fagos BAL (Garneau & Moineau 2011).

Como se indicó anteriormente, la leche y los ingredientes derivados de la leche pueden contener fagos virulentos y, por lo tanto, deben tratarse para reducir la carga viral. Siempre que sea posible, se deben usar ingredientes o medios estériles. En la industria láctea, la pasteurización de la leche es la práctica más común utilizada para reducir el crecimiento microbiano y el deterioro del producto.

El calentamiento puede reducir en gran medida la actividad de las partículas del fago ya que provoca la liberación del ADN y cambios en la morfología del fago. Sin embargo, está documentada la existencia de bacteriófagos que no se inactiva a las temperaturas normales de pasteurización (Binetti & Reinheimer 2000). Por ejemplo, los estudios han demostrado que algunos fagos de *Lactococcus* pueden resistir a temperaturas de hasta $97^\circ C$ durante 5 min. (Garneau & Moineau 2011).

La rotación de cepas es probablemente tan antigua como el uso de cultivos iniciadores definidos y sigue siendo la piedra angular de un sistema de control de fagos. Es una práctica especialmente utilizada en plantas de elaboración de queso, como forma de evitar la amplificación recurrente del mismo fago en procesos de fermentación consecutivos (Garneau & Moineau 2011). Aún así, la rotación de cepas requiere un seguimiento riguroso para detectar la aparición de nuevos fagos virulentos.. El rango de hospedadores puede identificar las cepas de BAL sensibles al mayor número de fagos en cultivos *starters* definidos y quizás, eventualmente, llevar a su reemplazo por cepas no relacionadas. El problema principal con este abordaje es que los métodos para identificar si las fallas son provocadas efectivamente por fagos, cuales cepas son sensibles de las disponibles en planta, etc., no se ajustan a los tiempos de producción de la planta, además de que rara vez hay personal entrenado o la infraestructura para llevar a cabo los ensayos. Esto resulta en que, en la práctica, la rotación es llevada a cabo a ciegas y no siempre resulta efectiva, ya que pueden introducirse nuevas cepas sensibles al esquema de producción, agravando los problemas.

Una aproximación más efectiva, es la utilización de cepas que sean naturalmente resistentes. Cuando se aísla un nuevo fago, las empresas proveedoras pueden intervenir reemplazando las cepas sensibles por otras que no lo sean, y reintroducir el producto en el mercado. La escasa disponibilidad de cepas para sustituir en las rotaciones, así como el acceso a colecciones de fagos hacen que esta estrategia sea poco sostenible en el tiempo. Aún así, es la estrategia que ha mostrado ser más exitosa hasta el momento. Esta estrategia de diseño se apoya principalmente en la existencia de diferentes sistemas de defensa naturales, como lo son el sistema restricción-modificación (R-M) y el sistema CRISPR-Cas.

Mecanismos de defensa

Todos los organismos vivos sin importar al dominio al que pertenezcan se encuentran en constant conflict con virus y otros patógenos. Como parte de la perpetua carrera armamentista entre patógenos y hospederos, dos enfoques estratégicos de inmunidad han surgido: la inmunidad innata y la inmunidad adaptativa. Mientras que la inmunidad innata implica sistemas preexistentes codificados genéticamente que reconocen características generales y comunes en patógenos, la inmunidad adaptativa involucra sistemas que son capaces de generar respuestas inmunes específicas y memoria inmunológica contra invasores previamente no detectados.

Los procariontes no son la excepción. Las presiones selectivas impuestas por los bacteriófagos tienen un profundo impacto en la composición y el comportamiento de las comunidades microbianas en cada entorno ecológico (Rodríguez-Valera et al. 2009), y los hospedadores microbianos han desarrollado varios mecanismos para evadir la infección (Sorek et al. 2013).

Inmunidad innata: sistemas de restricción-modificación (R-M)

Los sistemas de restricción-modificación (R-M) son los efectores de la inmunidad innata procarionte por excelencia. Los genes que codifican sistemas de restricción-modificación están presentes en aproximadamente el 90% de los genomas bacteriano secuenciados (Roberts et al. 2015). Estos sistemas pueden estar codificados en plásmidos o en el cromosoma, y su función general es reconocer y atacar el ADN extraño invasor con enzimas de restricción, mientras que simultáneamente protegen el ADN del huésped mediante la actividad de enzimas metiltransferasa (MTasas) (Murphy et al. 2013).

Actualmente se reconocen cuatro tipos de sistemas R-M (I, II, III y IV). Los sistemas R-M tipo II son el tipo más prevalente y se componen por dos proteínas con funciones definidas: una endonucleasa de restricción (REasa) que corta el ADN invasor en posiciones definidas dentro o cerca de su sitio de reconocimiento; y una metiltransferasa (MTasa) que protege al ADN propio mediante la adición de un grupo metilo en el sitio de reconocimiento (Murphy et al. 2013).

Históricamente, la apreciación del sistema inmune microbiano se restringía a mecanismos de defensa innata, pero este paradigma cambió hace algunos años, con el descubrimiento de un sistema inmune adaptativo procarionta (Horvath & Barrangou 2010).

Inmunidad adaptativa: el sistema CRISPR-Cas

Los *loci* CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) se observaron por primera vez en bacterias a fines de la década de 1980 y, a mediados de la década de 1990, en los genomas de varios linajes de archaeas. Su aparente amplia distribución sugería que podrían jugar un papel celular común y fundamental en ambos dominios, aunque su prevalencia en organismos extremófilos sugería que de alguna manera podrían facilitar la adaptación a ambientes extremos (Mojica et al. 2000). La estructura repetido-espaciador (tándems de secuencias cortas palindrómicas iguales llamadas repetidos, separadas entre sí por secuencias variables de igual longitud llamadas espaciadores) era distinta en arquitectura y secuencia de otros sistemas funcionales conocidos y, en consecuencia, no había precedentes para inferir su función biológica. La primera evidencia que documenta su funcionalidad surgió de estudios en las *haloarcheas* *Haloferax mediterranei* and *Haloferax volcanii*, donde se detectaron transcripciones de estos *loci* en cepas de tipo salvaje y se observaron defectos de crecimiento en células transformadas con plásmidos artificiales con arreglos de tipo espaciador-repetido (Mojica et al. 1995). Un descubrimiento importante fue la identificación de genes asociados a CRISPR, llamados *cas* (*CRISPR-associated*). Aunque análisis bioinformáticos iniciales de las secuencias de aminoácidos de estas proteínas Cas mostró que jugaban un papel en el metabolismo de ácidos nucleicos (Jansen et al. 2002), los primeros indicios del verdadero funcionamiento de este sistema CRISPR-Cas surgieron a partir del descubrimiento del origen de los espaciadores. Estudios detallados revelaron que existía una incompatibilidad entre secuencias espaciadoras y elementos genéticos extraños que contuvieran secuencias similares; es decir, que por ejemplo un plásmido que contuviera una secuencia homóloga a un espaciador en la bacteria huésped, era eliminado de la misma. A estas secuencias homólogas se las llamó protoespaciadores y su descubrimiento llevó a elaborar la teoría de que el sistema CRISPR-Cas es capaz de generar inmunidad contra ADN extraño de forma adaptativa (Bolotin et al. 2005; Pourcel et al. 2005). La evidencia experimental que apoya esta hipótesis surgió del trabajo de Barrangou y colaboradores en interacciones fago-hospedero en *Streptococcus thermophilus* (Barrangou

et al. 2007). Por tanto se puede definir al sistema CRISPR-Cas como un sistema inmune adaptativo y heredable procariota.

Funcionamiento

Si tuviéramos que resumirlo en una definición mínima, el sistema CRISPR-Cas se compone de un repertorio de pequeños ARNs que reconocen por homología secuencias de ADN invasor, y guían un complejo enzimático para degradarlo. El mecanismo por el cual se desencadena esta respuesta inmune no se encuentra esclarecido totalmente, sin embargo varios estudios afirman que algunas de las proteínas Cas son expresadas de forma constitutiva, mientras que otras son reguladas por la presencia de cAMP como respuesta a la invasión del hospedero por parte de un ADN extraño (Sorek et al. 2013). Las bacterias y las arqueas adquieren resistencia a elementos genéticos extraños como plásmidos y bacteriófagos integrando selectivamente fragmentos cortos del ácido nucleico extraño (que como se mencionó anteriormente se llaman espaciadores) en su propio cromosoma, los cuales luego son procesados en los pequeños ARNs mencionados anteriormente. La inmunidad adaptativa mediada por CRISPR se divide en tres etapas distintas: adquisición de ADN extraño, biogénesis de ARN CRISPR (crARN), e interferencia. Aunque estas tres etapas básicas parecen ser comunes a todos los sistemas CRISPR, los *loci* CRISPR y las proteínas que median en cada etapa de la inmunidad adaptativa son notablemente diversas (Figura 1.4) (Sorek et al. 2013).

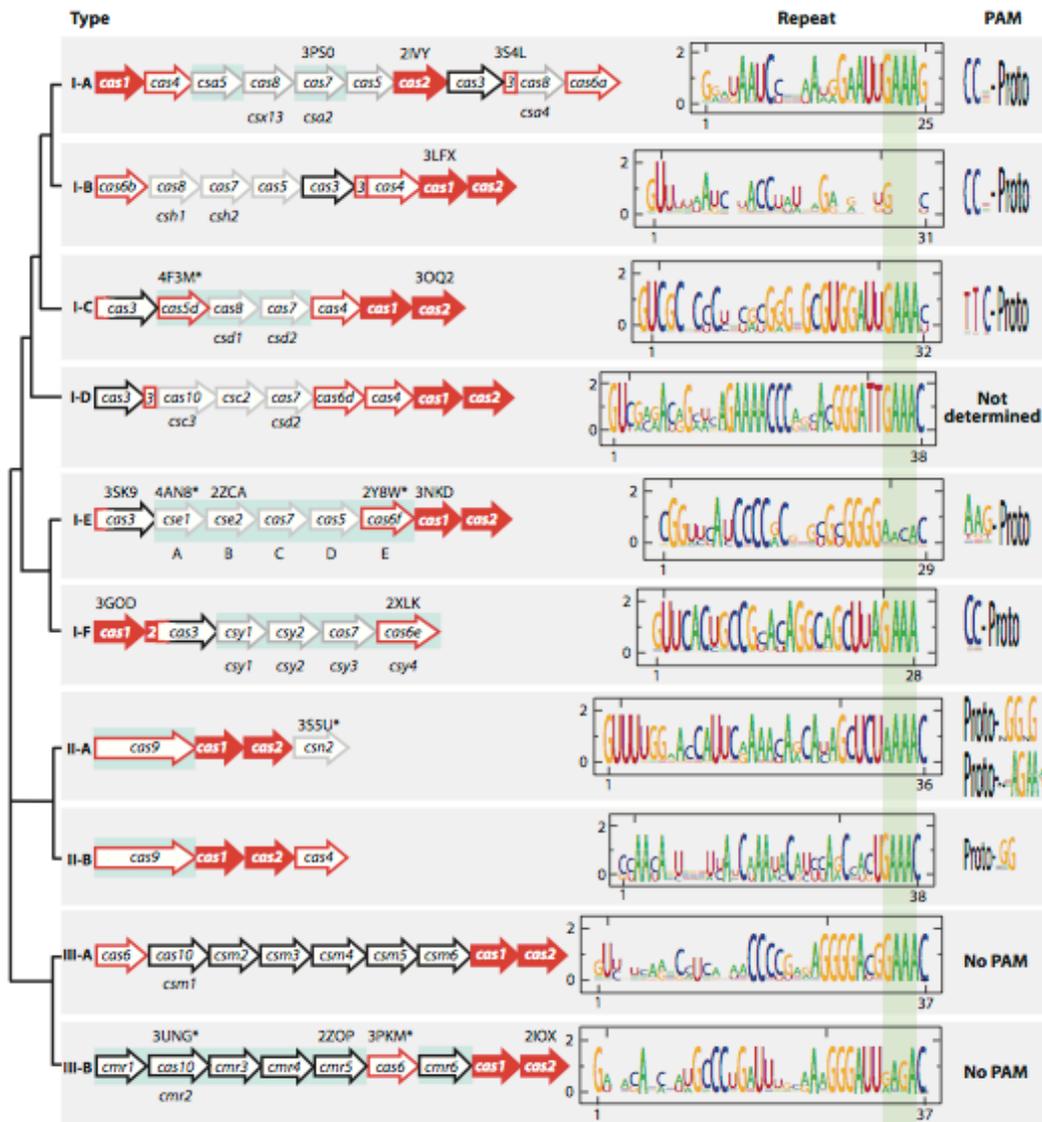


Figura 1.4. Estructura, espaciadores y PAMs de los diferentes tipos de sistema CRISPR (Sorek et al. 2013)

En la etapa de adaptación, las proteínas Cas identifican el ADN objetivo y adquieren una nueva secuencia espaciadora derivada de éste. Es esta colección de secuencias espaciadoras que se integran en el *locus* CRISPR las que forman la memoria inmunológica. La adaptación es un proceso complejo y de múltiples etapas en el que se necesita “extraer” un protoespaciador de un ADN extraño invasor y, posteriormente, almacenarlo en el *loci* CRISPR como espaciador (Figura 1.5). En primer lugar, el ADN extraño debe ser reconocido como un objetivo para la adquisición de espaciadores; segundo, una secuencia de un tamaño específico (típicamente 30-40 pb, dependiendo del subtipo del sistema) debe ser adquirido desde el ADN extraño; y finalmente, la secuencia adquirida debe estar integrada al *locus* CRISPR como un nuevo espaciador. Aunque los componentes de la maquinaria y los requisitos previos para la adquisición de un espaciador varían entre organismos y subtipos del sistema, varios de estos están conservados universalmente y son esenciales

entre todos los subtipos CRISPR-Cas. Estos componentes son las proteínas Cas1 y Cas2 y, dentro del locus CRISPR, la secuencia líder y el primer repetido (Amitai & Sorek 2016). Al ingresar un ADN invasor, el complejo formado por Cas1 y Cas2 lo reconoce y escinde un fragmento del mismo que se encuentre adyacente a un motivo conocido como PAM (*Protospacer Adjacent Motif*). La secuencia del PAM varía entre organismos y tipos de sistema, pero es esencial tanto para esta etapa como para la etapa efectora de la inmunidad. El fragmento de ADN es luego incorporado al locus en una reacción también catalizada por Cas1 y Cas2 en la primera posición del *locus* contando desde el extremo 5'. Esta integración está guiada por la secuencia líder, una región promotora ubicada inmediatamente en el extremo 5' del primer repetido del *locus*.

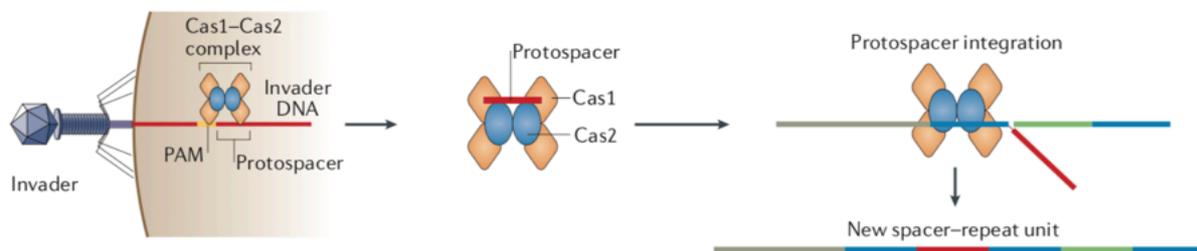


Figura 1.5. Esquema del proceso de adquisición de nuevos espaciadores mediado por Cas1 y Cas2 (Amitai & Sorek 2016)

La biogénesis de los crARNs es un proceso fundamental para la activación del sistema. En esta etapa, el *locus* completo es transcrito y procesado para formar fragmentos de ARN cortos. El mecanismo en que se procesan estos ARNs y la estructura final de los mismos depende de cada tipo de sistema. En la figura 1.6 se muestran ejemplos de biogénesis para algunos tipos de sistema CRISPR.

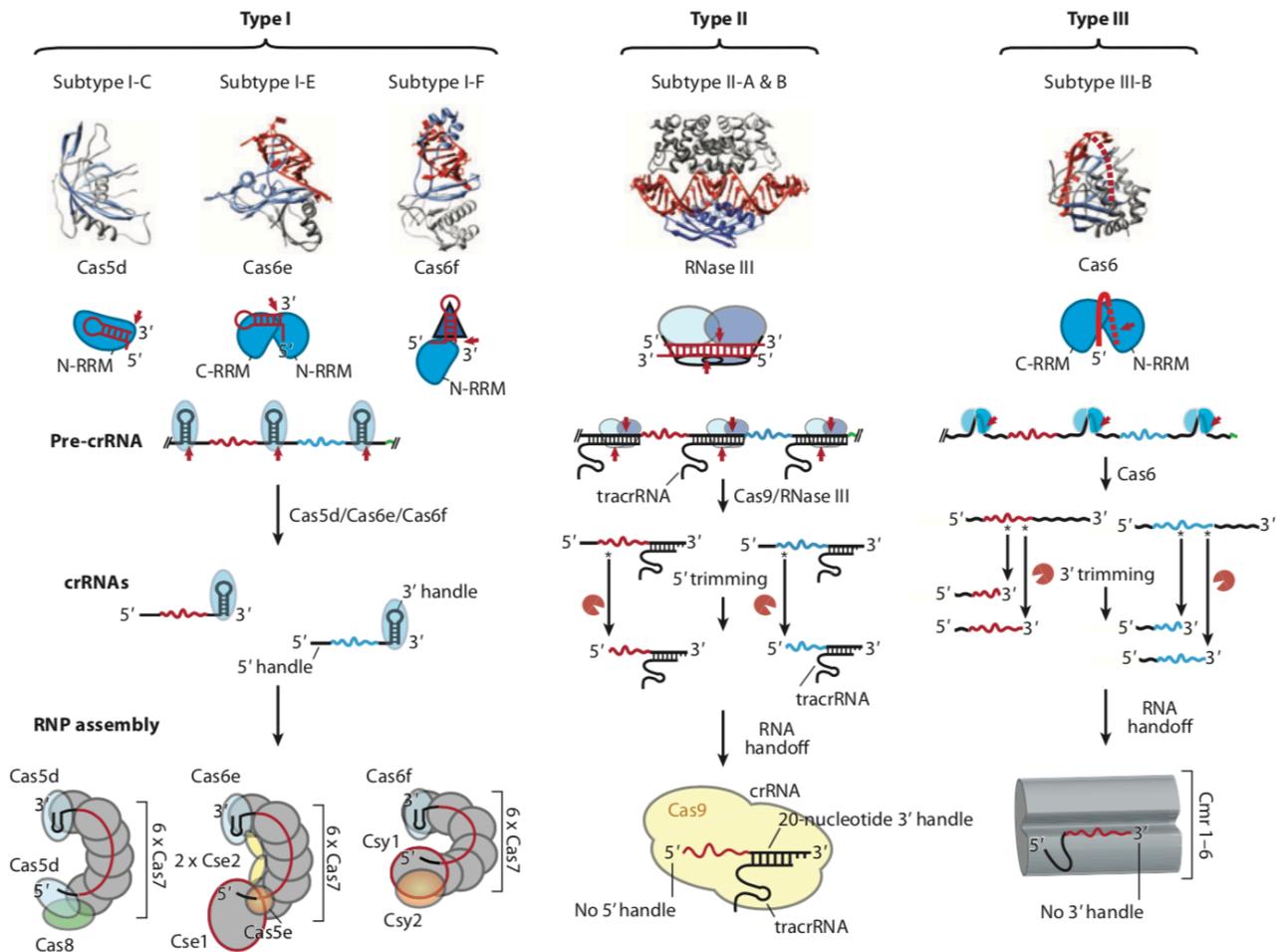


Figura 1.6. Diferentes tipos y mecanismos de síntesis de crARNs (Sorek et al. 2013)

Por último, en la etapa de interferencia, los crARNs guían a las proteínas Cas efectoras para reconocer por homología al ADN invasor y degradarlo. El mecanismo efector así como la maquinaria encargada del mismo es, al igual que la biogénesis, dependiente del tipo de sistema. Estos aspectos se abordan a continuación únicamente para el subtipo IIA ya que es el tipo de sistema funcional presente en *Streptococcus thermophilus*.

Sistema CRISPR-Cas de tipo IIA

El sistema CRISPR-Cas tipo II es el más simple en términos de organización y número de genes asociados. Este sistema consta de cuatro proteínas Cas: Cas9 que es la proteína *hallmark* del sistema, involucrada tanto en los procesos de maduración de los crARN como en la etapa de interferencia; Cas1 y Cas2, cuyas funciones fueron descritas anteriormente y; Csn2 en el caso del sub tipo IIA o Cas4 en el caso del sub tipo IIB. El rol de éstas últimas no ha sido esclarecido, sin embargo se ha observado que mutaciones en el gen *csn2* en *Streptococcus thermophilus* traen como consecuencia la imposibilidad de incorporar

nuevos espaciadores (Deveau et al. 2008). En *S. thermophilus* se conocen cuatro *loci* CRISPR, de los cuales dos se ha demostrado están activos; estos son los *loci* CRISPR1 y CRISPR3 y ambos pertenecen al subtipo IIA. La biogenética de los crARN en los sistemas de tipo II es única ya que requiere un crARN trans-activador llamado tracrARN. En el caso de *S. thermophilus*, este tracrARN tiene una longitud de 36 pares de bases; presenta en su extremo 3' una región de 18 bases que hibrida por homología a los repetidos del *locus*; y se encuentra codificado entre los genes *cas9* y *cas1*. La unión entre el tracrARN y la secuencia del repetido forma un ARN doble hebra, el cual es reconocido por una ARNasa bacteriana que ayuda a procesarlo. Finalmente el crARN maduro tiene una estructura compleja, con una región ssARN que corresponde a la secuencia espaciadora propiamente dicha, una región dsARN formada por la unión de la secuencia del repetido al tracrARN, y un *loop* formado por la región del tracrARN que no hibrida con el repetido. En la maduración también interviene Cas9, aunque su función no está esclarecida. Sin embargo, su papel en la etapa de interferencia ha sido ampliamente estudiado y documentado. El complejo ribonucleo proteico formado por Cas9 y el crARN reconocen al ADN invasor y ésta lo degrada. Para que ocurra el reconocimiento deben darse varias condiciones, siendo la primera que exista un PAM sin alteraciones. Como se mencionó anteriormente, este pequeño motivo es esencial para la adquisición de nuevos espaciadores catalizada por Cas1-Cas2, pero también es indispensable para que Cas9 pueda llevar a cabo su función. La segunda condición es que ocurra un reconocimiento entre la secuencia del crARN y el ADN invasor. La formación de un híbrido ARN-ADN dada por el reconocimiento entre espaciador y protoespaciador, es lo que brinda especificidad al sistema. Una vez que se ha dado esta hibridación, Cas9 corta ambas hebras del ADN invasor a una distancia aproximada de tres bases del PAM, gracias a que posee dos dominios catalíticos HNH y RuvC. Aunque el paradigma central establece que debe existir un reconocimiento completo entre espaciador y protoespaciador, algunos autores afirman que el reconocimiento se debe dar únicamente en una región denominada semilla de no más de ocho pares de bases (Semenova et al. 2011). La figura 1.7 muestra un esquema general de la estructura y el funcionamiento de este tipo de sistema

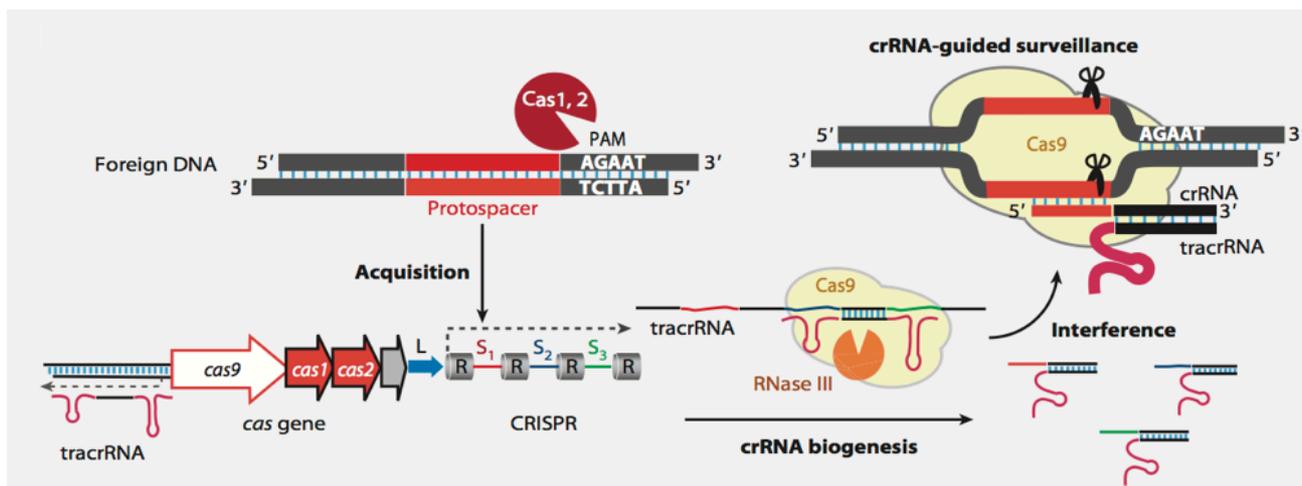


Figura 1.7. Estructura y esquema de funcionamiento del sistema CRISPR tipo IIA (Sorek et al. 2013)

Industria láctea nacional

La industria láctea es una de las de mayor facturación en Uruguay. Según datos del INALE, en el año 2017, los ingresos por concepto de exportación fueron de 591 millones de dólares, un 4% superior al año 2016. Este crecimiento se ha dado principalmente por un aumento en el valor de la tonelada de todos los productos lácteos exportados, lo que demuestra una creciente apreciación internacional por los lácteos uruguayos. La industria no se encuentra en su mejor momento; durante la última década se observó un crecimiento constante con un pico de facturación y exportaciones de queso en 2014, seguida por una rápida caída en el año 2015 (un 30% menos de volumen exportado sumado a una caída del 17% en el valor de la tonelada, ver figuras 1.8 y 1.9). Esto tiene su explicación en la crisis económica en Venezuela, país que desde el año 2008 hasta el 2014 fue destino de prácticamente el 60% de las exportaciones del sector, recibiendo en 2015 únicamente el 24% (Figura 1.10). Traducido en toneladas, la salida abrupta de Venezuela del mapa exportador uruguayo significó una diferencia negativa entre 2014 y 2015 de 14.000 toneladas de queso vendido. Sin embargo la industria ha ido recuperándose se encuentra nuevamente en crecimiento, gracias a mejores posicionamientos en el mercado brasileño y mexicano, pero por sobre todo por la mejora en la calidad de los productos que hacen que el valor por tonelada aumente (datos disponibles en www.inale.org).

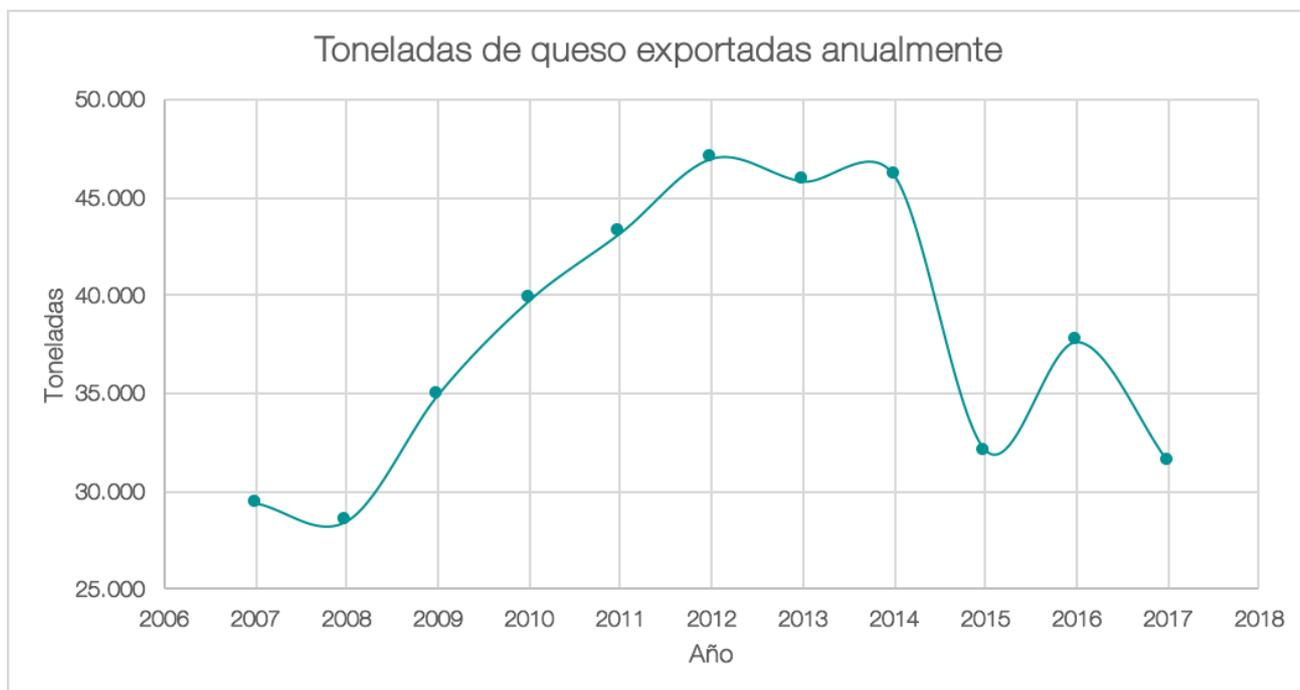


Figura 1.8. Evolución de las toneladas de queso exportadas por Uruguay (www.inale.org)

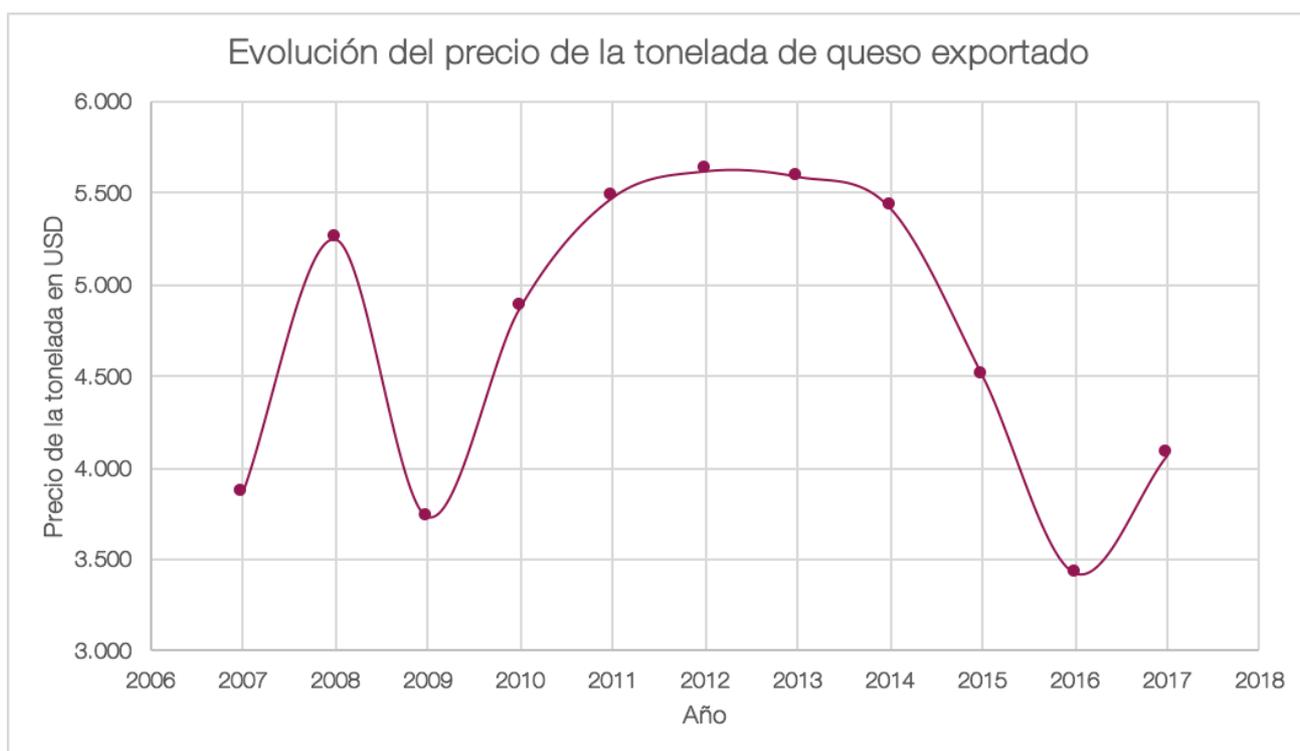


Figura 1.9. Variación anual del precio de la tonelada de queso exportado por Uruguay (www.inale.org)

La industria láctea uruguaya (como la mayoría de los sectores productivos nacionales), encuentra problemas para competir en volumen con sus rivales directos por mercados como son Argentina y Oceanía, por tanto debe apuntar a mejorar la eficiencia de sus procesos y aumentar la calidad de sus productos. Otro déficit que posee la industria es el acceso a insumos. En Uruguay no existen productores de cultivos *starters* y otros insumos

indispensables para la elaboración de quesos, los cuales deben ser importados, impactando negativamente en la balanza comercial de la industria. Es por esto que se hace imperativa la investigación y desarrollo en el área, teniendo como objetivos la generación de soluciones tecnológicas propias que permitan proveer a la industria nacional de insumos que la ayuden a mejorar sus procesos, generen un valor agregado en si mismas y reduzcan la necesidad de importación, por consiguiente, ayudando a inclinar favorablemente la balanza comercial del país.

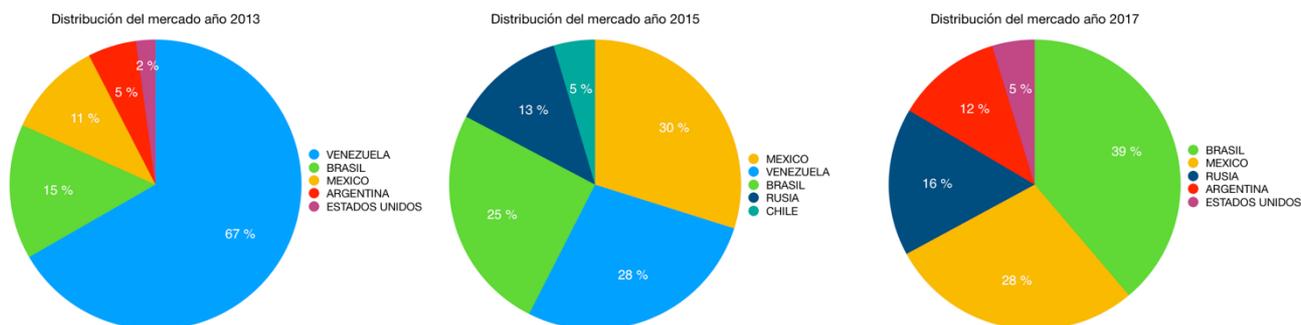


Figura 1.10. Distribución del mercado exportador años 2013, 2015 y 2017 (www.inale.org)

El diseño *starters* presenta una oportunidad de desarrollo tecnológico especialmente para una industria como la uruguaya que debe competir en calidad para ganar mercados de exportación, y que además es totalmente dependiente de la importación para abastecerse de insumos; insumos que no necesariamente han sido desarrollados según las necesidades puntuales de la industria local. En este trabajo nos proponemos el desarrollo de una metodología de obtención de cepas de *Streptococcus thermophilus* mejoradas para su uso en la industria láctea nacional, con especial énfasis en buscar soluciones a la problemática de las infecciones fágicas en planta, que como se mencionó en este capítulo constituyen el principal problema de productividad.

Objetivos

El objetivo general de este trabajo es contar con una estrategia de diseño de cultivos starters comerciales, que permita seleccionar cepas con buenos atributos tecnológicos y capaces de resistir al ataque de bacteriófagos.

Objetivos específicos

- Caracterizar genéticamente cepas de *Streptococcus thermophilus*, y estudiar atributos asociados con productividad.
- Caracterizar el banco de fagos mediante secuenciación completa y análisis de los genomas virales
- Buscar la presencia de sistemas CRISPR-Cas nativos y funcionales en las cepas de la colección
- Estudiar la presencia de sistemas de tipo R-M en las cepas de la colección
- Diseñar genéticamente derivados resistentes mediante el empleo del sistema CRISPR-Cas, evaluando diferentes variables en lo referente a los espaciadores, a fin de obtener las mejores resistentes posibles

Antecedentes y estrategia

Nuestro grupo comenzó a trabajar en la temática a fines del año 2010 en asociación con diferentes empresas nacionales y extranjeras. Desde entonces se ha monitoreado la situación del ambiente industrial uruguayo en lo referente al problema de fagos, desarrollando y optimizando técnicas de detección precoz. Se ha generado la primera colección de estos virus del país a partir del análisis de cientos de muestras y cientos de aislamientos, reduciéndose la diversidad en una primera instancia a 15 tipos diferentes de fagos (Achigar 2014).

En este trabajo proponemos desarrollar una estrategia innovadora de diseño de starters de *Streptococcus thermophilus*, dirigida a la obtención de los cultivos adaptados a la industria nacional, capaces de mantener la resistencia a bacteriófagos a lo largo del tiempo y por ende mejorando substancialmente los procesos de elaboración y calidad del producto final. Para ello, a partir del banco de cepas comerciales que poseemos, realizamos una primera etapa de evaluación de atributos tecnológicos y caracterización molecular en busca de marcadores asociados a alta productividad. La siguiente etapa consistió en la búsqueda de sistemas de defensa tipo restricción-modificación y evaluación del contenido de loci CRISPR. Posteriormente, se llevó a cabo el análisis del banco de fagos con especial énfasis en el contenido de protoespaciadores, que son el insumo fundamental para el desarrollo de derivados resistentes basados en el sistema CRISPR-Cas. Luego de estudiar cuales eran los eventos de adquisición óptimos, se evaluó el comportamiento natural de las cepas frente a fagos y la dinámica de adaptación. Por último, con la información generada en este trabajo, se procedió al diseño de un locus CRISPR artificial que condensara las características ideales para otorgar resistencia a largo plazo a las cepas de *Streptococcus thermophilus*.

El capítulo 2 del manuscrito aborda la caracterización de las cepas de *S. thermophilus* a nivel fenotípico y molecular, así como la evaluación de los sistemas de defensa. El capítulo 3 se centra en el análisis del banco de fagos. El capítulo 4, aborda la generación de mutantes fago resistentes y la evaluación de la calidad de la resistencia. Por último, en el capítulo 5 se presentan los avances obtenidos en la generación de cepas de *S. thermophilus* resistentes de manera artificial.

2

Caracterización de cepas de *Streptococcus thermophilus*

Resumen

La caracterización de cepas es un paso fundamental en el proceso de selección de cultivos *starters*. Como se mencionó en el capítulo anterior, el primer paso en la estrategia planteada en este trabajo es la evaluación de ciertas propiedades tecnológicas asociadas a la performance de las cepas para luego enfocarnos en la interacción con bacteriófagos. El material de partida consistió en una colección de 50 cepas comerciales de *Streptococcus thermophilus* proporcionadas por la empresa Biomás S.A., además de una colección de fagos aislados del ambiente industrial uruguayo por nuestro grupo previamente (Achigar 2014). En este capítulo abordamos el estudio de las cepas basado en cuatro características tecnológicas y su caracterización a nivel molecular. Esto último es importante por dos motivos: en primer lugar, para conocer su diversidad y también para desarrollar una metodología más rápida que la actual para clasificar nuevos aislamientos.

Materiales y métodos

Cepas bacterianas

La colección de cepas utilizada en este trabajo se compone de 50 cepas de *Streptococcus thermophilus* proporcionadas por las empresas Biomás S.A. y Diagramma S.A. Las cepas se cultivaron en caldo M17 suplementado con 5 g/L de lactosa (LM17) y a 42°C sin agitación y en agar M17 suplementado con 5 g/L de lactosa. Las cepas de *Escherichia coli* y *Proteus mirabilis* se cultivaron en medio TSB a 37°C durante 16 horas previo a su utilización en los ensayos correspondientes.

Bacteriófagos

Los bacteriófagos utilizados en este trabajo fueron aislados por nuestro grupo a partir de muestras industriales en un trabajo previo (Achigar 2014). Los fagos fueron propagados en caldo M17 suplementado con 5 g/L de lactosa y 10mM CaCl₂ y utilizando la cepa sensible correspondiente. Una vez obtenidos los lisados los stocks fueron titulados en placa y por qPCR según fue descrito anteriormente (Achigar 2014) y conservados a 4°C para su uso rutinario y a -80°C como respaldo.

Detección del gen PrtS

La detección del gen PrtS fue llevada a cabo según se describe en el trabajo de Dandoy y col. (Dandoy et al. 2011). Los primers utilizados fueron prtsF (ATACCTGCACCTTGTTGGCG) y prtsR (GGTTCTGTTGTTATTGCAGC). La enzima utilizada fue la Taq DNA Polymerase Recombinant (Invitrogen, CA, E.E.U.U.) según instrucciones del fabricante. Los productos de PCR se visualizaron en gel de agarosa al 1% en buffer TBE 0,5x y agregando un volumen conocido del marcador de peso molecular 1kb+ (Invitrogen, CA, E.E.U.U.)

Actividad proteolítica en placa

El medio utilizado fue leche descremada reconstituida al 1% suplementada con 15 g/L de agar. Volúmenes conocidos de cultivo con una densidad óptica a 600nm de 0,6 se inocularon sobre un círculo de papel de filtro de 0,45 µm esterilizado por autoclave y se

colocaron sobre el medio solidificado. Las placas se incubaron 16 horas a 37°C y posteriormente se observó la formación de halos de proteólisis.

Actividad Ureasa

La actividad ureasa fue evaluada siguiendo el protocolo y reactivos descritos por Mora y col (Mora et al. 2002). con modificaciones. 100 µL de cultivo puro bacteriano con una densidad óptica a 600nm equivalente a 10 se inocularon en 1 mL de solución de reacción. Luego de dos horas de incubación a 37°C se visualizó el cambio de color en el medio.

Fermentaciones en leche

Para los ensayos de fermentaciones, un pre inóculo con una densidad óptica a 600nm de 0,6 se inoculó al 5% en 50 mL de leche descremada en polvo estéril reconstituida al 10% y esterilizada mediante vapor fluente a 100°C por 15 minutos. Posteriormente la fermentación fue llevada a cabo durante tres horas a 42°C sin agitación y el pH monitoreado cada veinte minutos.

Sensibilidad a bacteriófagos

Los ensayos de sensibilidad/resistencia fueron llevados a cabo en mediante *spot test*. 500 µL de cultivo bacteriano con una densidad óptica a 600nm de 0,6 fueron inoculados en LM17CaCl₂ con 7 g/L de agar termostatzado a 45°C y vertidos en placas de Agar LM17CaCl₂. Luego de solidificar el medio, alícuotas de 10 µL de lisado fágico fueron añadidos en forma de gota a la placa. Luego de que el líquido fuera absorbido se incubaron las placas a 37°C durante 16 horas.

Electroforesis en campo pulsado (PFGE)

La preparación de las muestras fue realizada en bloques de agarosa especial para PFGE (Bio-Rad, CA, E.E.U.U.) según las instrucciones del fabricante. La digestión fue llevada a cabo con la enzima *Sma*I (Thermo Fisher Scientific, MA, E.E.U.U.) La electroforesis fue llevada a cabo en un equipo CHEF-DRII PFGE (Bio-Rad, CA, E.E.U.U.) en gel de agarosa al 1% en buffer TBE 0,5x. La rampa programada fue de 5 a 30 segundos durante 23 horas con un

voltaje constante de 6 V/cm. Se utilizó como marcador de peso molecular cromosomas de *Saccharomyces cerevisiae* (Bio-Rad, CA, E.E.U.U.).

Amplificación aleatoria de ADN polimórfico (RAPD)

Los ensayos de RAPD fueron llevados a cabo utilizando el primer OPI-02 (GGAGGAGAGG) descrito en Mora y col. (Mora et al. 2002). Se utilizó la enzima *Top Taq Master Mix* (Qiagen CA, E.E.U.U.) Los resultados se visualizaron en gel de agarosa al 1% en TBE 0,5x. Se utilizó el marcador de peso molecular SmartLadder (Thermo Fisher Scientific, MA, E.E.U.U.). La matriz binaria fue construida calculando los tamaños de banda obtenidos en todas las amplificaciones. Con estos tamaños se generó un repertorio de posibles bandas y se asignó una posición relativa dentro del repertorio según su tamaño. La matriz se generó asignado valores de 1 y 0 en la posición indicada según el perfil mostrara la banda o no. Los dendrogramas en formato NWK fueron obtenidos con el servidor DendroUPGMA (genomes.urv.cat) mediante el algoritmo de UPGMA y el estadístico de coeficiente de correlación de Pearson. La salida gráfica se obtuvo con el software *CLC Sequence Viewer* (Qiagen, CA, E.E.U.U.). Esta metodología también fue utilizada en la generación del dendrograma de sensibilidad resistencia.

Evaluación del contenido de los *loci* CRISPR

El ADN bacteriano fue obtenido según se describió anteriormente (Achigar 2014). La amplificación de los *loci* CRISPR fue llevada a cabo utilizando los primers CR1F (TGCTGAGACAACCTAGTCTCTC) CR1R (TAAACAGAGCCTCCCTATCC) CR3F (GGTGACAGTCACATCTTGTCTAAAACG) y CR3R (GCTGGATATTCGTATAACATGTC) y la enzima *High Fidelity Master Mix* (Thermo Fisher Scientific, MA, E.E.U.U.) según instrucciones del fabricante. Los productos se visualizaron en gel de agarosa al 1% en buffer TBE 0,5x. Posteriormente se enviaron a secuenciar a Macrogen (Seúl, Corea) El software utilizado para el análisis de las secuencias fue ApE (v2.0.49.10) y SnapGene (v4.1.7)

Extracción de proteínas bacterianas

10 mL de cultivo bacteriano fueron centrifugados durante 15 minutos a 4700 g. El pellet fue re suspendido en 1 mL de buffer Tris-HCl pH 7,0, se agregaron 10 mg de lisozima y se incubó durante 60 minutos a 37°C. Luego, las células fueron sometidas a disrupción

mediante sonicado 60 ciclos de 30 segundos añadiendo solución de inhibición de proteasas. Por último, el lisado fue centrifugado a 4700 g durante 15 minutos a 4°C para eliminar los restos celulares. El sobrenadante se transfirió a un tubo limpio y se guardó a 4°C hasta su uso.

Extracción de ADN viral

La extracción de ADN viral se llevó a cabo según se describe anteriormente (Achigar 2014). 50 mL de lisado fueron tratados con 200 µL de DNasa I (1 mg/ml) y 200 µL de RNasa (1 mg/ml) durante 30 minutos a 37°C. Posteriormente se agregó PEG 6000 a una concentración final de 10%. Se disolvió mediante agitación orbital y se dejó precipitando durante la noche a 4°C. Se centrifugó a 10.000 g durante 20 minutos a 4°C. El pellet se re suspendió en 1 mL de NaCl 1M, Se agregaron 100 µL de buffer de lisis (500 mM Tris-HCl, 250 mM EDTA, 2.5% SDS) y se incubó a 65°C 30 minutos. Se agregaron 200 µL de KAc 8M frío y se incubó 30 minutos en hielo. Luego de centrifugar a 16.000 g durante 30 minutos, se extrajo dos veces con fenol:cloroformo (1:1) y se precipitó con isopropanol frío. El pellet se lavó con etanol 70% frío y posteriormente se re suspendió en 25 µL de buffer TE.

Digestión con extractos celulares

300 µL de extracto de proteínas fueron mezclados con 5 µL de ADN viral más buffer Tango o buffer R según el caso. Las incubaciones se llevaron a cabo a 37°C y 42°C con el fin de identificar condiciones óptimas durante 16 horas. Posteriormente se agregó 1 mL de isopropanol frío y se centrifugó a máxima velocidad durante 5 minutos. El pellet se lavó con etanol 70% frío y posteriormente se re suspendió en 10 µL de buffer TE. Los 10 µL se cargaron en un gel de agarosa al 1% en buffer TBE 0,5x para visualizar los resultados. Como control se utilizó pUC19 metilado proveniente de una *E. coli* DH5α transformada previamente.

Digestión de ADN viral con DpnI

La digestión fue llevada a cabo con la enzima DpnI de Thermo Fisher Scientific (MA, E.E.U.U.) Según instrucciones del fabricante. Los resultados se visualizaron en gel de

agarosa al 1% en buffer TBE 0,5x utilizando como marcador de peso molecular el 1kb+
(Invitrogen, CA, E.E.U.U.)

Resultados

Atributos tecnológicos

La especie *Streptococcus thermophilus* es ampliamente utilizada para la preparación de varios productos lácteos, como leches fermentadas, yogurt y quesos. Su rol en la fermentación de la leche no solo está relacionado con la producción de ácido láctico y la consecuente disminución del pH del medio sino también con otras propiedades tecnológicas importantes que se deben tener en cuenta para seleccionar cepas *starters*. En particular, en este trabajo nos enfocamos en cepas de *S. thermophilus* con el objetivo de utilizarlas como *starters* en la elaboración de queso y no en cepas para elaboración de yogurt que requieren un set de características tecnológicas diferentes. Entre las propiedades más importantes a relevar en *S. thermophilus* para su uso como starter se encuentran la capacidad de proteólisis de la caseína, la actividad ureasa y la resistencia a bacteriófagos (Mora et al. 2002). También es importante la determinación de la tasa de crecimiento y el comportamiento de las cepas en leche dado que su crecimiento en esta matriz se ve afectado por las características metabólicas globales de las cepas como la utilización de lactosa, tolerancia al pH, metabolismo de la galactosa, etc. De forma de conocer su comportamiento global en leche se realizó la curva de acidificación para cada cepa.

Actividad proteolítica

Urshev y colaboradores afirman que las cepas de *Streptococcus thermophilus* pueden clasificarse en dos grandes grupos; las cepas del grupo “H” (grupo minoritario) y un segundo grupo que integra al resto (sin nombre definido) (Urshev et al. 2014). Este grupo H (nombrado así por *high acidification rate* o alta tasa de acidificación) es según los autores un grupo minoritario en la diversidad de cepas de *S. thermophilus* y está compuesto por cepas capaces de acidificar muy rápidamente la leche, y que además tienen como característica común la presencia del gen codificante para una proteasa de membrana *prtS* (las características de esta enzima se detallan en el capítulo 1) (Urshev et al. 2014), fenómeno que ya había sido reportado por Fernandez-Espla y colaboradores (Fernandez-Espla et al. 2000), y Dandoy y colaboradores (Dandoy et al. 2011). Tomando en cuenta estos

reportes, es que decidimos realizar un estudio cualitativo de la capacidad proteolítica de las cepas, mediante un *screening* de la colección en busca del gen codificante para PrtS mediante PCR. La metodología utilizada se detalla en materiales y métodos al final de este capítulo.

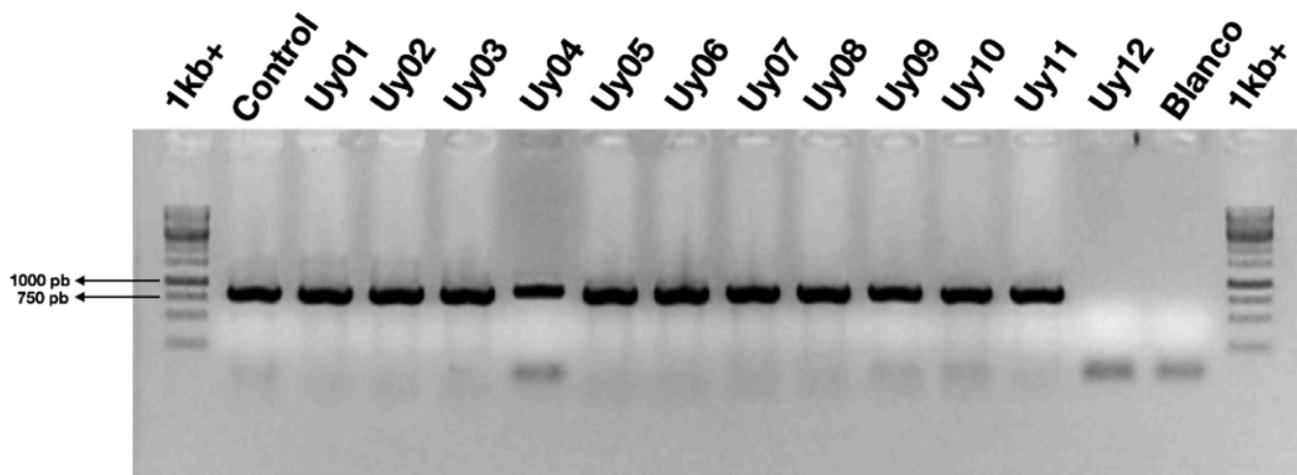


Figura 2.1: Electroforesis en gel de agarosa de los productos de amplificación para el gen *prtS*. En esta imagen ejemplo se observan los productos obtenidos para 12 cepas de la colección. El carril marcado como control es un amplicón de un ADN previamente obtenido y confirmado mediante secuenciación. Se observa la ausencia de amplificación para la cepa Uy12

El resultado obtenido fue la amplificación del fragmento del tamaño esperado para 49 de las 50 cepas de la colección.

Además, para determinar si existían diferencias en la capacidad de realizar proteólisis entre las cepas que confirmaron por PCR la presencia del gen para PrtS se realizó un ensayo semi-cuantitativo de forma de observar los niveles de proteólisis de caseína en agar leche (ver materiales y métodos). La leche es una suspensión coloidal donde la fase dispersa está constituida por las moléculas de caseína. Como resultado de la proteólisis se forman subproductos solubles, por lo que el medio de cultivo deja de ser opaco y pasa a ser transparente. Cuanto mayor sea el nivel de proteólisis, mayor será el diámetro transparente alrededor de la zona de siembra del cultivo. Por otra parte, se puede observar un halo opaco rodeando la zona de siembra; esto corresponde a proteínas precipitadas como consecuencia de la sobreproducción de ácido láctico. Los resultados obtenidos fueron bastante homogéneos haciendo difícil establecer diferencias significativas, sin embargo, se observaron algunos casos extremos.

En la figura 2.2, se observan las diferencias en los halos de proteólisis para tres cepas diferentes de la colección. En las regiones 1 y 2 se sembraron diferentes volúmenes de un cultivo puro de la cepa Uy05, en el 3 y 4 de la cepa Uy03 y en el 4 y 5 de la cepa Uy04. Como puede apreciarse, existen marcadas diferencias entre el diámetro obtenido para la cepa Uy05 comparada con las otras dos cepas. Esto refleja que, si bien la capacidad de consumo de caseína está vinculada a la presencia del gen *prtS*, el grado de proteólisis depende de otros factores, como pueden ser mutaciones, cambios en el promotor, contexto genómico (lugar del cromosoma donde esté integrada la isla) etc. En la figura 2.3 se observa el ensayo para la cepa Uy12; la flecha roja muestra la práctica inexistencia de halo de proteólisis confirmado el resultado del ensayo de PCR.

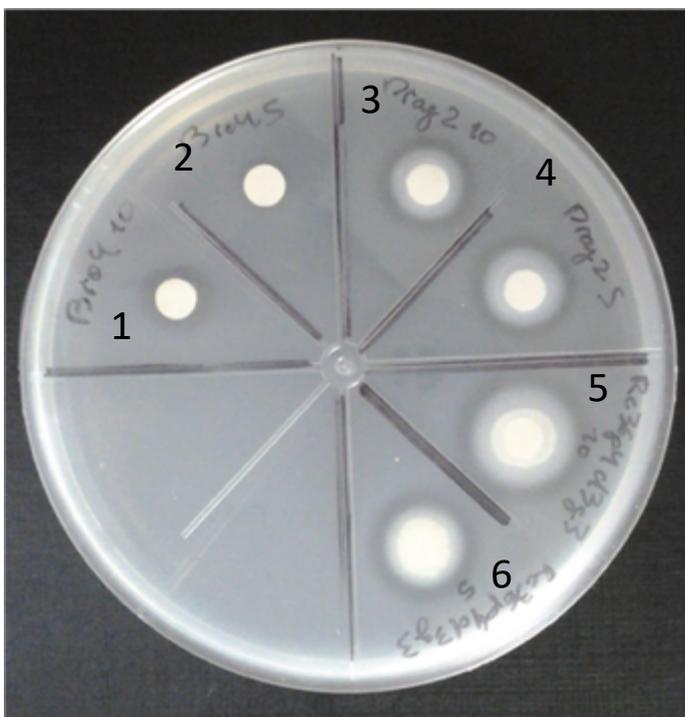


Figura 2.2. Ensayo semi-cuantitativo de actividad proteolítica. Los cuadrantes 1 y 2 corresponden a la siembra de 10 y 5 μ L respectivamente de un cultivo puro de Uy05, los cuadrantes 3 y 4 lo mismo para la cepa Uy03, y 5 y 6 para la cepa Uy04. Es posible observar los halos traslucidos de proteólisis de caseína, y para el caso de las cepas Uy03 y Uy04 se observa además el halo blanco de precipitación. También se aprecia que no hay diferencias significativas cuando se siembran diferentes volúmenes de cultivo a este nivel. Los nombres que aparecen rotulando las placas son los utilizados por la empresa para codificar las cepas.

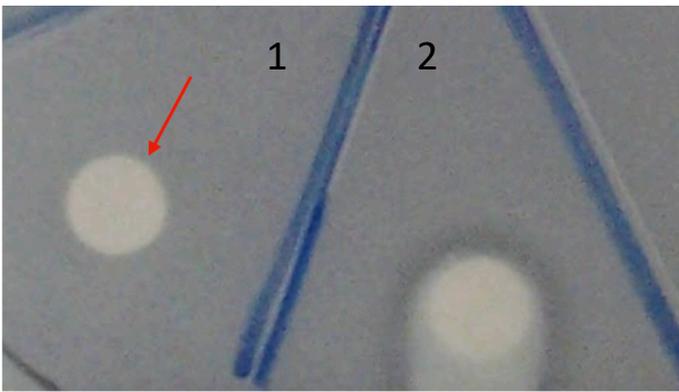


Figura 2.3. Ensayo semi-cuantitativo de actividad proteolítica para la cepa Uy12. En el cuadrante 1 se encuentra sembrada una alícuota de un cultivo puro de Uy12, y en el cuadrante 2 una alícuota de cultivo de Uy03 como control. La flecha roja marca la ausencia de halo de proteólisis.

Estos resultados muestran que si bien el *screening* por PCR es un primer paso para seleccionar las cepas con mayor potencial, es necesario llevar a cabo ensayos cualitativos para tener datos más confiables.

Actividad ureasa

Como se mencionó en el capítulo anterior *Streptococcus thermophilus* es la única BAL con una actividad ureasa significativa (Mora et al. 2004), y solo se han descrito unas pocas cepas ureolíticas lentas o negativas (Mora et al. 2002). La urea es el principal compuesto con nitrógeno no proteico presente en la leche, y su contenido puede alcanzar el 5% del nitrógeno total. Los altos niveles de urea en la leche son causados por multiplicidad de factores, siendo hoy en día el principal, la alimentación con raciones altas en proteínas proporcionadas al ganado lechero. La degradación de la urea por parte de la bacteria produce amonio, cuya acumulación durante el proceso de elaboración de quesos puede afectar negativamente el perfil de acidificación, ya que neutraliza la acción del ácido láctico. Se ha demostrado que mutantes espontáneos ureasa-deficientes (que pierden el *clúster* asociado), muestran perfiles de acidificación significativamente más rápidos que sus cepas madre (Urshev 2013).

La evaluación de la actividad ureasa fue llevada a cabo mediante un método colorimétrico según se describe en materiales y métodos. La figura 2.4 muestra un ejemplo de ensayo realizado a microescala, mientras que la figura 2.5 muestra los resultados del ensayo para la totalidad de las cepas.

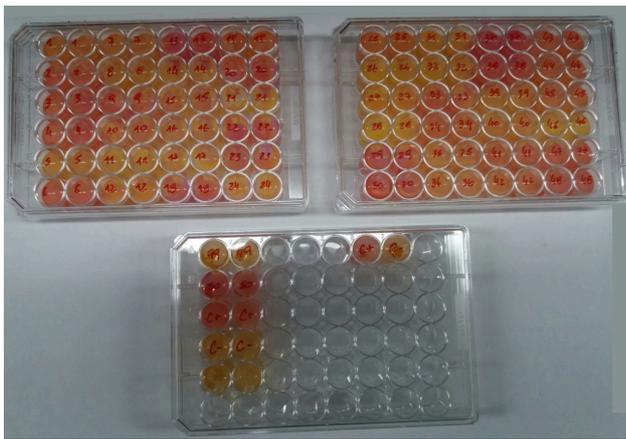


Figura 2.4. Ensayo a microescala de actividad ureasa. Luego de dos horas de incubación se observa un desarrollo de color rojizo para las cepas con actividad ureolítica positiva, mientras que las que no metabolizan urea no generan viraje de color en el medio.

Uy01	Uy02	Uy03	Uy04	Uy05
Uy06	Uy07	Uy08	Uy09	Uy10
Uy11	Uy12	Uy13	Uy14	Uy15
Uy16	Uy17	Uy18	Uy19	Uy20
Uy21	Uy22	Uy23	Uy24	Uy25
Uy26	Uy27	Uy28	Uy29	Uy30
Uy31	Uy32	Uy33	Uy34	Uy35
Uy36	Uy37	Uy38	Uy39	Uy40
Uy41	Uy42	Uy43	Uy44	Uy45
Uy46	Uy47	Uy48	Uy49	Uy50

Figura 2.5. Matriz esquemática del fenotipo mostrado por las cepas de la colección. Las celdas coloreadas con fondo amarillo corresponden a cepas que no mostraron actividad ureolítica, mientras que las coloreadas con fondo violáceo corresponden a cepas con actividad ureolítica.

Podemos observar que 21 cepas mostraron actividad ureasa negativa y 29 actividad positiva, lo que es una proporción muy buena dado la escasez de cepas con este fenotipo reportadas (Urshev 2013; Mora et al. 2002; Mora et al. 2004)

Curvas de fermentación

Como se mencionó anteriormente, la curva de fermentación en leche aporta información acerca del comportamiento metabólico global de las cepas. Las características tecnológicas básicas deseadas (como las estudiadas en este capítulo) aportan a una condición fundamental que deben poseer las cepas de *S. thermophilus* utilizadas como *starters*, poseer una alta velocidad de acidificación. La forma de comparar el desempeño de las cepas es mediante la visualización de sus perfiles de acidificación graficados juntos. Las fermentaciones fueron llevadas según se describe en la sección materiales y métodos. La figura 2.6 muestra las curvas de fermentación para las 50 cepas. Los ensayos fueron realizados por triplicado y los promedios graficados para cada punto. Por razones visuales se omiten en la figura las barras de error.

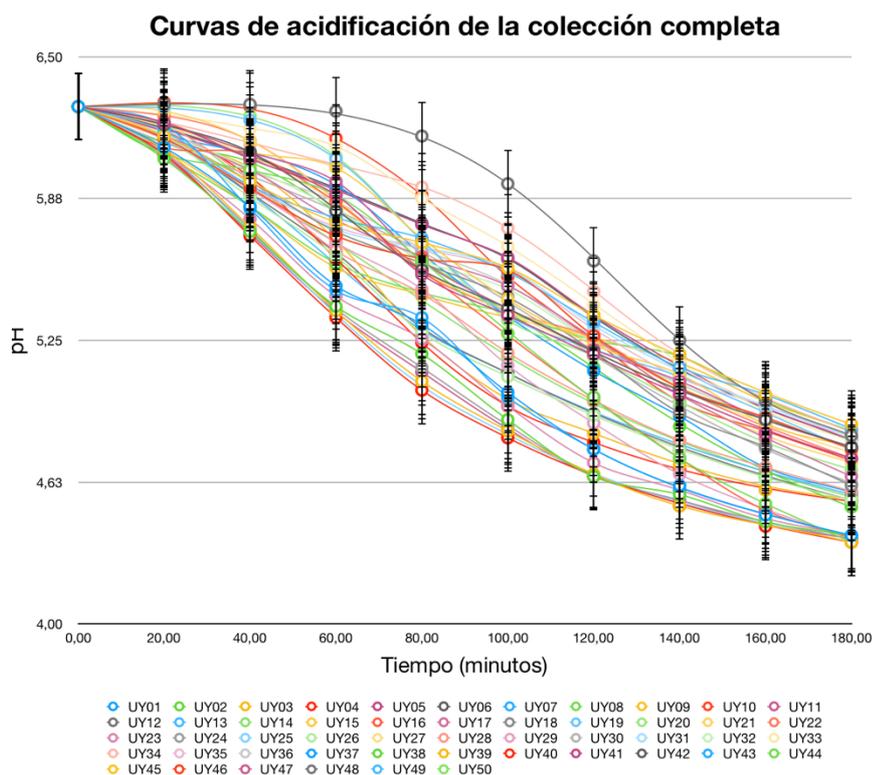


Figura 2.6. Curvas de fermentación para las 50 cepas de la colección.

Se puede observar que las curvas se encuentran agrupadas en un rango con poca dispersión con la marca excepción de UY12. Cabe recordar que esta cepa es la única que mostró resultados negativos para la amplificación del gen *prtS*, por lo que era esperable

que su comportamiento fuera diferente al de las otras cepas. También podemos observar que, a pesar de la poca dispersión de los puntos, hay diferencias notorias entre cepas PrtS positivas. De la misma forma que se vio en el ensayo semi-cuantitativo de actividad proteolítica, la mera presencia del gen no garantiza que el desempeño de todas las cepas sea similar. Para probar la relación entre el grado de proteólisis observado en placa y la velocidad de acidificación de las cepas, se realizaron los cálculos para las mismas cepas del ensayo anterior. En la figura 2.7 se muestran las curvas de fermentación para las cuatro cepas: las cepas con alta proteólisis Uy03 y Uy04, la cepa que mostró baja proteólisis Uy05, y la cepa no proteolítica Uy12. Puede observarse que existen diferencias importantes entre los tres fenotipos.

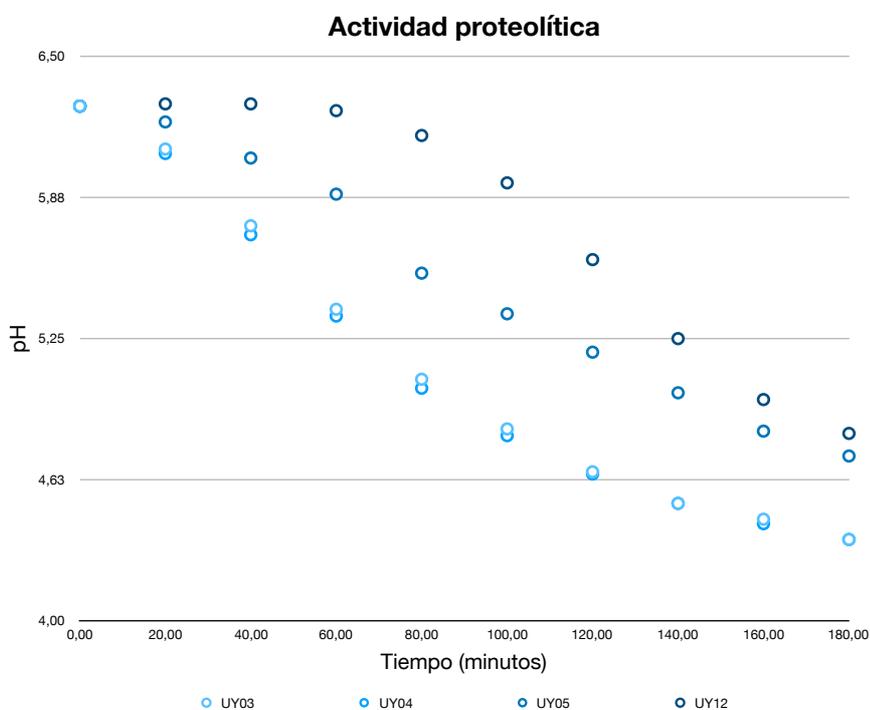


Figura 2.7. Curvas de fermentación para las cuatro cepas evaluadas en su grado de proteólisis. Pueden observarse diferencias entre Uy12 y las cepas PrtS positivas. A su vez se observan diferencias entre las PrtS positivas

Para cuantificar estas diferencias, se calculó la velocidad específica de acidificación como una función del pH (q_{pH}). En la figura 2.8 se grafican las velocidades específicas en cada intervalo de la fermentación, señalándose para los cuatro casos el valor máximo ($q_{pH\text{máx}}$) y en qué tiempo de fermentación se alcanzó. Es interesante observar que no solo $q_{pH\text{máx}}$ es

mayor cuanto mayor sea el grado de proteólisis, sino que el momento de mayor velocidad también se alcanza antes, por lo que estos efectos combinados hacen suponer que redundarán en un mayor rendimiento de la cepa, lo cual se puede traducir en mayores productividades como resultado de los procesos de elaboración.

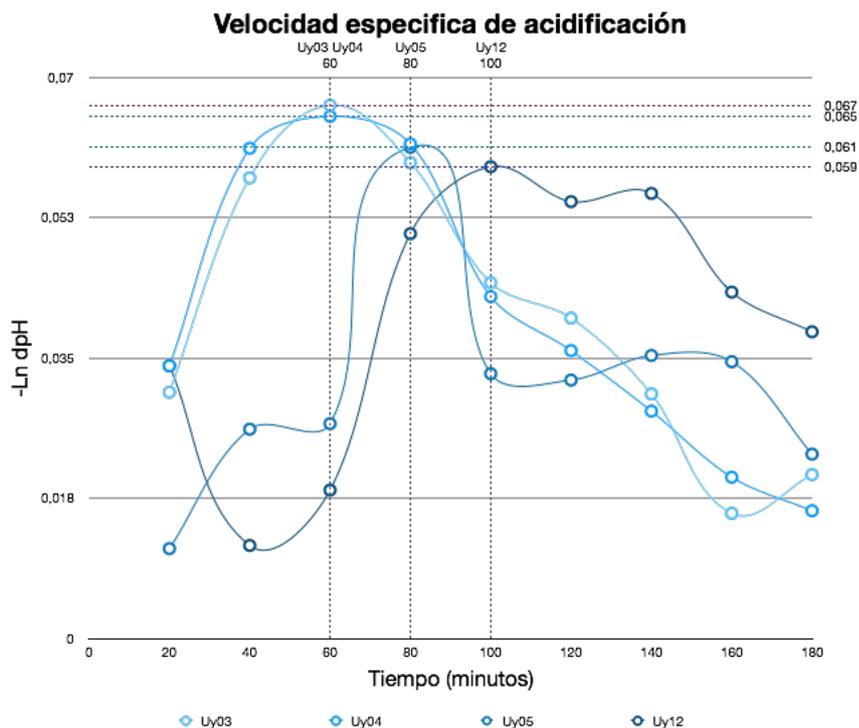


Figura 2.8. Gráfico de q_{pH} para las cepas evaluadas en su capacidad proteolítica. Puede observarse que el valor de $q_{pH\text{máx}}$ para Uy03 es el mayor (línea punteada roja) seguida con muy poca diferencia por Uy04 (línea punteada azul). En ambos casos este valor se alcanza a los 60 minutos de fermentación. En tercer lugar, se ubica la cepa Uy05, alcanzándose el pico a los 80 minutos. Por último, el valor de $q_{pH\text{máx}}$ de la cepa Uy12 es el más bajo (línea punteada violeta) y se alcanza a los 100 minutos de fermentación.

Por último, observamos que en el contexto estudiado no existe una relación entre los resultados de los ensayos de la actividad ureasa y las curvas de fermentación. En las figuras 2.9 y 2.10 se muestran las curvas de fermentación de las cepas según los resultados obtenidos para los ensayos de actividad ureolítica. Este resultado puede deberse a que, al tratarse de fermentaciones en leche en polvo reconstituida, los niveles de urea pueden ser muy bajos para generar un impacto. Datos manejados por el INTA (inta.gov.ar) muestran que las concentraciones máximas de urea en leche se reportan en invierno alcanzando valores de 0,069 g/100 mL, y que la leche en polvo obtenida de esta leche fresca y

reconstituida al 13% presenta valores máximos de 0,035 g/100 mL. Extrapolando estos datos a las condiciones de nuestro trabajo (las concentraciones a las que nosotros reconstituimos la leche), hablamos de una pérdida aproximada de 2/3 de la urea original en el proceso de elaboración.

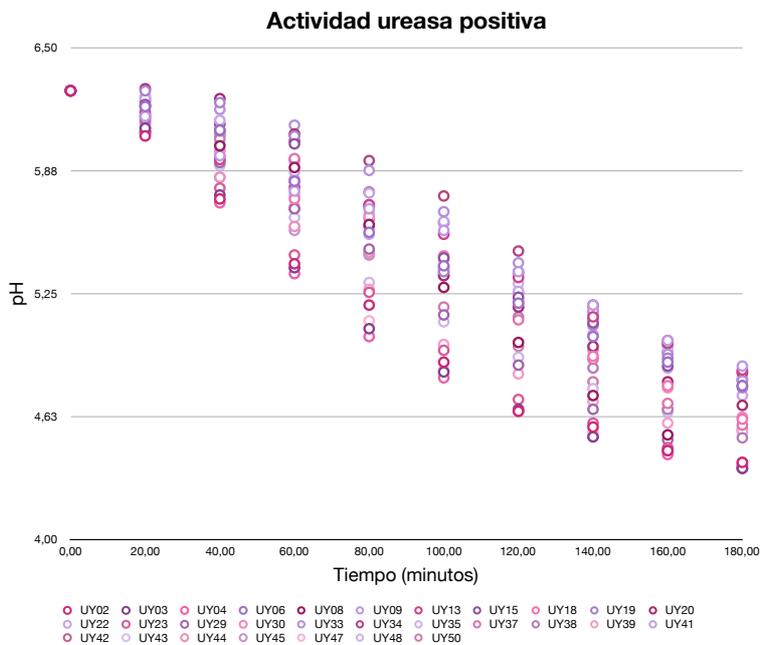


Figura 2.9: Curvas de fermentación de cepas con actividad ureasa positiva

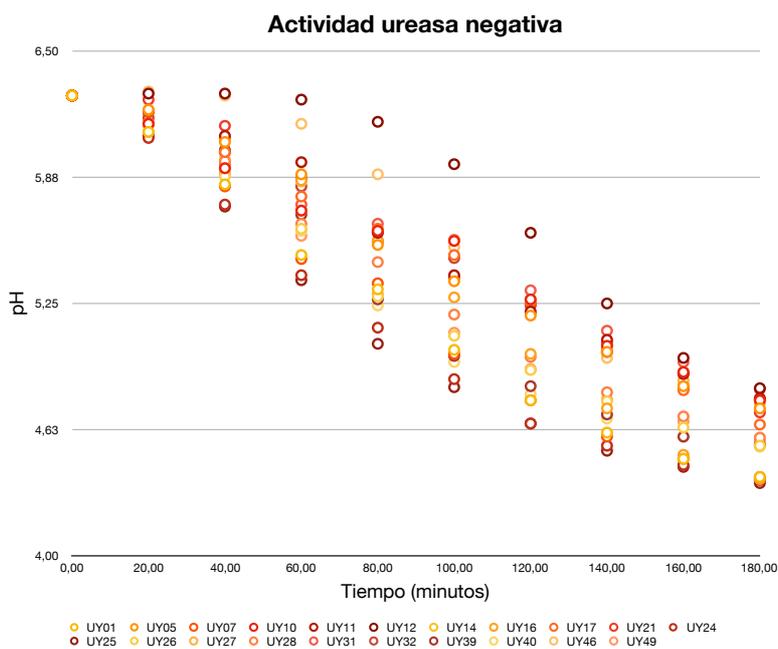


Figura 2.10: Curvas de fermentación de cepas con actividad ureasa negativa

Si bien los resultados no son concluyentes en este aspecto, la literatura es vasta en referencias acerca del papel de la actividad ureasa en el desempeño de las cepas. Aún así, para ajustar la estrategia a condiciones de reproducibilidad, queda pendiente la realización de ensayos de fermentación con concentraciones de urea diferentes que reproduzcan los hábitos de alimentación del ganado lechero de nuestro país.

Sensibilidad a bacteriófagos

Como se mencionó anteriormente, uno de los insumos de este trabajo es una colección de bacteriófagos obtenidos del ambiente industrial uruguayo. Estos fagos fueron aislados principalmente de sueros de fermentaciones fallidas, mediante una primera detección por qPCR y posterior propagación en las cepas utilizadas en el mismo proceso de fermentación (Achigar 2014). Si bien los perfiles de ataque contra cepas de la colección fueron realizados en un primer momento (Achigar 2014), en este trabajo se generó una nueva matriz de sensibilidad/resistencia tomando en cuenta los bajos títulos virales, acción de endolisinas en el medio, etc. Este trabajo utilizó un banco compuesto por varios fagos de tipo cosy un fago de tipo *Pac*. De los cientos de muestras positivas que analizamos, únicamente un fago tipo *Pac* fue detectado y aislado, por lo que los fagos tipo cosserían los más abundantes en el ambiente industrial uruguayo. Esto se encuentra en concordancia con el hecho de que la enorme mayoría de genomas de fagos de *S. thermophilus* secuenciados pertenecen a este tipo (McDonnell et al. 2017).

Una de las características principales a la hora de seleccionar un *starter* es la resistencia a bacteriófagos. Como se mencionó en el capítulo anterior, los bacteriófagos pueden causar efectos muy adversos en los procesos biotecnológicos, y son la principal causa de fermentaciones lácticas fallidas a nivel industrial (Bruttin et al. 1997). El ambiente de elaboración de quesos es sumamente propicio para que las poblaciones fágicas prosperen, se mantengan y diversifiquen. En general, los fagos de *S. thermophilus* aislados de plantas de elaboración de queso muestran mayor diversidad que aquellos derivados de elaboraciones de yogurt. Una explicación para esto radica en las diferencias de iniciadores y condiciones de fermentación. La propagación fágica es dependiente de la presencia de hospedadores adecuados y en las elaboraciones de yogurt se emplea un pequeño número de *starters* bien definidos, pero también las condiciones de tratamientos térmicos y diseños de planta y procesos (tinajas de fermentación cerradas versus abiertas), las plantas de elaboración de yogurt son diferentes, menos susceptibles

que las queserías a la invasión por fagos provenientes de fuentes ambientales. Es por esto que los patrones de sensibilidad/resistencia son mucho más importantes en las cepas utilizadas como *starters* para la elaboración de quesos, como es el caso de las cepas utilizadas en este trabajo. La tabla 2.1 muestra los resultados obtenidos de los ensayos de sensibilidad/resistencia. De las 50 cepas, 21 fueron resistentes a los 11 fagos de la colección, mientras que las 29 restantes mostraron diferentes grados de sensibilidad. También se observa que el fago 53 es el más agresivo de todos los aislados, infectando un total de 13 cepas. En el capítulo siguiente se aborda en profundidad la caracterización de los bacteriófagos de la colección mediante análisis genómicos.

De las características tecnológicas introducidas en este capítulo, la resistencia a bacteriófagos es la más compleja, variable y la que mayor impacto tiene sobre la producción. Esto se debe a que la sensibilidad a fagos, a diferencia de las otras características, puede impactar en la producción en forma repentina, grave e impredecible. A nivel de planta es complejo predecir la presencia de fagos en el ambiente o en la materia prima. Mientras que una cepa caracterizada con alta capacidad proteolítica mantendrá su fenotipo constante, la resistencia a fagos por parte de una cepa es relativa y dependiente de los desafíos con que se encontrará al momento de la producción; la variabilidad de los fagos y las altas tasas de mutación hacen difícil de garantizar si los fenotipos observados se mantendrán en el tiempo. Además, las estrategias de control de bacteriófagos a nivel de planta son muchas veces insuficientes y poco efectivas. Por estas razones el foco principal del trabajo está dirigido al desarrollo de cepas de *Streptococcus thermophilus* resistentes a bacteriófagos.

Tabla 2.1: matriz de sensibilidad resistencia: en las columnas se ubican los bacteriófagos de la colección mientras que en las filas las cepas de *S- thermophilus*. Las celdas coloreadas indican sensibilidad.

		Bacteriófagos de la colección										
		53	73	93	94	107	115	123	128	140	142	ST
Cepas de <i>Streptococcus thermophilus</i> de la colección	Uy01											
	Uy02											
	Uy03											
	Uy04											
	Uy05											
	Uy06											
	Uy07											
	Uy08											
	Uy09											
	Uy10											
	Uy11											
	Uy12											
	Uy13											
	Uy14											
	Uy15											
	Uy16											
	Uy17											
	Uy18											
	Uy19											
	Uy20											
	Uy21											
	Uy22											
	Uy23											
	Uy24											
	Uy25											
	Uy26											
	Uy27											
	Uy28											
	Uy29											
	Uy30											
	Uy31											
	Uy32											
	Uy33											
	Uy34											
	Uy35											
	Uy36											
	Uy37											
	Uy38											
	Uy39											
	Uy40											
	Uy41											
	Uy42											
	Uy43											
	Uy44											
	Uy45											
	Uy46											
	Uy47											
	Uy48											
	Uy49											
	Uy50											

Caracterización molecular

El objetivo de la caracterización molecular era conocer la diversidad de las cepas de la colección, más allá de sus características tecnológicas, fundamentalmente conocer si existía alguna correlación entre la base genética de las cepas y los fenotipos de resistencia a fagos.

Como primera aproximación, decidimos utilizar dos estrategias clásicas; la electroforesis en campo pulsado (PFGE) y la amplificación aleatoria de ADN polimórfico (RAPD). La primera consiste en la digestión del genoma bacteriano completo y observación del patrón de bandas obtenido, y la segunda en la amplificación aleatoria con *primers* especiales y también y la comparación de los patrones obtenidos. Nuestra estrategia consistió en comparar los resultados obtenidos mediante estas dos técnicas (es decir cómo se agrupaban las cepas con respecto a sus perfiles de restricción o amplificación dependiendo de la técnica) con los resultados obtenidos en los ensayos de sensibilidad/resistencia. Como primer paso y para obtener resultados que pudieran compararse de forma aproximada, decidimos transformar la matriz de sensibilidad/resistencia en un dendrograma y observar las agrupaciones que se formaban. Esto fue realizado convirtiendo en primer lugar la matriz de resultados en una matriz binaria, y luego utilizando esta última, se calcularon las distancias mediante el algoritmo UPGMA (ver la sección de materiales y métodos). El resultado se muestra en la figura 2.11. Si bien como se mencionó esto es solo una aproximación, reproduce los resultados observados en la matriz con bastante precisión, agrupando fenotipos similares. Además, en la figura se marcan los *clústeres* identificados, algunos de ellos se pueden explicar con solvencia, mientras que otros parecen agrupaciones más azarosas (a pesar de que se computaron 100 réplicas de *bootstrap* en la generación del dendrograma). El *clúster* número 1 agrupa a las cepas que fueron resistentes a todos los fagos; el número 2 está integrado únicamente por Uy05, la única cepa sensible al fago 93 (tipo *Pac*); el número 3 agrupa dos cepas que únicamente son sensibles a los fagos 128 y 142; el *clúster* 4 agrupa a las cepas que únicamente son sensibles al fago 73; el número 5a todas las cepas que son sensibles solamente al fago 53; el *clúster* 6 agrupa cepas que tienen en común la sensibilidad al fago 123, siendo algunas únicamente sensibles a este, y el resto sensibles además al 107; el número 7 agrupa dos cepas que tienen en común, únicamente que son sensibles al fago 140, siendo para una el único capaz de infectarla y para la otra se suma además el 123; y por último el *clúster* número 8 engloba al

subgrupo de cepas que mostró los perfiles de sensibilidad más grandes y por tanto, son las cepas más débiles, en oposición al *clúster* 1.

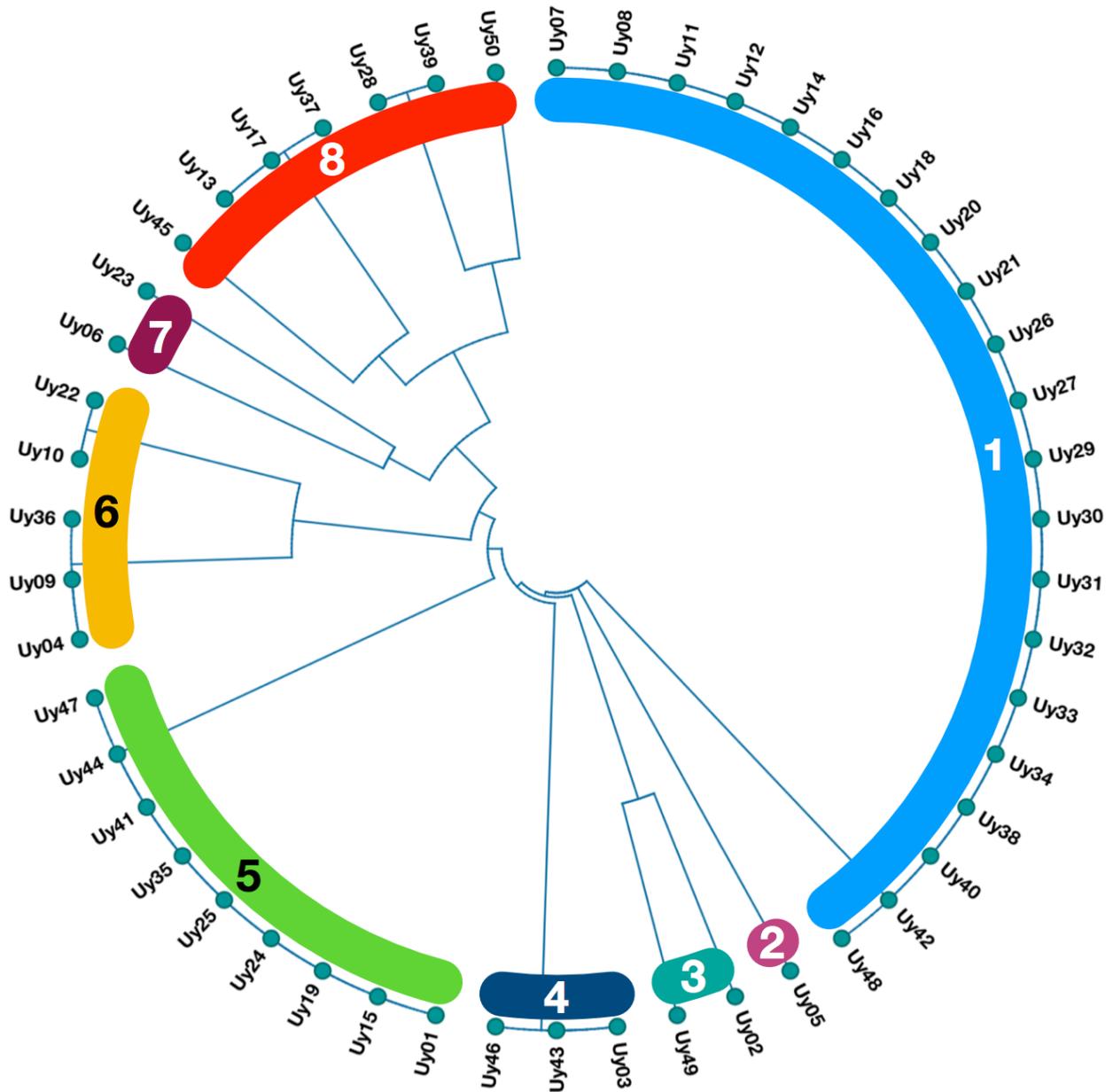


Figura 2.11. Dendrograma obtenido a partir de la matriz de sensibilidad resistencia.

Electroforesis en campo pulsado (PFGE)

Los ensayos de electroforesis en campo pulsado fueron realizados según se describe en materiales y métodos. La figura 2.12 muestra un ejemplo de los resultados obtenidos para los ensayos de PFGE.

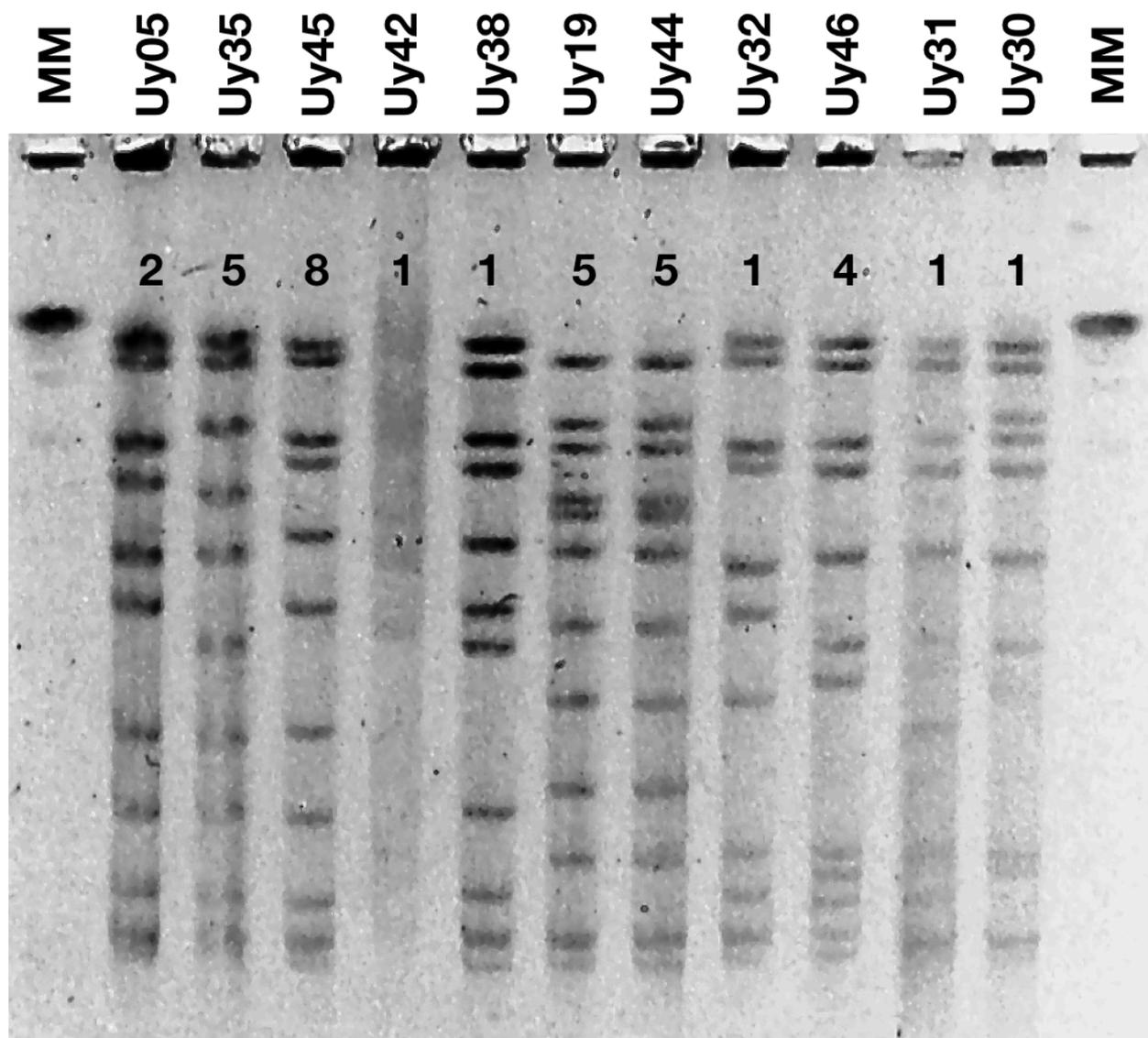


Figura 2.12. Electroforesis en campo pulsado para muestras de ADN digerida con la enzima *SmaI*. Los números en los carriles en el gel identifican al *clúster* de sensibilidad al que pertenece cada cepa (ver figura 2.11). El marcador utilizado son cromosomas de *Saccharomyces cerevisiae*.

En el ejemplo, puede observarse que las cepas Uy30, Uy31, Uy32 y Uy38 presentan diferentes perfiles de digestión, aunque presentan el mismo perfil de resistencia. Para el caso de Uy19 y Uy44, puede observarse que presentan el mismo perfil de restricción y de

resistencia, sin embargo, la cepa Uy35 que comparte la misma resistencia que las anteriores, presenta un perfil de digestión diferente. Estas inconsistencias hacen que esta metodología es ineficaz para predecir el comportamiento de nuevos aislamientos frente a los bacteriófagos aislados.

Amplificación aleatoria de ADN polimórfico (RAPD)

Esta técnica de *fingerprinting* se basa en la amplificación de regiones aleatorias del ADN utilizando *primers* cuyas secuencias se diseñan de forma arbitraria. A diferencia de los análisis tradicionales basados en PCR, esta técnica no requiere información previa de las secuencias de ADN genómico del organismo a evaluar. Las amplificaciones fueron llevadas a cabo según se describe en materiales y métodos. Una vez obtenidas las amplificaciones, se analizaron los patrones de bandas y se generó una matriz binaria a partir de los mismos, la cual fue el insumo para construir un dendrograma utilizando el algoritmo UPGMA. La figura 2.13 muestra una imagen de un gel de agarosa con la electroforesis realizada posterior para algunas cepas de la colección.

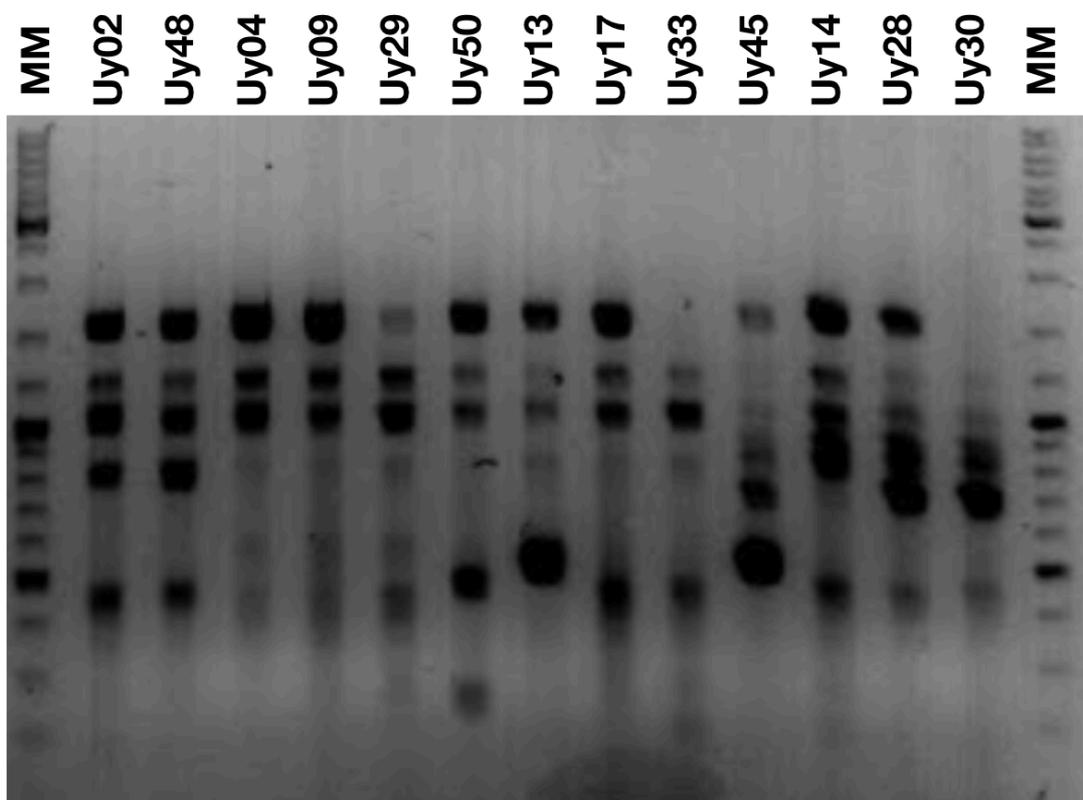


Figura 2.13. Patrones de RAPD obtenidos para alguna cepas de la colección utilizando el primer OPI-02 (Mora et al. 2002). El marcador utilizado fue el SmartLadder (Thermo Fisher Scientific, Waltham MA, E.E.U.U.)

La figura 2.14 Muestra el dendrograma obtenido a partir de los resultados de los perfiles de amplificación.

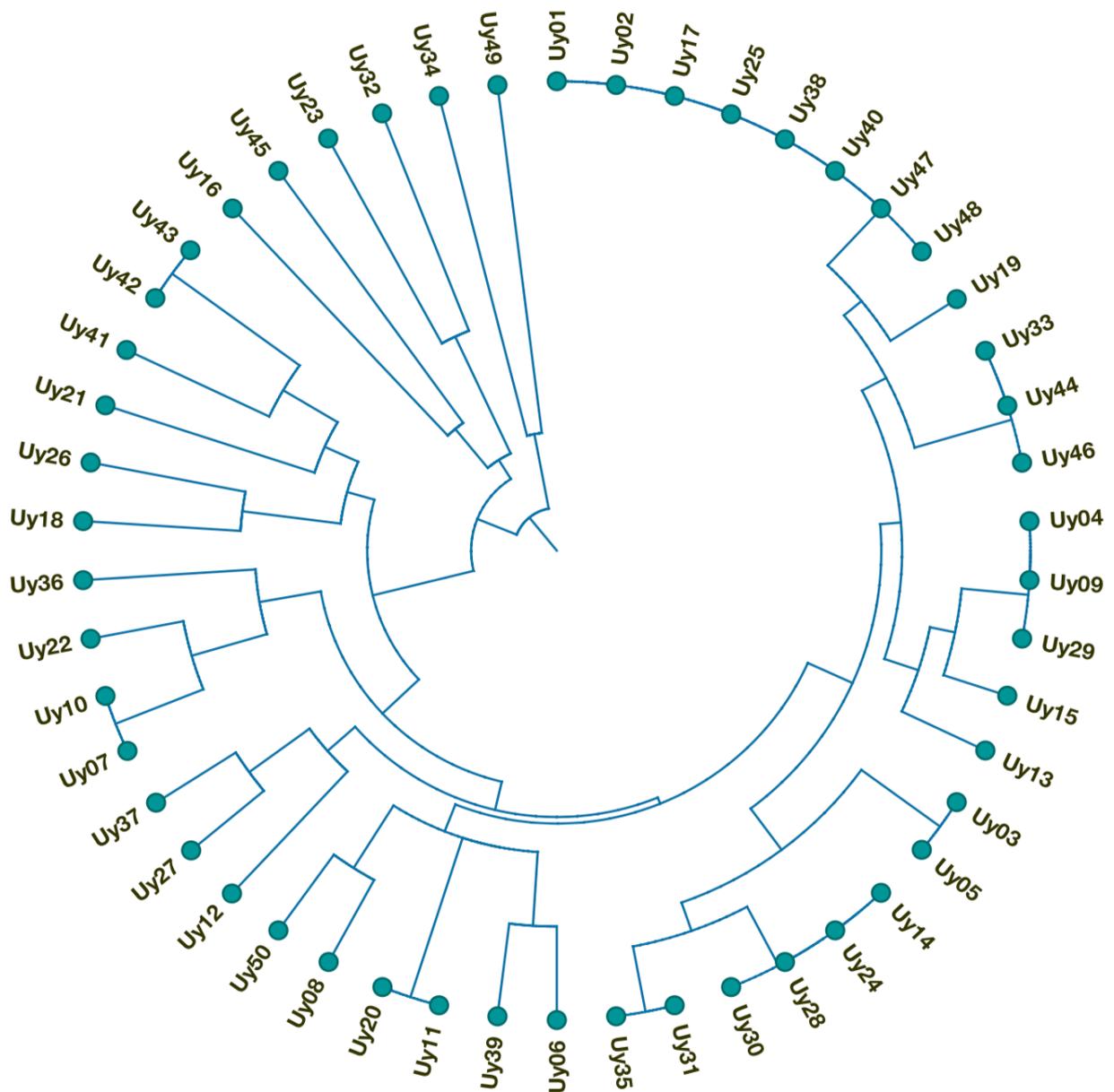


Figura 2.14. Dendrograma obtenido con los datos de RAPD de las cepas de la colección. Puede observarse una topología y agrupamiento de las cepas totalmente diferente al observado en la figura 2.11

Al comparar ambos dendrogramas, se puede observar que no muestran similitud alguna en su topología, por lo que no se pueden relacionar de forma alguna los patrones de amplificación a los patrones de sensibilidad/resistencia de las cepas. Si bien los datos obtenidos de estos ensayos son en sí mismos un resultado interesante desde el punto de

vista de la diversidad del banco de cepas, no aportan información alguna acerca del comportamiento de las cepas frente a los diferentes fagos de la colección, por lo que no se ajustan a lo esperado en este trabajo. La siguiente estrategia que adoptamos para buscar marcadores moleculares asociados a los patrones de sensibilidad /resistencia, fue el análisis de los *loci* CRISPR de las cepas de la colección.

Contenido de los *loci* CRISPR

Como se describió en el capítulo anterior, CRISPR-Cas es un sistema inmune adaptativo exclusivo de organismos procariontes. Su elemento central es el locus CRISPR, una estructura genética que se caracteriza por secuencias cortas palindrómicas repetidas, que se encuentran separadas por secuencias de igual longitud, pero de origen exógeno, como, por ejemplo, parte de genomas de bacteriófagos. Estos espaciadores son los componentes que le otorgan especificidad al sistema y juntos conforman un registro de los enfrentamientos que ha tenido que soportar el organismo a lo largo de su historia evolutiva.

Desde el punto de vista de los objetivos de este trabajo, el análisis del contenido de espaciadores en los *loci* CRISPR parece una mejor estrategia a adoptar para evaluar y predecir el comportamiento de las cepas frente a los bacteriófagos. La limitante que tiene este enfoque es por ser una estrategia de dos componentes: no solo se debe tener la información del contenido de espaciadores, sino que deben conocerse las secuencias de los genomas de los fagos a evaluar. El análisis de los genomas virales se convierte en un insumo fundamental a la hora de seleccionar y diseñar cepas resistentes. Es por esta razón que los genomas de los 11 fagos utilizados en este trabajo fueron completamente secuenciados y analizados. En los próximos capítulos se ahondará en el análisis de estas secuencias y en la información que brindan a la hora de planificar una estrategia de diseño de cepas resistentes. En este apartado nos enfocamos principalmente en el contenido de espaciadores de las cepas, la determinación de si existían protoespaciadores homólogos funcionales a estos en los genomas virales y si esta información genética era concordante con los fenotipos de sensibilidad/resistencia observados.

En el caso de *S. thermophilus*, se han reportado dos *loci* CRISPR activos llamados CRISPR1 y CRISPR3, ambos pertenecientes al sub tipo IIA (Horvath et al. 2008). A pesar de que pueden considerarse homólogos, ya que el funcionamiento global de ambos sistemas es

igual, estos dos loci presentan diferencias sustanciales: 1- las secuencias de sus repetidos y líder son diferentes; 2- el *locus* CRISPR1 presenta un repetido final en la posición 3' cuya secuencia difiere al resto; 3- el *locus* CRISPR3 posee una secuencia terminadora conservada en el extremo 3' del último repetido; 4- la secuencia del PAM de ambos *loci* es diferente.

A pesar de estas diferencias, existen similitudes fundamentales para ambos tipos de sistema: es condición excluyente la presencia de un PAM adyacente para el correcto funcionamiento del espaciador; las mutaciones pueden afectar el desempeño del mismo, especialmente en una secuencia de ocho pares de bases en el extremo 3' del espaciador conocida como secuencia semilla (Semenova et al. 2011). Este criterio introducido en el trabajo de Semenova y colaboradores de 2011 es el que adoptamos para guiarnos en este trabajo: básicamente luego de obtener las secuencias de los espaciadores, se buscaron secuencias homólogas en los genomas virales que cumplieran con el criterio de estar adyacente a un motivo PAM sin mutaciones, y que además presentara la menor cantidad de mutaciones puntuales en la secuencia, con especial cuidado de observar si había mutaciones en las secuencias semilla. Las secuencias homólogas que se encontraron, pero que no cumplieron con este criterio, no fueron consideradas protoespaciadores y por tanto no se les atribuyó actividad a dichos espaciadores.

Secuencias espaciadoras

Las secuencias de los *loci* se obtuvieron según se detalla en materiales y métodos. Únicamente en dos de las 50 cepas, no fuimos capaces de amplificar ningún *loci* CRISPR. En total, tres cepas poseen únicamente un *locus* CRISPR1, seis únicamente un *locus* CRISPR3 y las restantes 39 ambos *loci*.

El análisis se llevó a cabo como se mencionó anteriormente, mediante mapeo de las secuencias espaciadoras en busca de protoespaciadores homólogos en los genomas virales. De esta forma se pudo asignar funciones a cada uno de ellos, determinar que espaciadores otorgan resistencia, la proporción de espaciadores funcionales y en última instancia hacer suposiciones acerca de qué *loci* podrían encontrarse activos y cuáles no.

Luego de secuenciados, los espaciadores fueron analizados y se les asignó un número distintivo. En total, de las 50 cepas se obtuvieron un total de 416 secuencias diferentes, 228 espaciadores provenientes de *loci* CRISPR1 y 188 de *loci* CRISPR3. En total, los 45 *loci*

CRISPR3 analizados parecen estar activos, ya que al analizar el contenido de espaciadores de los mismos se encontraron en todos los casos secuencias con un protoespaciador homólogo en algún genoma viral de la colección. De los 42 *loci* CRISPR1, siete no presentaron ningún espaciador con una secuencia homóloga en el genoma de alguno de los fagos de la colección, por lo tanto, a los efectos del alcance de este trabajo, no los consideramos como activos (ver Tabla 2.2). En los anexos pueden encontrarse detallados la secuencia de cada espaciador y la resistencia que éste otorga.

Tabla 2.2: *Locis* CRISPR amplificadas en cada cepa. En verde las cepas en las que no se logró amplificación. La discriminación entre *locus* presentes y activos se estableció mediante análisis del contenido de espaciadores. Aquellos *locus* que no presentaron ninguna secuencia homóloga aun protoespaciador en los genomas de la colección, a los efectos de este trabajo se lo consideró como no activo.

	CRISPR1 presente	CRISPR3 presente	CRISPR1 activo	CRISPR3 activo
Uy01	x	x	x	x
Uy02	x			
Uy03	x	x	x	x
Uy04	x	x	x	x
Uy05	x	x		x
Uy06		x		x
Uy07	x	x	x	x
Uy08	x	x	x	x
Uy09	x	x	x	x
Uy10	x	x	x	x
Uy11	x	x	x	x
Uy12				
Uy13	x	x	x	x
Uy14	x	x		x
Uy15	x	x	x	x
Uy16		x		x
Uy17	x	x	x	x
Uy18	x	x	x	x
Uy19	x	x	x	x
Uy20	x	x	x	x
Uy21	x	x	x	x
Uy22	x	x	x	x
Uy23	x	x	x	x
Uy24	x	x		x
Uy25	x	x	x	x
Uy26		x		x
Uy27	x	x	x	x
Uy28	x	x		x
Uy29	x	x	x	x
Uy30	x	x	x	x
Uy31		x		x
Uy32	x	x		x
Uy33	x	x	x	x
Uy34	x	x	x	x
Uy35	x	x	x	x

Cepas de *Streptococcus thermophilus* de la colección

Uy36		x		x
Uy37	x	x	x	x
Uy38	x	x		x
Uy39		x		x
Uy40	x	x	x	x
Uy41	x	x	x	x
Uy42	x	x	x	x
Uy43	x	x	x	x
Uy44	x	x	x	x
Uy45	x		x	
Uy46	x	x	x	x
Uy47	x	x	x	x
Uy48	x	x	x	x
Uy49				
Uy50	x		x	

Otro dato interesante es la proporción de espaciadores que son efectivamente útiles en el contexto del ambiente industrial uruguayo. De los 416 espaciadores únicamente 87 (aprox. 21%) aportan resistencia contra algún fago de la colección; si lo discriminamos por *loci*, 35 de los 288 espaciadores CRISPR1 poseen homólogos en genomas virales de la colección (12%) y 52 de 188 para el caso de CRISPR3 (27%). En cualquier caso, los valores son muy bajos. Esto puede tener varias explicaciones. Dentro de las que consideramos de mayor sustento es que los espaciadores adquiridos ancestrales son muy estables, es decir que una vez adquiridos se mantienen a pesar de que no exista una presión selectiva, como es el caso de cepas *starters* mantenidas en laboratorios. Por otro lado, estas cepas, como todas las cepas de *S. thermophilus* que circulan en el ambiente industrial uruguayo (sacando las elaboraciones artesanales) son de origen extranjero (en este caso principalmente de origen argentino), por tanto no han estado necesariamente expuestas a lo largo de su historia evolutiva a los fagos aislados por nuestro grupo. Este punto acerca de la diversidad de los bacteriófagos de la colección y qué representatividad poseen se aborda en detalle en el capítulo 4.

Luego de analizar de forma individual cada espaciador, decidimos realizar una comparación entre los *loci* de las diferentes cepas con el fin de utilizar esta información para determinar relaciones evolutivas entre ellas. Las figuras 2.15 y 2.16 muestran cómo se agrupan las cepas de acuerdo al contenido de sus *loci* CRISPR1 y CRISPR3 respectivamente. Como podemos observar, para el caso de CRISPR1 las cepas con características comunes pueden agruparse en seis *clústeres*, mientras que para el caso de CRISPR3 se agrupan en 4. También es válido observar que hay cepas que no tienen nada en común con el resto de la colección, lo que evidencia historias evolutivas muy diferentes. Este es el caso de las

cepas Uy01 y Uy18, cuyos *loci* CRISPR1 como CRISPR3 no presentan similitud con ninguno de los *loci* en la colección; los casos parciales de las cepas Uy24 y Uy34, que poseen *loci* CRISPR1 exclusivos pero sus *loci* CRISPR3 poseen relación con otros de la colección; y el caso de Uy38 que posee un *locus* CRISPR3 único y un *locus* CRISPR1 con similitud a los de otras cepas. Para hacer más fácil la comparación, en la tabla 2.3 se muestran los *clústeres* en los que se ubica cada *loci* por cepa. Es interesante observar que existen *clústeres* relacionados y que se mantienen constantes en muchas cepas, pero también cierta variabilidad, como es el caso de: las cepas que poseen un *locus* CRISPR3 incluido en el *clúster* 1 y según se observa pueden poseer un *locus* CRISPR1 clasificado en los *clústeres* 1, 2 y 4; las cepas que poseen un *locus* CRISPR1 agrupado en el *clúster* 4 que pueden poseer un *locus* CRISPR3 incluido en los *clústeres* 1 y 3; las cepas con un *locus* CRISPR 3 del *clúster* 3 que pueden poseer un *locus* CRISPR1 de los *clústeres* 4 y 6. Estos cruces entre *loci* de diferentes *clústeres* en ambos tipos, muestra que estos sistemas pueden moverse en la población mediante eventos de transferencia horizontal, como ya había sido sugerido para los genes *cas* en el trabajo de Godde y colaboradores (Godde & Bickerton 2006).

En lo referente a la relación entre el contenido de los *loci* CRISPR y los fenotipos de sensibilidad/resistencia, pudimos observar una correlación casi total entre los resultados obtenidos en ambos ensayos. Aun así, se observaron algunas incongruencias, que se pueden observar en la tabla 2.6. Éstas fueron las siguientes: 1- las cepas Uy03, Uy43 y Uy46 poseen espaciadores que deberían conferirles resistencia al fago 73 y sin embargo son sensibles. Debemos destacar que no hay dudas acerca de este fenotipo ya que la cepa Uy03 fue la utilizada para aislar este fago); 2- las cepas Uy04 y Uy09 son sensibles al fago 123 aunque poseen espaciadores que deberían conferirles resistencia; 3- similar es el caso de la cepa Uy23 y el fago 140, con el añadido que esta cepa al igual que en el primer caso fue la utilizada para aislar el fago; 4- las cepas Uy28 y Uy29, son resistentes al fago 93 y no poseen espaciadores que expliquen esta resistencia; 5- por último, la cepa Uy24 muestra ser resistente a los fagos 123 y 140 y no posee espaciadores que expliquen su fenotipo. Estos últimos casos pueden ser causa de que existan otros mecanismos de defensa (como los sistemas R-M) que también pueden otorgar resistencia. Más adelante en este capítulo se aborda ese tema.

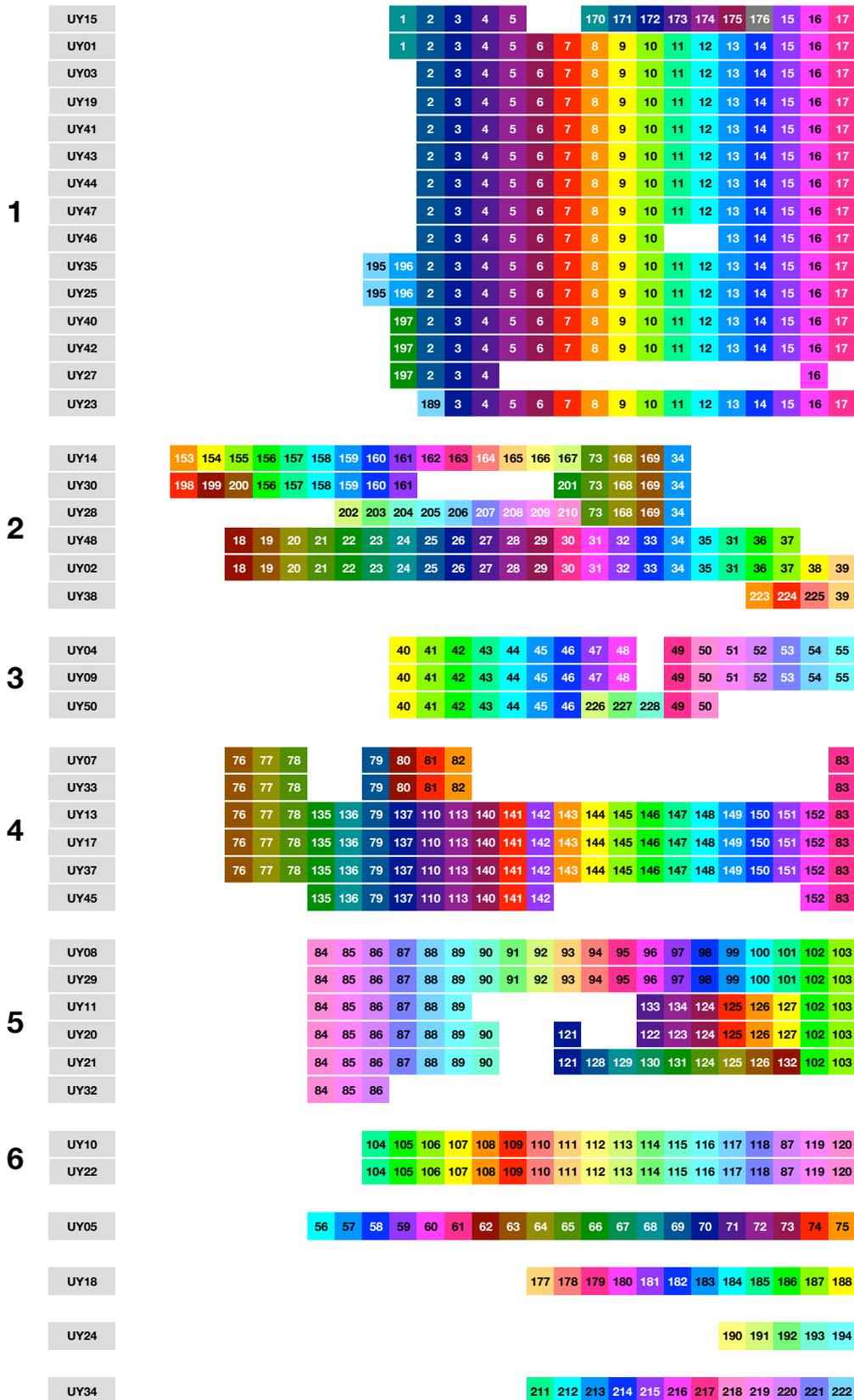


Figura 2.15. Alineamiento y agrupamiento en *clústeres* de *loci* CRISPR1 de las cepas de la colección. Cada espaciador es representado por un cuadrado con su número correspondiente. Los colores sirven como guía para visualizar las regiones donde los *loci* se alinean.

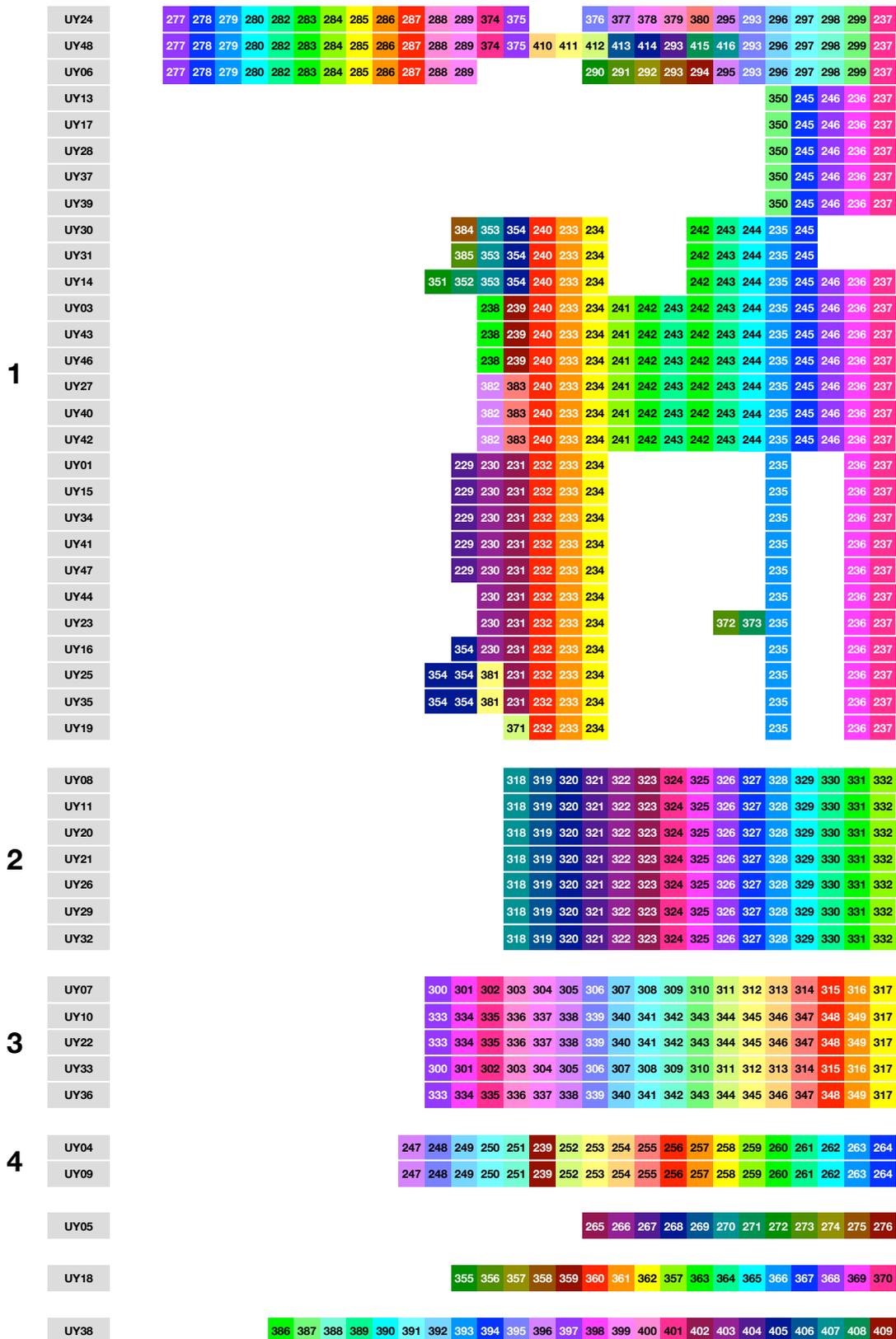


Figura 2.16. Alineamiento y agrupamiento en *clústeres* de *loci* CRISPR3 de las cepas de la colección. Cada espaciador es representado por un cuadrado con su número correspondiente. Los colores sirven como guía para visualizar las regiones donde los *loci* se alinean.

Tabla 2.3: Clústeres en los que se agruparon los *loci* CRISPR de cada cepa. La columna S/R indica el cluster según la matriz de sensibilidad-resistencia y la última columna agrupa todas las cepas que presentan las mismas características. Las celdas marcadas con No hacen referencia a que no pudo amplificarse el *locus*, mientras las marcadas con Si, indican que el *locus* fue secuenciado pero no presenta similitudes con ningún otro de los estudiados.

	Clúster CRISPR1	Clúster CRISPR3	S/R	Clúster global
Uy14	2	1	1	A
Uy30	2	1	1	
Uy48	2	1	1	
Uy27	1	1	1	B
Uy40	1	1	1	
Uy42	1	1	1	
Uy03	1	1	4	C
Uy43	1	1	4	
Uy46	1	1	4	
Uy01	1	1	5	D
Uy15	1	1	5	
Uy19	1	1	5	
Uy25	1	1	5	
Uy35	1	1	5	
Uy41	1	1	5	
Uy44	1	1	5	
Uy47	1	1	5	
Uy16	No	1	1	E
Uy31	No	1	1	
Uy13	4	1	8	F
Uy17	4	1	8	
Uy37	4	1	8	
Uy07	4	3	1	G
Uy33	4	3	1	
Uy10	6	3	6	H
Uy22	6	3	6	
Uy08	5	2	1	
Uy11	5	2	1	
Uy20	5	2	1	
Uy21	5	2	1	
Uy29	5	2	1	
Uy32	5	2	1	
Uy04	3	4	6	I
Uy09	3	4	6	
Uy02	2	No	3	
Uy38	2	Si	1	
Uy28	2	1	8	
Uy23	1	1	7	
Uy06	No	1	7	
Uy34	Si	1	1	
Uy24	Si	1	5	
Uy39	No	1	8	
Uy45	4	No	8	
Uy36	No	3	6	
Uy26	No	2	1	
Uy50	3	No	8	
Uy05	Si	Si	2	

Cepas de *Streptococcus thermophilus* de la colección

Uy18	Si	Si	1
Uy12	No	No	1
Uy49	No	No	3

Por último, una observación importante es que algunos espaciadores aparecen con muy alta frecuencia en la población. De hecho, estos espaciadores son los que ayudan en cierto modo al agrupamiento en los diferentes *clústeres*. En la tabla 2.4 listamos los espaciadores que aparecen con mayor frecuencia y cuál es su aporte a la resistencia de los fagos de la colección.

Tabla 2.4: espaciadores de mayor frecuencia en la colección de cepas

Espaciador	Cepas en la que aparece	Locus	Resistencia
237	24	CRISPR3	-
236	21	CRISPR3	-
233	20	CRISPR3	Fagos 128 y 143
234	20	CRISPR3	-
235	20	CRISPR3	Fagos 73, 94, 107, 115, 123, 128, 140, 142, ST
3	15	CRISPR1	Fago 93
4	15	CRISPR1	-
16	15	CRISPR1	-
242	15	CRISPR3	-
243	15	CRISPR3	Fago 53
2	14	CRISPR1	-
5	14	CRISPR1	-
15	14	CRISPR1	-
17	14	CRISPR1	-
6	13	CRISPR1	-
7	13	CRISPR1	-
8	13	CRISPR1	-
9	13	CRISPR1	-
10	13	CRISPR1	-
13	13	CRISPR1	-

14	13	CRISPR1	-
11	12	CRISPR1	-
12	12	CRISPR1	-
245	12	CRISPR3	Fagos 53, 115, 128, 142
232	11	CRISPR3	Fago 115
231	10	CRISPR3	-
246	10	CRISPR3	-

Como se puede observar la mayoría de los espaciadores no están correlacionados con la resistencia a fagos. Con el fin de determinar su función y evaluar cuál es su importancia real (que justifique su amplia distribución), realizamos comparaciones contra la base de datos de GenBank mediante BLAST. La tabla 2.5 muestra los resultados obtenidos. Como es posible observar, una baja proporción de los espaciadores presentan alguna homología con genomas de fagos presentes en la base de datos. Por tanto, el resultado de estos estudios es que la función de estos espaciadores tan frecuentes permanece desconocida.

Tabla 2.5: Resultado del BLAST realizado a los espaciadores de alta frecuencia que no presentaron protoespaciadores homólogos en los genomas de fagos uruguayos. Se obviaron de los resultados los alineamientos con espaciadores de *loci* CRISPR de otras cepas con genoma anotado en GenBank.

Espaciador	Cepas en <i>Isa</i> que aparece	Locus	Resultado del BLAST
excluyendo237	24	CRISPR3	-
236	21	CRISPR3	-
234	20	CRISPR3	Fago TP-J34
4	15	CRISPR1	-
16	15	CRISPR1	-
242	15	CRISPR3	-
2	14	CRISPR1	-
5	14	CRISPR1	Fago P7133
15	14	CRISPR1	-
17	14	CRISPR1	Fago vB_SthS_VA460
6	13	CRISPR1	Fago vB_SthS_VA460
7	13	CRISPR1	-

8	13	CRISPR1	-
9	13	CRISPR1	-
10	13	CRISPR1	-
13	13	CRISPR1	-
14	13	CRISPR1	Fago Sfi19
11	12	CRISPR1	-
12	12	CRISPR1	-
231	10	CRISPR3	-
246	10	CRISPR3	-

Tabla 2.6: Incongruencias en la matriz sensibilidad resistencia. Las celdas marcadas con una O identifican eventos de sensibilidad en cepas que poseen espaciadores que deberían hacerlas resistentes. Las celdas marcadas con ? muestran eventos de resistencia que no pueden explicarse observando sus espaciadores.

		Bacteriófagos de la colección										
		53	73	93	94	107	115	123	128	140	142	ST
Uy01												
Uy02												
Uy03		O										
Uy04												
Uy05												
Uy06												
Uy07												
Uy08												
Uy09												
Uy10												
Uy11												
Uy12												
Uy13												
Uy14												
Uy15												
Uy16		?										
Uy17												
Uy18												
Uy19												
Uy20												
Uy21												
Uy22												
Uy23												
Uy24												
Uy25												
Uy26												
Uy27												
Uy28			?									
Uy29												
Uy30												
Uy31												
Uy32												
Uy33												
Uy34												
Uy35												
Uy36												
Uy37												
Uy38												
Uy39			?									
Uy40												
Uy41												
Uy42												
Uy43		O										
Uy44												
Uy45												
Uy46		O										
Uy47												
Uy48												
Uy49												
Uy50												

Inmunidad innata: sistemas de restricción-modificación (R-M)

Como se mencionó anteriormente, uno de los objetivos de este trabajo es investigar la presencia de sistemas tipo R-M en las cepas de la colección. La combinación de sistemas de defensa diferentes como CRISPR-Cas y R-M, otorga un efecto sinérgico (Dupuis et al. 2013). Por otro lado, como se observó en la sección anterior, algunos fenotipos resistentes no pudieron ser explicados mediante el análisis de los *loci* CRISPR de las cepas, por lo que estudiar la presencia de este tipo de sistemas se presentó como una oportunidad atractiva.

En el capítulo uno se detalla el funcionamiento de este tipo de sistema inmune. Básicamente consiste en que la bacteria metila su propio ADN de forma que el ADN foráneo que entra en la célula sin metilar es reconocido y degradado por endonucleasas. Para llevar a cabo la búsqueda de sistemas R-M, seleccionamos a las 21 cepas que mostraron ser resistentes a todos los virus, ya que eran las que presentaban mejores perspectivas de poseerlos (ver tabla 2.1). La metodología empleada se detalla en materiales y métodos. La figura 2.17 esquematiza el proceso.

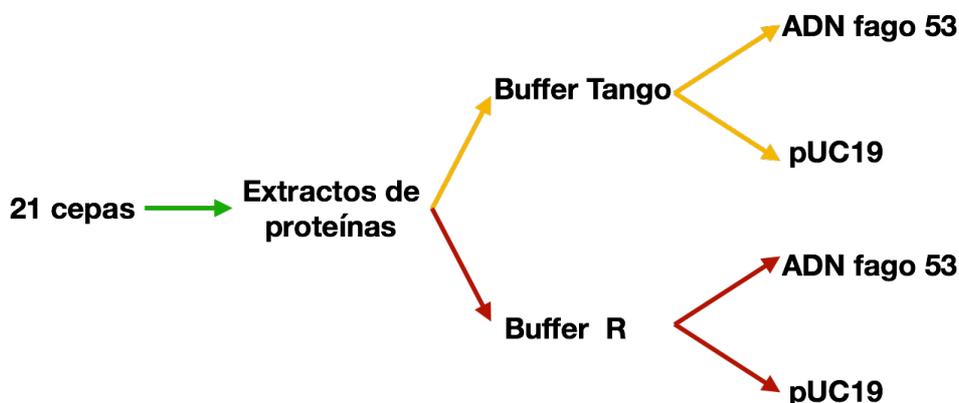


Figura 2.17. Esquema del procedimiento utilizado en la búsqueda sistemas de defensa tipo R-M en las cepas de la colección

Los resultados de los ensayos fueron en todos los casos negativos, no pudiéndose observar degradación del ADN tratado con los extractos de proteínas en ninguna de las condiciones ensayadas. Esto puede deberse a que efectivamente ninguna de las cepas posea un sistema de restricción modificación, o que el método utilizado no haya funcionado. En cualquier caso, queda planteado como perspectiva el seguir ahondando en este aspecto.

Una particularidad que observamos al analizar los genomas virales (tema que se desarrollará en el capítulo siguiente) fue la presencia de una ADN adenina metiltransferasa en el genoma del fago 73 (Figura 2.18). Ha sido reportado que algunos fagos de bacterias ácido lácticas

que incorporan a su ADN genes bacterianos capaces de expresar estas enzimas, logran evadir los sistemas de defensa R-M (Murphy et al. 2014; McDonnell et al. 2017). Si esta enzima es funcional en este bacteriófago, significaría que el virus es capaz de evadir al sistema inmune primario bacteriano.

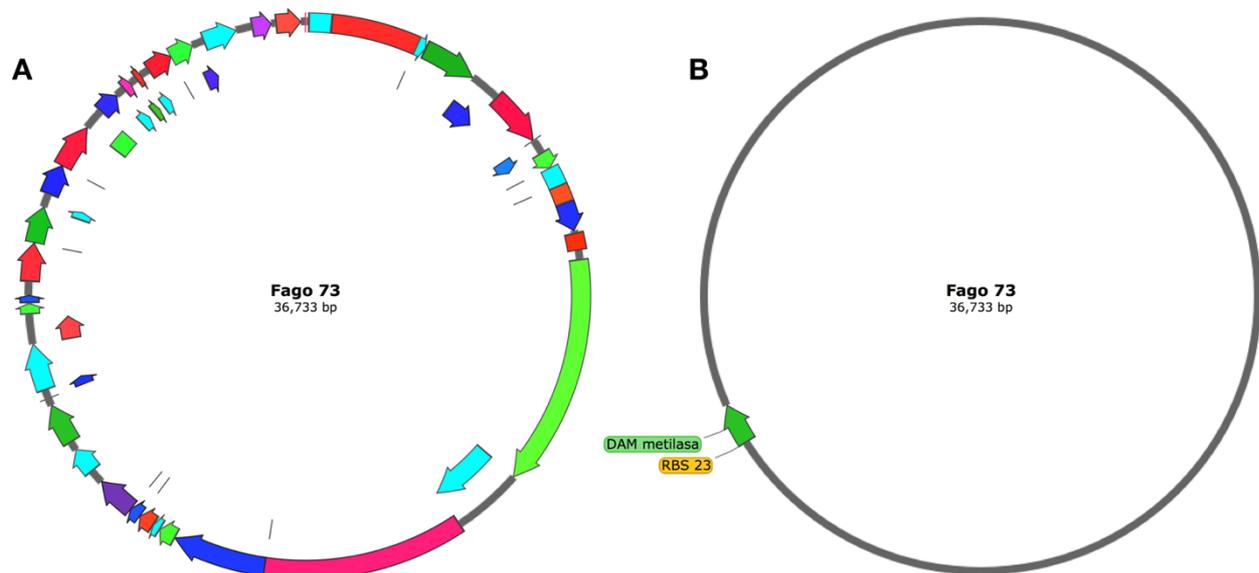


Figura 2.18. Mapa del genoma del fago 73. En la figura A se muestra la disposición de todos los genes del bacteriófago. En la figura B se observa la posición del ORF 23 que corresponde al gen codificante para la enzima metiltransferasa.

Con el fin de probar su actividad, se decidió comprobar si el ADN del bacteriófago se encontraba metilado luego de un proceso de infección. El procedimiento utilizado fue el siguiente: se realizó una propagación en medio líquido del bacteriófago utilizando la cepa Uy03 y posteriormente se realizó una extracción de ADN del lisado. Este ADN fue sometido a una digestión con la enzima DpnI y se observó si el mismo era cortado (ver sección materiales y métodos). La enzima DpnI corta de forma específica solamente si el ADN se encuentra metilado. Analizando la secuencia del genoma del fago 73 encontramos que existen 26 sitios de corte para la enzima (Figura 2.19), y mediante simulación *in silico* pudimos determinar el patrón de banda esperado si la digestión tenía lugar.

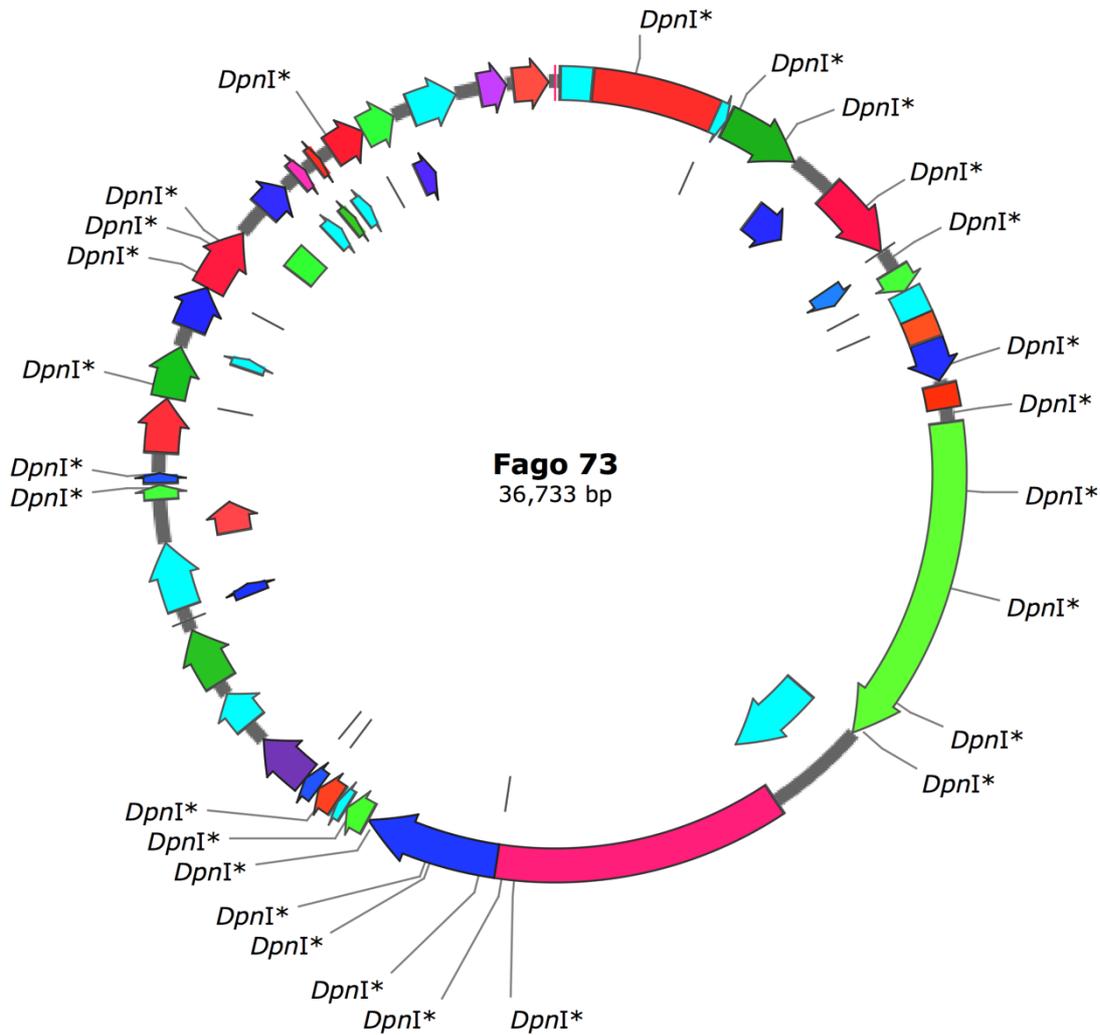


Figura 2.19. Mapa del genoma del fago 73 ubicando los 26 sitios de corte para la enzima DpnI

En la figura 2.20 puede observarse la simulación *in silico* y la electroforesis en gel de agarosa para la digestión real. Los resultados muestran que efectivamente ocurrió digestión y por lo tanto la enzima es capaz de proteger al bacteriófago. Este resultado es significativo, ya que demuestra que en el ambiente industrial uruguayo circulan bacteriófagos capaces de evadir este tipo de respuesta inmune, por lo que hacen que su estudio no resulte tan prometedor en el marco de los objetivos de este trabajo. Desde este punto en adelante el trabajo se enfoca en el estudio y mejoramiento basado únicamente en el sistema de defensa CRISPR-Cas

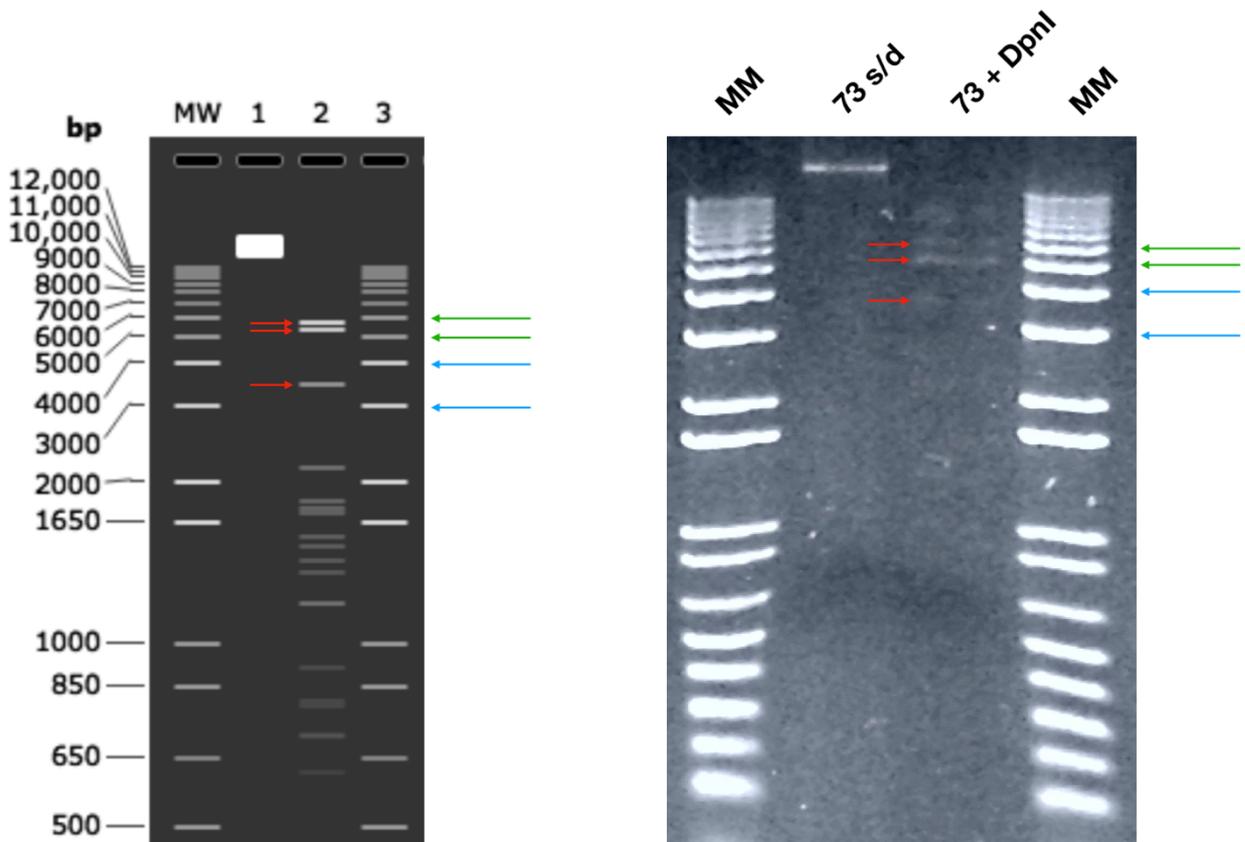


Figura 2.20. Digestión de ADN genómico del fago 73 con la enzima DpnI. A la izquierda puede observarse la simulación *in silico* de la digestión y a la derecha la digestión real. En ambos casos se muestra un carril sin digerir y otro digerido, flanqueados por dos carriles con el marcador de peso molecular 1kb+ (Invitrogen, CA, E.E.U.U.). Las flechas sirven de guía para visualizar mejor las bandas obtenidas: en rojo se señalan las bandas que fue posible visualizar en el gel de agarosa y sus homólogas en la simulación. En azul las bandas de 3000 y 4000 pb del marcador y en verde las bandas de 5000 y 6000 pb del marcador. Los resultados muestran que el patrón de bandas es coincidente en lo que puede observarse por la calidad de la imagen obtenida.

3

Análisis de genomas virales

Resumen

En este capítulo nos centraremos en el análisis de las características generales a nivel genómico de los bacteriófagos de la colección. Como se mencionó anteriormente, los bacteriófagos son el insumo principal para la generación de derivados resistentes y por tanto resulta de interés conocer sus características.

Antecedentes

Aproximadamente 150 aislamientos virales fueron realizados por nuestro grupo a partir de muestras del ambiente industrial uruguayo. La detección de los mismos fue llevada a cabo mediante PCR en tiempo real, y su aislamiento fue realizado utilizando cepas de la colección (Achigar 2014). Estos fagos fueron caracterizados en el mismo proceso de detección según el tipo de mecanismo de empaquetamiento de ADN, dividiéndolos en fagos tipo *cos* y fagos tipo *Pac*. Posteriormente, se realizaron perfiles de restricción que, junto con los perfiles de ataque observados, llevaron a realizar una primera clasificación de la colección en catorce fagos de tipo *cos* y un único fago tipo *Pac* (Achigar 2014). En este trabajo realizamos la secuenciación completa de los genomas de estos fagos y, posterior al análisis, el grupo se reorganizó en 10 fagos de tipo *cos* y un fago tipo *Pac*, cuyas características se presentan a continuación. Si bien recientemente han sido identificados dos nuevos grupos minoritarios de fagos de *S. thermophilus* (el grupo 5093 y 987) (McDonnell et al. 2017), al momento de generar la colección presentada en este trabajo no se encontraban aún reportados, por tanto ningún miembro de éstos aparece representado.

Materiales y métodos

Microscopía electrónica

Los fagos fueron observados mediante microscopía electrónica de transmisión como fue descrito anteriormente (Fortier & Moineau 2007). Brevemente, 1,5mL de lisado fue centrifugado a 23.500g por una hora a 4°C y el pellet lavado dos veces con NH₄Ac 0,1M pH 7,0. La preparación fágica resultante fue teñida con ácido fosfotúngstico 2% pH 7,0 y observada en un microscopio JEOL 1230 en la Université Laval.

Secuenciación de genomas virales y análisis

El ADN viral fue extraído según se explica en el capítulo 2. Las librerías de secuenciación fueron preparadas con el kit Nextera XT DNA Library Preparation Kit y secuenciadas en un MiSeq system (Illumina). El ensamblaje de los *reads* fue realizado utilizando Ray Assembler. Para completar la secuenciación de los *gaps* en los genomas se utilizó un ABI Prism 3100 Genetic Analyzer en la *Plateforme de séquençage et de génotypage des génomes du Centre Hospitalier de l'Université Laval*. El ensamblaje de los contigs se realizó con PreGap y Gap4 de la suite Staden Package (Bonfield et al. 1995). El análisis fue llevado a cabo utilizando el software *SnapGene* (GSL Biotech LLC, IL, E.E.U.U.) para predecir los marcos abiertos de lectura (ORFs), realizar las anotaciones y determinar las características de las proteínas. Un ORF fue considerado tal solamente cuando el codón de inicio era ATG, TTG o GTG y contaba con al menos 30 aminoácidos. Para comparar y predecir la función de las secuencias de aminoácidos se utilizó BLASTp (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Los sitios de unión al ribosoma (RBS) debían ser similares al de *Streptococcus thermophilus* (5'AAAGGAGGTGA3'). El criterio que se estableció para aceptar un posible RBS fue que la secuencia se encontrara a no más de 10 bases aguas arriba del codón de inicio de la transcripción, y que debía poseer un mínimo de cinco bases similares a la secuencia consenso.

Resultados

Morfología

Todos los fagos de *S. thermophilus* identificados hasta el momento pertenecen a la familia *Siphoviridae*. Estos virus se caracterizan por poseer una cápside icosaédrica y una cola larga no contráctil, pudiendo o no poseer espículas o fibras. Algunos bacteriófagos de la colección fueron seleccionados para ser analizados mediante microscopía electrónica. La figura 3.1 muestra micrografías de los fagos tipo *cos* 53, 73 y 128 y del fago tipo *Pac* 93. Puede observarse que todos poseen la morfología esperada típica de este tipo de virus, e incluso es posible observar que los fagos 53 y 93 presentan una larga fibra al final de la cola. En la tabla 3.1 se presentan las dimensiones de los fagos tipo *Cos*.

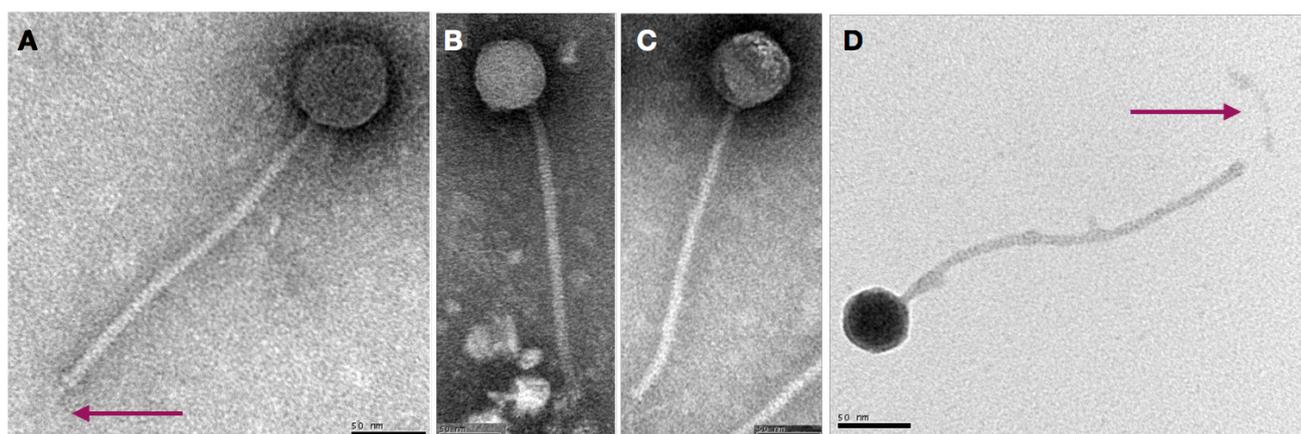


Figura 3.1. Micrografías de los fagos 53 (A), 73 (B), 128 (C) y 93 (D). Las flechas señalan la presencia de fibras en la parte final de la cola de los fagos 53 y 93

Tabla 3.1. Dimensiones de las partículas virales

Fago	Dimensiones (nm)		
	Cápside	Cola (largo)	Cola (ancho)
53	65,9 +/- 2,7	249,4 +/- 6,1	11,0 +/- 0,6
73	65,5 +/- 1,9	244,9 +/- 9,3	11,2 +/- 0,5
128	67,2 +/- 2,9	251,7 +/- 4,7	10,9 +/- 0,6

Características genómicas

Los genomas de los fagos fueron secuenciados y analizados según se detalla en materiales y métodos. La tabla 3.2 resume algunas características generales. Los tamaños de los genomas se encuentran entre 34.239 pares de bases (fago 53) y 42.017 pares de bases (fago 115). Este último presenta una región duplicada de 6.044 pares de bases, la cuál comprende parte del receptor viral formándose un ORF (marco abierto de lectura) más pequeño y otros seis ORFs completos entre los cuales se encuentran tres proteínas hipotéticas, una proteína estructural y dos enzimas involucradas en la lisis del hospedero. La cantidad de ORFs se encuentra entre 40 y 56, y el contenido GC se ubica en el entorno de 39%, lo cuál es esperable ya que coincide con el porcentaje del genoma de *Streptococcus thermophilus*.

Tabla 3.2. Características generales de los fagos de la colección

Fago	Hospedero (cepa tipo)	Tamaño (pb)	ORFs	Codificante (%)	%GC	Grupo	Replicación	Regulador transcripcional	Atributos especiales
53	Uy01	34.239	41	91,3	39	Cos	Tipo Sfi21 (I)	Ltrb	Endonucleasa HNH
73	Uy03	36.733	46	93,7	39	Cos	Tipo 7201 (II)	Ltrb	ADN adenina metiltransferasa / Proteína tipo ERF / Endonucleasa HNH
93	Uy05	40.563	56	90	39	Pac	Tipo 7201 (II)	ArpU	Proteína tipo ERF / Proteína tipo YopX / Endonucleasa HNH
94	Uy45	37.611	46	92,5	38	Cos	Tipo Sfi21 (I)	Ltrb	ADN citosina metiltransferasa / Endonucleasa HNH
107	Uy09	36.938	50	93,2	39	Cos	Tipo Sfi21 (I)	Ltrb	Recombinasa tipo RecT / Endonucleasa Rus A / Endonucleasa HNH
115	Uy05	42.017	56	91,1	39	Cos	Tipo Sfi21 (I)	Ltrb	Endonucleasa HNH
123	Uy28	36.129	47	92	39	Cos	Tipo 7201 (II)	Ltrb	Recombinasa tipo RecT / Endonucleasa Rus A / Endonucleasa HNH
128	Uy02	34.539	40	90,7	39	Cos	Tipo Sfi21 (I)	Ltrb	Endonucleasa HNH
140	Uy23	36.064	43	91,4	38	Cos	Tipo 7201 (II)	Ltrb	Endonucleasa HNH (X2)
142	Uy02	34.639	42	91,7	39	Cos	Tipo Sfi21 (I)	Ltrb	Endonucleasa HNH
ST	Uy45	35.581	44	93,9	39	Cos	Tipo Sfi21 (I)	Ltrb	ADN citosina metiltransferasa (x2) / Endonucleasa HNH

Otra característica que es transversal a los diferentes tipos de fagos, es la existencia de dos clases de módulos de replicación, conocidos como tipo Sfi21 (o grupo 1) y tipo 7201 (o grupo 2) por los nombres de los fagos tipo que los poseen (Stanley et al. 2000) (Figura 3.2).

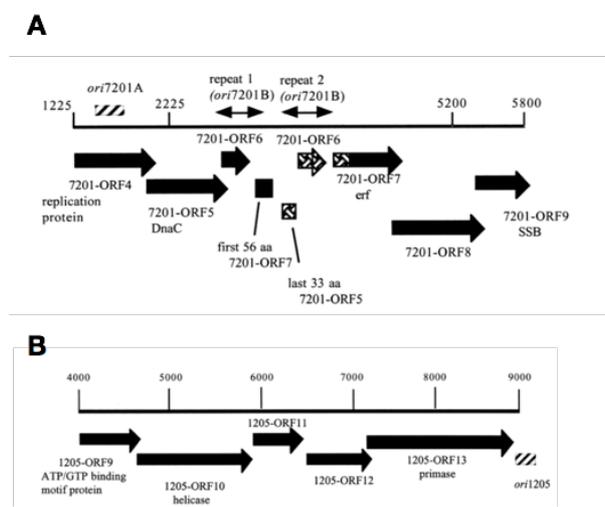


Figura 3.2. A: organización del módulo de replicación tipo 7201. B: organización del módulo de replicación tipo Sfi21 (Stanley et al. 2000)

Todos los fagos secuenciados hasta el momento (con la excepción de ALQ132 (Guglielmotti et al. 2009) y 4761 (McDonnell et al. 2017)) poseen un regulador transcripcional tardío, que pueden clasificarse en dos tipos, tipo Ltrb y tipo ArpU. Los diez fagos tipo *cos* mostraron poseer secuencias codificantes para reguladores tipo Ltrb mientras que el fago 93 tipo *Pac* presentó una secuencia codificante para un regulador tipo ArpU. Esto es concordante con observaciones previas que proponen que el tipo de regulador está relacionado con el mecanismo de empaquetamiento del ADN por parte del fago. De manera similar, la presencia de endonucleasas tipo HNH en la región aguas abajo del regulador transcripcional se encuentra relacionada con la forma de empaquetar el genoma por parte de los virus. Este tipo de enzimas parece ser especialmente importante para el procesamiento post replicación del genoma de los fagos tipo *cos* (McDonnell et al. 2017), por tanto es coherente la observación de un ORF codificante para este tipo de proteína en la misma posición en todos los genomas de este tipo de la colección (ver tablas 3.3 a 3.13). También es posible observar la presencia de otros atributos especiales codificados en los genomas de estos fagos. Por lo general estas secuencias provienen del genoma del hospedero y se localizan en la región aguas abajo del módulo de replicación que actúa como un *hotspot* de eventos de recombinación (McDonnell et al. 2017). El más sobresaliente de éstos es la presencia de metiltransferasas, las cuales como se mostró en el capítulo anterior, ayudan al virus a escapar del sistema de defensa R-M. Tres de los virus de la colección poseen

metiltransferasas en su genoma; en el caso del fago 73 se demostró que es funcional y sería capaz de proteger al virus de la degradación de su genoma por parte del hospedero. También encontramos en cuatro genomas enzimas involucradas en proceso de recombinación, aunque no podemos identificar a ciencia cierta cual es la ventaja de poseer este tipo de atributos. Por último el ORF 52 del fago 93 codifica una proteína tipo YopX, común en fagos de otras bacterias Gram positivas como *Enterococcus faecium* (Yasmin et al. 2010). Estas proteínas son de origen plasmídico y han demostrado ser factores de virulencia; en ciertas bacterias como *Yersinia pestis* estas proteínas son transportadas al citosol de las células del hospedero mediante sistemas de secreción de tipo III (Yasmin et al. 2010). La presencia de este ORF podría ser evidencia de un pasado promiscuo de este fago, pudiendo algún antepasado haber infectado a otros huéspedes y adquiriendo material genético en el proceso.

La organización genómica general, y la conservación de aminoácidos en grupos definidos de fagos, es comparable con la de los ejemplos publicados anteriormente (Guglielmotti et al. 2009). Esto evidencia que la estructura de organización modular se encuentra ampliamente conservada en estos virus. En la figura 3.3 se presentan los mapas de los diez genomas virales, señalando los diferentes ORFs y módulos, y los grados de homología a nivel de aminoácidos que comparten los diferentes virus.

Por último, se llevó a cabo el análisis de todos los ORF de cada fago, comparándolos con la base de datos de GenBank. Las tablas 3.3 a la 3.13 contienen los resultados obtenidos de este análisis, además de otras características como la posición, el contenido GC el tamaño y el sitio de unión al ribosoma (RBS) que precede a cada uno de los genes. Cabe aclarar que en algunos casos, el resultado del BLAST dio como resultado ORFs de los fagos 53, 73 y 128. Esto se debe a que éstos han sido ingresados a la base de datos como resultado de publicaciones recientes de nuestro grupo (Achigar et al. 2017).

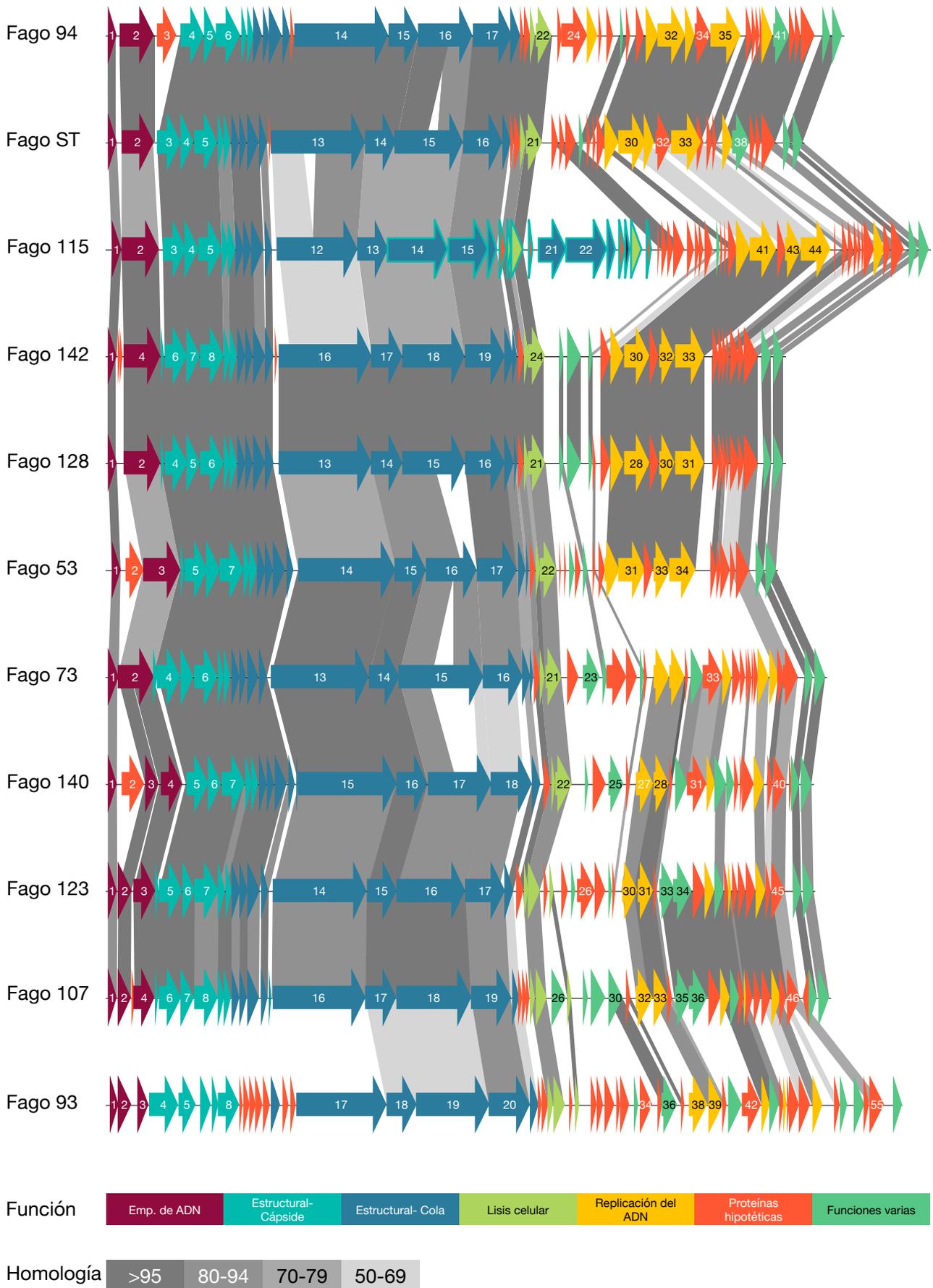


Figura 3.3. Representación gráfica y comparación de los genomas de fagos de la colección. Las flechas representan diferentes ORFs. Los colores de las flechas hacen referencia a su función o módulo al que pertenecen. Las sombras en gris señalan regiones de homología a nivel de aminoácidos entre ORFs

Tabla 3.3. Características generales del fago 53

Fago 53 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	296	757	40	153	17,42	AACGGAGAGGAGTAATGATGAGTG	Terminasa, subunidad menor	P7154	149/153(97%)	ARU13711.1
2	1.030	1.938	35	302	35,9	AAAGGGCAAAAATG		P7151	301/302(99%)	ARU13621.1
3	1.919	3.793	40	624	71,9	AAAGGAGGTGCTTG	Terminasa, subunidad mayor	P7151	616/620(99%)	ARU13622.1
4	3.797	3.976	37	59	6,6	AGAGGAGTATTAATATATG	Proteína de unión cápside-cola	DT1	59/59(100%)	NP_049393.1
5	3.994	5.154	41	386	42,7	AAAGGAGGTGATAACAATTG	Proteína Portal	DT1	383/386(99%)	NP_049394.1
6	5.141	5.809	38	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de andamiaje	phiAbc2	222/222(100%)	YP_003347414.1
7	5.824	7.017	38	397	44,1	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	P5651	396/397(99%)	ARU13408.1
8	7.032	7.346	44	104	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
9	7.346	7.696	40	116	13,3	GAAAGAGGTGACTAATG	Proteína de cola	DT1	116/116(100%)	NP_049398.1
10	7.703	8.125	43	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	73	140/140(100%)	ALJ99545.1
11	8.130	8.501	32	123	14,1	AAGGGAGGGGAGTAATTAAGTG	Proteína de cola	73	122/123(99%)	ALJ99546.1
12	8.520	9.128	40	202	21,9	AAAGGAGAAAATATATATG	Cola, subunidad mayor	73	202/202(100%)	ALJ99547.1
13	9.202	9.555	37	117	13,5	AAAGGAGTAAAGACCACAATG	Proteína de cola	73	117/117(100%)	ALJ99548.1
14	9.774	14.744	42	1656	182,8	AAAGGAGGGAATATAACATG	Proteína de cola	73	1631/1656(98%)	ALJ99549.1
15	14.741	16.300	40	519	58,5	TTAGGAGGTCAAATTATTTG	Proteína de cola	phiAbc2	517/519(99%)	YP_003347424.1
16	16.300	18.885	39	861	97,8	GAAGGAGCGCTTTGTTTAATG	Receptor viral	MD1	780/861(91%)	AAK83243.1
17	18.886	20.934	42	682	75	GTAGGAGGTTTTTAATTG	Proteína de cola	P9901	664/682(97%)	ARU14710.1
18	20.960	21.370	34	136	15,7	AAAGGAATAATTATG	-	ALQ13.2	135/136(99%)	YP_003344866.1
19	21.390	21.536	30	48	5,6	AAAGGATAAAAAGATATG	-	DT1	48/48(100%)	NP_049410.1
20	21.554	21.877	36	107	12,3	ATAGGAGGGATGTGTTATG	-	P7573	103/107(96%)	ARU13865.1
21	21.874	22.116	37	80	8,9	AGAGGATAATAATAAAATG	-	DT1	78/80(98%)	NP_049412.1
22	22.118	22.963	41	281	31,3	AAAGGAGAAAATAAAATG	Lisina	phiAbc2	257/281(91%)	YP_003347431.1
23	23.057	23.179	35	40	4,7	GAAGCCTCAGCATTATG	-	P7574	36/37(97%)	ARU13916.1
24	23.348	23.445	32	31	3,7	ATAAGTGGTAATATAATTG	-	P7132	29/30(97%)	ARU13510.1
25	23.650	23.877	38	75	8,6	AAAGGAGATAACCTATG	Represor tipo Cro	P7571	74/75(99%)	ARU13783.1

26	23.895	24.185	36	96	11,2	AAAGGAACAATATG	-	P7133	90/94(96%)	ARU13554.1
27	24.328	24.531	35	67	7,6	AGAGGAGGAACAAAAATG	Regulador tipo Cro	9872	67/67(100%)	YP_009289378.1
28	24.783	24.917	36	44	5,2	AAAGGAATTTAAAAAATG	-	P5641	43/44(98%)	ARU13382.1
29	25.119	25.445	34	108	12,8	AAAGTATCAACTATG	-	DT1	104/104(100%)	NP_049419.1
30	25.449	26.150	43	233	26,4	AAAGGAAGAAATAACGGATG	-	DT1	233/233(100%)	NP_049420.1
31	26.125	27.456	40	443	50,4	AAATTTGGTGATTTAGATG	Helicasa	DT1	439/443(99%)	NP_049421.1
32	27.463	27.918	36	151	17,3	TATGGAGATAAAAAACTATG	-	DT1	150/151(99%)	NP_049422.1
33	27.921	28.736	40	271	30,5	ACCTCCGTTCTAATTATG	Proteína de replicación	DT1	271/271(100%)	NP_049423.1
34	28.717	30.240	36	507	59,2	AAATAAGGAGGATTG	Primasa	P7632	500/505(99%)	ARU14127.1
35	30.777	31.100	38	107	12,2	AAAGGAGATGTATG	-	P7152	106/107(99%)	ARU13703.1
36	31.081	31.329	37	82	9,7	CTATGAGGATAGTTGATG	-	DT1	82/82(100%)	NP_049427.1
37	31.314	31.790	39	158	18,5	GAAAGAGATGGTAGAACTATG	Proteína de unión al ADN	2972	103/109(94%)	YP_238525.1
38	31.759	32.067	41	102	11,2	AAAGGAAAGATGGTAAATG	-	P9854	89/102(87%)	ARU14689.1
39	32.064	32.771	37	235	27,7	AAAGGAAGAGGGCAATG	-	phiAbc2	235/235(100%)	YP_003347454.1
40	33.105	33.524	38	139	16,5	AAATTATTATACCATG	-	DT1	132/132(100%)	NP_049433.1
41	33.597	34.145	46	182	21,2	TCAGAAGGACACAGTAGTG	Endonucleasa HNH	DT1	180/185(97%)	NP_049434.1

Tabla 3.4. Características generales del fago 73

Fago 73 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	87	459	39	152	17	AACGGAGAGGAGTAATGATGAGTG	Terminasa, subunidad menor	Sfi21	145/152(95%)	NP_049966.1
2	567	2.438	41	623	71,5	AAAGGAGCAACAGTG	Terminasa, subunidad Mayor	P9902	621/623(99%)	ARU14742.1
3	2.442	2.621	37	59	6,6	ATTAGAGGAGTATTAATATATG	Proteína de unión cápside-cola	phiAbc2	59/59(100%)	YP_003347412.1
4	2.639	3.799	41	368	42,7	AAAGGAGGTGATAACAAATG	Proteína portal	phiAbc2	384/386(99%)	YP_003347413.1
5	3.786	4.454	38	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de anamiae	DT1	222/222(100%)	NP_049395.1
6	4.469	5.662	38	397	44,2	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	P7132	396/397(99%)	ARU13493.1
7	5.677	5.991	44	104	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
8	5.991	6.341	40	116	13,5	GAAAGAGGTGACTAATG	Proteína de unión cápside-cola	phiAbc2	115/116(99%)	YP_003347417.1
9	6.348	6.770	43	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	P7633	139/140(99%)	ARU14147.1
10	6.775	7.146	32	123	14	AAGGGAGGGGAGTGATTAAGTG	Proteína de cola	P7132	121/123(98%)	ARU13497.1
11	7.165	7.773	40	202	21,8	AAAGGAGAAAATATATATG	Cola, subunidad mayor	P7132	201/202(99%)	ARU13498.1
12	7.847	8.200	37	117	13,5	AAAGGAGTAAAGACCACAATG	Proteína de cola	P5641	115/117(98%)	ARU13361.1
13	8.419	13.389	41	1656	183,2	AAAGGAGGGAATATAACATG	Proteína de cola	P9902	1647/1656(99%)	ARU14754.1
14	13.386	14.945	40	519	58,4	TTAGGAGGTCAAATTATTTG	Proteína de cola	phiAbc2	514/519(99%)	YP_003347424.1
15	14.924	19.234	41	1436	159,6	CCAACAATTGAAATTTTCATG	Receptor viral	P9901	404/1429(98%)	ARU14709.1
16	19.235	21.241	42	668	73	GTAGGAGGTTTTTAATTG	Cola, subunidad menor	P7571	629/668(94%)	ARU13772.1
17	21.257	21.604	33	115	13,3	AAAGAAGGAAAATTCATG		TP-778L	114/115(99%)	YP_008772103.1
18	21.624	21.770	34	48	5,4	GAAAGAGGAAAAAGATATG		P4761	47/48(98%)	ARU13316.1
19	21.788	22.111	35	107	12,5	ATAGGAGGGATGTGTTATG		ALQ13.2	106/107(99%)	YP_003344868.1
20	22.119	22.361	36	80	8,9	TGAGAGGATGAAGAATAAATG	Holina	ALQ13.2	79/80(99%)	YP_003344869.1
21	22.363	23.208	42	281	31,2	AAAGGAGAAATAAAATG	Lisina	P9902	276/281(98%)	ARU14762.1
22	23.526	24.077	38	183	21,4	GGGAGAGGTAAACAAAATG		DT1	176/183(96%)	NP_049416.1
23	24.284	25.156	32	290	33,5	AGAGAGGGATTTAATG	ADN adenina metiltransferasa	<i>Streptococcus thermophilus</i>	285/290(98%)	CAB46541.1
24	25.333	25.536	38	67	7,7	GAAGGAGGAACAAAATG	Regulador tipo Cro	CHPC1151	64/67(96%)	APC45891.1
25	25.526	26.521	37	331	38,7	AAATCGTCTGATTTGTATG		P9901	312/314(99%)	ARU14719.1

26	26.514	27.053	35	179	20,7	GAAGGAGAAATCATCAATG		P9901	177/177(100%)	ARU14720.1
27	27.180	27.389	28	69	8,2	AAAGGAGAAACGAATG	Represor tipo Cro	ALQ13.2	68/69(99%)	YP_003344877.1
28	27.422	27.565	31	47	5,5	TAGAGAGGAATCAAAAATG		P0092	46/47(98%)	ARU13031.1
29	27.877	28.686	33	269	31,5	AAAGAGAGGGATAAGATTAATG		P4761	265/269(99%)	ARU13332.1
30	28.699	29.481	39	260	30,5	CTAAGAGGTTCTTTATATG	Proteína tipo DnaC	CHPC926	236/260(91%)	APC45852.1
31	29.478	29.660	36	60	7,3	CAAGAGGATGATGTTATG		phiAbc2	60/60(100%)	YP_003347441.1
32	29.788	30.444	37	218	25	AAAAGAGGATATGACATG	Proteína tipo ERF	phiAbc2	217/218(99%)	YP_003347442.1
33	30.447	31.418	39	323	37,7	AACGGAAGGGATAAATATG		phiAbc2	323/323(100%)	YP_003347443.1
34	31.415	31.867	42	150	17	AAAGGAGAAAACAAATG	Proteína de unión a ssADN	phiAbc2	149/150(99%)	YP_003347444.1
35	31.877	32.338	38	153	18	TAAGGTGAAACTATG		phiAbc2	153/153(100%)	YP_003347445.1
36	32.336	32.571	37	78	9,2	CAAGGAGTTGGAATG		P5641	73/78(94%)	ARU13391.1
37	32.562	32.735	41	57	6,5	GAAAGAGATGATAGAACTATG		858	53/57(93%)	YP_001686836.1
38	32.732	32.887	28	51	6,3	GTAGGAGATTAGTAGAGTTATG		P9901	50/51(98%)	ARU14732.1
39	32.898	33.011	37	37	4	ATAGATGGCAAGATATG		TP-J34	20/25(80%)	YP_007392267.1
40	33.016	33.228	34	70	8	GAAGGGATAGAATAATG		DT1	69/70(99%)	NP_049429.1
41	33.215	33.763	36	182	33,7	CTAGGAGAAGAAAATG		TP-J34	131/181(72%)	YP_007392269.1
42	33.764	34.270	38	168	19,1	GACAGAGGTGGAATAGATG	Proteína de unión a ADN	P5641	162/168(96%)	ARU13395.1
43	34.245	34.547	39	100	11,1	AAAGGAATAATGATTGATG		DT1	99/100(99%)	NP_049431.1
44	34.544	35.251	38	235	27,5	AAAGGAAGAGGGCAATG		P3681	217/235(92%)	ARU13247.1
45	35.611	36.009	39	132	15,5	AAAGGAAAGACAATTTATG		DT1	132/132(100%)	NP_049433.1
46	36.121	36.639	47	172	20	AGAGGAGGGAAGCCAATG	Endonucleasa HNH	phiAbc2	172/172(100%)	YP_003347457.1

Tabla 3.5. Características generales del fago 93

Fago 93 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	183	635	39	150	16,8	TTTGCGGATGAATACTTG	Terminasa, subunidad menor	P7952	146/150(97%)	ARU14233.1
2	622	1.293	43	223	25,3	AAAGGAGCTGTAAGCGATG	Terminasa, subunidad mayor	2972	220/223(99%)	YP_238486.1
3	1.582	2.169	43	195	22,6	CTTTGCGATAACAAATAATG	Terminasa, subunidad mayor	P7571	189/194(97%)	ARU13755.1
4	2.175	3.683	41	502	57,5	AATTAGGAGGAATG	Proteína portal	ALQ132	471/502(94%)	YP_003344849.1
5	3.680	4.573	42	297	34,3	TGAGAGGGTTTATGAATG	Proteína de cápside	TP-778L	292/297(98%)	YP_008772087.1
6	4.765	5.346	43	193	21,2	ATAGGAGAACAAAATG	Proteína de andamiaje	TP-778L	185/193(96%)	YP_008772088.1
7	5.366	5.725	40	119	12,7	AAAGGAAATTTTAAATG	Cápside, subunidad mayor	Sfi11	113/119(95%)	NP_056679.1
8	5.744	6.790	41	384	37,5	AAAGGAGGAATATTTAAACATG	Proteína de cápside	TP-778L	345/348(99%)	YP_008772090.1
9	6.802	6.963	38	53	6	TTAAGAGGTACTGATATG		TP-778L	52/53(98%)	YP_008772091.1
10	6.978	7.316	37	112	13	TAGTGAGGTATGGCGTGATG		TP-778L	111/112(99%)	YP_008772092.1
11	713	7.627	42	104	11,5	GGGTGAGGTGCTATTTCTATG		P7571	100/104(96%)	ARU13763.1
12	7.629	7.973	40	114	12,9	CAAGGTGATGAAATAACATG		P7952	108/114(95%)	ARU14244.1
13	7.970	8.356	37	128	14,8	AAATGGGATGAAACATG		ALQ132	124/128(97%)	YP_003344858.1
14	8.369	8.878	41	169	18,5	ACAGGAGGAAAAAATG	Proteína de cola	ALQ132	164/169(97%)	YP_003344859.1
15	8.953	9.306	44	117	13,2	ATAGGAGTTAAAAACAATG		ALQ132	116/117(99%)	YP_003344860.1
16	9.357	9.674	42	105	12,7	CGAGGAATTAATCACTAGTGCTATG		O1205	98/105(93%)	NP_695117.1
17	9.664	14.301	44	1545	156,7	AAAGAGGGTTGCTAGATG	Proteína de cola	ALQ132	1458/1534(95%)	YP_003344862.1
18	14.301	15.836	40	511	58	AGAGGAATTAATAATG	Proteína de cola	P9853	484/511(95%)	ARU14618.1
19	15.836	19.540	41	1243	138	AGGAGAGGTATCTATAATG	Receptor viral	ALQ132	993/1237(80%)	YP_003344864.1
20	19.540	21.603	41	687	76,4	GTAGGAGGTGCATAATG	Proteína estructural	P7132	577/687(84%)	ARU13504.1
21	21.625	22.041	38	138	15,7	AAGAAAGGAAAAATCTATG		P7954	123/138(89%)	ARU14344.1
22	22.048	22.194	30	48	5,5	AAAGGATAAAAAAATATG		P3684	48/48(100%)	ARU13269.1
23	22.210	22.533	36	107	12,5	AAAGGAGGTATGTCGTG		P3681	99/107(93%)	ARU13227.1
24	22.542	22.784	39	80	8,8	TTGAGAGGTTAAATAACAATATG	Holina	P3684	73/80(91%)	ARU13271.1
25	22.786	23.388	42	200	21,8	AAAGGAAGGAAAATAGTATG	Lisina	CHPC926	190/200(95%)	APC45843.1
26	23.595	23.771	32	58	6,6	AAAGATGGTGCATAAGATG		DT1	52/55(95%)	NP_049414.1

27	23.889	24.116	43	75	8,6	AATGGAATCCCAGTTTCAATG	Lisina	D4276	73/75(97%)	ASD50985.1
28	24.670	25.005	38	111	12,3	TAGAGAGGTAAACAAAATG		5093	101/111(91%)	YP_002925119.1
29	25.034	25.324	33	98	11,1	AAAGGAGAGAATAATATG		P7602	96/96(100%)	ARU14017.1
30	25.352	25.903	38	183	21,3	AAAGGAAGTAAACAAAATG		Sfi11	165/183(90%)	NP_056701.1
31	25.987	26.202	30	71	8,8	AAAGGTTTAACTGATATG		Sfi11	66/71(93%)	NP_056703.1
32	26.224	26.661	25	145	17,4	AGAAGAAATGGTATTTTAAATATG		O1205	144/145(99%)	NP_695131.1
33	26.913	27.116	37	67	7,6	AAAGGAGGAACAAAATG	Represor tipo Cro	P3681	67/67(100%)	ARU13230.1
34	27.157	27.831	38	224	26	AAAGGAACAAGATG		9871	216/219(99%)	YP_009286854.1
35	28.115	28.324	44	69	8,1	AAAGGAGAGAATCTCATG		CHPC577	69/69(100%)	APC45801.1
36	28.337	29.083	33	248	29	AAAGGGTAATACATG	Antirepresos	CHPC1151	247/248(99%)	APC45894.1
37	29.275	29.466	35	64	8	AAAGGAAAATACTATG		CHPC1151	63/63(100%)	APC45895.1
38	29.735	30.640	37	301	34,7	GGAGAGGGTGAGTCTAAAATG	Componente del primosoma	P0095	283/301(94%)	ARU13186.1
39	30.652	31.434	40	260	30,3	TGAGGGGATTTCCCTATG	Proteína tipo DnaC	9871	256/260(98%)	YP_009286859.1
40	31.431	31.613	36	60	7,3	ATAAGAGGATGATGTTATG		phiAbc2	60/60(100%)	YP_003347441.1
41	31.741	32.397	37	218	25	AAAAGAGGATATGACATG	Proteína tipo ERF	phiAbc2	218/218(100%)	YP_003347442.1
42	32.400	33.371	38	323	37,7	AACGGAAGGGATAAATATG		P9901	323/323(100%)	ARU14728.1
43	33.368	33.814	42	148	16,6	AAAGGAGAAAACAAATG	Proteína de unión a ssADN	P9901	142/150(95%)	ARU14729.1
44	33.824	34.285	40	153	17,9	TTCTAAGGTGAAACTATG		TP-J34	150/153(98%)	YP_007392263.1
45	34.282	34.518	35	78	9,1	AAAGGAGCTGGAATG		phiAbc2	75/78(96%)	YP_003347446.1
46	34.509	34.679	40	56	6,3	GAAAGAGATGATAGAACTATG		20617	51/55(93%)	YP_009003362.1
47	34.676	34.831	30	51	6,4	CAAGGAGATTAGTAGATTTATG		5093	48/50(96%)	YP_002925091.1
48	34.832	35.314	36	160	19,7	GGAAGAGGTAAATTAATG		7201	127/149(85%)	NP_038316.1
49	35.348	35.872	35	174	20,8	TAAAGAGGTGACAGAATG		Sfi11	81/177(46%)	NP_056719.1
50	35.985	36.497	40	170	19,4	TTTGGAGGGAATAGCAAGTG	Proteína de unión a ADN	5093	161/170(95%)	YP_002925093.1
51	37.099	37.386	42	95	10,5	TTGCGAGGTATTTATG		P0091	87/95(92%)	ARU12999.1
52	37.383	37.769	34	128	14,7	AAAGGAAGAGGACAATG	Proteína YopX	phiS10	60/127(47%)	AGF87672.1
53	38.068	38.634	36	188	21,9	AAATGAGGTTGGCTAATTG	Endonucleasa HNH	Macrococcus caseolyticus	133/186(72%)	WP_101151978.1
54	38.592	38.876	32	94	11	AATTGCAATAGCAAAAATG		20617	77/80(96%)	YP_009003371.1
55	38.934	39.641	37	235	27,5	TCGAGGAGTGAGAACATG		P7955	200/235(85%)	ARU14417.1
56	40.147	40.563	44	138	16,2	GAGAGGGCAGAACAATG	Regulador transcripcional	5093	134/138(97%)	YP_002925096.1

Tabla 3.6. Características generales del fago 94

Fago 94 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	88	546	41	152	17,2	AGAGGAGAAACGATGAGTG	Terminasa, subunidad menor	P9901	151/152(99%)	ARU14694.1
2	681	2.447	38	588	67,8	AAAGACGCTGGAACACGATATTGTTT TGATGTTCTG	Terminasa, subunidad mayor	P7133	573/585(98%)	ARU13532.1
3	2.635	3.585	33	316	37,2	ATACGTGAATGTTTATAATTG	-	53	106/295(36%)	ALJ99584.1
4	3.819	4.979	41	386	42,7	AAAGGAGGTGATAACAATTG	Proteína portal	P7601	385/386(99%)	ARU13948.1
5	4.966	5.634	38	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de andamiaje	Abc2	222/222(100%)	YP_003347414.1
6	5.649	6.842	38	397	44,2	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	73	397/397(100%)	ALJ99542.1
7	6.857	7.171	43	104	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
8	7.171	7.521	40	116	13,5	GAAAGAGGTGACTAATG	Proteína de unión cápside-cola	Abc2	116/116(100%)	YP_003347417.1
9	7.528	7.950	43	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	73	140/140(100%)	ALJ99545.1
10	7.955	8.326	32	123	14,2	AAGGGAGGGGAGTGATTAAGTG	Proteína de cola	128	122/123(99%)	ALJ99633.1
11	8.345	8.953	40	202	22	AAAGGAGAAAATATATATG	Cola, subunidad mayor	P7601	202/202(100%)	ARU13955.1
12	9.027	9.380	37	117	13,5	AAAGGAGTAAAGACCACAATG	Proteína de cola	P7151	116/117(99%)	ARU13633.1
13	9.428	9.571	38	47	5,6	AATGGCATAGAGAATTTG	-	P5641	46/47(98%)	ARU13362.1
14	9.599	14.380	40	1593	174,7	AAAGGAGGGAATATAACATG	Proteína de cola	vB_SthS_VA698	1531/1593(96%)	AUN43331.1
15	14.377	15.942	40	521	58,5	TTAGGAGGTAGCATTATTTG	Cola, subunidad menor	P3681	499/521(96%)	ARU13222.1
16	15.921	18.668	38	915	103,1	AAAGACCCAACAATTGAAATTTTCATG	Receptor viral	P7151	833/908(92%)	ARU13637.1
17	18.668	20.680	41	670	74,9	GTAGGAGGTGCATAATG	Proteína de cola	P7601	641/670(96%)	ARU13961.1
18	20.650	21.105	35	151	17,4	AAGGGTGAATTTTGTG	-	P7134	125/136(92%)	ARU13591.1
19	21.125	21.268	31	48	5,6	AAAGGATAAAAAGATATG	-	P5651	46/48(96%)	ARU13421.1
20	21.289	21.612	34	107	12,5	ATAGGAGGGATGCGTTATG	-	TP-J34	99/107(93%)	YP_007392299.1
21	21.620	21.862	35	80	8,8	TGAGAGGATGAAAAATAAATG	Holina	TP-J34	73/79(92%)	YP_007392300.1
22	21.864	22.709	43	281	31	CAAGGAAGGAATATAGTATG	Lisina	53	253/281(90%)	ALJ99604.1
23	23.000	23.239	26	79	9,5	AAACGAGGAAAACAAAAATG	-	-	-	-
24	23.960	24.521	36	441	52,1	AAAGGAATTAATAGTTG	-	-	-	-

25	24.539	25.051	39	170	19,5	CAAGGAGAAAAATTATG	Proteína de particionamiento el cromosoma	D4276	144/167(86%)	ASD50992.1
26	25.071	25.250	33	59	7,3	AAAGTTGGTTAACATATG	-	-	-	-
27	25.516	25.833	41	105	12,2	AAAGAGAGGGAATAAATAATG	-	P7631	105/105(100%)	ARU14072.1
28	26.153	26.356	38	67	7,6	AGAGGAGGAACAAAATG	Regulador tipo Cro	P7633	67/67(100%)	ARU14166.1
29	26.515	26.706	35	63	7,8	AAAGAAAGGAAAATACTATG	-	9873	63/63(100%)	AMQ65822.1
30	26.982	27.455	37	157	18,1	TAGGGAGGGTAGGAATTAATATG	-	vB_SthS_VA460	143/157(91%)	AUN43437.1
31	27.452	28.141	40	229	25,7	AAAGGAGAAACCTTAACATAAGATG	Proteína de unión al ADN	TP-778L	228/229(99%)	YP_008772070.1
32	28.110	29.522	37	470	54,1	AAAGGGGTGTAAGGTAGATG	Helicasa	TP-778L	468/470(99%)	YP_008772071.1
33	29.529	30.002	34	157	18,2	TTTGAGATAAAAAACATG	Proteína de unión a ssADN	858	156/157(99%)	YP_001686832.1
34	30.007	30.822	38	271	30,6	GAAGTTTCAGATCTAGATTTACCATT CTAAGACTATG	-	858	265/271(98%)	YP_001686833.1
35	30.791	32.347	37	518	60,4	AAAGGAGTTGGACACTAAACATG	Primasa	TP-778L	513/518(99%)	YP_008772074.1
36	32.591	32.911	36	106	12,1	AAAGGTAAATTTTAAGTG	-	858	103/106(97%)	YP_001686835.1
37	32.895	33.131	38	78	9,2	TTTGATATAGTAAATG	-	vB_SthS_VA214	73/78(94%)	AUN43361.1
38	33.122	33.295	36	57	6,5	GAAAGAGATGATAGAACTATG	-	128	57/57(100%)	ALJ99657.1
39	33.292	34.470	26	51	6,3	GAAGGAGATTAGTAGATTTATG	-	128	51/51(100%)	ALJ99658.1
40	33.448	33.960	37	170	19,6	GGTTGAGGTAGAATAAATG	Proteína de unión al ADN	P9902	156/170(92%)	ARU14785.1
41	33.957	34.820	41	287	31,9	TAAGGAGGAAGACCTTTG	ADN citosina metiltransferasa	CHPC1151	286/287(99%)	APC45910.1
42	34.817	35.110	44	97	10,6	TTGCGAGGTATTTATG	-	P0091	92/95(97%)	ARU12999.1
43	35.107	35.385	34	92	10,3	AAAGGGAGAGGACGAGGAAAATG	-	IPP66	75/92(82%)	APD24440.1
44	35.398	36.105	38	235	27,7	TAAAGAGGTGTAGATATG	-	9874	209/235(89%)	YP_009280373.1
45	36.477	36.896	38	139	16,5	GAACTATTTGACAACATTATTATACC ATG	-	53	139/139(100%)	ALJ99622.1
46	36.969	37.517	46	182	21,3	TCAGAAGGACACAGTAGTG	Endonucleasa HNH	53	176/182(97%)	ALJ99623.1

Tabla 3.7. Características generales del fago 107

Fago 107 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	88	549	41	153	17,2	AGAGGAGAAACGATGAGTG	Terminasa, subunidad menor	P5652	151/153(98%)	ARU13444.1
2	590	1.279	38	229	26,3	AAGGGAGGTGATTAATAGTAAATG	Terminasa, subunidad menor	P5641	228/229(99%)	ARU13350.1
3	1.276	1.497	41	73	8,7	GAGAGTGGGCAATATGATAGTTTTC ATTTTACTTG	-	-	-	-
4	1.440	2.471	42	343	39,5	TTTAGGGAAGACCAAAGATTTTG	Terminasa, subunidad mayor	P7152	338/343(99%)	ARU13666.1
5	2.475	2.654	37	59	6,6	AGAGGAGTATTAATATATG	Proteína de unión cápside-cola	DT1	59/59(100%)	NP_049393.1
6	2.672	3.832	41	386	42,7	AAAGGAGGTGATAACAATTG	Proteína portal	P7132	385/386(99%)	ARU13491.1
7	3.819	4.487	38	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de andamiaje	DT1	222/222(100%)	NP_049395.1
8	4.502	5.695	38	397	44,2	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	P8921	396/397(99%)	ARU14427.1
9	5.710	6.024	44	104	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
10	6.024	6.374	44	116	13,5	GAAAGAGGTGACTAATG	Proteína de unión cápside-cola	73	116/116(100%)	ALJ99544.1
11	6.381	6.803	43	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	73	140/140(100%)	ALJ99545.1
12	6.808	7.179	33	123	14,2	GAGGGGAGTGATTAAGTG	Proteína de cola	7201	118/123(96%)	NP_038331.1
13	7.198	7.806	40	202	21,7	AAAGGAGAAAATATATATG	Cola, subunidad mayor	7201	193/203(95%)	NP_038332.1
14	7.879	8.232	38	117	13,4	AAAGGAGTAAGACCATAATG	Proteína de cola	DT1	112/117(96%)	NP_049402.1
15	8.280	8.420	37	47	5,6	AATCGCATTGAAGATTTG	-	P0091	46/47(98%)	ARU12971.1
16	8.451	13.229	42	1592	175,4	AAAGGAGGGAATATAACATG	Cola, subunida menor	P0091	1519/1593(95%)	ARU12972.1
17	13.226	14.785	39	519	58,5	CTAGGGGGTAGCATAATTTG	Proteína de cola	P7134	495/519(95%)	ARU13588.1
18	14.764	18.618	40	1284	143,4	CAAGCTGGATAGCCAAAGACCCAAC AATTGAAATTTTCATG	Receptor viral	DT2	874/1321(66%)	AAK83245.1
19	18.618	20.648	42	676	75,1	GTAGGAGGTGCATAATG	Proteína estructural	ALQ132	608/676(90%)	YP_003344865.1
20	20.664	21.050	34	128	14,6	AAAGAAGGAAAATTCATG	-	P5641	118/128(92%)	ARU13367.1

21	21.007	21.207	35	66	7,8	AATGGCTTAATCAATCAAATGTATCC TG	-	P5641	40/42(95%)	ARU13368.1
22	21.229	21.453	39	74	8,7	TTAGGAAGTGGTTGATTTATG	-	825	59/74(80%)	YP_001686819.1
23	21.450	21.581	37	44	5	AAGTGAGGTTAAAAATG	-	858	43/44(98%)	YP_001686820.1
24	21.618	21.860	37	80	8,8	GTTGGAGGAAATCAAATG	Holina	CHPC577	75/80(94%)	APC45793.1
25	21.862	22.464	40	200	22	TAAGGAAGGAAAATAGTATG	Lisina	53	187/198(94%)	ALJ99604.1
26	22.671	23.432	28	253	29,8	AAAGATGGTGTCATAAGATG	Endonucleasa HNH	9872	250/253(99%)	YP_009289375.1
27	23.482	23.721	45	79	9,1	AATAGTGGGTAGTAATGAGAAAACG GAATC	Lisina	ALQ132	71/75(95%)	YP_003344872.1
28	24.331	24.558	38	75	8,5	-	Represor tipo Cro	5093	71/75(95%)	YP_002925127.1
29	24.747	25.439	35	230	25,2	CTGGGAGGAGAACAAAAATG	Proteasa	128	230/230(100%)	ALJ99646.1
30	25.590	26.336	34	248	28,8	TGAGGAGTAGAAAGGGTAATACATG	Antirepresor	CHPC1151	247/248(99%)	APC45894.1
31	26.528	26.719	35	63	7,8	CAGAAAGGAAAATACTATG	-	CHPC1151	63/63(100%)	APC45895.1
32	27.043	27.849	34	268	31,5	AAAGAGAGAGATAAGATTAATG	Proteína de replicación de ADN	9873	267/269(99%)	AMQ65823.1
33	27.862	28.644	41	260	30,4	CTAAGAGGTTCTTTATATG	Proteína de replicación de ADN	9873	252/260(97%)	YP_009289383.1
34	28.641	28.838	37	65	7,8	ACAAGAGGATGATGCTATG	-	9874	59/65(91%)	YP_009280360.1
35	28.963	29.781	40	272	30,4	AAAGGAAGACAAGACATG	Recombinasa tipo RecT	9873	270/272(99%)	YP_009289386.1
36	29.720	30.619	38	299	35	GTAGACGCTGAAACCGGTG	Endonucleasa	9871	270/278(97%)	YP_009286862.1
37	30.659	31.303	34	214	25,9	AAATGAGGTAAGCAATG	-	9872	212/214(99%)	YP_009289388.1
38	31.318	31.764	45	148	16,5	AAAGGAGAAAATAAATG	Proteína de unión a ssADN	9871	144/148(97%)	YP_009286864.1
39	31.782	32.243	40	153	18	GAAGTTGATGAAGTTATG	Endonucleasa tipo RusA	P4761	150/153(98%)	ARU13338.1
40	32.240	32.482	36	80	9,5	CAAGGAGCTGGAATG	-	P9851	71/80(89%)	ARU14548.1
41	32.467	32.625	29	52	6,4	GAAAGAGATGATAGAACTATG	-	DT1	48/51(94%)	NP_049428.1
42	32.622	32.939	29	105	12,2	GGTTGAGGTAGATGAATG	-	phiAbc2	105/105(100%)	YP_003347449.1
43	32.968	33.351	31	127	14,4	ATTGGAGAAATTAATG	-	phiAbc2	126/127(99%)	YP_003347450.1
44	33.363	33.863	35	166	20,2	TAAGGAGAACTAGAATG	-	P5641	150/166(90%)	ARU13394.1
45	33.864	34.376	38	170	19,5	TGAAGAGGTAGATAGATG	Proteína de unión a ADN	9871	161/170(95%)	YP_009286871.1
46	34.345	34.647	40	100	11,1	AAAGGGAAAAATGATAAATG	-	P9854	91/98(93%)	ARU14689.1
47	34.651	35.358	37	235	27,7	TGAGGAGTTATTAAGACATG	-	SFi18	226/235(96%)	AAF63098.1
48	35.501	35.785	38	94	10,7	TTTTATGGTATAATAATATTG	-	P7952	56/77(73%)	ARU14279.1

49	35.818	36.216	39	132	15,5	AAAGGAAAGACAATTTATG	-	53	131/132(99%)	ALJ99622.1
50	36.326	36.844	47	172	20	AAAGGAGGCATGCCAATG	Endonucleasa HNH	7201	166/172(97%)	NP_038321.1

Tabla 3.8. Características generales del fago 115. En violeta se señalan los ORFs duplicados.

Fago 115 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	326	787	39,39	153	17,3	CGGAGAGGAGAAATGATGAGTG	Terminasa, subunidad menor	53	147/151(97%)	ALJ99583.1
2	828	2.699	40,87	623	71,3	AAGGGAGGTGATTAATAGTAAATG	Terminasa, subunidad mayor	P7152	608/623(98%)	ARU13666.1
3	2.900	4.060	41,34	386	42,7	AAAGGAGGTGATAACAATTG	Proteína portal	73	385/386(99%)	ALJ99540.1
4	4.047	4.715	39,01	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de anamiaeje	vB_SthS_VA460	222/222(100%)	AUN43411.1
5	4.730	5.923	38,78	397	44,2	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	P0091	389/397(98%)	ARU12964.1
6	5.938	6.255	40,88	105	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	vB_SthS_VA214	105/105(100%)	AUN43396.1
7	6.248	6.598	35,04	116	13,6	TAGCGAGGTAGTGGCTAATG	Proteína de unión cápside-cola	vB_SthS_VA214	116/116(100%)	AUN43395.1
8	6.605	7.027	43,03	140	15,7	AAAGTTGGGTGATGGCTTATG	Proteína de cola	P7602	139/140(99%)	ARU14001.1
9	7.032	7.403	31,99	123	14,2	AAGGGAGGGGAGTAATTAAGTG	Proteína de cola	P0091	119/123(97%)	ARU12968.1
10	7.422	8.030	40,23	202	21,6	AAAGGAGAAAATATATATG	Cola, subunidad mayor	P3681	197/202(98%)	ARU13218.1
11	8.101	8.454	37,01	117	13,4	AAAGGAGTAAAGACAAATG	Proteína de cola	73	113/117(97%)	ALJ99548.1
12	8.672	12.820	40,83	1382	151,0	AAAGGAGGAAAATAATATATG	Proteína de cola	D4276	887/1455(61%)	ASD50974.1
13	12.817	14.373	39,37	518	58,4	TTAGGAGGTCAAATTTTGG	Proteína de cola	P7633	493/518(95%)	ARU14153.1
14	14.352	17.393	39,94	1013	113,8	CCAACAATTGAAATTTTCATG	Receptor viral	phiAbc2	743/1017(73%)	YP_003347425.1
15	17.393	19.453	41,15	686	76,5	GTAGGAGGTGCATAATG	Proteína de placa basal	vB_SthS_VA460	602/687(88%)	AUN43455.1
16	19.450	19.893	34,91	147	16,9	TACGGACGGATATCCAGAAAATTG	-	P3681	134/138(97%)	ARU13225.1
17	20.062	20.385	35,19	107	12,5	AAAGGAGGTATGTCGTG	-	P7133	102/107(95%)	ARU13550.1
18	20.393	20.635	37,04	80	8,9	TTGAGAGGATAATAATAAAAATG	Holina	DT1	80/80(100%)	NP_049412.1
19	20.637	21.239	42,62	200	21,8	AAAGGAGAAAATAAAAATG	Lisina	P7571	192/200(96%)	ARU13777.1
20	21.479	21.727	44,98	82	8,9	CCACTAGGAATAGTG	-	-	-	-
21	22.019	23.437	41,44	472	52,8	CTCGTAGCTCTTTTGTGTTG	Receptor viral	vB_SthS_VA460	428/467(92%)	AUN43456.1
22	23.437	25.497	41,15	686	76,5	GTAGGAGGTGCATAATG	Proteína de placa basal	vB_SthS_VA460	602/687(88%)	AUN43455.1
23	25.494	25.937	34,91	147	16,9	TACGGACGGATATCCAGAAAATTG	-	P3681	134/138(97%)	ARU13225.1
24	26.106	26.429	35,19	107	12,5	AAAGGAGGTATGTCGTG	-	P7133	102/107(95%)	ARU13550.1
25	26.437	26.679	37,04	80	8,9	TTGAGAGGATAATAATAAAAATG	Holina	DT1	80/80(100%)	NP_049412.1

26	26.681	27.283	42,62	200	21,8	AAAGGAGAAATAAAATG	Lisina	P7571	192/200(96%)	ARU13777.1
27	27.523	27.771	44,98	82	8,9	CCACTAGGAATAGTG	-	-	-	-
28	28.080	28.193	28,95	37	4,5	TTTGTTGCTATTTTCATATAATTG	-	P4761	27/35(77%)	ARU13323.1
29	28.251	28.586	36,9	111	12,3	GAGAGAGGTAAACAAAATG	-	vB_SthS_VA214	111/111(100%)	AUN43379.1
30	28.612	28.890	32,62	92	10,7	AAAGGAGGAAATAACATG	-	P7602	90/96(94%)	ARU14017.1
31	28.918	29.469	37,14	183	21,5	AAAGGAAGTAAACAAAATG	-	9874	165/183(90%)	YP_009280354.1
32	29.558	29.956	36,34	132	15,0	AGAGGAAGTAAATGACATG	-	-	-	-
33	30.027	30.221	32,82	64	7,4	GAAGGATGTGATCTAAAATG	-	O1205	57/60(95%)	NP_695135.1
34	30.237	30.431	34,36	64	7,8	ATTGGTGGATTAATATG	-	Sfi11	56/64(88%)	NP_056703.1
35	30.474	30.908	25,06	144	17,3	GAAATGGTATTTTAATATG	-	O1205	140/144(97%)	NP_695131.1
36	31.067	31.273	31,88	68	7,8	AAAGGAGAACTTAATG	Represor tipo Cro	SFi18	67/69(97%)	AAF63080.1
37	31.293	31.415	26,02	40	5,0	AAAGGTATTTAAAATG	-	Sfi11	40/40(100%)	NP_056706.1
38	31.463	31.600	36,23	45	5,2	CCTCGATGTGCTGAGTG	-	-	-	-
39	31.680	32.153	37,13	157	18,0	GTGGGAGGGTAGGAATTAATATG	-	vB_SthS_VA460	145/157(92%)	AUN43437.1
40	32.150	32.824	44,59	224	25,4	AAAGGAGAAACCTTAACATAAGATG	Proteína de unión al ADN	128	224/224(100%)	ALJ99650.1
41	32.814	34.145	40,54	443	50,4	AGAAGAGGTCTTCAATTTTG	Helicasa	vB_SthS_VA698	430/443(97%)	AUN43317.1
42	34.152	34.607	36,18	151	17,3	TATGGAGATAAAAACTATG	-	P3681	148/151(98%)	ARU13238.1
43	34.610	35.416	40,77	268	30,1	ACCTTCCGTTCTAATTATG	Primasa	P7631	267/271(99%)	ARU14079.1
44	35.397	36.920	37,14	507	59,2	AAATAAGGAGGATTG	Primasa	53	500/507(99%)	ALJ99616.1
45	37.067	37.180	48,25	37	4,4	AGAGAGAGTCAAAAAATAAGGTG	-	P9852	30/34(88%)	ARU14593.1
46	37.517	37.771	36,08	84	10,0	GAAGCAGCAACAGTTATAGATG	-	DT1	58/58(100%)	NP_049425.1
47	37.753	38.076	40,43	107	12,2	TTGCGAGCATTTATAAGGAATG	-	DT1	107/107(100%)	NP_049426.1
48	38.054	38.299	36,59	81	9,6	CTATGAGGATAGTTG	-	128	80/80(100%)	ALJ99656.1
49	38.290	38.460	38,6	56	6,3	GAAAGAGATGATAGAACTATG	-	TP-J34	53/56(95%)	YP_007392265.1
50	38.457	38.615	33,33	52	6,5	CAAGGAGATTAGTAGAGTTATG	-	P7574	49/51(96%)	ARU13933.1
51	38.612	39.142	34,65	176	20,7	GGAAGAGGTGGAAGAATG	-	phiAbc2	96/167(57%)	YP_003347451.1
52	39.143	39.655	36,45	170	19,6	GAAAGAGGTTAAATAAATG	Proteína de unión al ADN	vB_SthS_VA214	140/170(82%)	AUN43356.1
53	39.624	39.956	45,05	110	12,3	AAAGGAAAATGATAAATG	-	P9902	105/110(95%)	ARU14786.1
54	39.953	40.660	38,42	235	27,9	AAAGGAACAGGGCAATG	-	P0095	206/235(88%)	ARU13202.1
55	40.909	41.307	39,35	132	15,6	AAAGGAAGACAATTTATG	-	DT1	125/132(95%)	NP_049433.1

Tabla 3.9. Características generales del fago 123

Fago 123 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	87	548	40,69	153	17,3	AACGGAGAGGAGTAATGATGAGTG	Terminasa, subunidad menor	DT1	145/153(95%)	NP_049390.1
2	629	1.318	38,26	229	26,2	AAGGGAGGTGATTAATAGTAAATG	Terminasa, subunidad mayor	vB_SthS_VA214	229/229(100%)	AUN43402.1
3	1.401	2.510	41,26	369	42,5	TGGGCAAGTAAAAGTG	Terminasa, subunidad mayor	P7132	362/369(98%)	ARU13489.1
4	2.514	2.693	35	59	6,6	ATTAGAGGAGTATTAATATATG	Proteína de andamiaje	vB_SthS_VA214	59/59(100%)	AUN43400.1
5	2.711	3.871	40,91	386	42,7	AAAGGAGGTGATAACAATTG	Proteína portal	P7151	376/386(97%)	ARU13624.1
6	3.858	4.526	38,27	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de andamiaje	phiAbc2	222/222(100%)	YP_003347414.1
7	4.541	5.734	39,11	397	44,1	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	P0091	393/397(99%)	ARU12964.1
8	5.749	6.066	40,88	105	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	vB_SthS_VA214	105/105(100%)	AUN43396.1
9	6.059	6.409	40,46	116	13,5	TAGCGAGGTAGTGGCTAATG	Proteína de unión cápside cola	phiAbc2	116/116(100%)	YP_003347417.1
10	6.416	6.838	43,74	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	73	139/140(99%)	ALJ99545.1
11	6.843	7.214	32,53	123	14,1	AGGGAGGGGAGTAATTAAGT	Proteína de cola	DT1	122/123(99%)	NP_049400.1
12	7.236	7.844	39,41	202	21,8	AAAGGAGAAAATATATATG	Cápside, subunidad mayor	P7602	198/202(98%)	ARU14003.1
13	7.917	8.270	37,57	117	13,4	AAAGGAGTAAAGACAAATG	Proteína de cola	vB_SthS_VA214	117/117(100%)	AUN43391.1
14	8.488	13.266	41,87	1592	175,6	AAAGGAGGAAAATAATACATG	Proteína de cola	vB_SthS_VA214	1562/1592(98%)	AUN43390.1
15	13.263	14.819	40,08	518	58,3	TTAGGAGGTCAAATTATTTG	Proteína de cola	DT1	515/518(99%)	NP_049405.1
16	14.819	18.304	39,76	1161	129,5	GAAGGAGCGCTTTGTTTAATG	Receptor viral	vB_SthS_VA460	1013/1161(87%)	AUN43456.1
17	18.304	20.334	40,57	676	75,2	GTAGGAGGTGCATAATG	Proteína de la placa basal	P8921	592/678(87%)	ARU14438.1
18	20.350	20.760	34,31	136	15,5	AAAGGAGATATAAACATG	-	ALQ132	133/136(98%)	YP_003344866.1
19	20.944	21.267	35,49	107	12,5	ATAGGAGGGATGTGTTATG	-	vB_SthS_VA698	106/107(99%)	AUN43325.1
20	21.275	21.517	38,27	80	8,8	GAGAGGATGAAGAATAAATG	Holina	D4276	76/80(95%)	ASD50982.1
21	21.519	22.121	41,29	200	21,9	CAAGGAAGGAAAATAGTATG	Lisina	P7151	196/200(98%)	ARU13643.1
22	22.328	22.513	28,49	61	7,0	AAAGATGGTGCATAAGATG	-	DT1	60/61(98%)	NP_049414.1
23	22.580	22.807	42,54	75	8,7	AATGGAATCCCGTTGCAATG	Lisina	P5651	75/75(100%)	ARU13426.1
24	23.151	23.528	37,57	125	14,7	AAATTAATGCCATG	-	P5641	121/125(97%)	ARU13374.1
25	23.715	23.918	36,27	67	7,6	AGAGGAGGAACAAAAATG	Resgulador tipo Cro	9872	64/67(96%)	YP_009289378.1

26	23.968	24.912	37,35	314	36,5	AAAGGAACTATATG	-	P4761	308/314(98%)	ARU13328.1
27	24.905	25.444	35,56	179	20,9	GAAGGAGAAATCATAAATG	-	P7951	174/179(97%)	ARU14214.1
28	25.571	25.780	28,57	69	8,2	AAAGGAGAGTTTGATG	Represor tipo Cro	5093	66/69(96%)	YP_002925130.1
29	25.814	25.951	38,41	45	5,4	TAGAGAGGAACCAAAAATG	-	CHPC926	44/46(96%)	APC45850.1
30	26.272	27.078	33,33	268	31,6	AAGAGAGATAAGATTAATG	Prroteína de replicación de ADN	P5641	261/268(97%)	ARU13383.1
31	27.091	27.873	39,85	260	30,4	CTAAGAGGTTCTTTATATG	Prroteína de replicación de ADN	P5641	256/260(98%)	ARU13384.1
32	27.870	28.067	38,38	65	7,8	ACAAGAGGATGATGTTATG	-	9874	64/65(98%)	YP_009280360.1
33	28.192	29.010	40,54	272	30,4	AAAGGAAGACAAAGACATG	Recombinasa RecT	9872	271/272(99%)	YP_009289386.1
34	28.949	29.848	38	299	35,0	AAAAGCGTAGACGCTGAAACCGGTG	Nucleasa	9871	270/278(97%)	YP_009286862.1
35	29.888	30.532	33,64	214	25,9	AAATGAGGTAAGCAATG	-	9872	212/214(99%)	YP_009289388.1
36	30.547	30.993	41,39	148	16,7	AAAGGAGAAAATAAATG	Proteína de unión a ssADN	7201	144/148(97%)	NP_038310.1
37	31.011	31.472	40,26	153	18,0	GAAGTTGATGAAGTTATG	Endonucleasa tipo RusA	P4761	149/153(97%)	ARU13338.1
38	31.469	31.711	35,39	80	9,5	CAAGGAGCTGGAATG	-	P9851	72/80(90%)	ARU14548.1
39	31.696	31.854	31,45	52	6,5	GAAAGAGATGGTAGAACTATG	Proteína de unión al ADN	P9854	48/50(96%)	ARU14686.1
40	31.851	32.168	28,93	105	12,2	TGAGGAGGTAGATGAATG	-	phiAbc2	105/105(100%)	YP_003347449.1
41	32.197	32.580	31,25	127	14,4	ATTGGAGAAATTAATG	-	phiAbc2	126/127(99%)	YP_003347450.1
42	32.592	33.092	35,13	166	20,2	TAAGGAGAACTAGAATG	-	P5641	150/166(90%)	ARU13394.1
43	33.093	33.605	38,4	170	19,5	TGAAGAGGTAGAATAGATG	Proteína de unión al ADN	9871	162/170(95%)	YP_009286871.1
44	33.574	33.876	40,26	100	11,1	AAGGGAAAAATGATAAATG	-	P9854	91/98(93%)	ARU14689.1
45	33.880	34.587	38,14	235	28,0	ATTTGAGGAGTTATTAAGACATG	-	ALQ132	231/235(98%)	YP_003344889.1
46	35.016	35.414	38,85	132	15,5	AAAGGAAAGACAATTTATG	Regulador transcripcional	53	132/132(100%)	ALJ99622.1
47	35.487	36.035	45,9	182	21,2	GAAGGACACAGTAGTG	Endonucleasa HNH	53	169/182(93%)	ALJ99623.1

Tabla 3.10. Características generales del fago 128

Fago 128 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	88	549	42	153	17,3	AGAGGAGAAACGATGAGTG	Terminasa, subunidad menor	P7631	150/153(98%)	ARU14044.1
2	902	2.773	41	623	71,5	AAAGGGGGTGATTAATAGTAAATG	Terminasa, subunidad mayor	P7633	612/623(98%)	ARU14140.1
3	2.777	2.956	37	59	6,6	AGAGGAGTATTAATATATG	Proteína de unión cápside-cola	phiAbc2	59/59 (100%)	YP_003347412.1
4	2.974	4.134	41	386	42,7	AAAGGAGGTGATAACAAATG	Proteína portal	73	386/386(100%)	ALJ99540.1
5	4.121	4.789	38	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de andamiaje	phiAbc2	222/222 (100%)	YP_003347414.1
6	4.804	5.997	39	397	44,1	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	P5651	394/397(99%)	ARU13408.1
7	6.012	6.326	44	104	11,5	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
8	6.326	6.676	40	116	13,4	GAAAGAGGTGACTAATG	Proteína de unión cápside-cola	73	116/116(100%)	ALJ99544.1
9	6.683	7.105	43	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	73	140/140(100%)	ALJ99545.1
10	7.110	7.482	32	123	14	GAGGGGAGTGATTAAGTG	Proteína de cola	P0091	120/123(98%)	ARU12968.1
11	7.500	8.111	39	203	22	AAAGGAGAAAATATATATG	Cola, subunidad mayor	P7601	202/203(99%)	ARU13955.1
12	8.144	8.539	37	131	15,2	CAAAGAGGTCAGGCTTATG	Proteína de cola	P7133	117/117(100%)	ARU13542.1
13	8.758	13.533	42	1591	175	AAAGGAGGGAATATAACATG	Cola, subunidad menor	P5651	1572/1591(99%)	ARU13416.1
14	13.530	15.095	38	521	58,7	TTAGGAGGTCAAATTATTTG	Cola, subunidad menor	P7154	508/521(98%)	ARU13727.1
15	15.091	18.274	39	1059	119,1	GAAGGAGCGTTTTGTATAATG	Receptor viral	P7631	1050/1059(99%)	ARU14061.1
16	18.274	20.322	41	682	76,3	GTAGGAGGTGCATAATG	Proteína estructural	P7631	664/682(97%)	ARU14062.1
17	20.343	20.771	35	142	16,4	AAATGAGGAATGAAAAAATATG	Proteína estructural	P5652	139/142(98%)	ARU13462.1
18	20.797	20.943	35	48	5,5	AAAGGGAAAAAGATATG		P5652	6/48(96%)	ARU13463.1
19	20.961	21.284	35	107	12,5	ATAGGAGGGATGTGTTATG		P9902	106/107(99%)	ARU14760.1
20	21.292	21.534	38	80	9,8	TTGAGAGGATAATAATAAAAATG	Holina	P7631	79/80(99%)	ARU14066.1
21	21.536	22.381	41	281	31,2	AAAGGAAGGAAAATAGTATG	Lisina	P7133	71/281(96%)	ARU13552.1
22	23.103	23.306	36	67	7,6	GAAGGAGGAACAAAATG	Regulador tipo Cro	phiAbc2	67/67 (100%)	YP_003347432.1
23	23.518	24.210	35	230	25,1	CTGGGAGGAGAACAAAAAATG	Proteasa	P0095	229/230(99%)	ARU13179.1
24	24.577	24.786	28	69	8,2	AAAGGAGAAACGAATG	Represor tipo Cro	ALQ132	69/69 (100%)	YP_003344877.1
25	24.820	24.954	35	44	5,2	TAGAGAGGAACCAAAAATG		P3681	41/44(93%)	ARU13234.1

26	25.228	25.701	36	157	18,4	AGGGTAGGAATTAATATG	-	P7602	50/157(96%)	ARU14024.1
27	25.698	26.372	45	224	25,4	AAAGGAGAAACCTTAACATAAGATG	Proteína con motivo de unión NTP	DT1	214/222 (96%)	NP_049420.1
28	26.362	27.693	40	443	50,2	AGAAGAGGTCTTCAATTTTG	Helicasa	P7601	441/443(99%)	ARU13976.1
29	27.700	28.155	36	151	17,3	TATGGAGATAAAAACTATG	-	P3681	150/151(99%)	ARU13238.1
30	28.158	28.973	41	271	30,5	AATTGACCTTCCATTCTAATTATG	Proteína de replicación de ADN	P3684	269/271(99%)	ARU13282.1
31	28.960	30.471	37	503	58,8	TAAGGAGGATTGGACTTG	Primasa	P8921	500/503(99%)	ARU14456.1
32	30.898	31.221	38	107	12,1	TTGCGAGCATTTATAAGGAATG	-	P7631	106/107(99%)	ARU14082.1
33	31.202	31.444	37	80	9,5	ACTGGAGATAGTTGATG		P7633	78/80(98%)	ARU14177.1
34	31.435	31.608	36	57	6,5	GAAAGAGATGATAGAACTATG		9871	52/57(91%)	YP_009286868.1
35	31.605	31.760	26	51	6,3	GAAGGAGATTAGTAGATTTATG		P5651	48/51(94%)	ARU13438.1
36	31.761	32.258	37	165	19	GGTTGAGGTAGAATAAATG	Proteína de unión a ADN	P7133	162/165(98%)	ARU13567.1
37	32.230	32.511	39	93	10,4	AAAGGAGTAATGATTG	-	P7133	91/93(98%)	ARU13568.1
38	32.508	33.215	40	235	28	AAAGGGAGGGGACAATG		P7134	234/235(99%)	ARU13615.1
39	33.482	33.880	37	132	15,4	AAAGGAAAGACAATTTATG	Apaf-1 related killer DARK	Sfi19	131/132 (99%)	NP_049923.1
40	33.990	34.499	48	169	19,5	AAAGGAGGCATGCCAATG	Endonucleasa HNH	P7631	168/169(99%)	ARU14091.1

Tabla 3.11. Características generales del fago 140

Fago 140 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	88	549	40,91	153	17,3	AGAGGAGAAACGATGAGTG	Terminasa, subunidad menor	P5652	145/153(95%)	ARU13444.1
2	810	1.943	26,46	377	44,6	CAAGGATAAAGTATG	-	P3681	375/377(99%)	ARU13207.1
3	1.982	2.671	37,68	229	26,2	AAAGGGGGTGATTAATAGTAAATG	Terminasa, subunidad mayor	vB_SthS_VA214	228/229(99%)	AUN43402.1
4	2.754	3.863	41,17	369	42,6	ATAGTTTCTGGGCAAGTAAAAGTG	Terminasa, subunidad mayor	P7132	363/369(98%)	ARU13489.1
5	4.064	5.215	41,06	383	42,3	AAAGGAGGTGATAACAAATG	Proteína portal	P7151	376/386(97%)	ARU13624.1
6	5.202	5.870	38,27	222	24,5	AAGGAGGTGAGATAAATG	Proteína de andamiaje	phiAbc2	222/222(100%)	YP_003347414.1
7	5.885	7.078	37,86	397	44,2	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	73	397/397(100%)	ALJ99542.1
8	7.093	7.407	43,17	104	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
9	7.407	7.757	40,17	116	13,5	GAAAGAGGTGACTAATG	Proteína de unión cápside-cola	phiAbc2	116/116(100%)	YP_003347417.1
10	7.764	8.186	43,5	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	73	140/140(100%)	ALJ99545.1
11	8.191	8.562	32,8	123	14,1	AAGGGAGGGGAGTAATTAAGTG	Proteína de cola	DT1	121/123(98%)	NP_049400.1
12	8.585	9.193	39,74	202	21,9	AAAAGGAGAAAATATATATG	Proteína de cola	P7133	202/202(100%)	ARU13541.1
13	9.267	9.620	36,72	117	13,5	AAAGGAGTAAAGACCACAATG	Proteína de cola	73	117/117(100%)	ALJ99548.1
14	9.668	9.808	38,3	46	5,6	TTAGAAAATGGCATAGAGAATTTG	-	P5641	46/47(98%)	ARU13362.1
15	9.839	14.809	41,46	1656	183,0	AAGGAGGGAATATAACATG	Proteína de cola	P9903	1644/1656(99%)	ARU14806.1
16	14.806	16.365	39,23	519	58,5	TTAGGAGGTCAAATTATTTG	Proteína de cola	phiAbc2	509/519(98%)	YP_003347424.1
17	16.365	19.643	38,91	1092	123,4	AAGGAGCGTTTTGTCTAATG	Receptor viral	P5652	937/1098(85%)	ARU13460.1
18	19.643	21.700	40,77	685	76,6	GTAGGAGGTGCATAATG	Proteína de la placa basal	P7631	652/684(95%)	ARU14062.1
19	21.726	22.136	34,06	136	15,7	AAGAAAGGAATAATTATG	-	53	135/136(99%)	ALJ99600.1
20	22.320	22.643	35,19	107	12,4	ATAGGAGGGATGTGTTATG	-	phiAbc2	105/107(98%)	YP_003347429.1
21	22.651	22.893	37,86	80	8,9	TTGAGAGGATAATAATAAAAATG	Holina	P7631	80/80(100%)	ARU14066.1
22	22.895	23.740	40,66	281	31,1	AAAGGAAGGAAAATAGTATG	Lisina	P7633	275/281(98%)	ARU14160.1
23	24.352	24.579	37,72	75	8,6	AAAGGAGATAACCTATG	Represor tipo Cro	53	75/75(100%)	ALJ99607.1
24	24.768	25.460	34,92	230	25,2	CTGGGAGGAGAACAAAAAATG	Proteasa	128	230/230(100%)	ALJ99646.1
25	25.611	26.357	33,6	248	28,8	TGAGGAGTAGAAAGGGTAATACATG	Antirrepresor	CHPC1151	247/248(99%)	APC45894.1

26	26.549	26.695	38,1	48	5,7	AAAGGAAAATACTATG	-	CHPC577	40/44(91%)	APC45802.1
27	27.008	27.880	34,02	290	34,2	AAGAGAGATAAGATTAATG	Proteína de replicación de ADN	P5641	264/290(91%)	ARU13383.1
28	27.893	28.675	40,74	260	30,4	CTAAGAGGTTCTTTATATG	Proteína de replicación de ADN	P9901	258/260(99%)	ARU14725.1
29	28.672	28.854	35,52	60	7,4	ACAAGAGGATGATGTTATG	.	phiAbc2	59/60(98%)	YP_003347441.1
30	28.982	29.614	36,81	210	24,0	GAGGATATGACATG	Proteína de unión a ssADN	vB_SthS_VA214	209/210(99%)	AUN43365.1
31	29.617	30.588	38,48	323	37,9	AAAGGGTAACTAGACATG	-	vB_SthS_VA214	321/323(99%)	AUN43364.1
32	30.585	31.031	44,07	148	16,4	TTAGGAGAAAACAAATG	Proteína de unión a ssADN	P4761	144/148(97%)	ARU13337.1
33	31.046	31.600	38,56	184	21,2	TGAGGAGGGTTTATG	Endonucleasa HNH	P0092	183/184(99%)	ARU13038.1
34	31.597	32.007	40,88	136	15,4	GTGGTAGGTGAAATATG	-	P0091	136/136(100%)	ARU12993.1
35	31.991	32.227	36,71	78	9,1	CATTGAGGAAACGTTAAATG	-	P5641	78/78(100%)	ARU13391.1
36	32.269	32.424	32,69	51	6,3	GAAGGAGATTGGTAGAGTTATG	-	P5641	48/49(98%)	ARU13392.1
37	32.425	33.006	35,57	193	23,3	GGTTGAGGTGGAATAAGTG	-	CHPC1151	149/193(77%)	APC45908.1
38	33.007	33.519	37,04	170	19,5	GGAAGAGGTGGAATAGATG	Proteína de unión al ADN	P3684	170/170(100%)	ARU13289.1
39	33.659	33.964	38,24	101	11,2	AAAGGAATAATGATTG	-	phiAbc2	100/100(100%)	YP_003347453.1
40	33.961	34.668	38,56	235	27,7	AAAGGAAGAGGGCAATG	-	phiAbc2	223/235(95%)	YP_003347454.1
41	34.775	34.948	35,06	57	6,7	AAACCCAGTAAATATTTTATAGAAAT G	-	P7601	43/44(98%)	ARU13990.1
42	34.935	35.333	38,35	132	15,5	AAAGGAAAGACAATTTATG	-	53	132/132(100%)	ALJ99622.1
43	35.406	35.954	46,45	182	21,2	TCAGAAGGACACAGTAGTG	Endonucleasa HNH	53	180/182(99%)	ALJ99623.1

Tabla 3.12. Características generales del fago 142

Fago 142 ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	88	549	41,99	153	17,3	TAGAGAGGAGAAACGATGAGTG	Terminasa, subunidad menor	128	152/153(99%)	ALJ99624.1
2	617	730	26,32	37	4,6	ATTATACGTGAATG	-	P9903	37/38(97%)	ARU14792.1
3	743	844	30,39	33	4,1	AGATTCCGGCGTATG	-	-	-	-
4	902	2.773	40,92	623	71,5	AAAGGGGGTGATTAATAGTAAATG	Terminasa, subunidad mayor	128	623/623(100%)	ALJ99625.1
5	2.777	2.956	36,67	59	6,6	ATTAGAGGAGTATTAATATATG	Proteína de unión cápside-cola	phiAbc2	59/59(100%)	YP_003347412.1
6	2.974	4.134	41	386	42,8	AAAGGAGGTGATAACAAATG	Proteína portal	73	386/386(100%)	ALJ99540.1
7	4.121	4.789	38,27	222	24,5	AAAGGAGGTGAGATAAATG	Proteína de andamiaje	phiAbc2	222/222(100%)	YP_003347414.1
8	4.804	5.997	38,61	397	44,1	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	128	397/397(100%)	ALJ99629.1
9	6.012	6.326	43,81	104	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
10	6.326	6.676	40,46	116	13,5	GAAAGAGGTGACTAATG	Proteína de unión cápside-cola	73	116/116(100%)	ALJ99544.1
11	6.683	7.105	42,55	140	15,6	AAGTTGGGTGATAGCTTATG	Proteína de cola	73	140/140(100%)	ALJ99545.1
12	7.110	7.481	31,99	123	14,2	AAGGGAGGGGAGTGATTAAGTG	Proteína de cola	128	122/123(99%)	ALJ99633.1
13	7.500	8.111	39,38	203	22,1	AAAGGAGAAAATATATATG	Cola, subunidad mayor	128	203/203(100%)	ALJ99634.1
14	8.144	8.539	36,87	131	15,2	CAAAGAGGTCAGGCTTATG	Proteína de cola	128	131/131(100%)	ALJ99635.1
15	8.587	8.727	38,3	46	5,6	AATGGCATAGAGAATTTG	-	P5641	46/47(98%)	ARU13362.1
16	8.758	13.533	42,32	1591	175,1	AAAGGAGGGAATATAACATG	Cola, subunidad menor	128	1591/1591(100%)	ALJ99636.1
17	13.530	15.095	38,31	521	58,7	TTAGGAGGTCAAATTATTTG	Cola, subunidad menor	128	520/521(99%)	ALJ99637.1
18	15.095	18.274	39,37	1059	119,2	GAAGGAGCGTTTTGTATAATG	Receptor viral	128	1059/1059(100%)	ALJ99638.1
19	18.274	20.322	40,9	682	76,5	GTAGGAGGTGCATAATG	Proteína estructural	128	682/682(100%)	ALJ99639.1
20	20.343	20.771	34,97	142	16,4	AAATGAGGAATGAAAAAATATG	Proteína estructural	128	142/142(100%)	ALJ99640.1
21	20.725	20.943	34,7	72	8,5	AAAGGGATGCTAAAGGATG	-	128	48/48(100%)	ALJ99641.1
22	20.961	21.284	35,49	107	12,5	ATAGGAGGGATGTGTTATG	-	128	107/107(100%)	ALJ99642.1
23	21.292	21.534	37,86	80	8,8	TTGAGAGGATAATAATAAAATG	Holína	128	80/80(100%)	ALJ99643.1
24	21.536	22.381	40,66	281	31,2	AAAGGAAGGAAAATAGTATG	Lisina	128	281/281(100%)	ALJ99644.1

25	23.103	23.306	37,75	67	7,6	GAAGGAGGAACAAAATG	Regulador tipo Cro	phiAbc2	67/67(100%)	YP_003347432.1
26	23.518	24.210	34,92	230	25,2	GGAGGAGAACAAAAATG	Proteasa	128	230/230(100%)	ALJ99646.1
27	24.577	24.786	27,62	69	8,2	AAAGGAGAAACGAATG	Represor tipo Cro	ALQ132	69/69(100%)	YP_003344877.1
28	25.228	25.701	35,86	157	18,4	GTGGGAGGGTAGGAATTAATATG	-	128	157/157(100%)	ALJ99649.1
29	25.698	26.372	44,59	224	25,4	AAAGGAGAAACCTTAACATAAGATG	Proteína de unión a ADN	128	224/224(100%)	ALJ99650.1
30	26.362	27.693	39,94	443	50,2	AGAAGAGGTCTTCAATTTTG	Helicasa	128	442/443(99%)	ALJ99651.1
31	27.700	28.155	35,75	151	17,3	TATGGAGATAAAAACTATG	-	P3681	150/151(99%)	ARU13238.1
32	28.158	28.973	40,56	271	30,5	AATTGACCTTCCATTCTAATTATG	Proteína de replicación de ADN	128	271/271(100%)	ALJ99653.1
33	28.954	30.471	36,96	505	59,1	CAAAGAGATAAGGAGGATTG	Primasa	128	502/503(99%)	ALJ99654.1
34	30.898	31.221	38,27	107	12,1	TTGCGAGCATTTATAAGGAATG	-	128	107/107(100%)	ALJ99655.1
35	31.199	31.444	36,99	81	9,6	ACTGGAGATAGTTG	-	128	80/80(100%)	ALJ99656.1
36	31.435	31.608	36,21	57	6,5	GAAAGAGATGATAGAACTATG	-	128	57/57(100%)	ALJ99657.1
37	31.605	31.760	26,28	51	6,3	GAAGGAGATTAGTAGATTTATG	-	128	51/51(100%)	ALJ99658.1
38	31.761	32.258	37,15	165	18,9	GGTTGAGGTAGAATAAATG	Proteína de unión a ADN	128	165/165(100%)	ALJ99659.1
39	32.230	32.511	39,01	93	10,4	AAAGGAGTAATGATTG	-	128	2/93(99%)	ALJ99660.1
40	32.508	33.215	39,83	235	27,9	AAAGGGAGGGGACAATG	-	128	235/235(100%)	ALJ99661.1
41	33.362	33.880	35,84	172	20,2		Regulador transcripcional	P3681	143/172(83%)	ARU13248.1
42	33.990	34.499	47,65	169	19,5	AAAGGAGGCATGCCAATG	Endonucleasa HNH	128	169/169(100%)	ALJ99663.1

Tabla 3.13. Características generales del fago ST

Fago ST ORFs y atributos generales										
ORF	Inicio	Fin	%GC	Tam (aa)	MW (kDa)	RBS AAAGGAGGTGA.....Start	Función	BLAST con fagos de <i>Streptococcus</i>	% identidad	No de acceso
1	88	546	40,74	152	17,2	TTGAGAGGAGAAACGATGAGTG	Terminasa, subunidad menor	P9901	151/152(99%)	ARU14694.1
2	807	2.441	38,59	544	62,6	TCAGTTGATGCATTTAACCGTTTCTT G	Terminasa, subunidad mayor	P9901	523/544(96%)	ARU14695.1
3	2.642	3.802	40,4	386	42,7	AAAGGAGGTGATAACAATTG	Proteína portal	73	377/386(98%)	ALJ99540.1
4	3.789	4.457	38,57	222	24,6	AAGGAGGTGAGATAAATG	Proteína de andamiaje	P3681	221/222(99%)	ARU13212.1
5	4.472	5.665	37,94	397	44,2	AAAGGAAAATAATTAATG	Cápside, subunidad mayor	73	397/397(100%)	ALJ99542.1
6	5.680	5.994	43,49	104	11,6	TTAGGAGGTAAGCTATG	Proteína de empaquetamiento de ADN	P7574	99/100(99%)	ARU13901.1
7	5.994	6.344	39,89	116	13,5	GAAAGAGGTGACTAATG	Proteína de unión cápside-cola	phiAbc2	116/116(100%)	YP_003347417.1
8	6.351	6.770	43,33	139	15,6	AAGTTGGGTGATGGCTTATG	Proteína de cola	P7602	138/139(99%)	ARU14001.1
9	6.778	7.149	32,53	123	14,2	AAGGGAGGGTAGTAATTAAGTG	Proteína de cola	vB_SthS_VA214	122/123(99%)	AUN43393.1
10	7.168	7.776	40,39	202	21,8	AAGGAGAAAATATATATG	Proteína de cola	P3681	201/202(99%)	ARU13218.1
11	7.850	8.203	38,7	117	13,5	AAAGGAGTAAAGACCACAATG	Proteína de cola	P7151	115/117(98%)	ARU13633.1
12	8.251	8.391	36,17	46	5,6	AATCGCATTGAAGATTTG	-	P7151	45/47(96%)	ARU13634.1
13	8.422	13.203	40,44	1593	174,7	AAAGGAGGGAATATAACATG	Proteína de cola	vB_SthS_VA698	1531/1593(96%)	AUN43331.1
14	13.200	14.765	40,23	521	58,5	TTAGGAGGTAGCATTATTTG	Proteína de cola	P3681	499/521(96%)	ARU13222.1
15	14.744	18.184	39,76	1146	128,0	CCAACAATTGAAATTTTCATG	Receptor viral	P3681	1063/1152(92%)	ARU13223.1
16	18.184	20.196	40,83	670	74,7	GTAGGAGGTGCATAATG	Cola, subunidad menor	P7601	626/670(93%)	ARU13961.1
17	20.166	20.621	35,53	151	17,4	AAAGGGTGAATTTTGTG	-	ALQ132	125/136(92%)	YP_003344866.1
18	20.641	20.784	31,25	47	5,6	ATAGGAGGGATGCGTTATG	-	P5651	46/48(96%)	ARU13421.1
19	20.805	21.128	33,64	107	12,5	GAGAGGATGAAAAATAAATG	-	TP-J34	99/107(93%)	YP_007392299.1
20	21.136	21.378	34,98	80	8,8	TGAGAGGATGAAAAATAAATG	Holína	TP-J34	73/79(92%)	YP_007392300.1
21	21.380	22.225	43,14	281	31,1	CAAGGAAGGAATATAGTATG	Lisina	53	261/281(93%)	ALJ99604.1
22	22.705	23.040	36,9	111	12,3	GAGAGAGGTAACAAAATG	-	vB_SthS_VA214	111/111(100%)	AUN43379.1
23	23.066	23.356	33,68	96	11,1	AAAGGAGGAAAATAACATG	-	P7602	93/96(97%)	ARU14017.1
24	23.384	23.935	37,86	183	21,5	AAAGGAAGTAAACAAAATG	-	P5641	168/183(92%)	ARU13374.1

25	24.122	24.322	37,81	66	7,6	AGAGGAGGAACAAAATG	Regulador tipo Cro	P7633	67/67(100%)	ARU14166.1
26	24.484	24.672	35,45	62	7,8	AAAGGAAAATACTATG	-	9873	63/63(100%)	AMQ65822.1
27	24.830	24.946	34,19	38	4,5	CATCGAGGGCTGTGTG	-	-	-	-
28	24.952	25.425	36,71	157	18,1	TAGGGAGGGTAGGAATTAATATG	-	vB_SthS_VA460	143/157(91%)	AUN43437.1
29	25.422	26.111	40,43	229	25,7	AAAGGAGAAACCTTAACATAAGATG	Proteína de unión a ADN	TP-778L	228/229(99%)	YP_008772070.1
30	26.080	27.492	37,01	470	54,2	AAAGGGGTGTAAGGTAGATG	Helicasa	TP-778L	468/470(99%)	YP_008772071.1
31	27.499	27.972	33,97	157	18,2	TTTGGAGATAAAAAACATG	-	858	156/157(99%)	YP_001686832.1
32	27.986	28.792	37,67	268	30,3	TAAGACTATGGTAGGGATG	Primasa	858	262/268(98%)	YP_001686833.1
33	28.761	30.317	37,44	518	60,5	AAAGGAGTTGGACACTAAACATG	Primasa	TP-778L	513/518(99%)	YP_008772074.1
34	30.142	30.429	32,64	95	11,5	TACTAAATTGACAAAACGCAAGTTTG	-	-	-	-
35	30.561	30.881	36,45	106	12,1	AAAGGTAAATTTAAGTG	Proteína de unión a ADN	858	103/106(97%)	YP_001686835.1
36	30.865	31.101	37,55	78	9,3	TAAGGAAGCTTTGGATATAGTAAATG	-	vB_SthS_VA214	73/78(94%)	AUN43361.1
37	31.418	31.930	37,23	170	19,6	GGTTGAGGTAGAATAAATG	Proteína de unión a ADN	P9902	156/170(92%)	ARU14785.1
38	31.927	32.790	40,62	287	31,9	TAAGGAGGAAGACCTTTG	ADN citosina metiltransferasa	CHPC1151	286/287(99%)	APC45910.1
39	32.078	32.497	42,38	139	15,8	GAAGGAGAAATAGAGCTACATG	ADN citosina metiltransferasa	P0091	139/139(100%)	ARU12998.1
40	32.787	33.080	43,54	97	10,6	TGCGAGGTATTTATG	-	P0091	92/95(97%)	ARU12999.1
41	33.077	33.355	34,05	92	10,3	AAAGGGAGAGGACGAGGAAAATG	-	IPP66	75/92(82%)	APD24440.1
42	33.368	34.075	37,99	235	27,7	TAAAGAGGTGTAGATATG	-	9874	209/235(89%)	YP_009280373.1
43	34.468	34.866	38,85	132	15,5	AAAGGAAAGACAATTTATG	-	53	132/132(100%)	ALJ99622.1
44	34.978	35.487	46,47	169	19,7	CAAAGAGGAGGGAAGCCAATG	Endonucleasa HNH	P7601	168/169(99%)	ARU13992.1

En este capítulo se han descrito las características principales de los fagos de la colección. Como se ha venido mencionando, la información almacenada en los genomas virales es el insumo fundamental para el desarrollo de derivados resistentes. En el capítulo siguiente se abordará el análisis de esa información específica no contemplada en el alcance de éste.

4

Generación de mutantes fago-resistentes

Resumen

La selección de derivados resistentes es el objetivo central de este proyecto. La estrategia implicó analizar la diversidad y conservación de secuencias protoespaciadoras en los genomas de los fagos aislados, evaluar si un *locus* CRISPR promete mejores resultados; establecer una hipótesis de cuál es el mejor resultado esperado; y por último evaluar la generación de resistentes en un escenario real. Nos enfocamos en los fagos de tipo *cos* ya que, dentro de los límites de nuestro trabajo, son el grupo dominante en nuestro país.

Materiales y métodos

Alineamientos

El alineamiento de los genomas virales y demás secuencias de ADN fue llevado a cabo con el software *CLC Sequence Viewer* (Qiagen, CA, E.E.U.U.) utilizando el paquete MUSCLE (v3.8.31).

Mapeo de protoespaciadores

El mapeo de protoespaciadores fue realizado mediante la búsqueda de los motivos PAM para cada sistema en los genomas virales con la ayuda del software ApE (v2.0.49.10). Cada motivo PAM más la secuencia adyacente 5´ de 30 pares de bases (el protoespaciador) fue recuperado y utilizados para crear la base de datos de los análisis. La frecuencia de aparición de cada espaciador se calculó utilizando la función CONTAR.SI en Microsoft Excel (Microsoft Inc.,WA, E.E.U.U.). A cada protoespaciador se le asignó un número según su posición en el genoma relativa al sitio Cos, teniendo el número más bajo aquellos más próximos a este por el extremo 3´.

Cálculo del CAI

El índice de adaptación de codones para los genes del fago 53 fue calculado utilizando la suite CAIcal (genovés.urv.es)

Selección de mutantes fago resistentes

La selección de mutantes fago resistentes fue llevada a cabo en Agar LM17 como se describe a continuación. Una alícuota de título viral conocido se pre incubó con 0,5 mL de cultivo fresco de la cepa de interés previamente ajustada su concentración a una OD₆₀₀ de 0,6 durante 15 minutos a temperatura ambiente. Posteriormente se transfirió a un tubo con medio LM17 10mM CaCl₂ semisólido (7 g/L de agar) termostatzado a 45°C. Se aplicó vortex durante 5 segundos y se vertió sobre una placa con medio Agar LM17 10mM CaCl₂ y se dejó solidificara temperatura ambiente por cinco minutos. Posteriormente se incubaron las placas durante 48-72 horas a 42°C hasta observar la aparición de colonias aisladas. Las colonias aisladas fueron recuperadas y cultivadas durante 16 horas en caldo LM17 a 42°C.

Posteriormente fueron congeladas para preservarlas hasta la realización de los subsiguientes ensayos.

Cálculo de la eficiencia de plaqueo (E.O.P.)

El cálculo del E.O.P. se realizó siguiendo el protocolo para la evaluación de la sensibilidad explicado en la sección materiales y métodos del capítulo 2 con las siguientes modificaciones: una vez solidificado el agar con el cultivo bacteriano, se sembraron 10 μ L de diluciones del *stock* de fagos titulado previamente a 10^{10} ufp/mL, de forma de asegurarnos sembrar hasta la extinción.

Resultados

Protoespaciadores

La inmunidad mediada por CRISPR-Cas, implica el reconocimiento por homología del ADN invasor por una pequeña molécula de ARN que guía un complejo enzimático encargado de degradarlo. Como se explica en el capítulo 2, *S. thermophilus* posee dos *loci* CRISPR funcionales que presentan ciertas diferencias, entre las que se encuentra la secuencia del PAM. Esto determina que, para cada genoma viral existe un *subset* de potenciales protoespaciadores “adquiribles” por el sistema CRISPR1, y otro diferente para el sistema CRISPR3. La pregunta central a la hora de seleccionar derivados resistentes es cuál espaciador es el mejor.

Mapeo de protoespaciadores

El contenido de protoespaciadores de los genomas virales de la colección, es el insumo fundamental para el diseño de derivados resistentes. Conocer cuál o cuáles presentan mejores perspectivas de otorgar resistencia duradera hace la diferencia entre una selección azarosa y una selección eficiente. El criterio principal para categorizar un espaciador como “bueno” o “malo”, es cuan probable es el evento de que el protoespaciador homólogo vaya a sufrir una mutación inactivante. Teóricamente una única mutación en la secuencia protoespaciadora en el fago es suficiente para que éste pueda eludir la resistencia, aunque hay autores que afirman que existe cierto grado de tolerancia para las mutaciones en algunas regiones del protoespaciador, y hablan de una secuencia “semilla” fundamental para el funcionamiento de éste (Semenova et al. 2011). En cualquier caso, lo que es fundamental para las etapas de adaptación e interferencia, es que exista un PAM sin mutaciones al que reconocer. Los genomas fágicos son sumamente plásticos y están sujetos a constantes cambios ocasionados por recombinación y mutación, por lo que es sumamente importante encontrar aquellas secuencias que sean fundamentales para el funcionamiento del fago y por ende tengan baja probabilidad de mutación.

Para encontrar estas regiones, nuestra primera aproximación fue realizar un alineamiento de todos los genomas virales tipo cos de la colección, a fin de ver que partes de los genomas se encuentran más conservadas. En la figura 4.1 se observa el alineamiento de

los genomas, pudiendo apreciarse que la región más conservada es la que corresponde a los genes estructurales (entre las 5 y 20 kb aprox.). Este resultado en un principio fue sorprendente, ya que la mayor diferencia entre los diferentes grupos de fagos radica en sus proteínas estructurales (fagos *Cos*, *Pac* y el recientemente descubierto grupo 987 todos presentan proteínas estructurales totalmente diferentes) y por lo tanto se podría suponer que estos genes no estarían sujetos a una presión selectiva tan estricta. Por el contrario, los módulos de replicación del ADN y lisis celular (de las 20kb en adelante) mostraron gran variabilidad, lo que también sorprende ya que en estos módulos se encuentran codificadas muchas enzimas con dominios catalíticos sensibles que podrían causar la inactivación del virus si sufrieran una mutación.

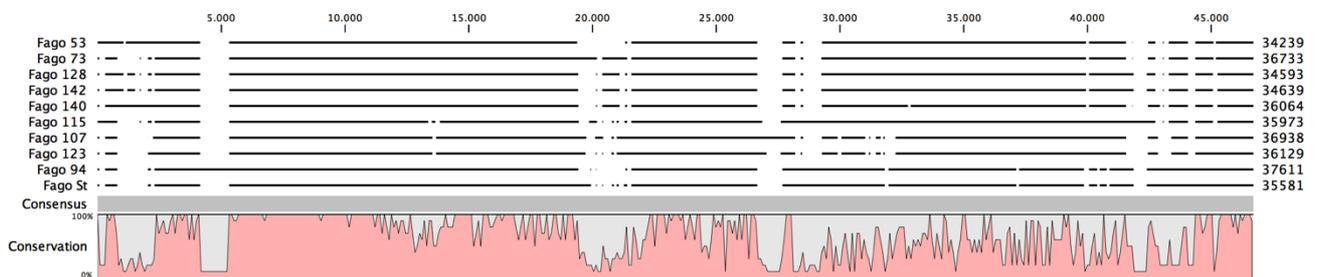


Figura 4.1. Alineamiento de los genomas tipo *cos* de la colección. El gráfico en color rosa muestra el nivel de identidad en cada región del alineamiento, asignando para cada punto un puntaje de 0 a 100% de similitud.

Si bien el alineamiento de los genomas da una pista de cuáles son las regiones más conservadas, no necesariamente brindan la información completa acerca de qué protoespaciadores se presentan como mejores candidatos. Para esto, la comparación debe darse exclusivamente en el contenido de protoespaciadores.

El objetivo es seleccionar los mejores espaciadores posibles. En este sentido, debemos entender al contenido de protoespaciadores de un genoma dado, como una parte de la diversidad total en un *pool* de posibles espaciadores. La frecuencia con que un protoespaciador dado aparece en la población, no solo es una medida de su conservación, si no de su utilidad. Un protoespaciador con una alta frecuencia en la población es de gran valor ya que será más estable (si está muy conservado puede entenderse que es menos probable que sufra alguna mutación en el tiempo) y aporta un amplio espectro de resistencia (un evento de adquisición de un espaciador conservado puede otorgar inmunidad contra otros virus con los que la bacteria no ha tenido contacto antes). Un espaciador que aparezca con baja frecuencia en la población será menos útil ya que aportará resistencia contra menor

cantidad de virus, además de ser una señal de una posible alta variabilidad lo que acotaría su utilidad temporal.

Para llevar a cabo este análisis de frecuencia y conservación de protoespaciadores, se mapearon en los genomas de los fagos todas las posibles secuencias plausibles de ser adquiridas por ambos sistemas CRISPR1 y CRISPR3. Estas secuencias, para el caso de *S. thermophilus* son fragmentos de 30 pares de bases que se encuentren en el extremo 5' adyacente a un motivo PAM NNAGAA(A/T) para CRISPR1 y NGGNG para CRISPR3. Una vez mapeados todos los posibles protoespaciadores (en el orden de 250 protoespaciadores para CRISPR1 y 450 para CRISPR3 en cada genoma), se calculó la frecuencia individual de cada uno en la población. La figura 4.2 muestra las frecuencias de protoespaciadores CRISPR1 y la figura 4.3 la frecuencia de protoespaciadores CRISPR3. En ambas figuras, los protoespaciadores se encuentran ordenados por orden de aparición en el genoma tomando como punto de partida (protoespaciador número 1) aquel que se encontrara más cerca del sitio *cos*. La frecuencia expresada en escala del uno al diez según el número de veces que aparece en la población fue codificada en una escala de color para buscar patrones de conservación. El resultado a la vista para ambos sistemas es que parecen existir regiones donde se agrupan espaciadores de alta frecuencia. La existencia de estas zonas es un aspecto muy favorable, ya que hablan de un contexto genómico estable de donde se pueden tomar espaciadores de una frecuencia dada con mayor confianza que espaciadores con frecuencias similares de regiones menos conservadas. En otras palabras, los protoespaciadores con frecuencia alta pero en un entorno genómico con protoespaciadores predominantemente de baja frecuencia podrían ser menos estables en el tiempo, a no ser que se tratara de una secuencia ubicada en un motivo crítico. Para estar seguros de esto se debería estudiar caso por caso, pero como regla general, tenderemos a confiar en los espaciadores adquiridos de regiones con alta conservación.

Protoespaciadores CRISPR1

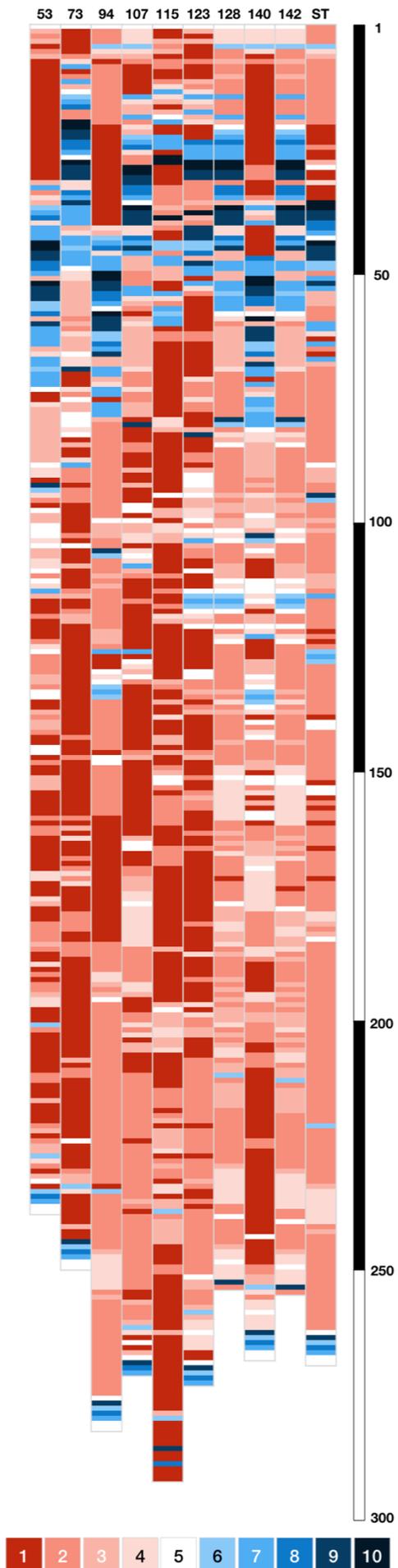


Figura 4.2. Representación gráfica de los protoespaciadores CRISPR1 de los diez genomas de fagos tipo cos de la colección. Los protoespaciadores se encuentran ordenados desde arriba hacia abajo comenzando por el protoespaciador más próximo al sitio cos. Los colores representan la frecuencia con la que aparecen en la población. En la parte inferior de la figura se encuentra detallado el esquema de color desde el uno al diez.

Protoespaciadores CRISPR3

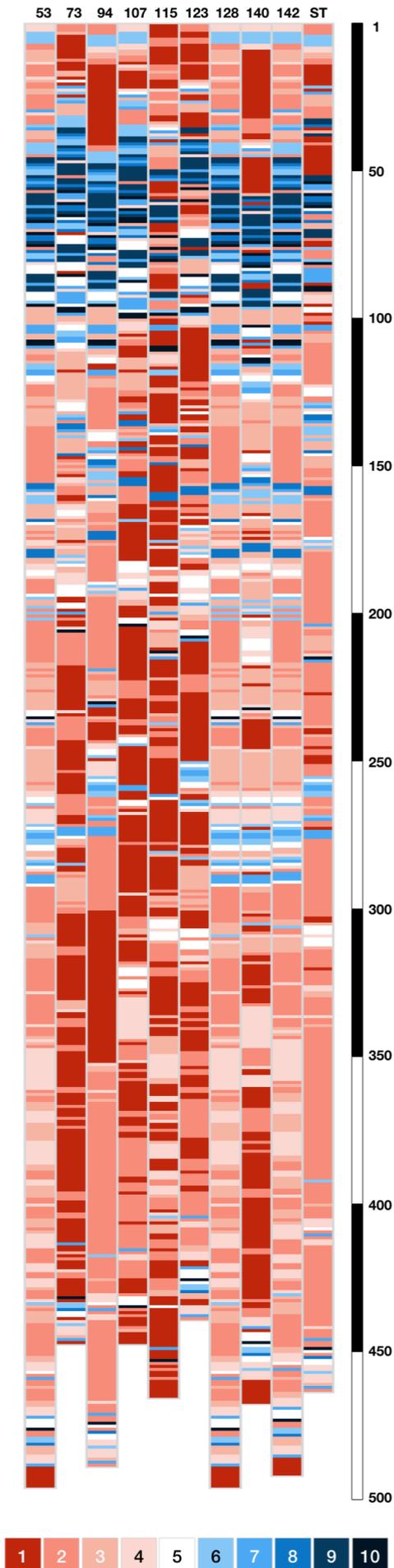


Figura 4.3. Representación gráfica de los protoespaciadores CRISPR3 de los diez genomas de fagos tipo *cos* de la colección. Los protoespaciadores se encuentran ordenados desde arriba hacia abajo comenzando por el protoespaciador más próximo al sitio *cos*. Los colores representan la frecuencia con la que aparecen en la población. En la parte inferior de la figura se encuentra detallado el esquema de color desde el uno al diez.

Con el fin de evaluar la presencia de regiones conservadas con acumulación de protoespaciadores de alta frecuencia, se graficaron las frecuencias de cada protoespaciador en función de su posición en el genoma, y se añadieron líneas de tendencia. Para evitar sesgos en la tendencia dados por la sobreestimación del peso de pocos protoespaciadores particularmente frecuentes en regiones de baja conservación, se graficó la media móvil en cada caso, tomando como ventana diez protoespaciadores. De este modo cada punto en la línea de tendencia surge del promedio de 10 frecuencias y se va desplazando de a un punto por vez, eliminando así el peso excesivo que pudiera otorgar un espaciador muy frecuente en un contexto de baja frecuencia. Se observó con claridad para prácticamente todos los casos que existe una región muy favorable donde se da un pico en la acumulación de espaciadores de alta frecuencia. Esta zona coincide para todos los genomas y se encuentra aproximadamente comenzando al final del gen codificante para la subunidad grande de la terminasa, y se extiende hasta el comienzo del gen que codifica el componente más grande de la cola. En esta zona quedan comprendidos todos los genes codificantes para las proteínas de la cápside y algunos de los genes codificantes para proteínas de la cola. Este resultado termina de confirmar que los genes más conservados y estables a nivel de ADN para estos fagos son, justamente genes estructurales. En las figuras 4.4 a 4.13 se muestran las gráficas de frecuencia de protoespaciadores en función de la posición para cada uno de los genomas. En gris se encuentra sombreada la región ideal para la adquisición de nuevos espaciadores.

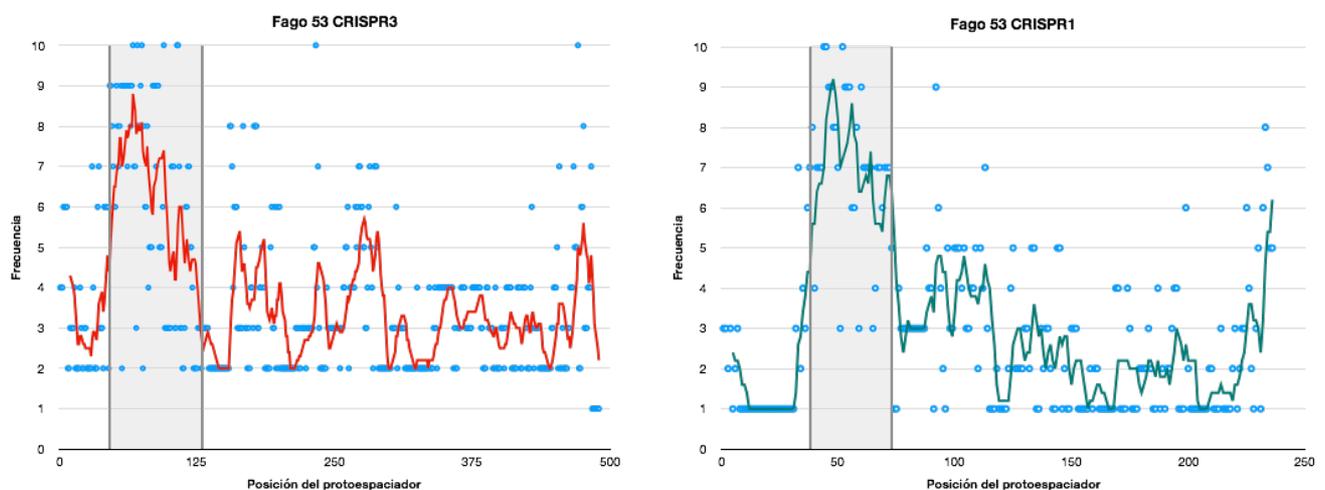


Figura 4.4. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 53. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

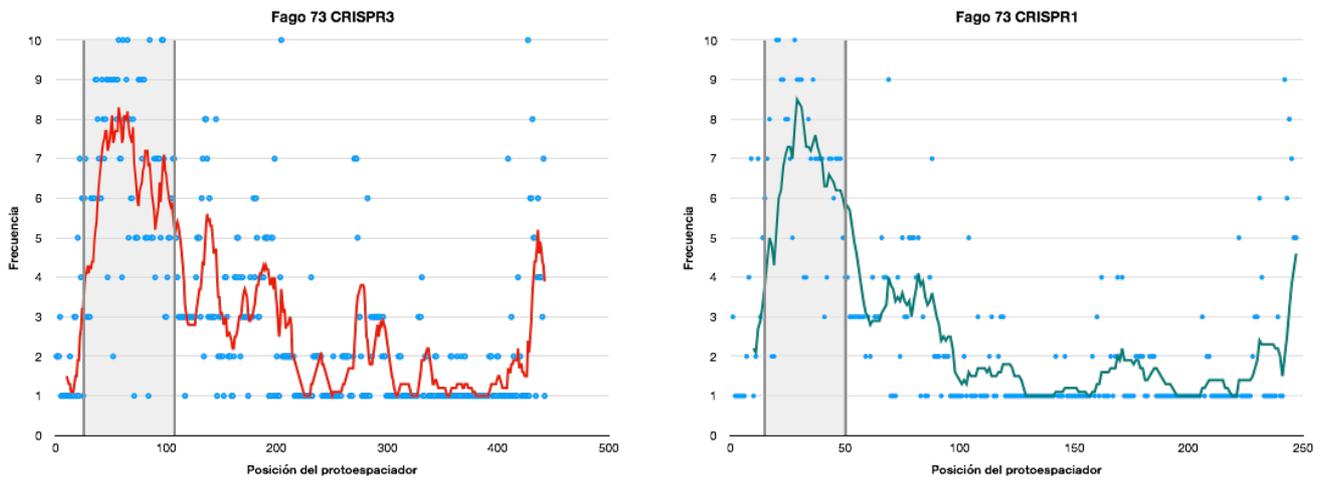


Figura 4.5. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 73. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

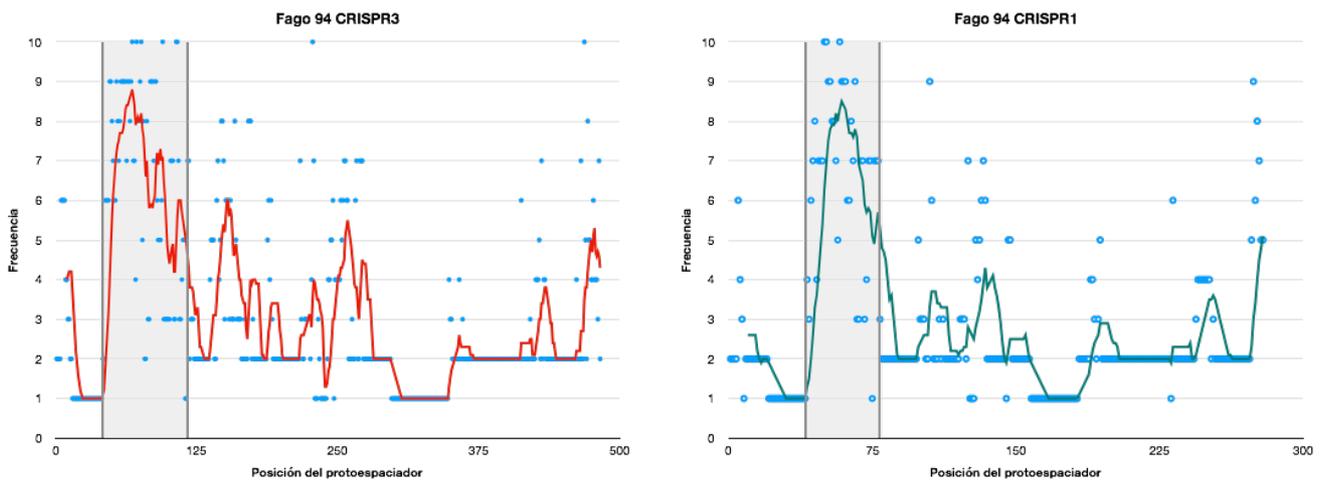


Figura 4.6. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 94. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

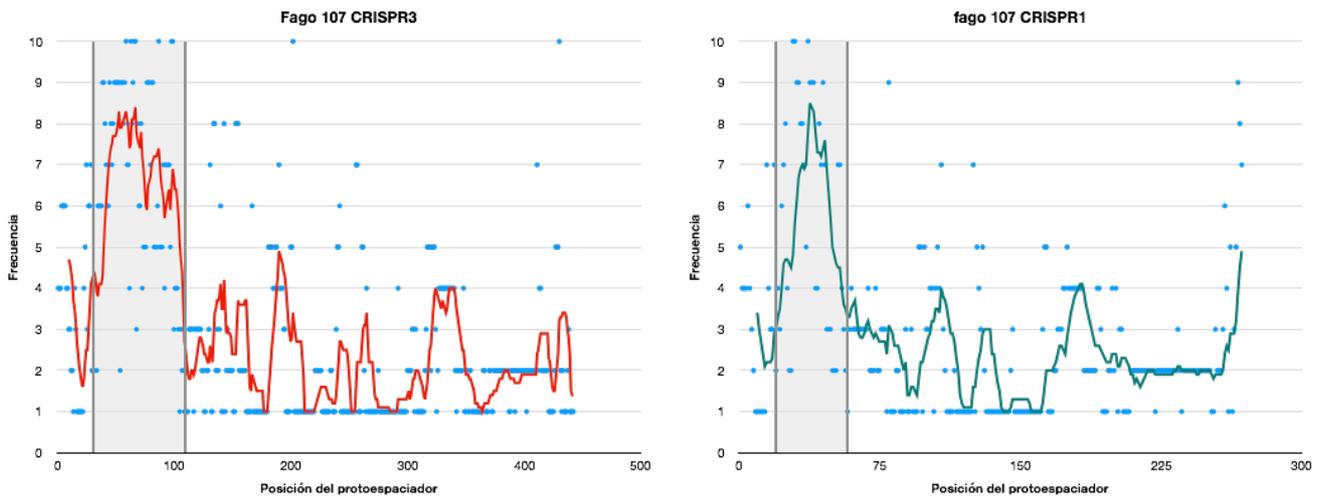


Figura 4.7. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 107. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

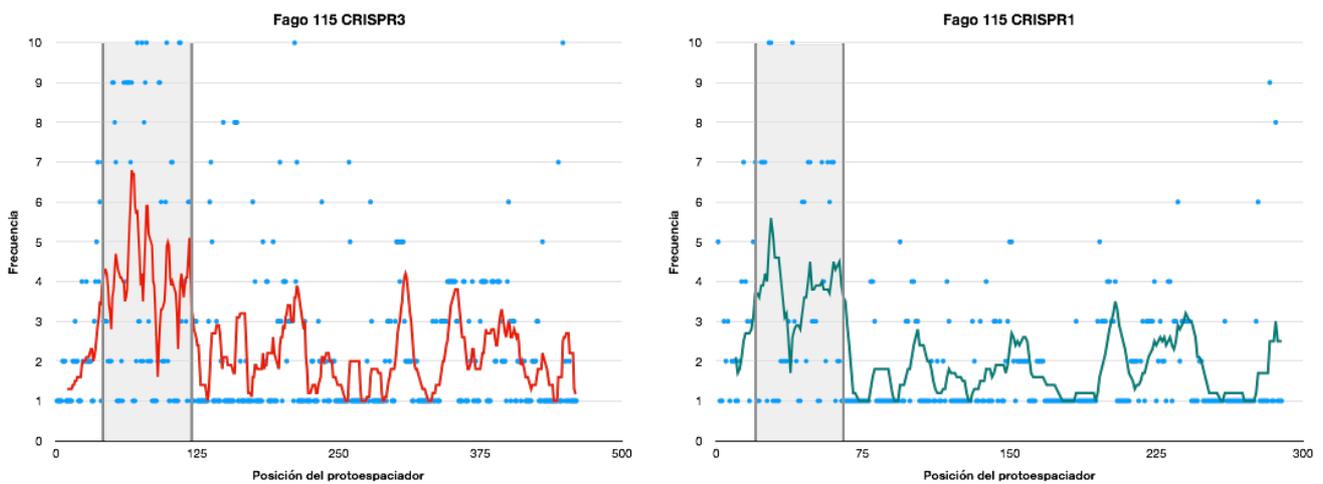


Figura 4.8. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 115. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

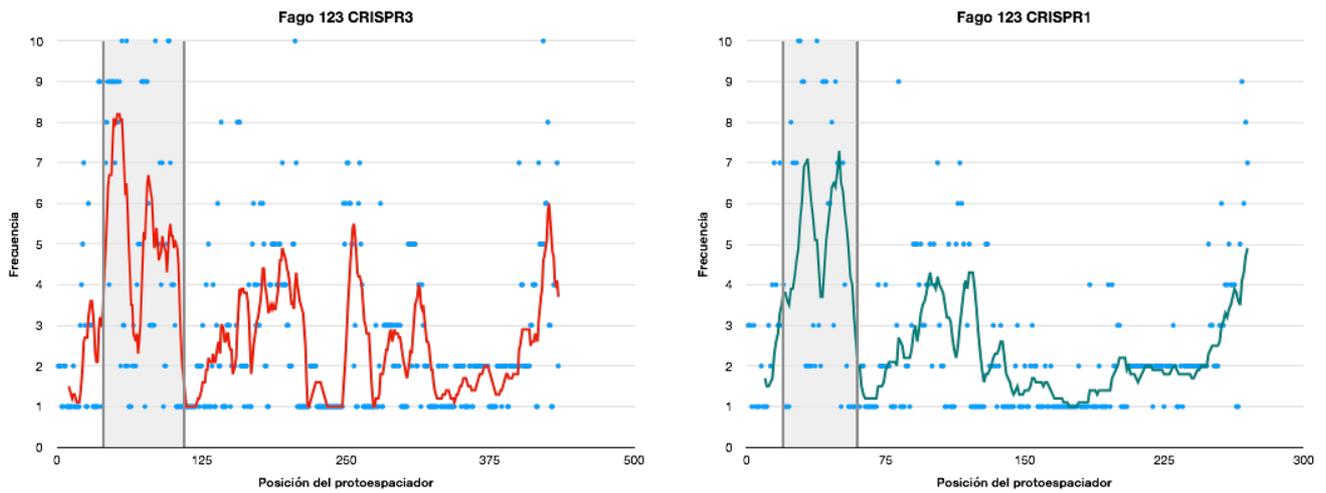


Figura 4.9. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 123. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

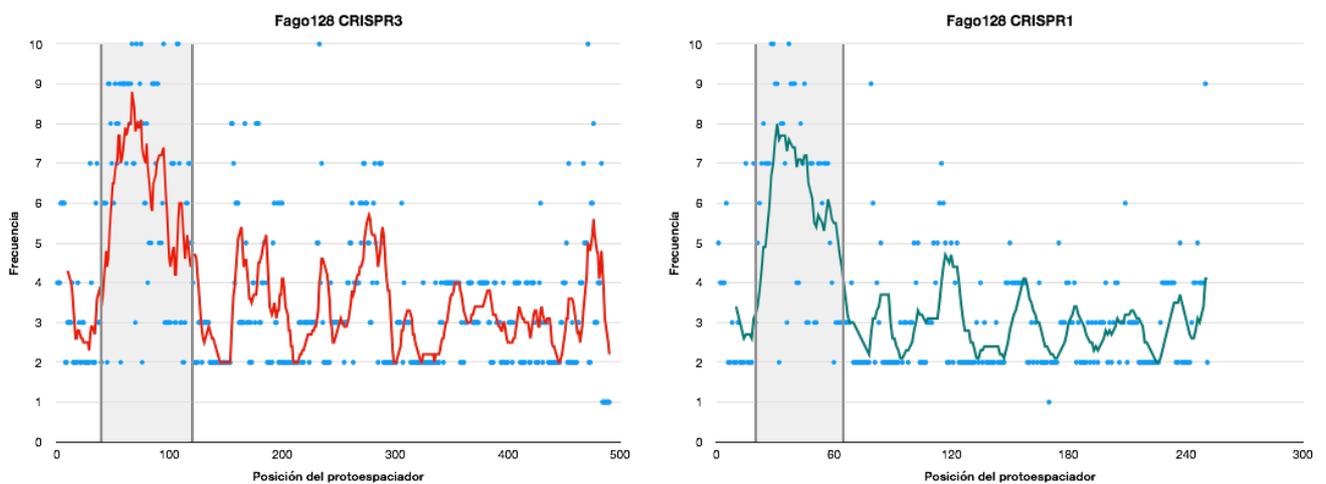


Figura 4.10. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 128. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

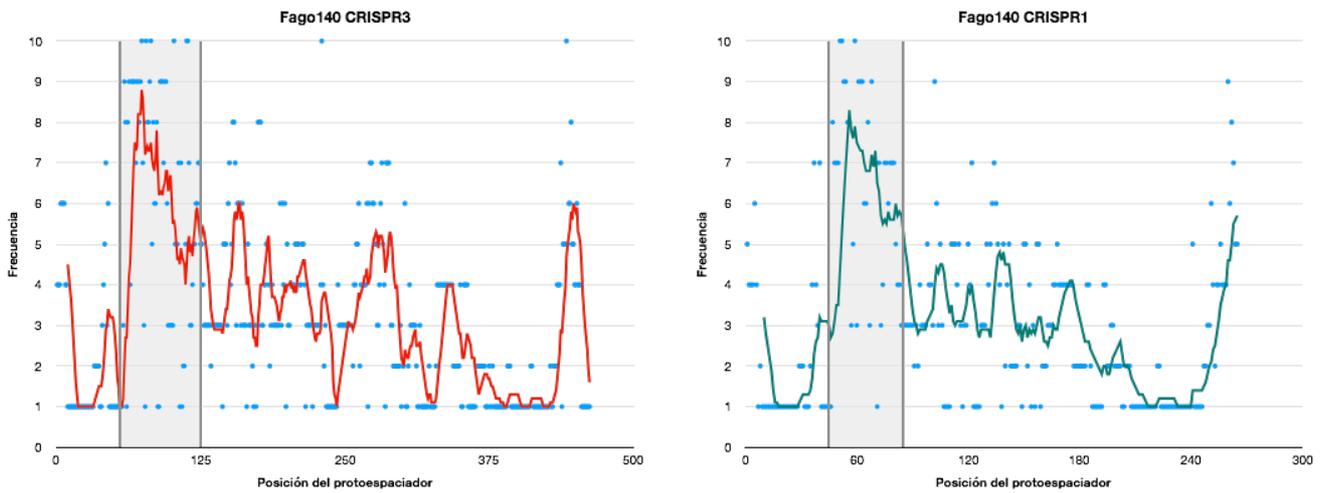


Figura 4.11. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 140. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

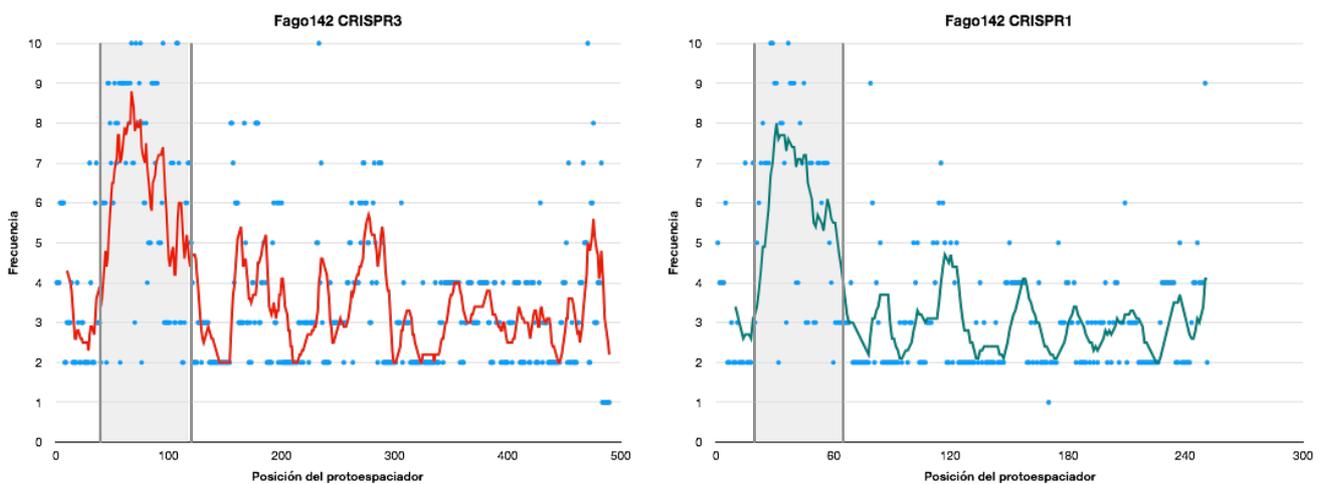


Figura 4.12. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago 142. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

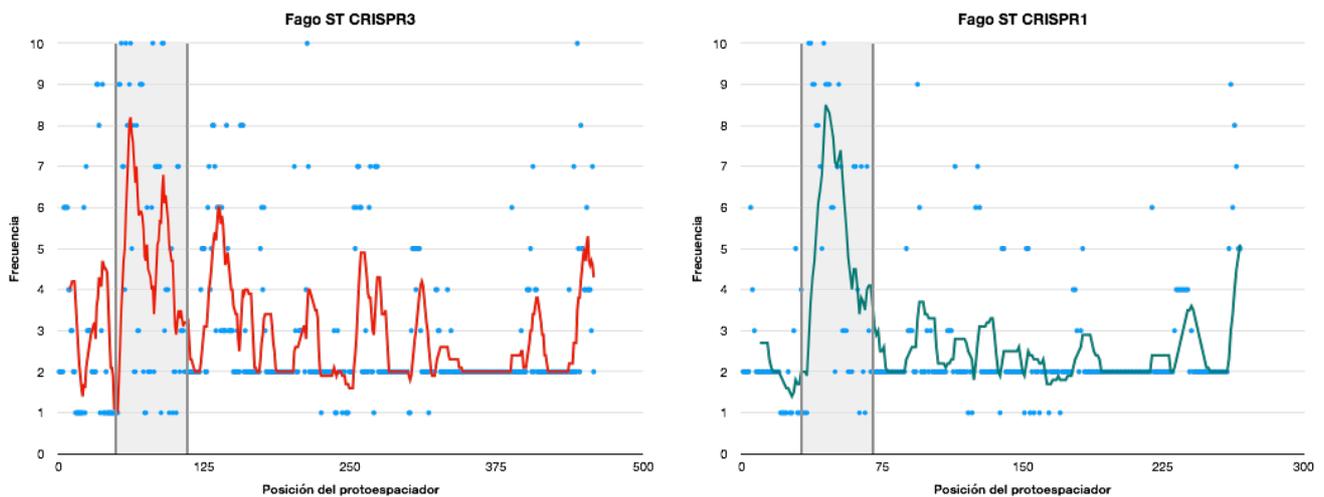


Figura 4.13. Frecuencia de protoespaciadores CRISPR 3 (izquierda) y CRISPR1 (derecha) para el fago ST. Las líneas roja y verde muestran la media móvil para cada tipo de sistema respectivamente. En gris se encuentra sombreada la región donde la densidad de protoespaciadores de alta frecuencia es mayor.

Búsqueda de sesgos en la adaptación

La siguiente pregunta es, como es el comportamiento natural con respecto a la adaptación. Como se demostró anteriormente en este capítulo, los protoespaciadores de ambos sistemas se encuentran distribuidos de manera uniforme en los genomas virales, pero existen regiones donde éstos se encuentran más conservados y por ende presentan mejores perspectivas. Existen múltiples trabajos que intentan explicar el mecanismo de adaptación desde el punto de vista de los componentes que intervienen (Amitai & Sorek 2016; Sorek et al. 2013), pero no así de si existen regiones preferenciales de donde obtener los espaciadores. Un punto importante que si se aborda en el trabajo de Amitai y colaboradores, es como el complejo Cas1-Cas2 obtiene el material necesario para generar el nuevo espaciador. En este trabajo se elaboran dos teorías de cómo puede funcionar el sistema utilizando como modelo el genoma de *Escherichia coli*; una en la cuál la fuente de ADN son restos producidos por errores en la replicación y la otra es que se une a fragmentos de ssADN y una polimerasa de origen desconocido intervenga y forme el intermediario dsADN. Este punto es de menor importancia (y no existe aún evidencia experimental que soporte a ninguna de las dos teorías), pero lo que si encontraron es que existía un sesgo en la adquisición de espaciadores en regiones del genoma donde la velocidad de replicación es menor, es decir, donde la horquilla de replicación se enlentece (Amitai & Sorek 2016). Esto da una pista sumamente interesante, ya que si extrapolamos estos hallazgos al modelo de fagos de *S. thermophilus*, se podría predecir si existe un posible sesgo.

Además de la horquilla de replicación, el ADN es desnaturalizado en el proceso de transcripción, por lo que, siguiendo el mismo principio mencionado el párrafo anterior, la transcripción también podría exponer el material necesario para que el complejo Cas1-Cas2 lleve a cabo la adquisición de espaciadores, y, aquellos genes que sean transcritos de forma lenta favorecerían este proceso. Para buscar si existe algún sesgo de este tipo, se calculó el índice de adaptación de codones (CAI por sus siglas en inglés) de todos los genes codificados en los genomas virales de la colección. El CAI es una medida del sesgo en el uso de codones; es calculado mediante la comparación de los codones del gen a estudiar con una matriz de uso de codones generada a partir de genes de alta expresión. El fundamento de esta aproximación radica en que los genes deben “competir” por recursos celulares a la hora de expresarse, siendo los más importantes los ARNs de transferencia. Los genes de alta expresión estarán formados por codones cuyos tARNs complementarios serán los más abundantes en la célula. El cálculo del CAI por lo tanto, es un reflejo de cuan próximo se encuentra el contenido de codones del gen en cuestión con el óptimo del organismo. El resultado es un número entre 0 y 1, siendo los valores más próximos al 1 los que reflejan los mejores resultados. Para el caso de los fagos, es importante que sus genes tengan un uso de codones similar a los genes de alta expresión de su bacteria hospedera, de manera de llevar a cabo su ciclo vital de la forma más eficiente posible.

En la figura 4.14, se muestran a modo de ejemplo los valores de CAI para los ORFs del fago 53. Puede observarse que no hay grandes diferencias en los valores para cada gen, manteniéndose en promedio en el orden de 0,69. Como referencia aparecen los valores de dos genes de alta expresión de *S. thermophilus*, la ADN polimerasa I y la ADN polimerasa III con valores de 0,676 y 0,725 respectivamente. Este resultado muestra que: los genes virales tienden a expresarse como genes de alta expresión de su hospedero, y más importante para este estudio, no existen regiones diferenciadas que puedan llegar a enlentecer la horquilla a causa de la transcripción.

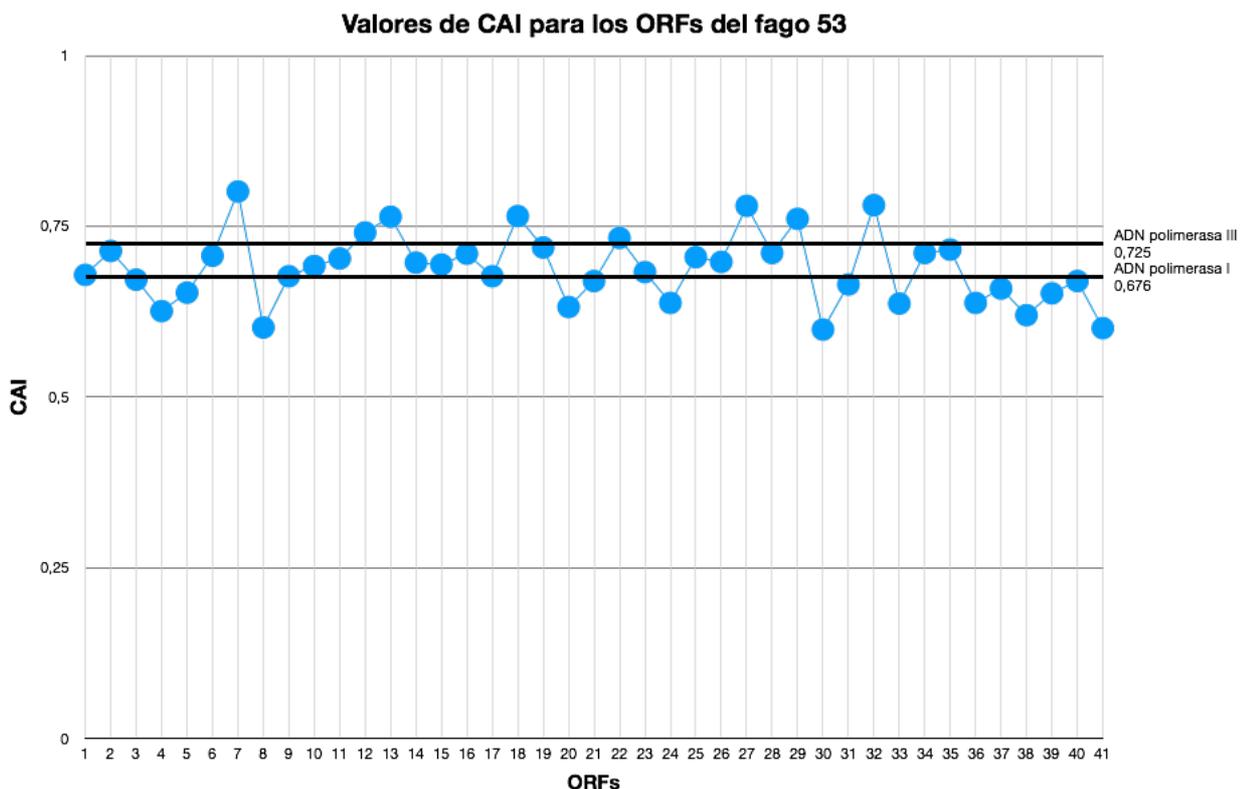


Figura 4.14. Valores de CAI para cada gen del fago 53. Las líneas negras sirven como referencia de genes de alta expresión y representan el valor de CAI para la ADN polimerasa I y la ADN polimerasa III de *Streptococcus thermophilus*.

El siguiente elemento que podría impactar sobre la adquisición de espaciadores es el factor temporal. La horquilla de replicación/transcripción comienza en una región específica del genoma viral y avanza conforme al ciclo viral. El trabajo de Duplessis y colaboradores de 2005, determinó para el fago modelo DT1 el marco temporal de expresión de sus genes (Duplessis et al. 2005). Los diferentes ORFs se agruparon según el momento en que se daba su expresión, en tempranos, medios y tardíos. Como es la característica general de los *Caudovirales* (os fagos de *S. thermophilus* demostraron no ser la excepción) la organización espacial del genoma está directamente relacionada con el marco temporal de expresión de los genes, por lo que los genes tempranos, medios y tardíos se encuentran agrupados. Un hallazgo importante de este trabajo fue determinar la región por la cual comienza la expresión de los genes. Datos experimentales mostraron que ésta comienza por el ORF 28, una proteína hipotética ubicada entre el gen codificante para un endolisina y el gen codificante del represor *cro*. Por tanto, si la transcripción jugara un rol en la adaptación, debería observarse una densidad de espaciadores mayor en los genes tempranos, y disminuyendo conforme se avanza a los medios y tardíos.

Por último, podría observarse un sesgo causado por la replicación del genoma. El modelo de replicación para los *Siphoviridae* es el fago lambda. En este virus, la replicación se da en dos etapas con estrategias diferentes, una etapa llamada *theta* y una llamada *sigma* o de círculo rodante. En las primeras etapas, el ADN circularizado comienza a replicarse de forma bidireccional desde el origen de replicación, formando una estructura tipo *theta*. Esta forma de replicación se da en las etapas tempranas de la infección. Luego, por razones no del todo establecidas, la estrategia cambia a la de círculo rodante, mecanismo por el cual se forman múltiples copias del genoma unidas en una superestructura llamada concatémero (Figura 4.15). Este concatémero comienza a empaquetarse en las cápsides y es cortado en los sitios *Cos*, garantizando de esta forma que una copia del genoma sea empaquetado por cápside y generando los extremos cohesivos típicos de este tipo de fago.

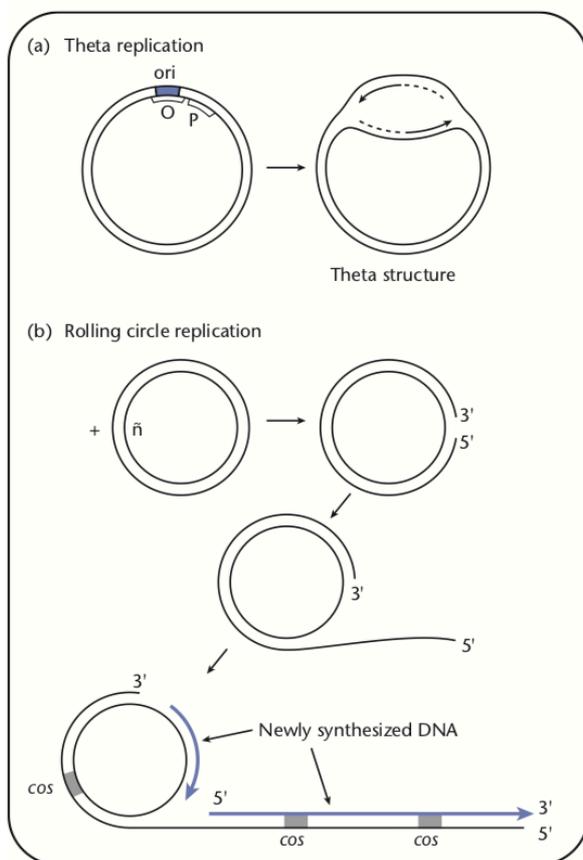


Figura 4.15. Tipos de estructuras formadas durante la replicación del genoma de fagos tipo *Cos*. En una primera etapa se formará una estructura tipo *theta* y la replicación se dará de forma bidireccional. En una segunda etapa, la replicación adopta una estrategia de círculo rodante formándose la estructura de concatémero (Trun & Trempey 2009).

Siguiendo este modelo, y asumiendo que la adaptación debe ocurrir en las primeras etapas de la infección (pues en las últimas etapas implicaría que ya no habría genoma donde

incorporar los nuevos espaciadores) entonces el complejo Cas1-Cas2 tendría disponible para interactuar la región donde se abre la horquilla bidireccional y por ende se debería observar una densidad de espaciadores adquiridos superior en esta región que en el resto del genoma. El trabajo de Lamothe y colaboradores de 2005 caracterizó esta región *ori* para el fago DT1, ubicándola entre los ORFs 36 y 37, más específicamente en un fragmento de unas 750 pares de bases entre dos sitios de corte de la enzima *EcoRI* (Lamothe et al. 2005), la cuál, por homología, logramos ubicar en el genoma del fago 53.

Por tanto, teniendo en cuenta los datos de los trabajos de Amitai, Duplessis y Lamothe (y teniendo en cuenta que el cálculo del CAI para los genes virales no mostró ninguna diferencia entre ellos), podemos predecir que regiones del genoma viral serán las que con mayor probabilidad, el complejo Cas1-Cas2 obtenga su material si existiera un sesgo provocado por la horquilla de replicación o por la transcripción. En la figura 4.16 se muestra un esquema de dichas regiones para el fago 53.

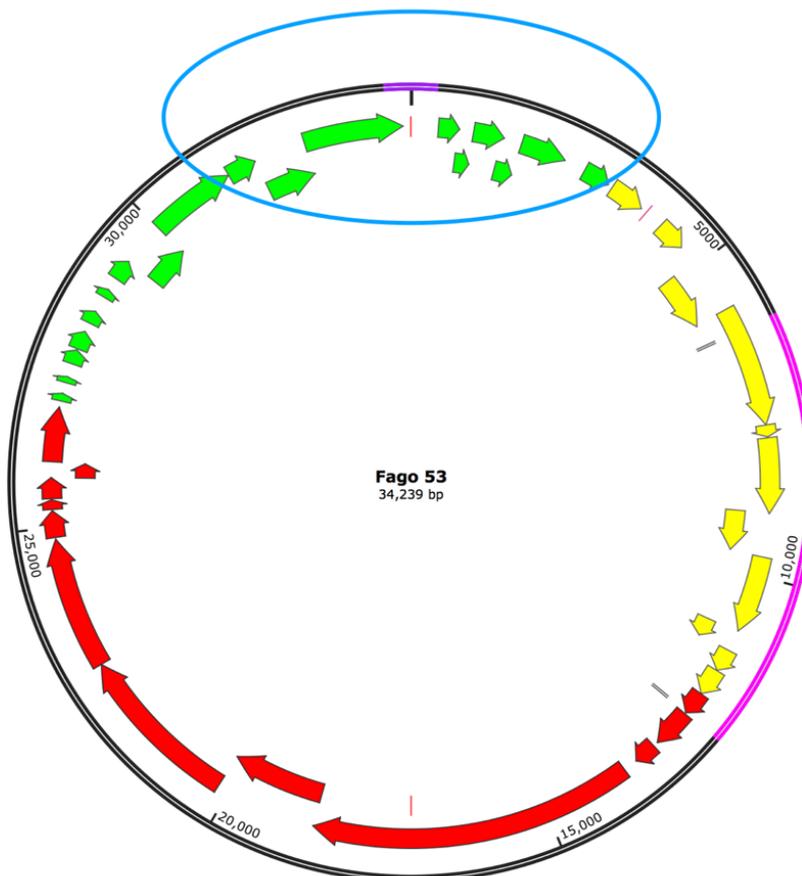


Figura 4.16. Esquema circular del genoma del fago 53. En verde se encuentran los genes tempranos, de donde deberían obtenerse más cantidad de espaciadores si existiese un sesgo dado por la apertura del ADN durante la transcripción. En amarillo se marcan los genes medios y en rojo los tardíos. El círculo azul muestra la región donde comenzará a formarse la horquilla de replicación en la estructura *theta*, de donde deberían

adquirirse mayor cantidad de espaciadores si existiera un sesgo dado por la replicación del genoma. Por último en rosado se encuentra señalada la región donde se acumulan los protoespaciadores de alta frecuencia.

De ser cierto alguno de estos escenarios, la región ideal para la adquisición de espaciadores estaría por fuera de las regiones de mayor densidad.

Generación de mutantes resistentes

Con el fin de probar como es el comportamiento natural en el proceso de adaptación, se planteó llevara cabo un ensayo de generación de mutantes fago-resistentes (BIMs por sus siglas en inglés) a gran escala. El objetivo de este ensayo era obtener en un solo experimento un número suficientemente grande de mutantes que permitiera observar la distribución de probabilidades de adquisición de espaciadores de diferentes regiones.

Utilizamos la cepa Uy01 y el fago 53 como modelo de trabajo, utilizando la metodología descrita en Materiales y Métodos. El objetivo establecido fue obtener al menos 50 cepas resistentes, de modo de analizar un mínimo de 100 eventos probables de adaptación (dado por la actividad de los *loci* CRISPR1 y CRISPR3).

Luego de repetidos intentos, se lograron generar 56 mutantes en un solo experimento, lo que nos permitió evaluar 112 posibles eventos de adaptación. En la tabla 4.1 se detallan los espaciadores incorporados por cada resistente. Los 112 *locus* CRISPR fueron secuenciados y analizados en busca de nuevos espaciadores. El primer resultado llamativo fue que todas las adquisiciones de nuevos espaciadores ocurrieron en el *locus* CRISPR3. El análisis bioinformático de los genes *cas* del *locus* CRISPR1 de Uy01 así como de la región líder, no mostraron mutaciones aparentes que pudieran afectar su funcionamiento. Se decidió hacer una prueba con la cepa Uy47 que posee el mismo *locus* CRISPR3 que Uy01 pero difieren en el *locus* CRISPR1 (ver capítulo 3) Se obtuvieron siete mutantes fago-resistentes en el ensayo control, y en todos los casos al igual que con Uy01, los nuevos espaciadores fueron adquiridos en CRISPR3 (ver Anexo). La otra opción era utilizar la cepa Uy34 que también posee el mismo *locus* CRISPR3 y diferente CRISPR1, pero ésta es resistente al fago 53 probablemente a causa de dos espaciadores en su *locus* CRISPR1 (no se pudo ensayar con un virus diferente ya que Uy34 presenta resistencia a todos los de la colección). Esto es a su vez interesante y desconcertante, ya que brindaría la prueba de que

en este subgrupo de cepas el *locus* CRISPR1 si es activo, no habiendo una explicación para lo observado con las otras dos cepas.

De los 56 clones obtenidos, 42 adquirieron espaciadores y 14 no. Para comprobar que esto no se debió a que adquirieron resistencia por otra vía, se calculó la EOP (ver Materiales y Métodos) con el fago 53 (titulado a 1×10^{10} ufp/ml) para cada uno de los clones. Los 42 clones que adquirieron espaciadores no mostraron lisis en ninguna de las diluciones (EOP=-8), mientras que las 14 que adquirieron espaciadores mostraron lisis en todas las diluciones probadas al igual que el control salvaje, dándonos la pauta de que la aparición de esos 14 clones sin cambios genotípicos ni fenotípicos fue en realidad un artefacto del ensayo.

El siguiente dato interesante, es que a pesar de que se generaron 42 clones resistentes, la cantidad de espaciadores diferentes adquiridos fue de 22, únicamente siete de los clones adquirieron espaciadores únicos, presentando las demás espaciadores repetidos entre sí. También se observaron múltiples adquisiciones ectópicas. La adquisición ectópica es la incorporación de un espaciador en una posición del locus diferente a la adyacente a la secuencia líder. Este fenómeno fue reportado en *Streptococcus thermophilus* por nuestro grupo en 2017 para el locus CRISPR1 de la cepa Uy02 (Achigar et al. 2017). En este trabajo, atribuimos como posible causa de este fenómeno una mutación puntual en la secuencia líder. En este caso, (el primero que conocemos para CRISPR3), la secuencia líder del *loci* de Uy01 no presenta ninguna mutación. Sin embargo, todas las adquisiciones ectópicas y pérdidas parciales observadas se dan en la misma posición, entre el espaciador 3 y el espaciador 4. El espaciador 4 es de mayor tamaño que el resto (32 pares de bases), pero no identificamos en principio otra característica particular. En la figura 4.17 se muestra un esquema del *locus* CRISPR3 salvaje de Uy01 y la secuencia del espaciador 4. Al realizar un alineamiento entre el espaciador y el extremo 3' de la secuencia líder encontramos que algunos residuos quedan alineados (Figura 4.18), pudiendo significar esto que tienen alguna significancia biológica, como ser sitios de apoyo para el complejo Cas1-Cas2. Sin embargo no tenemos evidencia experimental que soporte esta teoría por el momento.

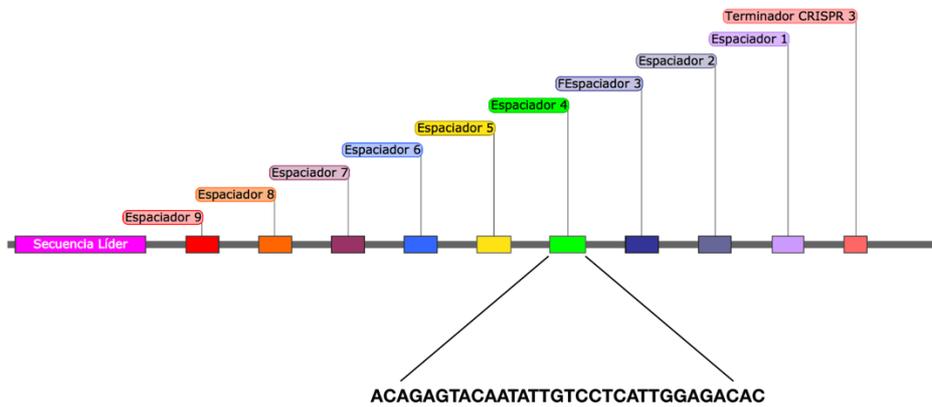


Figura 4.17. Esquema del *locus* CRISPR3 de la cepa Uy01.

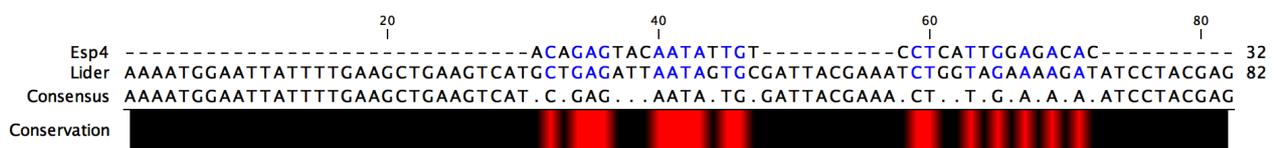


Figura 4.18. Alineamiento entre la secuencia del espaciador cuatro y el extremo 3' de la secuencia líder del *locus* CRISPR3.

Por último, llevamos a cabo el mapeo de los espaciadores obtenidos en el genoma del fago 53. En la figura 4.19 se muestra un esquema de la posición de los espaciadores. Como podemos observar, existe una polarización de la adquisición en dos regiones del genoma, una en la región cercana al origen de replicación, y otra más importante aún en la región opuesta. La polarización en la región cercana al *ori* era esperable si se cumplía el modelo explicado anteriormente, sin embargo no hay una explicación para la acumulación en la región opuesta. Serían necesarios más datos para determinar si esta tendencia se mantiene, pero aún así, lo que observamos es que como predijimos, la densidad de espaciadores adquiridos de la región más conservada del genoma del fago fue muy baja, adquiriéndose únicamente un espaciador de esta zona (Figura 4.20).

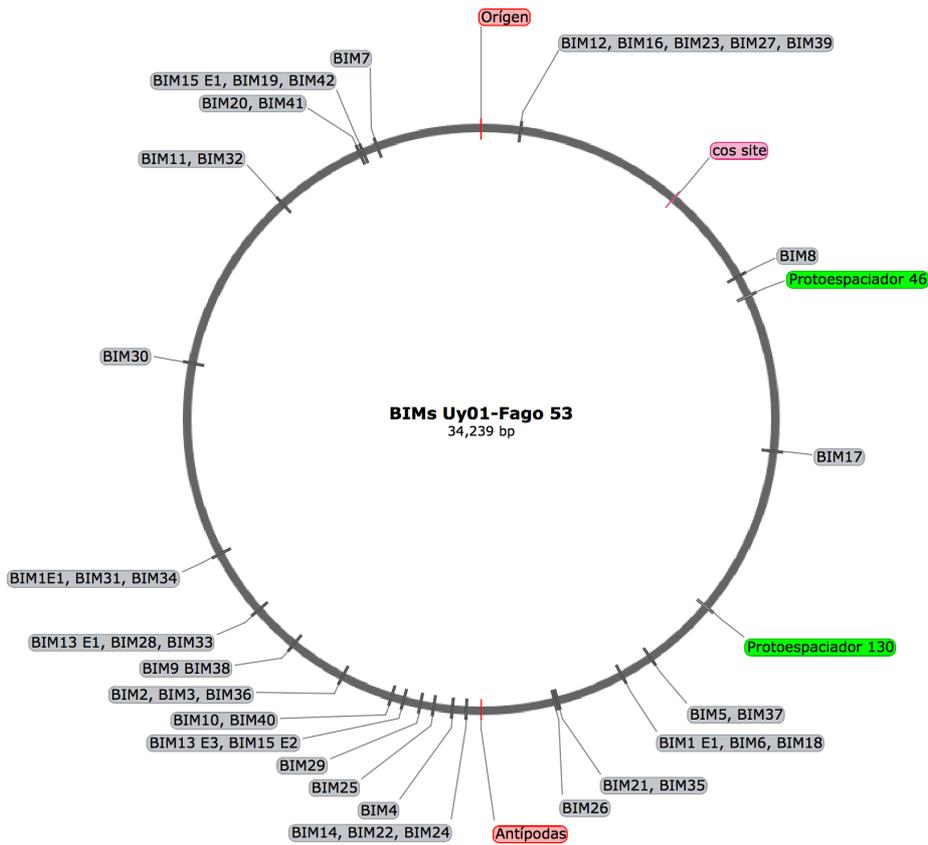


Figura 4.19. Ubicación de los espaciadores adquiridos por las BIMs de Uy01 en el genoma del fago. En verde se marcan los protoespaciadores que delimitan la región ideal donde se encuentran los protoespaciadores más conservados.

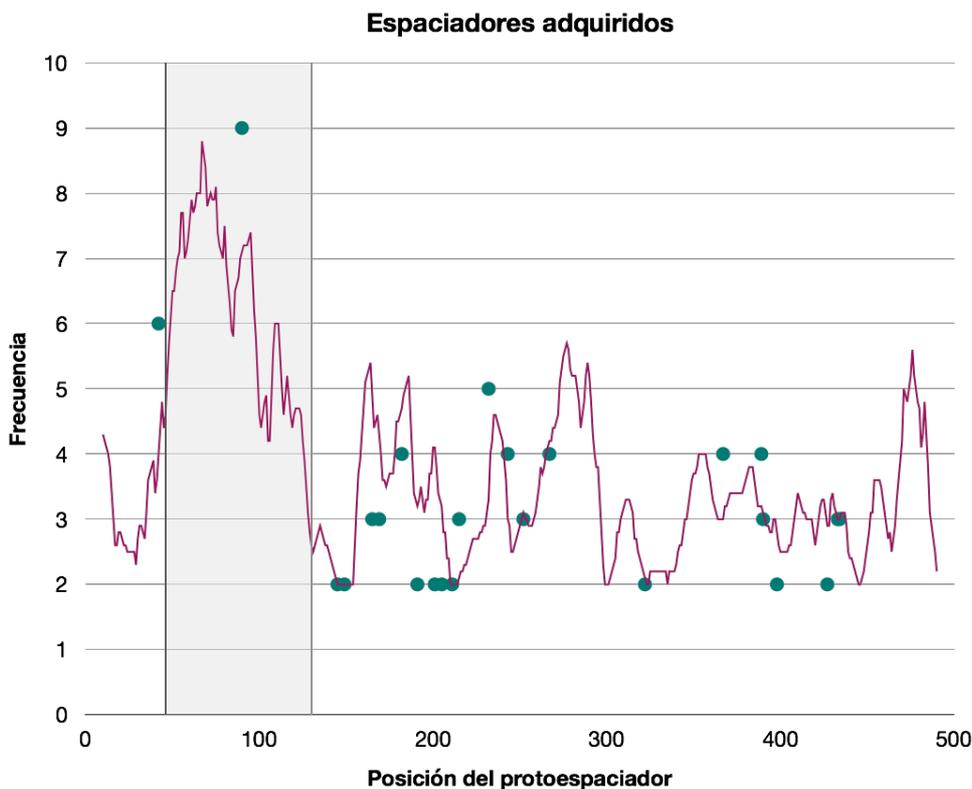


Figura 4.20. Frecuencia de los espaciadores adquiridos por las BIMs de Uy01

Tabla 4.1. Espaciadores obtenidos en el ensayo de generación de resistentes de Uy01 versus el fago 53

Nombre	Posición en CRISPR3	Secuencia	Número de Protoespaciador	Observaciones
BIM1	11	TTTATCTTTTTTCAGCGCAGTTTAACGGGTC	268	Ídem BIM31 y BIM34
	10	TGTTTCAGTATCGTCGACTTCATTCCCCAAA	150	Ídem BIM6 y BIM18
BIM2	10	ATGCGGTTTGATGATGTCGTTTCAGTACTAC	232	Ídem BIM3 y BIM36
BIM3	10	ATGCGGTTTGATGATGTCGTTTCAGTACTAC	232	Ídem BIM2 y BIM36
BIM4	10	TACTTATGATGGTGAAGATTACAATAA	191	-
BIM5	10	GAAGGCTCAAGAAAAGTCAGATGCGAAGCGC	146	Ídem BIM37
BIM6	4	TGTTTCAGTATCGTCGACTTCATTCCCCAAA	150	Ídem BIM1 y BIM36. Pérdida de espaciadores salvajes desde el 4 al 9
BIM7	10	AAGAGGTGCTTCTGTTATGCTTCCATCTTT	398	-
BIM8	10	TCATTCGTATCAGCTATATGGATGATGTAA	42	-
BIM9	10	TTTCAACGCTGGTAATTTTCATTGGACAAGA	243	Ídem BIM38
BIM10	10	TAAACGATACTTGAATCTGTTAACTGATA	215	Ídem BIM40
BIM11	10	TGAGCAGGTAGTCAATACATTCAAAGAAA	367	Ídem BIM32
BIM12	10	AGTTGATGGTAAAACGGTGAATGACCATA	434	Ídem BIM16, BIM23, BIM27 y BIM39
BIM13	12	GAAGACAGCACAAACCCAGCAGATTATACA	252	Ídem BIM28 y BIM33
	11	TCAGAATGGCTGATGGACGTTGGTTTGATAC	428	-
BIM14	10	CTACCGCTTCAAAGATAAAGCGAAAGACTC	182	Ídem BIM22 y BIM24
	10	CTCATATTCGTTTGTGGCTTTTGAATAAAA	211	Ídem BIM15
BIM15	2	TACATTCTTTGTCATTGACGTGGACATGCA	390	Ídem BIM19 y BIM42. Pérdida de todos los espaciadores salvajes
	1	CTCATATTCGTTTGTGGCTTTTGAATAAAA	211	Ídem BIM13. Pérdida de todos los espaciadores salvajes.
BIM16	10	AGTTGATGGTAAAACGGTGAATGACCATA	434	Ídem BIM12, BIM23, BIM27 y BIM39
BIM17	10	TGAAGAAATTACGCTTAATATTGCTTCAAA	90	-
BIM18	4	TGTTTCAGTATCGTCGACTTCATTCCCCAAA	150	Ídem BIM1 y BIM6. Inserción ectópica
BIM19	10	TACATTCTTTGTCATTGACGTGGACATGCA	390	Ídem BIM15 y BIM42
BIM20	10	CCACACCCTCCGAATGTCGTTTTAGTCAT	388	Ídem BIM41
BIM21	4	AAACGTCAAAAAAGCTGGTAGTAAGGTCAA	165	Ídem BIM34. Pérdida de todos los espaciadores salvajes del 4 al 9
BIM22	10	CTACCGCTTCAAAGATAAAGCGAAAGACTC	182	Ídem BIM14 y BIM24

BIM23	10	AGTTGATGGTAAAACGGTGGAAATGACCATA	434	Ídem BIM12, BIM16, BIM27 y BIM39
BIM24	10	CTACCGCTTCAAAGATAAAGCGAAAGACTC	182	Ídem BIM14 y BIM22
BIM25	10	TAAGATTTTGGCGACTATACAGAAGGAATC	201	-
BIM26	10	TCGTGGAGAAGACGCAGTAAATAATAAACT	168	-
BIM27	4	AGTTGATGGTAAAACGGTGGAAATGACCATA	434	Ídem BIM12, BIM16, BIM23 y BIM39. Pérdida de espaciadores salvajes del 5 al 9
BIM28	10	GAAGACAGCACAAACCCAGCAGATTATACA	252	Ídem BIM13 y BIM33
BIM29	4	TTATGCAAACGGTGGCCTAGTCCACAAGAA	206	Adquisición ectópica
BIM30	4	CATTACAGACACAGGAGAAGGCGGCTATTA	322	Adquisición ectópica
BIM31	10	TTTATCTTTTTCAGCGCAGTTTAACGGGTC	268	Ídem BIM1 y BIM34
BIM32	10	TGAGCAGGTAGTCAATACATTCAAAGAAA	367	Ídem BIM11
BIM33	10	GAAGACAGCACAAACCCAGCAGATTATACA	252	Ídem BIM13 y BIM18
BIM34	10	TTTATCTTTTTCAGCGCAGTTTAACGGGTC	268	Ídem BIM1 y BIM31
BIM35	10	AAACGTCAAAAAAGCTGGTAGTAAGGTCAA	165	Ídem BIM21
BIM36	10	ATGCGGTTTGATGATGTCGTTCAGTACTAC	232	Ídem BIM2 y BIM3
BIM37	10	GAAGGCTCAAGAAAGTCAGATGCGAAGCGC	146	Ídem BIM5
BIM38	10	TTTCAACGCTGGTAATTTTATTGGACAAGA	243	Ídem BIM9
BIM39	10	AGTTGATGGTAAAACGGTGGAAATGACCATA	434	Ídem BIM12, BIM16, BIM23 y BIM27
BIM40	10	TAAACGATACTTGAATCTGTTTAACTGATA	215	Ídem BIM10
BIM41	10	CCACACCCTCCGAATGTGTTTTTCAGTCAT	388	Ídem BIM20
BIM42	10	TACATTCTTTGTCATTGACGTGGACATGCA	390	Ídem BIM15 y BIM19

Diversidad de protoespaciadores

Hasta hace muy poco tiempo, existían únicamente 20 genomas de bacteriófagos completamente secuenciados en la base de datos del NCBI: O1205 (Stanley et al. 1997), Sfi19 and Sfi21 (Desiere et al. 1998), DT1 (Denise M Tremblay & Moineau 1999), Sfi11 (Lucchini et al. 1999), 7201 (Stanley et al. 2000), 2972 (Lévesque et al. 2005), 858 (Deveau et al. 2008), ALQ13.2 and Abc2 (Guglielmotti et al. 2009), 5093 (Mills et al. 2011), TP-J34L and TP-778L (Ali et al. 2014), 9871, 9872, 9873, and 9874 (McDonnell et al. 2016), CHPC577, CHPC926, y CHPC1151 (Szymczak et al. 2017). No fue hasta el presente año que el trabajo de McDonnell y colaboradores de la Universidad de Cork en Irlanda, que se amplió el repertorio de genomas disponibles a 60 genomas (McDonnell et al. 2017). En este trabajo no están incluidos los genomas de los fagos uruguayos 53, 73 y 128 publicados en 2017 (Achigar et al. 2017) ni, por supuesto, el resto de los genomas que se presentan en este trabajo.

Las preguntas centrales con respecto a este punto son las siguientes: ¿La colección de bacteriófagos generada es una muestra representativa de la diversidad existente en Uruguay?, ¿La información genética disponible en esta colección es suficiente como para confiar en un desarrollo efectivo de cepas resistentes?

Como se mencionó anteriormente, el contenido genético y, específicamente el contenido de potenciales protoespaciadores, son el insumo fundamental para el correcto desarrollo y evaluación de cepas resistentes. En este capítulo hemos estudiado el contenido de protoespaciadores de los fagos de tipo cos ya que representan el grupo mayoritario en nuestro país. De este análisis hemos obtenido el grado de conservación de los protoespaciadores, la distribución de los mismos a en las diferentes regiones de los genomas, etc., pero no se responde la pregunta de si esto representa una foto puntual de un tiempo y espacio acotado, o si esta información realmente representa en un alto grado a la diversidad fágica que existe en el ambiente industrial de Uruguay. El responder esta pregunta es fundamental ya que significa la diferencia entre un monitoreo constante en pos de obtener nuevos fagos para analizar y generar resistentes, o utilizar la información que hemos generado hasta este momento para el desarrollo actual y futuro de cepas resistentes sin que sea crítico aislar nuevos bacteriófagos. Lo que si podemos confirmar, es que es el set de genomas de fagos de tipo cos más grande secuenciado de un único lugar.

Para intentar contestar esta pregunta, decidimos utilizar un acercamiento pangénomico.

El concepto de pangenoma aparece explicado en el trabajo de Medini y colaboradores de 2005 (Medini et al. 2005). En este trabajo, se define como genoma *core* a todos aquellos genes que son compartidos entre los genomas de los organismos secuenciados de la misma especie, y como genoma *dispensable* o descartable a la suma de genes que no son compartidos entre las mismas. A su vez, se distinguen dos tipos de genes descartables, aquellos que son únicos para cada cepa y aquellos compartidos por dos o más pero no por todas. La suma de todos ellos es lo que conforma el pangenoma de la especie. Lo más interesante del concepto de pangenoma, no es solamente la posibilidad de conocer el repertorio genético de una especie, si no que conocer su tamaño aporta una enorme cantidad de información. Más aún, el tamaño del pangenoma es una medida discreta que poco aporta comprender la diversidad, pero saber cómo evoluciona el tamaño del pangenoma a medida que nuevos genomas son añadidos al análisis (a este tipo de análisis se conoce como curva de pangenoma) nos permite saber de forma aproximada si se ha alcanzado un nivel de información que comprenda la diversidad de la especie. Dicho de otro modo, los autores afirman que es prácticamente imposible conocer el repertorio genético de una especie ya que aparecerán nuevos genes constantemente según se añadan genomas al análisis, pero el ritmo al que aparecen eventualmente se aproximará al cero estableciéndose el límite del pangenoma, y por tanto, perdiendo relevancia la información que puedan aportar nuevos genomas. Este límite puede aparecer en cualquier momento, no necesariamente se debe esperar a secuenciar enormes cantidades de genomas para obtener esta información. Como se observa en la figura 4.21, los autores muestran que para el caso de *Bacillus anthracis* unos pocos genomas son suficientes para describir su diversidad genética (pangenoma cerrado), mientras que para *Streptococcus agalactiae* aún se está lejos de tener disponible toda la información que describa la diversidad de la especie, por tanto cada nuevo genoma que se secuencie aportará información valiosa (pangenoma abierto). Lo que podemos concluir de este tipo de análisis es que a medida que la pendiente de la curva se aproxima al cero, más cerca estaremos de cerrar el pangenoma y por tanto se volverá irrelevante la secuenciación de nuevos genomas a efectos de conocer la diversidad del conjunto a estudiar.

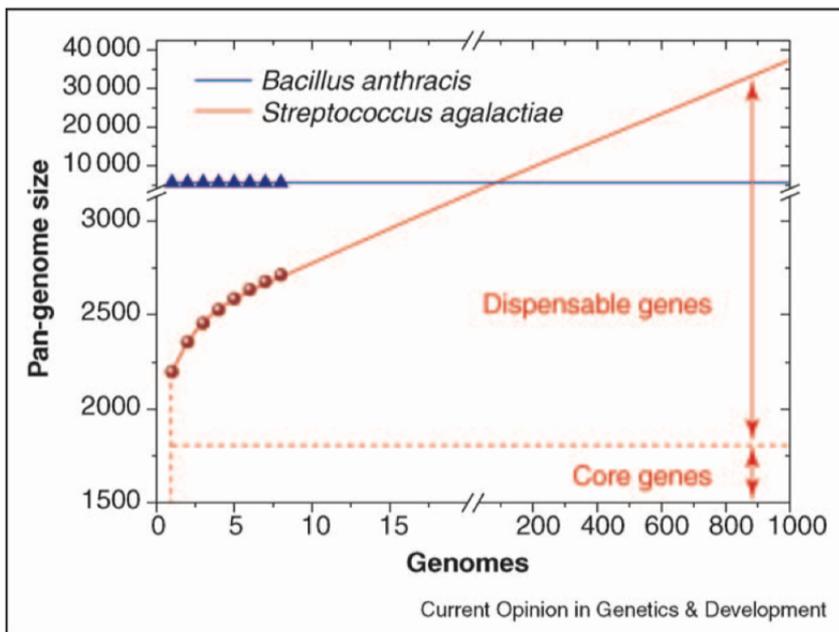


Figura 4.21. Curvas de pangomas para *Bacillus anthracis* y *Streptococcus agalactiae* (Medini et al. 2005).

Aplicando estos conceptos es que se llevó a cabo un análisis pangénomico de los bacteriófagos tipo *cos* de la colección. Ya que la pregunta que intentamos responder es si la información de que disponemos es suficiente en el contexto de desarrollar cepas resistentes, en lugar de utilizar los ORFs de los genomas como insumo, fueron utilizadas las secuencias de potenciales protoespaciadores para ambos tipos de sistema CRISPR. Es así que se llevó a cabo una curva analizando el contenido de protoespaciadores CRISPR1 y una curva analizando los protoespaciadores CRISPR3, las cuales llamaremos panviomas.

La curva se construyó de la siguiente forma: el contenido de protoespaciadores de los genomas fue analizado agrupándolos empezando por tres genomas hasta llegar a diez agregando de a un genoma por ronda. Se contaron de forma individual los espaciadores que aparecían una vez en la matriz, dos veces, etc., y así hasta el máximo posible de apariciones en cada iteración, siendo la primera tres veces y la última diez veces. Una vez sumados los grupos individualmente, fueron divididos por el número de apariciones correspondiente, con el fin de no sobreestimar la cantidad de espaciadores. Por ejemplo, para la última iteración de la curva de protoespaciadores CRISPR1, la suma de espaciadores que aparece en los diez fagos fue de 30, lo que quiere decir que únicamente 3 protoespaciadores están presentes en los tres genomas. Una vez realizada esta operación, se sumaron cada uno de los valores obtenidos para cada grupo de espaciadores y esa suma corresponde al tamaño del panviroma para cada iteración. Las tablas 4.2 y 4.3 muestran las sumas de estos grupos y el tamaño del panviroma obtenido en cada iteración.

Tabla 4.2. Panviroma calculado con los protoespaciadores tipo CRISPR1

Número de apariciones del protoespaciador	Tres genomas	Cuatro genomas	Cinco genomas	Seise genomas	Siete genomas	Ocho genomas	Nueve genomas	Diez genomas
10	-	-	-	-	-	-	-	3
9	-	-	-	-	-	-	3	8
8	-	-	-	-	-	3	10	5
7	-	-	-	-	3	11	13	20
6	-	-	-	4	12	17	13	11
5	-	-	18	20	20	15	25	28
4	-	24	12	15	18	42	61	62
3	30	21	36	48	69	88	111	117
2	76	90	134	167	219	211	254	407
1	492	674	767	908	930	973	871	722
Tamaño del panviroma	598	809	967	1162	1271	1360	1361	1383

Tabla 4.3. Panviroma calculado con los protoespaciadores tipo CRISPR3

Número de apariciones del protoespaciador	Tres genomas	Cuatro genomas	Cinco genomas	Seise genomas	Siete genomas	Ocho genomas	Nueve genomas	Diez genomas
10	-	-	-	-	-	-	-	8
9	-	-	-	-	-	-	16	17
8	-	-	-	-	-	16	12	15
7	-	-	-	-	18	12	18	29
6	-	-	-	20	15	22	40	38
5	-	-	46	31	32	53	54	53
4	-	58	23	33	60	70	111	116
3	75	43	66	102	121	162	204	220
2	181	223	271	308	371	377	420	640
1	823	1086	1267	1459	1508	1562	1405	1182
Tamaño del panviroma	1079	1410	1673	1953	2125	2274	2280	2318

El tamaño de los panviromas fue graficado en función de la cantidad de genomas utilizados para su cálculo y se trazó una línea de tendencia. Las figuras 4.22 y 4.23 muestran la curva para protoespaciadores CRISPR1 y protoespaciadores CRISPR3 respectivamente. Como se puede apreciar, en ambos casos la curva parece alcanzar su límite superior, sugiriendo

que los panviromas están en proceso de cerrarse. Más aún, al calcular el porcentaje de información que se agrega al añadir un nuevo genoma al análisis, encontramos que para el caso de CRISPR1 pasar de tres a cuatro genomas introduce un 35% de nueva información (el tamaño aumenta de 598 a 809 protoespaciadores), y pasar de nueve a diez genomas introduce un 1,6% de nueva información (aumento en el tamaño de 1361 a 1383 protoespaciadores). En el caso de CRISPR3, vemos que el pasaje de tres a cuatro genomas introduce un 30% de nueva información (aumento en el tamaño de 1079 a 1401 protoespaciadores), mientras que el pasaje de nueve a diez introduce un 1,7% de nueva información (aumento en el tamaño de 2280 a 2318 protoespaciadores). Esto quiere decir que en este escenario secuenciar más genomas de fagos uruguayos no aportaría nueva información sustancial en lo referente a la diversidad de protoespaciadores de la población.

Estos resultados nos permiten pensar que, en el contexto de nuestro trabajo y de las preguntas planteadas, la información genética generada es representativa de la diversidad fágica de Uruguay y, por tanto, sería suficiente para el desarrollo confiable y efectivo de cepas resistentes.

Para evaluar si nuestra colección de fagos nos permitiría el desarrollo de cepas resistentes no solamente específicas para el ambiente industrial uruguayo, si no con posibilidades de ser introducidas en otros mercados, decidimos introducir al análisis el genoma de un bacteriófago que estuviera suficientemente distanciado en espacio y tiempo. Para esto elegimos el genoma del fago de referencia DT1 (D M Tremblay & Moineau 1999), el cuál fue aislado hace veinte años en Canadá.

Los resultados muestran que la adición de este genoma aporta nueva información, aunque su impacto no es significativo como para revertir la tendencia de cierre de los panviromas que observamos anteriormente (Figuras 4.24 y 4.25). Si bien este es un muy buen indicio, deberíamos incluir más bacteriófagos de lugares distantes para corroborar que la tendencia se mantiene.

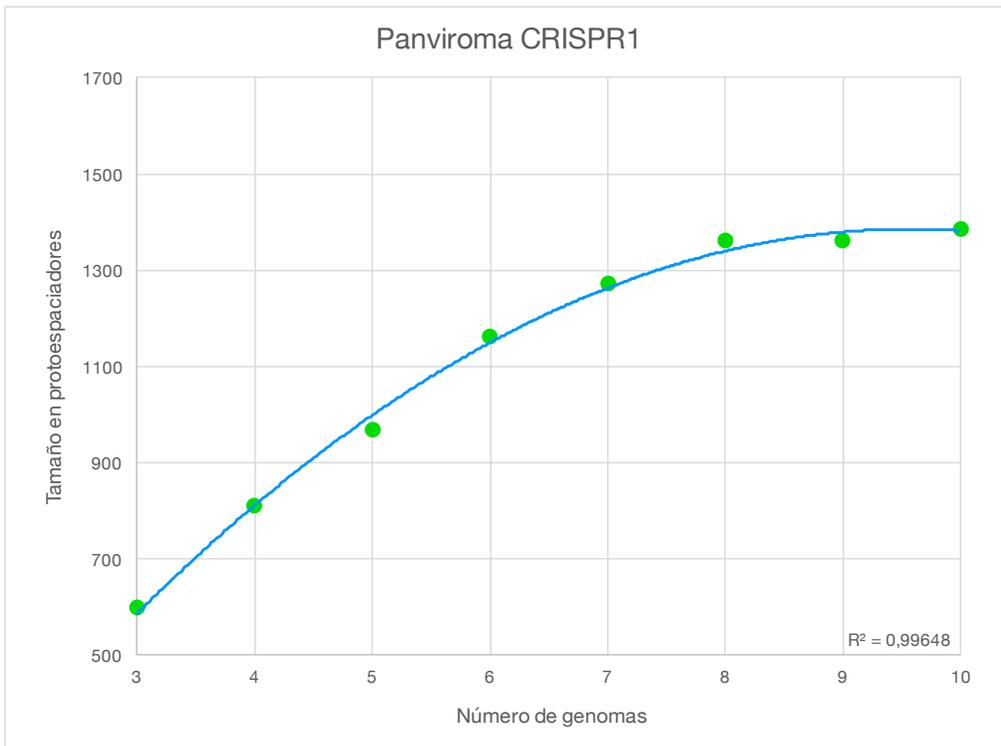


Figura 4.22. Curva construida con los protoespaciadores CRISPR1. La línea de tendencia muestra que el panviroma se está cerrando.

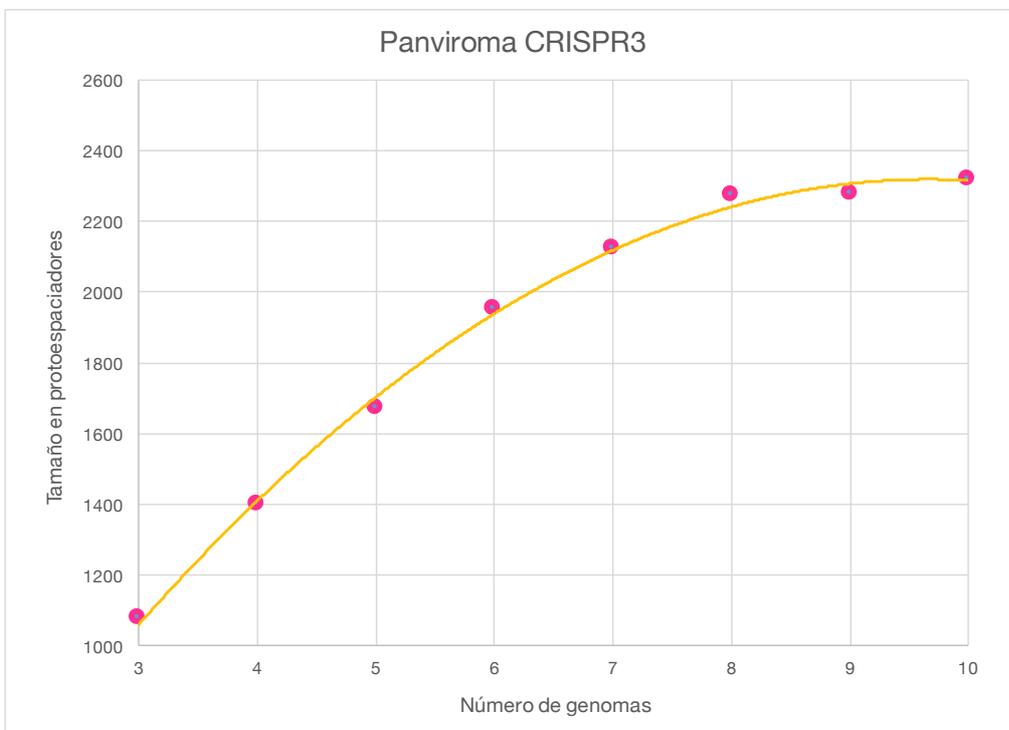


Figura 4.23. Curva construida con los protoespaciadores CRISPR3. La línea de tendencia muestra que el panviroma se está cerrando.

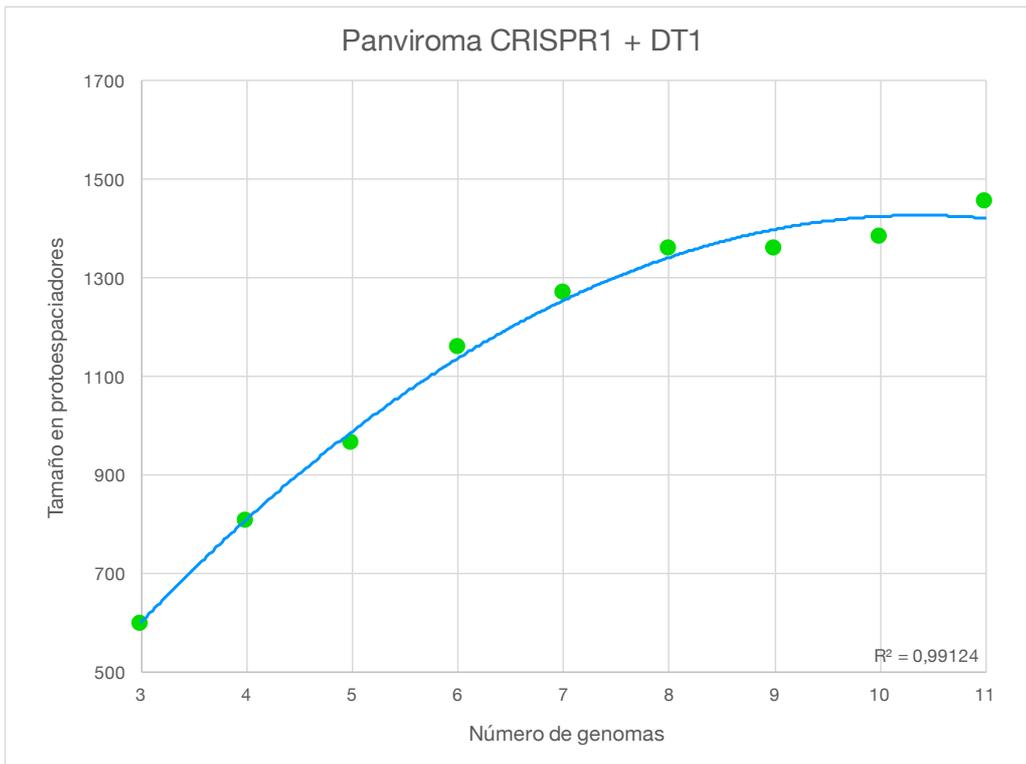


Figura 4.24. Curva construida con los protoespaciadores CRISPR1 de los fagos cos de la colección más DT1.

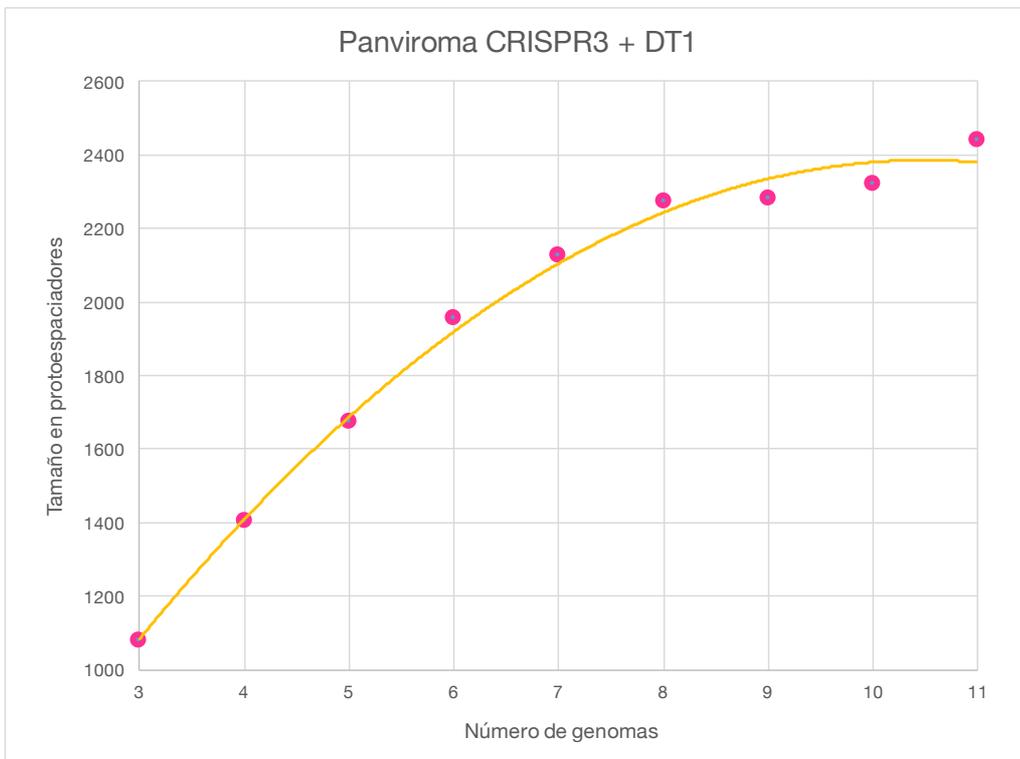


Figura 4.25. Curva construida con los protoespaciadores CRISPR3 de los fagos cos de la colección más DT1.

En este capítulo logramos observar el grado de conservación de los protoespaciadores en la población de fagos cos de nuestra colección, así como identificar las regiones ideales para la adquisición de nuevos espaciadores. También observamos que la cantidad de información que aporta la colección sería suficiente para generar derivados resistentes con confianza para introducirse al mercado local, e incluso, en menor grado, en el mercado mundial. Sin embargo, la estrecha ventana de protoespaciadores conservados y la baja probabilidad de que estos sean incorporados en eventos de adaptación, hacen que esta estrategia de generación de resistentes sea bastante limitada. En el capítulo siguiente exploramos la alternativa de generar derivados resistentes artificiales aprovechando todo este conocimiento adquirido.

5

Diseño de cepas resistentes

Resumen

Este capítulo final aborda la estrategia adoptada para obtener cepas multiresistentes sin pasar por un proceso de selección al azar, más específicamente, el uso de ingeniería genética para el diseño de cepas a la medida de las necesidades. Se generó un locus CRISPR artificial y se ensayaron varias estrategias para introducirlo en el genoma de *S. thermophilus*. Por último se evaluó la capacidad del mismo de conferir resistencia frente a fagos de la colección.

Antecedentes

En 2012, el grupo liderado por Jennifer Doudna de la U.C. Berkley demostró que el sistema CRISPR-Cas podía programarse para cortar DNA *in vitro* (Jinek et al. 2012). En 2013, el grupo de investigación de Feng Zhang del Instituto Broad del MIT y Harvard, adaptan con éxito el sistema CRISPR-Cas9 para la edición del genoma en células eucariotas (Cong et al. 2013). Para ello diseñan dos genes ortólogos diferentes Cas9 y demuestran la escisión específica del genoma en células humanas y de ratón. Desde entonces se han multiplicado exponencialmente los trabajos en los que se aplica el sistema CRISPR de tipo II como herramienta de edición de genomas *in vivo*, generándose múltiples variantes de Cas9 con diferentes características. Aún así, la variante más utilizada es la Cas9 salvaje de *Streptococcus pyogenes*, principalmente porque el motivo PAM que ésta reconoce es NGG, el cuál abre un abanico casi ilimitado de regiones donde pueden ser mutados los genes de interés. A grandes rasgos, la edición de DNA mediada por CRISPR requiere de dos componentes: una Cas9 y un ARN guía el cual debe poseer o imitar la estructura de horquilla característica del crARN de tipo II, con una secuencia homóloga a la región target de interés.

Esta secuencia espaciadora debe estar localizada inmediatamente en el extremo 5' de un PAM.

Existen dos tipos de ediciones que pueden llevarse a cabo utilizando CRISPR-Cas9, que se diferencian entre si por el uso o no de un DNA donante y por el tipo de sistema de reparación propio del organismo que se activa luego del corte:

1. NHEJ: *Non Homologous End Joining* es un tipo de mecanismo de reparación del DNA capaz de introducir errores en la doble hebra. Una vez que se produce el corte mediado por Cas9, la célula vuelve a unir los extremos de la doble hebra cortada y en el proceso inserta o elimina bases al azar. Estas inserciones o deleciones pueden causar el silenciamiento del gen objetivo.
2. HDR: El mecanismo de reparación conocido como *Homology Directed Repair*, necesita de un DNA donante para cerrar la doble hebra que se ha cortado. Para esto, el DNA donante debe poseer regiones de homología a la región del corte de al menos 50 pares de bases en cada extremo. Al co-transformar el sistema el sistema CRISPR-Cas9 junto con un DNA donante, pueden realizarse inserciones dirigidas a cualquier lugar del genoma con cualquier tipo de marcador.

Estrategia

A partir del conocimiento adquirido previamente, se adoptó una metodología de generación de resistentes basada en el diseño y construcción de un sistema artificial y posterior transformación en cepas de *S. thermophilus*. La estrategia ideada, consistió en la síntesis de un *locus* CRISPR1, mediante la selección de espaciadores conservados para fagos tipo *cos* y fagos tipo *Pac*, y su integración al cromosoma bacteriano en un lugar específico utilizando un sistema CRISPR1-Cas9 de *S. thermophilus*.

La estrategia se basa en que, a diferencia de en células eucariotas, no es posible realizar modificaciones genéticas basadas en NHEJ en procariontes ya que la acción de Cas9 sobre secuencias genómicas en bacterias causa la muerte del organismo. (Jiang et al. 2013). Sin embargo las bacterias poseen sistemas de reparación del ADN basados en recombinación homóloga (Wigley 2013). Por tanto, propusimos utilizar un sistema CRISPR-Cas9 que realice un corte dirigido a un sitio específico del genoma bacteriano, junto con un ADN donante capaz de reparar el corte de esta doble hebra y que contenga un locus CRISPR

artificialmente diseñado. El sistema CRISPR1-Cas9 clonado en el vector suicida junto con el ADN lineal donante serían co-transformados en las cepas de interés. Mediante selección con el antibiótico correspondiente se recuperarían los clones transformados, y dado que el corte en la doble hebra es deletéreo, únicamente crecerían aquellas bacterias que hubieran podido incorporar el ADN donante. El último paso sería eliminar el vector de la cepa y por tanto eliminar los marcadores de selección y devolver el nivel GRAS.

Materiales y métodos

Cepas bacterianas y plásmidos

Las cepas de *Streptococcus thermophilus* utilizadas en este capítulo pertenecen a la colección de nuestro laboratorio y fueron propagadas y mantenidas según se describe en el apartado Materiales y métodos del capítulo 2. La cepa de *Lactococcus lactis* NZ9000 y los vectores pNZ8148 y pNZ124 fueron adquiridos a la empresa MoBiTec GmbH (Goettingen, Alemania). La cepa fue cultivada en caldo M17 suplementado con 5 g/L de glucosa (GM17) y a 30°C sin agitación y en agar M17 suplementado con 5 g/L de glucosa. La cepa de *E. coli* utilizada fue adquirida a la empresa New England BioLabs Inc. (MA, E.E.U.U.) y propagada en medio LB a 37°C con agitación y LB Agar. Las cepas de *Lactobacillus paracasei* Lpc12777 y Lpc14676 y de *Lactobacillus rhamnosus* Lr5366 y Lr12697 pertenecen a la colección de la empresa Christian Hansen A/S y fueron cultivadas en caldo MRS a 37°C o en MRS Agar a 37°C en condiciones de microaerofilia. Los vectores pGh8 y pGh9 fueron proporcionados por la empresa Christian Hansen A/S. La síntesis de los genes sintéticos fue llevada a cabo por el servicio GeneArt de Invitrogen inc. (CA, E.E.U.U.). El diseño bioinformático, las simulaciones de ensamblaje y los mapas de las construcciones fueron realizados con el software *SnapGene* (GSL Biotech LLC, IL, E.E.U.U.)

Transformación de *Lactococcus lactis* por electroporación

Realizar un pre-cultivo en 10mL de GM17 a 30°C durante una noche a partir de una colonia pura del cultivo en placa. Inocular 10mL de GM17 + 0,5M sacarosa + 1% glicina e incubar a 30°C hasta que la OD₆₀₀ alcance un valor de 0,2-0,5. Centrifugar las células a 8.000 rpm durante 10 minutos y re suspender el *pellet* en 1mL de solución 0,5M sacarosa + 10% glicerol (fría). Repetir este paso cuatro veces. Al final de la última centrifugación, re suspender las células en 100µL de solución 0,5M sacarosa+10% glicerol (fría). Para la electroporación utilizar alícuotas de 40-50µL, pudiendo almacenar las competentes a -80°C durante un mes.

Tomar la alícuota de competentes y mezclar con 1-5µL de ADN plasmídico (0,2-0,5µg total) en un tubo de 1,5mL. Tomar la mezcla y transferirla a una cubeta de electroporación previamente enfriada en hielo. Realizar un pulso con el electroporador seteado en las

siguientes condiciones 25µF, 200 ohms, 2,5kV. Inmediatamente añadir 1mL de GM17 + 0,5M sacarosa + 20mM MgCl₂ + 2mM CaCl₂. Re suspender las células, transferir a un tubo de 1,5mL limpio e incubar durante dos horas a 30°C con agitación. Sembrar el cultivo en placas de GM17 (10µL, 100µL y resto), con el correspondiente antibiótico (Eritromicina 5µg/mL, Tetraciclina 2µg/mL, Cloranfenicol 5µg/mL). Incubar a 30°C durante dos días. Realizar el cálculo de eficiencia de transformación como la relación entre clones obtenidos y ADN utilizado para transformar.

Transformación de *Streptococcus thermophilus* por electroporación

Realizar un pre-cultivo en 10mL de LM17 a 37°C durante una noche. Inocular 5mL de LM17 e incubar a 42°C hasta que la OD₆₀₀ alcance un valor de 0,5. Añadir 5mL de LM17 + 1M sacarosa + 20% glicina e incubar a 42°C durante una hora. Centrifugar las células a 8.000 rpm durante 10 minutos y re suspender el *pellet* en 1mL de solución 0,5M sacarosa + 10% glicerol (fría). Repetir este paso cuatro veces. Al final de la última centrifugación, re suspender las células en 100µL de solución 0,5M sacarosa+10% glicerol (fría). Para la electroporación utilizar alícuotas de 40-50µL, pudiendo almacenar las competentes a -80°C durante un mes.

Tomar la alícuota de competentes y mezclar con 1-5µL de ADN plasmídico (0,2-0,5µg total) en un tubo de 1,5mL. Tomar la mezcla y transferirla a una cubeta de electroporación previamente enfriada en hielo. Realizar un pulso con el electroporador seteado en las siguientes condiciones 25µF, 200 ohms, 2,5kV. Inmediatamente añadir 1mL de LM17 + 0,5M sacarosa + 20mM MgCl₂ + 2mM CaCl₂. Re suspender las células, transferir a un tubo de 1,5mL limpio e incubar durante dos horas a 42°C con agitación. Sembrar el cultivo en placas de LM17 (10µL, 100µL y resto), con el correspondiente antibiótico (Eritromicina 5µg/mL, Tetraciclina 2µg/mL, Cloranfenicol 5µg/mL). Incubar a 42°C durante dos días. Realizar el cálculo de eficiencia de transformación como la relación entre clones obtenidos y ADN utilizado para transformar.

Transformación de *Lactobacillus spp.* por electroporación

Realizar un pre-cultivo en 10mL de MRS a 37°C durante una noche. Inocular 200mL de MRS + 1% glicina pre calentado a 37°C. Incubar a 37°C sin agitación hasta alcanzar una OD₆₀₀ de 0,6. Centrifugar las células a 7.000 rpm a 4°C durante 10 minutos, re suspender el *pellet* en

200mL de ddH₂O y repetir la centrifugación. Re suspender el *pellet* en 3mL de ddH₂O, separar en alícuotas de 1mL en tubos de 1,5mL. Centrifugar a 15.000 rpm por dos minutos. Lavar dos veces con 1mL de ddH₂O y una tercera con 1mL de PEG8000 30%. Re suspender las células en 0,5mL de PEG8000 30% y conservar a -80°C.

Añadir 900µL de ddH₂O o 750µL de solución de acetato de litio (200mM acetato de litio, 1,2M sacarosa, 2mM Tris pH 7,5) MÁS 150µL de 100mM DTT e incubar durante 30 minutos. Centrifugar durante tres minutos y lavar el *pellet* con 1mL de PEG8000 30%. Resuspender el *pellet* en 0,5mL de PEG8000 30%. Tomar 100µL de células, mezclar con 200µg de ADN plasmídico y transferir a una cubeta de electroporación previamente enfriada en hielo. Realizar un pulso con el electroporador seteado en las siguientes condiciones 25µF, 400 ohms, 2,5kV. Inmediatamente añadir 900µL de MRS + 0,5M sacarosa. Re suspender las células, transferir a un tubo de 1,5mL limpio e incubar durante cuatro horas a 37°C sin agitación. Sembrar el cultivo en placas de MRS (10µL, 100µL y resto). con el correspondiente antibiótico (Eritromicina 5µg/mL, Tetraciclina 2µg/mL). Incubar a 37°C durante cuatro días en condiciones de microaerofilia. Realizar el cálculo de eficiencia de transformación como la relación entre clones obtenidos y ADN utilizado para transformar.

Transformación de *E. coli.* por electroporación

Realizar pre-cultivo en 5mL de medio LB e incubar durante la noche a 37°C con una agitación de 200rpm. Inocular 200mL de medio LB e incubar a 37°C a 200rpm hasta alcanzar una OD₆₀₀ de 0,4. Colocar el cultivo en hielo para frenar el crecimiento Alicuotar el cultivo en tubos de 50 mL, centrifugar a 2.400rpm durante 20 minutos a 4°C. Descartar el sobrenadante y re suspender cada *pellet* en 50mL de ddH₂O fría. Volver a centrifugar a 2.400rpm durante 20 minutos a 4°C. Descartar el sobrenadante y re suspender cada *pellet* en 25mL de ddH₂O fría. Combinar el cultivo re suspendido de dos tubos en uno. Centrifugar a 2.400rpm durante 20 minutos a 4°C. Volver a descartar el sobrenadante y re suspender cada *pellet* con 25 mL de glicerol 10% frío. Centrifugar a 2.400rpm durante 20 minutos a 4°C. Con cuidado retirar el sobrenadante. Re suspender cada *pellet* con 500µL de glicerol 10% frío. Alicuotar en tubos de 1.5 mL previamente enfriados, aproximadamente 40µL de cultivo bacteriano y almacenar a -80°C.

Tomar un alícuota de células electrocompetentes y mezclar con 2ng del plásmido de interés. Dejar en hielo durante 5 minutos. Transferir a una cubeta de electroporación previamente

enfriada en hielo. Realizar un pulso con el electroporador seteado en las siguientes condiciones 25µF, 200 ohms, 1,7kV. Inmediatamente añadir 900µL de medio LB. Re suspender las células, transferir a un tubo de 1,5mL limpio e incubar durante una hora a 37°C con agitación. Sembrar el cultivo en placas de LB (10µL, 100µL y resto). con el correspondiente antibiótico. Incubar a 37°C durante 24-48 horas. Realizar el cálculo de eficiencia de transformación como la relación entre clones obtenidos y ADN utilizado para transformar.

Ensamblaje de construcciones en pGh9

El ensamblaje de los distintos componentes fue realizado mediante la metodología de *Gibson assembly* utilizando el kit NEBuilder® HiFi DNA Assembly Master Mix de New England BioLabs Inc. (MA, E.E.U.U.) según especificaciones del fabricante. Brevemente, se diseñaron primers para linealizar el vector por PCR y primers que amplificaran los diferentes elementos a ensamblar con colas de homología con los extremos del vector linealizado. En la tabla 5.4 se detallan todos los primers utilizados en este trabajo. Las amplificaciones fueron llevadas a cabo con la mix Phusion High Fidelity DNA Polymerase de New England BioLabs Inc. (MA, E.E.U.U.) según especificaciones del fabricante. Luego, los productos de amplificación fueron incubados juntos a 50°C con la mix de ensamblaje. Una vez culminada la reacción de ensamblaje, los productos fueron transformados en *Lactococcus lactis* NZ9000. Los vectores fueron recuperados utilizando el kit GenElute Plasmid Miniprep Kit de Sigma-Aldrich Co. LLC (MO, E.E.U.U.) según instrucciones del fabricante.

Tabla 5.4. Primers utilizados para el ensamblaje de los distintos plásmidos generados en este trabajo

Nombre	Secuencia 5´- 3´	Función
pGh9 R	GGGGGCCCGGTACCCAATTC	Abrir pGh9 para <i>Gibson Assembly</i>
pGh9 F	GGATCCACTAGTTCTAGAGCGGCCG	Abrir pGh9 para <i>Gibson Assembly</i>
pUYCR1target	ATTGGGTACCGGGCCCCCGTTTCAGCTGTTTAGATA AAAATGATATTAATTG	Clonar el sistema CRISPR1-Cas9 pGh9
pUYtarCR1 R2	CTAGAACTAGTGGATCCGGCCCATGAGGCCAGC A	Clonar el sistema CRISPR1-Cas9 pGh9
pGh9doF	GGGTACCGGGCCCCCAGGTGCCACCTTGGTCG	Clonar el ADN donante en pGh9
pGh9doR	CTAGAACTAGTGGATCCGGAAGTCTAAGCAAGAT GCATG	Clonar el ADN donante en pGh9
pGh9BetaF	GGGTACCGGGCCCCCATATTTAGTAAAAGTTT AGTAAAAACC	Clonar la construcción Promotor+Beta recombinasa en pGh9
pGh9BetaR	AGAACTAGTGGATCCTCATGCTGCCACCTTCTG	Clonar la construcción Promotor+Beta recombinasa en pGh9

Modificación del fenotipo *gal+* en *Lactobacillus* mediante *recombineering*

La modificación fue llevada a cabo mediante la co-transformación del vector pGh9Beta y oligos especialmente diseñados en las cepas *Lactobacillus paracasei* Lpc12777 y Lpc14676. La transformación se realizó según se describe en este mismo apartado con algunas modificaciones: a la mezcla de transformación se añadieron 100µg de oligo además de los 200µg de plásmido; los clones recuperados en MRS + Eritromicina fueron aislados y plaqueados en un medio mínimo con glucosa como única fuente de carbono para corroborar el cambio fenotípico. Luego se calculó la eficiencia de mutación como la relación entre clones incapaces de crecer en galactosa y transformantes totales obtenidos. La composición del medio mínimo par *Lactobacillus* así como las secuencias de los oligos utilizados son propiedad de la empresa Christian Hansen A/S y permanecen en reserva.

Resultados

Diseño del *locus* CRISPR donante

Para generar el locus CRISPR artificial debíamos primero decidir si diseñar un locus CRISPR1 o uno CRISPR3. En principio no habría motivos para elegir uno sobre otro, ya que según los resultados experimentales de este trabajo y de trabajos previos de nuestro grupo (Achigar 2014) ambos tipos de sistema son activos en las cepas de la colección. Tanto el tamaño de los panviromas de protoespaciadores como las regiones con mayor frecuencia de protoespaciadores mostraron ser similares para ambos tipos. Una característica a evaluar entre ambos sistemas fue la probabilidad de que los motivos PAM sufrieran una mutación y por tanto, dejaran obsoletos los espaciadores. Como observamos en el capítulo anterior, existen sesgos en el grado de conservación en diferentes regiones del genoma viral que sin embargo son similares para ambos tipos de sistema. Por tanto, se decidió eliminar esta variable y enfocarse solamente en mutaciones sinónimas (mutaciones que no cambian el contenido de aminoácidos de las proteínas). Para evaluar esto se eligió la estrategia de determinar las configuraciones de tripletes que se pueden formar con las secuencias de los PAMs; calcular la probabilidad de sufrir una mutación sinónima para cada codón formado (calculada como el número de mutaciones sinónimas silenciosas sobre la cantidad de colones que pueden formarse); calcular la probabilidad de mutación acumulada de cada configuración y por último, determinar la probabilidad promedio para cada tipo de sistema CRISPR.

En el caso de CRISPR1, la secuencia del PAM es NNAGAA(A/T). Esto quiere decir que puede presentarse en una secuencia codificante en las siguientes configuraciones: NNA GAA ANN; NNA GAA TNN; NAG AAA; NAG AAT; AGA AAN; AGA ATN. Dado que también pueden encontrarse PAMs en la secuencia complementaria ((A/T)TTCTNN), puede presentarse también en las siguientes configuraciones: NNA TTC TNN; NNT TCC TNN; NAT TCT; NTC TCT; ATT CTN; TTT CTN.

Para el caso del sistema CRISPR3 cuyo PAM es NGGNG, las configuraciones en las que puede presentarse son las siguientes: NNG GNG; NGG NGN; GGN GNN; NNC NCC; NCN CCN; CNC CNC.

Tras evaluar la probabilidad de cada codón y cada configuración, calculamos la probabilidad promedio de sufrir una mutación sinónima para los dos PAMs. En ambos casos ésta se encuentra en el entorno del 6%, por tanto este punto tampoco es influyente en la decisión a tomar. Todos los cálculos y detalles relacionados con este punto pueden encontrarse en los anexos. La elección final de diseñar un locus CRISPR1 terminó siendo influenciada por el hecho de haber seleccionado un sistema CRISPR1-Cas9 como herramienta de edición genética y simplemente para continuar trabajando sobre un único tipo de sistema. En el apartado siguiente se fundamenta la elección del sistema CRISPR1-Cas9.

El locus CRISPR1 artificial diseñado (que de ahora en adelante llamaremos donante) está compuesto por seis espaciadores diferentes (cuatro para fagos tipo *cos* y dos para fagos tipo *Pac*). La selección de espaciadores fue realizada tomando en cuenta la información obtenida en el capítulo anterior (esto únicamente para los espaciadores de fagos *Cos*) y mediante alineamiento con genomas de fagos presentes en la base de datos de GenBank. Estos espaciadores fueron elegidos de una lista mayor de secuencias candidatas con la intención de que el locus no fuera innecesariamente grande. Ya que el diseño y síntesis del ADN donante (y el resto de los componentes del sistema de modificación) fue realizado en colaboración y con el apoyo de la empresa Christian Hansen A/S (Dinamarca), las secuencias espaciadoras elegidas permanecen en reserva.

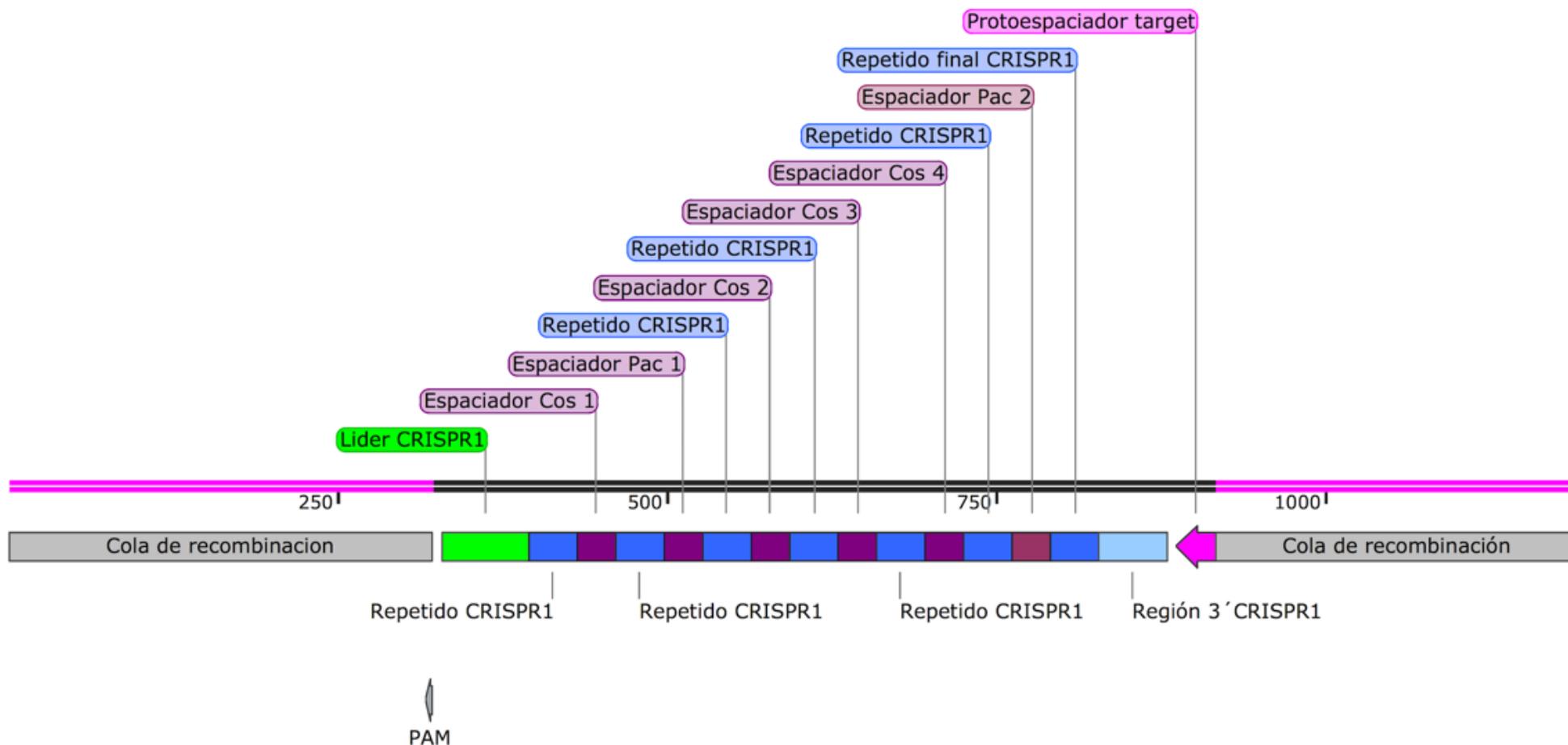


Figura 5.1. Estructura del *locus* CRISPR1 diseñado en este trabajo. Pueden observarse en los extremos regiones de recombinación flanqueando el *locus* propiamente dicho, compuesto por la secuencia líder y 6 espaciadores. Puede observarse también como el PAM y el protoespaciador *target* quedan separados previniendo de esta forma eventos de corte post recombinación.

Además de las secuencias espaciadoras (y los repetidos que las separan) los otros componentes agregados en el ADN donante fueron una secuencia líder completa, un repetido final (para el caso de CRISPR1 posee una secuencia con algunas diferencias al resto, que marca el final del locus) y dos regiones de homología flanqueantes necesarias para la integración en el genoma. La naturaleza de estas secuencias flanqueantes se detalla más adelante. En la figura 5.1 puede observarse la estructura del ADN donante diseñado, con un tamaño final de 1189 pares de bases.

Sistema de modificación CRISPR1-Cas9

La proteína Cas9 de *Streptococcus pyogenes* (SpCas9) es utilizada por excelencia en la edición genética, y pertenece a un sistema de tipo II al igual que los sistemas activos en *S. thermophilus* (Cong et al. 2013). Por tanto, la estrategia para diseñar un ensayo de modificación genética basado en CRISPR1-Cas9 es similar a los reportados en bibliografía. Existe una búsqueda constante de nuevas enzimas Cas9 o similares con diferentes propiedades que puedan mejorar algunos de los defectos de SpCas9. En el caso de CRISPR1-Cas9, presenta como ventajas el ser una enzima pequeña (1122 aa vs 1369 aa) y poseer un PAM más grande y "complejo" que el de SpCas9 (NNAGAA(A/T) vs NGG), lo cuál podría ayudar a disminuir el problema de los cortes inespecíficos observados en algunos casos con SpCas9 (O'Geen et al. 2015). La herramienta de edición está compuesta por un mini locus CRISPR1 compuesto por la secuencia líder, dos repetidos, un espaciador y la secuencia codificante y promotor del tracrARN para la correcta síntesis del ARN guía (gARN); también está incluida la secuencia codificante de la Cas9 con su promotor correspondiente. La justificación para añadir esta Cas9 en este diseño fue aumentar el número de copias del gen en la célula y de esta forma aumentar sus niveles de expresión. Se decidió utilizar el propio promotor de Cas9 ya que, según se menciona en el capítulo 1, los genes Cas presentan cierto grado de expresión constitutiva y su nivel se ve aumentado con la presencia de ADN foráneo en un mecanismo mediado por cAMP (Sorek et al. 2013), y por tanto el propio evento de transformación con el sistema de edición desencadenaría la expresión fuerte de Cas9.

La secuencia espaciadora fue seleccionada para guiar el corte de Cas9 a unas 60 pares de bases aguas arriba del promotor del gen cas9 del locus CRISPR1 salvaje. Esta región fue seleccionada para que ambos *loci* se encontraran en el mismo contexto genómico y su

Uno de los puntos claves en la estrategia fue que las cepas mantuvieran su nivel GRAS una vez culminado el proceso de modificación genética. Por lo tanto se hizo necesaria la elección de vectores capaces de ser removidos una vez que el locus CRISPR artificial fuera adquirido. La elección recayó en dos vectores suicidas de *Lactococcus lactis*, pGh8 y pGh9. La diferencia esencial entre ambos vectores radica en el marcador de resistencia que posee cada uno, eritromicina para el caso de pGh9 y tetraciclina para pGh8. Estos vectores fueron obtenidos por Maguin y colaboradores (Maguin et al. 1992) y tienen la particularidad de ser termosensibles, es decir que a altas temperaturas (en este caso a 42°C) su replicación se ve bloqueada y, por tanto, las bacterias crecidas a esta temperatura lo pierden. Estos vectores son específicos para *Lactococcus lactis*, sin embargo dada la cercanía filogenética entre estas BAL, es que decidimos probar si *S. thermophilus* era capaz de aceptarlos. Tres cepas de la colección fueron seleccionadas al azar para ser transformadas. Como control se utilizó la cepa de *L. lactis* NZ9000 y otros dos vectores también de lactococos, el pNZ124 y pNZ8148 (resistencia a cloranfenicol). En la tabla 5.1 se muestran las eficiencias de transformación para cada cepa y plásmido.

Tabla 5.1. Eficiencia de transformación para los diferentes vectores (ufc/ μ g)

	pGh9	pGh8	pNZ124	pNZ8148
Uy40	1,10E+02	n/c	n/c	n/c
Uy09	6,20E+01	n/c	n/c	n/c
Uy01	6,70E+01	n/c	n/c	n/c
NZ9000	8,48E+04	1,21E+04	1,28E+05	2,38E+05

Como puede observarse, únicamente el vector pGh9 fue aceptado por las cepas de la colección, por lo tanto este fue el elegido. La figura 5.3 muestra el sistema CRISPR1-Cas9 clonado en el vector pGh9 para formar el plásmido pGh9Target. El ensamblaje de las piezas fue realizado por fusión isotérmica según se describe en materiales y métodos.

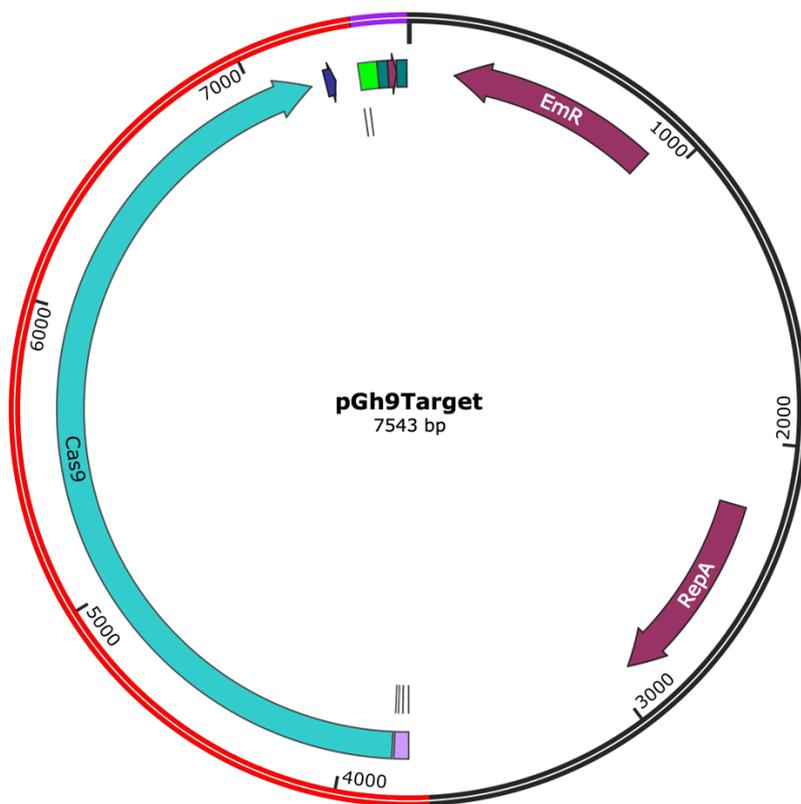


Figura 5.3. Mapa del vector pGh9Target. En rojo se encuentra marcada la región codificante para Cas9 con su promotor y el tracrARN. En violeta la secuencia del mini *locus* CRISPR1 que sirve como guía a Cas9.

Las construcciones fueron clonadas en las mismas tres cepas utilizadas para probar los vectores incorporando además las cepas Uy03 y Uy05 (esta última por ser la única sensible al fago *Pac* de la colección). En contraposición a los ensayos con la cepa control, en ninguno de los ensayos con las cinco cepas seleccionadas se logró obtener transformantes al utilizar pGh9Target (tabla 5.2). De esta forma quedó demostrado que el corte con Cas9 fue deletéreo para las bacterias y el ADN donante no fue capaz de repararlo. Este resultado demuestra la efectividad de sistema CRISPR1-Cas9 y su potencialidad para ser utilizado como herramienta de edición genética en otros contextos (más específicamente podría probarse en modelos eucariotas), pero en este caso el inconveniente fue no lograr integrar el locus artificial al genoma ni probar su eficacia. Para lograr estos objetivos planteamos otras estrategias que se describen a continuación.

Tabla 5.2. Eficiencia de transformación para los diferentes vectores (ufc/ μ g)

	pGh9Target + donante	pGh9Target	pGh9
Uy40	n/c	n/c	1,00E+01
Uy09	n/c	n/c	5,8E+01
Uy01	n/c	n/c	6,40+01
Uy03	n/c	n/c	7,30E+01
Uy05	n/c	n/c	5,20E+01
NZ9000	6,43E+04	5,56E+04	1,20E+05

Integración y silenciamiento por *recombineering*

La siguiente estrategia que se planteó fue la de clonar el ADN donante en el vector pGh9 e intentar que el nuevo plásmido llamado pGh9Donor (figura 5.4) se integrase al cromosoma bacteriano. El fundamento de esta estrategia es el siguiente: si existen regiones de homología entre un vector y alguna secuencia del genoma bacteriano, puede darse una integración por cruzamiento simple (*single crossover recombination*). Existen registros en la literatura del uso de vectores termosensibles para la integración de genes en genomas de algunas bacterias Gram positivas (Monk & Foster 2012; Zhang et al. 2014). Sin embargo, en nuestro conocimiento, esta estrategia no ha sido realizada en cepas de *S. thermophilus*.

En estos experimentos, los clones con el vector integrado serían seleccionados en base a la resistencia correspondiente y aumentando la temperatura a una no-permisiva para el vector. De este modo aquellos clones que mantengan el plásmido libre lo perderán y no podrán crecer en el medio con el antibiótico. El ADN donante posee dos secuencias de homología con el genomas de *S. thermophilus*, lo que hace posible que ocurriera la integración por recombinación (figura 5.5). El problema con esta estrategia es que los clones que se obtuvieran perderían su nivel GRAS al mantener la resistencia a eritromicina.

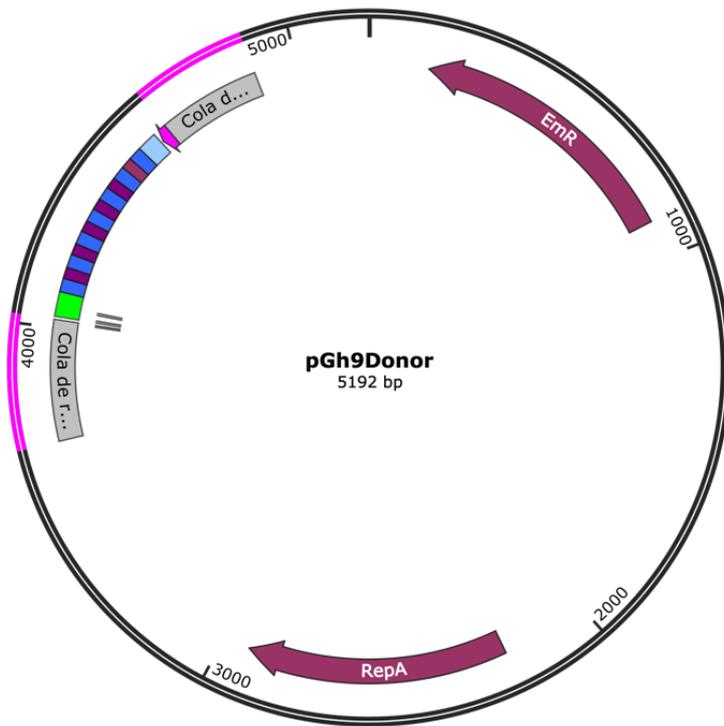


Figura 5.4. Mapa del vector pGh9Donor

El trabajo de Monk y colaboradores (Monk & Foster 2012) postula que al disminuir la temperatura hasta una permisiva para el vector, éste puede escindir y arrastrar consigo el gen de interés. Por tanto decidimos aplicar un método que nos permitiera simultáneamente eliminar la resistencia a eritromicina y bloquear la posibilidad de escisión del vector, mediante mutación dirigida al gen *emR* (gen que confiere resistencia a eritromicina) y *repA* (proteína termosensible responsable de la replicación y número de copias del vector). Para llevar a cabo esta doble mutación se diseñó una nueva herramienta basada en el sistema lambda red o *recombineering*.

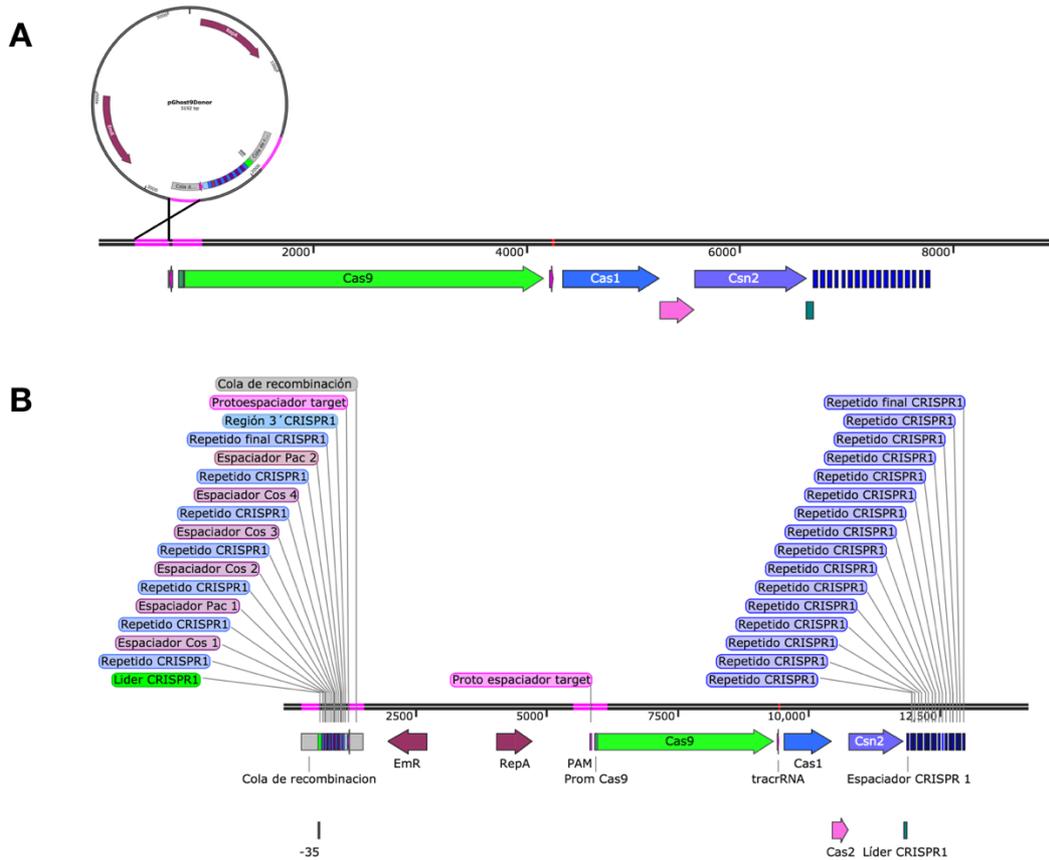


Figura 5.5. evento de integración esperado. A: cruzamiento con una de las regiones de homología en pGh9Donor. De ocurrir un doble cruzamiento solo se integraría el vector vacío sin inserto, manteniendo la cepa el fenotipo salvaje en lo que respecta a la resistencia fagos. B: esquema de cómo quedaría el plásmido integrado.

Tras la identificación del gen *recA* de *E. coli*, se encontró que el bacteriófago λ podía llevar a cabo recombinación eficientemente en huéspedes *recA* deficientes, revelando que λ codifica su propio sistema de recombinación. Este sistema está conformado por tres componentes principales, las proteínas Exo, Beta y Gam. Los genes *exo*, *bet* y *gam* se encuentran localizados dentro del operón PL del fago λ y su expresión se da durante las etapas tempranas de la infección del fago y posterior a la inducción del profago. Estas tres proteínas cumplen diferentes roles específicos en el proceso de recombinación. Mientras que la proteína Exo es una exonucleasa 5'→3' dsADN dependiente, Beta protege el ssDNA generado por acción de Exo y cataliza el alineamiento con regiones de homólogas de ssADN del hospedero y Gam tiene como función bloquear la degradación del ADN lineal mediada por los sistemas endógenos RecBCD y SbcCD. Al conjunto de estos genes se los llamó *red*, ya que su identificación fue llevada a cabo mediante ensayos de mutagénesis aleatoria y

los fagos derivados de estos ensayos eran deficientes a la hora de recombinar (recombination deficient = red). La figura 5.6 muestra el funcionamiento general del sistema.

Las características de este sistema han impulsado el desarrollo de diversas herramientas de manipulación genética que se engloban bajo el nombre de *recombineering*, palabra que surge de la combinación de los términos ingeniería y recombinación (*recombination* y *engineering*). Si bien ha sido mayormente utilizado en *E. coli* existen reportes en la bibliografía de su uso en bacterias Gram positivas, especialmente en algunos géneros de bacterias ácido lácticas (Van Pijkeren & Britton 2012). De los tipos de modificación que pueden llevarse a cabo con este sistema, en este trabajo nos enfocamos en el uso de la proteína Beta la cual cataliza la recombinación de ADN de cadena simple. Utilizando esta enzima, pueden utilizarse pequeños oligos con secuencias homólogas a los genes de interés, pero introduciéndoles mutaciones que ocasionen como resultado el silenciamiento de los mismos. Esta estrategia había sido puesta a punto para *Lactobacillus rhamnosus* y *Lactobacillus paracasei* en los laboratorios de Christian Hansen A/S al momento de nuestra colaboración, por tanto nos propusimos replicar el sistema pero para *S. thermophilus*.

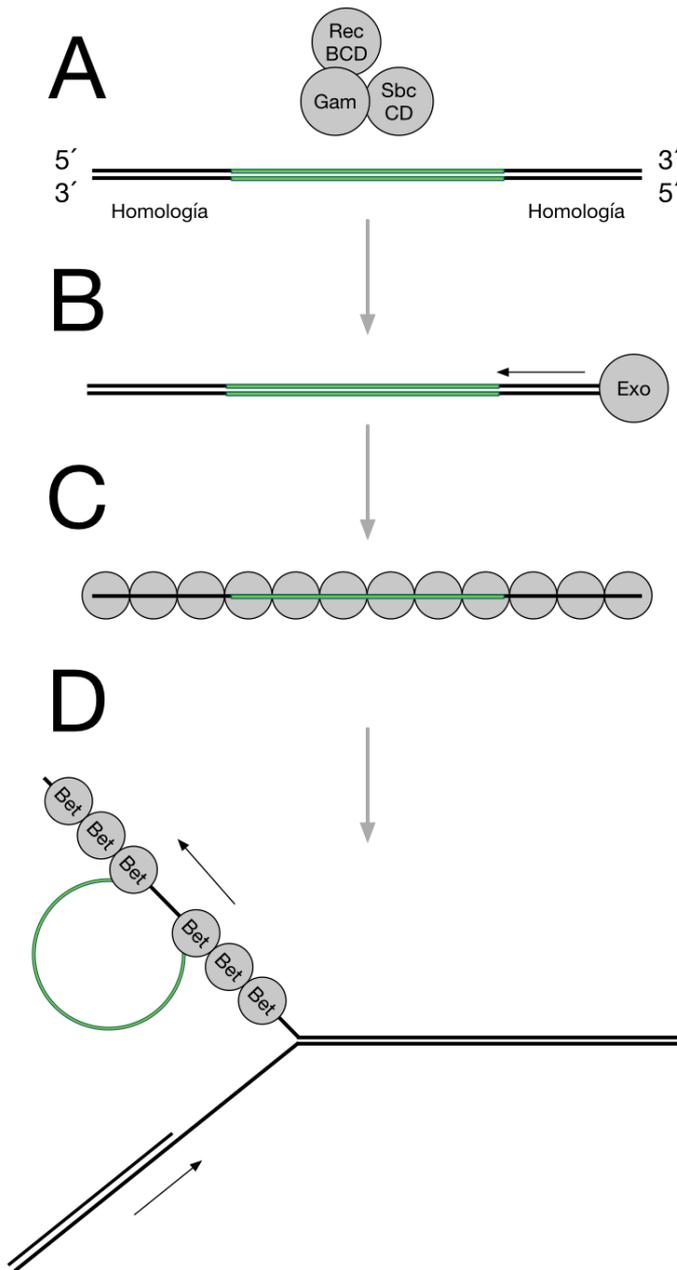


Figura 5.6. Esquema de funcionamiento del sistema Lambda Red. A: La proteína Gam se une a las endonucleasas propias del hospedero bloqueando su acción. B: La proteína Exo se une a un extremo del DNA doble hebra y degrada completamente una de ellas desde el extremo 5´ C: Una vez formado el ssDNA la proteína Beta se une a él protegiéndolo y a su vez catalizando la recombinación entre las regiones homólogas del DNA lineal y el DNA receptor. Como se observa en el esquema, la recombinación se da por una “invasión” del DNA lineal a la horquilla de replicación. Beta facilita la recombinación estabilizando la unión entre las pequeñas regiones de homología y el DNA receptor, haciendo que estas actúen como si se trataran de fragmentos de Okasaki a partir de los cuales se elonga la hebra *lag*.

Además de sintetizar un gen *bet* con uso de codones específico para *S. thermophilus*, se sintetizó un promotor fuerte que dirigiera su expresión. Este promotor fue diseñado a partir del promotor del gen *lacS* del operón *gal-lac* de *Streptococcus salivarius*. El promotor posee una modificación en la caja -35 que baja la toxicidad que tiende a generarse cuando se expresan proteínas recombinantes con promotores constitutivos fuertes (figura 5.7). Para probar la eficacia del promotor diseñado, fue sintetizado en fase con una proteína reportera mCherry en un vector pUC57 y transformado en *E.coli* DH5 α . En la figura 5.8 se puede apreciar el crecimiento color rosado de los transformantes lo que es prueba el funcionamiento del promotor.

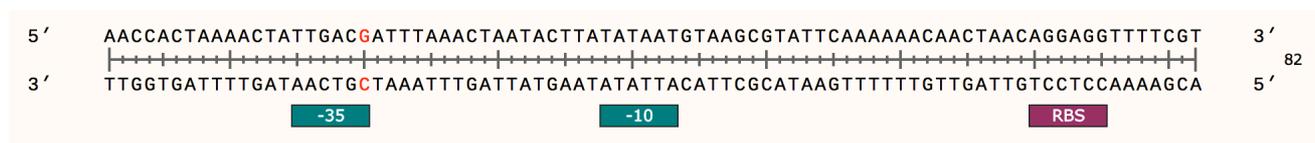


Figura 5.7. Secuencia del promotor utilizado. En rojo se encuentra marcada la mutación en la caja -35.

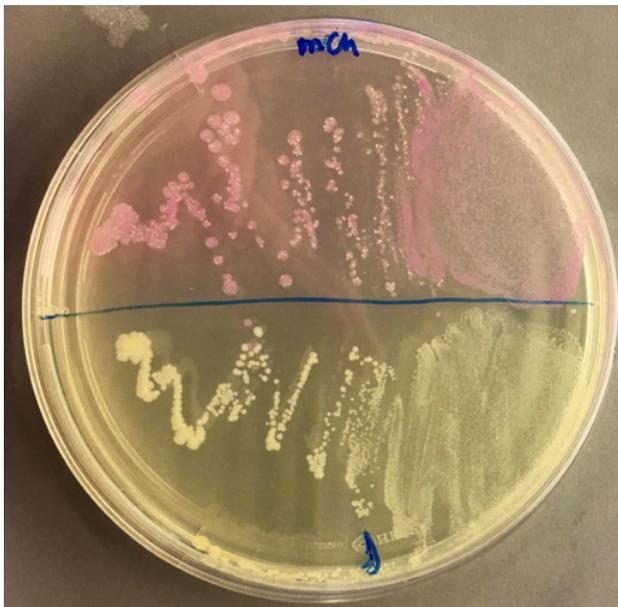


Figura 5.8. Expresión de la proteína reportera mCherry dirigida por el promotor seleccionado. En la parte superior de la imagen se observan las colonias rosadas esperadas. En la parte inferior se encuentra el control con pUC57 sin inserto.

La construcción fue clonada en pGh9 (con el objetivo de eliminar posteriormente el vector mediante incremento de temperatura) y puesta a prueba en las cepas de *Lactobacillus paracasei* Lpc12777 y Lpc14676 replicando ensayos previamente realizados por el equipo

de Christian Hansen A/S (ver materiales y métodos). Estos ensayos consistieron en el uso de oligos modificados que introdujeran mutaciones en el promotor del operón *gal*, generando así cepas incapaces de crecer en medio mínimo con galactosa como fuente de carbono.

Por último, se diseñaron oligos para noquear a los dos genes de interés. Los oligos fueron diseñados con un doble sistema de interrupción, en primer lugar introducen un codón stop temprano en los genes, y además producen un corrimiento en el marco de lectura en el región 3' al codón introducido (tabla 5.3). Este doble sistema garantiza que no ocurra una expresión parcial de las proteínas que puedan tener algún tipo de actividad. Otra consecuencia esperada de este diseño, es que estos oligos actúen sobre el vector pGh9Beta y por tanto produzcan la auto eliminación del sistema. En caso de que esto no ocurriese, pGh9Beta se eliminaría mediante el aumento de temperatura del cultivo.

En este caso, tampoco tuvimos éxito en el objetivo de introducir el gen en el cromosoma. A pesar de reiterados intentos en ningún caso logramos que el vector pGh9Donor se integrara al genoma de alguna de las cepas de la colección, por lo que la efectividad del sistema pGh9Beta no pudo ser evaluada en *S. thermophilus*.

Tabla 5.3. oligos diseñados para noquear los genes *repA* y *emR*. En mayúscula se marcan las bases a introducir para generar la interrupción de la expresión.

Nombre	Secuencia
orecRepA	tttctttccaatcattaggaattgagtcaggatataatCTCaaatccaaaatttctagctttagtatttttaatagccat
orecEmR	aagaaggagtgattacatgaacaaaaataaaaaatTGAtcaaaaacttttaacgagtgaaaaagtactcaaccaaata

Resistencia adquirida

Por último, decidimos probar si el *locus* CRISPR1 diseñado funcionaba otorgando resistencia en *trans*. Para esto, las cepas Uy01, Uy03 y Uy05 transformadas con el plásmido pGh9Donor fueron desafiadas correspondientemente mediante *spot test* (ver sección materiales y métodos del capítulo 2) con los fagos 53, 73 y 93 titulados a una concentración de 1×10^{10} ufp/mL. En ningún caso se observó lisis en ninguna de las diluciones para los clones transformados, mientras que los controles mostraron lisis en todas las ilusiones. Este

resultado es sumamente alentador, ya que confirma que el *locus* CRISPR1 artificial diseñado en este trabajo, es capaz de conferir resistencia en *trans*. Cabe destacar que los espaciadores que confieren resistencia a los fagos 53 y 73 son capaces de conferir resistencia al resto del espectro de los fagos de la colección.

Como conclusiones de este capítulo, podemos afirmar que diseñamos un *locus* CRISPR1 artificial capaz de conferir resistencia a fagos del ambiente industrial uruguayo. A pesar de que no alcanzamos el objetivo de introducir el *locus* en el genoma de las cepas de la colección, se logró el diseño de dos nuevas herramientas de edición genética con el potencial de tener aplicación a futuro.

Conclusiones finales y perspectivas

En este trabajo el objetivo planteado fue el desarrollo de una estrategia de diseño de cultivos starters para la industria láctea que se ajustaran a las necesidades de los productores nacionales.

En primer lugar pusimos a punto metodologías de evaluación de atributos tecnológicos relevantes a nivel industrial, utilizando como material de partida las cepas de *Streptococcus thermophilus* de nuestra colección. La causa principal de fallas a nivel industrial son las infecciones con fagos, y, debido a las características de estos (altas tasas de mutación, persistencia en los ambientes y resistencia a los tratamientos de desinfección), la resistencia a bacteriófagos es uno de los atributos más importantes y frágiles que posee una cepa. Por esta razón este es el atributo principal abordado a lo largo de este trabajo.

El primer paso para evaluar la resistencia fue la búsqueda de sistemas funcionales en las cepas de la colección, específicamente sistemas R-M y CRISPR-Cas. No fue posible en el marco de nuestra investigación identificar sistemas R-M funcionales, lo que pudo deberse a que efectivamente no estuvieran presentes o a limitaciones en la metodología elegida para evaluarlo.

Se lograron identificar sistemas CRISPR-Cas funcionales en prácticamente todas las cepas de la colección, por lo que sobre éstos profundizamos en nuestro estudio.

A la hora de seleccionar cepas resistentes basados en el sistema CRISPR-Cas, un insumo fundamental es conocer los fagos que circulan en el ambiente industrial a nivel genómico, particularmente el contenido de protoespaciadores de los mismos. Realizamos la caracterización de once fagos uruguayos, diez de tipo cos y uno de tipo Pac previamente aislados. Se evaluó el contenido de protoespaciadores para ambos sistemas CRISPR activos en *S. thermophilus*, CRISPR1 y CRISPR3, pudiendo determinar que existen regiones de conservación diferencial y, además, mediante una aproximación pangénómica se

concluyó que la cantidad de información que aporta nuestra colección sería en principio suficiente para seleccionar cepas resistentes con confianza.

El paso siguiente fue evaluar el fenómeno de adaptación (adquisición de nuevos espaciadores) de forma natural, para determinar si existía algún sesgo favorable o desfavorable a nuestra estrategia. Pudimos observar que, en los límites de nuestro trabajo, podría existir un sesgo en la adquisición de espaciadores, aunque dado que estos resultados son preliminares y no podemos afirmar con seguridad que esto ocurra.

Aún así, observamos que de seguir una estrategia de generación de resistentes confinada a la adquisición natural, la probabilidad de que la cepa adquiriera los espaciadores que a nuestro juicio serían los mejores candidatos en cuanto a distribución y estabilidad sería muy baja. Es por esto que decidimos enfocarnos en el diseño de un locus CRISPR artificial.

Utilizando la información y conocimientos adquiridos, diseñamos un locus CRISPR1 que fue capaz de otorgar resistencia a las cepas en las cuales fue probado. Aún así y a pesar de reiterados intentos y diferentes estrategias, no logramos introducir el locus de forma estable en las cepas sin que éstas perdieran su nivel GRAS. Como resultados paralelos, logramos determinar que nuestro sistema CRISPR1-Cas9 es capaz de realizar modificaciones y probamos que el sistema de modificación basado en la enzima beta-recombinasa que sintetizamos funciona en bacterias ácido-lácticas.

Las perspectivas que quedan planteadas son en principio realizar un ensayo pangenómico con fagos de diferentes partes del mundo para determinar si realmente ya contamos con la información suficiente para el desarrollo de cepas resistentes que sean útiles en cualquier mercado, sin la necesidad de realizar constantes aislamientos virales.

También queda planteada como perspectiva el empleo de nuevas estrategias para intentar introducir el locus CRISPR1 en el genoma bacteriano. Entre las que evaluamos como más prometedoras, son el uso de enzimas Cas9 con un único dominio funcional, lo que permitiría realizar el corte en una sola hebra potenciando la recombinación sin ser deletéreo para la bacteria. Otra estrategia que planteamos es el diseño de un sistema de edición basado en recombineering pero compuesto por las tres enzimas, lo que permitiría no solo realizar mutaciones, sino evaluar la posibilidad de introducir el locus completo en el genoma. Por último, la tercera estrategia que planteamos es el uso de variantes de los vectores

termosensibles que contengan secuencias de inserción, lo que aumentaría las probabilidades de que ocurriera la integración del vector en el cromosoma.

Los resultados de esta tesis darán lugar al menos a dos publicaciones en revistas científicas, enfocadas en el estudio pormenorizado de los genomas virales y el mapeo de protoespaciadores; y la dinámica de la adaptación en *S thermophilus* y estrategias de diseño de resistentes. Este es, según nuestro conocimiento, el primer trabajo que aborda la problemática de los fagos a nivel industrial en nuestro país enfocado hacia el mejoramiento de cepas como potencial solución.

Bibliografía

- Achigar, R., 2014. *Mejoramiento para resistencia a bacteriófagos de cepas de Streptococcus thermophilus de uso en la industria láctea*. Uruguay: Universidad de la Republica.
- Achigar, R. et al., 2017. Phage-host interactions in *Streptococcus thermophilus*: Genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. *Scientific Reports*, 7(January), p.43438. Available at: <http://www.nature.com/articles/srep43438>.
- Ajdić, D. & Ferretti, J.J., 1998. Transcriptional regulation of the *Streptococcus mutans* gal operon by the GalR repressor. *Journal of Bacteriology*, 180(21), pp.5727–5732.
- Ali, Y. et al., 2014. Temperate *Streptococcus thermophilus* phages expressing superinfection exclusion proteins of the Ltp type. *Frontiers in Microbiology*, 5(MAR).
- Amitai, G. & Sorek, R., 2016. CRISPR-Cas adaptation: Insights into the mechanism of action. *Nature Reviews Microbiology*, 14(2), pp.67–76. Available at: <http://dx.doi.org/10.1038/nrmicro.2015.14>.
- Anon, 2005. The Science Underpinning Food Fermentations. In *Food, Fermentation and Micro-organisms*. Blackwell Publishing Ltd, pp. 1–39. Available at: <http://dx.doi.org/10.1002/9780470995273.ch1>.
- Awad, S., Hassan, A.N. & Muthukumarappan, K., 2005. Application of Exopolysaccharide-Producing Cultures in Reduced-Fat Cheddar Cheese: Texture and Melting Properties. *Journal of Dairy Science*, 88(12), pp.4204–4213. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022030205731064>.
- Barrangou, R. et al., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), pp.1709–1712.
- Binetti, a G. & Reinheimer, J. a, 2000. Thermal and chemical inactivation of indigenous *Streptococcus thermophilus* bacteriophages isolated from Argentinian dairy plants. *Journal of food protection*, 63(4), pp.509–515.

- Van Den Bogaard, P.T.C. et al., 2000. Control of lactose transport, β -galactosidase activity, and glycolysis by CcpA in *Streptococcus thermophilus*: Evidence for carbon catabolite repression by a non-phosphoenolpyruvate-dependent phosphotransferase system sugar. *Journal of Bacteriology*, 182(21), pp.5982–5989.
- Van Den Bogaard, P.T.C. et al., 2004. Sugar Utilisation and Conservation of the gal-lac Gene Cluster in *Streptococcus thermophilus*. *Systematic and applied microbiology*, 27(1), pp.10–17.
- Bolotin, A. et al., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151(8), pp.2551–2561.
- Bolotin, A. et al., 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nature Biotechnology*, 22(12), pp.1554–1558.
- Bonfield, J.K., Smith, K. f & Staden, R., 1995. A new DNA sequence assembly program. *Nucleic acids research*, 23(24), pp.4992–4999.
- Broadbent, J.R. & Steele, J.L., 2007. Proteolytic enzymes of lactic acid bacteria and their influence on bitterness in bacterial-ripened cheeses. In ACS Publications.
- Broome, M.C., Krause, D.A. & Hickey, M.W., 1990. The use of non-starter lactobacilli in Cheddar cheese manufacture. *Australian Journal of Dairy Technology*, 45(2), pp.67–73.
- Bruttin, A. et al., 1997. Molecular ecology of bacteriophage infections in a cheese factory. *Applied and environmental microbiology*, 63(8), pp.3144–3150.
- Cai, H. et al., 2009. Genome sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-associated evolution. *Genome Biology & Evolution*, 2009.
- Calendar, R. & Abedon, S.T., 2005. *The bacteriophages*, Oxford University Press.
- Cogan, T.M. et al., 2007. Invited Review: Advances in Starter Cultures and Cultured Foods. *Journal of Dairy Science*, 90(9), pp.4005–4021. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022030207718581>.

- Cole, S.T. et al., 2001. Massive gene decay in the leprosy bacillus. *Nature*, 409(6823), pp.1007–1011.
- Cong, L. et al., 2013. Multiplex genome engineering using CRISPR-Cas systems. *Science (New York, N.Y.)*, 339(6121), pp.819–23. Available at: <http://science.sciencemag.org/content/339/6121/819.abstract> [Accessed July 9, 2014].
- Courtin, P., Monnet, V. & Rul, F., 2002. Cell-wall proteinases PrtS and PrtB have a different role in *Streptococcus thermophilus*/Lactobacillus bulgaricus mixed cultures in milk. *Microbiology*, 148(11), pp.3413–3421.
- Crow, V., Curry, B. & Hayes, M., 2001. The ecology of non-starter lactic acid bacteria (NSLAB) and their use as adjuncts in New Zealand Cheddar. *International Dairy Journal*, 11(4–7), pp.275–283.
- Dandoy, D. et al., 2011. The fast milk acidifying phenotype of *Streptococcus thermophilus* can be acquired by natural transformation of the genomic island encoding the cell-envelope proteinase PrtS. *Microb Cell Fact*, 10(Suppl 1), p.S21.
- Desiere, F., Lucchini, S. & Brüssow, H., 1998. Evolution of *streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. *Virology*, 241(2), pp.345–356.
- Deveau, H. et al., 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), pp.1390–1400.
- Dudley, E.G. & Steele, J.L., 2005. Succinate production and citrate catabolism by Cheddar cheese nonstarter lactobacilli. *Journal of applied microbiology*, 98(1), pp.14–23.
- Duplessis, M. et al., 2005. Global gene expression analysis of two *Streptococcus thermophilus* bacteriophages using DNA microarray. *Virology*, 340(2), pp.192–208.
- Dupuis, M.E. et al., 2013. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun*, 4, p.2087.
- Ercolini, D. et al., 2005. Sequence heterogeneity in the lacSZ operon of *Streptococcus thermophilus* and its use in PCR systems for strain differentiation. *Research in*

Microbiology, 156(2), pp.161–172.

- Farrow, J.A. & Collins, M.D., 1984. DNA base composition, DNA-DNA homology and long-chain fatty acid studies on *Streptococcus thermophilus* and *Streptococcus salivarius*. *J Gen Microbiol*, 130(2), pp.357–362. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6726177.
- Fernandez-Espla, M.D. et al., 2000. *Streptococcus thermophilus* cell wall-anchored proteinase: Release, purification, and biochemical and genetic characterization. *Applied and Environmental Microbiology*, 66(11), pp.4772–4778.
- Floch, M.H., Ringel, Y. & Walker, W.A., 2016. *The microbiota in gastrointestinal pathophysiology: implications for human health, prebiotics, probiotics, and dysbiosis*, Academic Press.
- Fortier, L.-C. & Moineau, S., 2007. Morphological and genetic diversity of temperate phages in *Clostridium difficile*. *Applied and environmental microbiology*, 73(22), pp.7358–7366.
- Fox, P.F. et al., 2000. *Fundamentals of Cheese Science*, Aspen, Gaithersburg, MD. *Conversion of milk to curd*, 69.
- Franz, C.M.A.P., Holzappel, W.H. & Stiles, M.E., 1999. Enterococci at the crossroads of food safety? *International journal of food microbiology*, 47(1–2), pp.1–24.
- Garault, P. et al., 2000. Branched-chain amino acid biosynthesis is essential for optimal growth of *Streptococcus thermophilus* in milk. *Applied and Environmental Microbiology*, 66(12), pp.5128–5133.
- Garneau, J.E. & Moineau, S., 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. *Microb. Cell Fact*, 10 Suppl 1(Suppl 1), p.S20.
- Godde, J.S. & Bickerton, A., 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of Molecular evolution*, 62(6), pp.718–729.
- Guglielmotti, D.M. et al., 2009. Genome analysis of two virulent *Streptococcus*

thermophilus phages isolated in Argentina. *International Journal of Food Microbiology*, 136(1), pp.101–109. Available at:
<http://www.sciencedirect.com/science/article/pii/S016816050900470X>.

Gunnewijk, M.G.W. & Poolman, B., 2000. HPr (His~ P)-mediated phosphorylation differently affects counterflow and proton motive force-driven uptake via the lactose transport protein of *Streptococcus thermophilus*. *Journal of Biological Chemistry*, 275(44), pp.34080–34085.

Hassan, A.N., Awad, S. & Muthukumarappan, K., 2005. Effects of Exopolysaccharide-Producing Cultures on the Viscoelastic Properties of Reduced-Fat Cheddar Cheese. *Journal of Dairy Science*, 88(12), pp.4221–4227. Available at:
<http://linkinghub.elsevier.com/retrieve/pii/S0022030205731088>.

Hatfull, G., 2010. Bacteriophages: Genes and Genomes. *iBioSeminars*.

Hayacibara, M.F. et al., 2004. The influence of mutanase and dextranase on the production and structure of glucans synthesized by streptococcal glucosyltransferases. *Carbohydrate research*, 339(12), pp.2127–2137.

Hendrix, R.W., 2003. Bacteriophage genomics. *Current opinion in microbiology*, 6(5), pp.506–511.

Hernández Magadán, A., 2007. Aislamiento, caracterización y detección precoz de bacteriófagos de *Streptococcus thermophilus* en la industria láctea.

Høier, E. et al., 2010. The production, application and action of lactic cheese starter cultures. *Technology of Cheesemaking, Second Edition*, pp.166–192.

Hols, P. et al., 2005. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS microbiology reviews*, 29(3), pp.435–463.

Horvath, P. et al., 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), pp.1401–1412.

Horvath, P. & Barrangou, R., 2010. CRISPR-Cas, the immune system of bacteria and archaea. *Science (New York, N.Y.)*, 327(5962), pp.167–170.

- Iyer, R. et al., 2010. Streptococcus thermophilus strains: Multifunctional lactic acid bacteria. *International Dairy Journal*, 20(3), pp.133–141. Available at: <http://dx.doi.org/10.1016/j.idairyj.2009.10.005>.
- Jansen, R. et al., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 43(6), pp.1565–1575. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11952905>.
- Jiang, W. et al., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, 31(3), pp.233–239. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23360965> <http://www.nature.com/nbt/journal/v31/n3/pdf/nbt.2508.pdf>.
- Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), pp.816–821.
- Kunji, E.R.S. et al., 1996. The proteolytic system of lactic acid bacteria. *Genetics and biotechnology of lactic acid bacteria*, 70, pp.187–221. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/8879407>.
- Kurtboke, I., 2012. *Bacteriophages*, Available at: <http://www.intechopen.com/books/bacteriophages>.
- Kutter, E. & Sulakvelidze, A., 2004. *Bacteriophages: biology and applications*, Available at: http://books.google.com/books?hl=en&lr=&id=6sTBd2EJpmYC&oi=fnd&pg=PP2&dq=BACTERIOPHAGES+Biology+and+Applications&ots=j_fXYdh4lu&sig=JG1NNIDLPGS7aP84vH_ymHFsQBk%5Cnhttp://books.google.com/books?hl=en&lr=&id=6sTBd2EJpmYC&oi=fnd&pg=PP2&dq=Bacteriophages:+bio.
- Laht, T.M. et al., 2002. Role of arginine in the development of secondary microflora in Swiss-type cheese. *International Dairy Journal*, 12(10), pp.831–840.
- Lahtinen, S. et al., 2011. *Lactic acid bacteria: microbiological and functional aspects*, CRC Press.
- Lamothe, G. et al., 2005. Characterization of the cro-ori region of the Streptococcus thermophilus virulent bacteriophage DT1. *Appl Environ Microbiol*, 71(3), pp.1237–1246.

- Leiman, P.G. et al., 2010. Morphogenesis of the T4 tail and tail fibers. *Virology Journal*, 7.
- Letort, C. & Juillard, V., 2001. Development of a minimal chemically-defined medium for the exponential growth of *Streptococcus thermophilus*. *Journal of Applied Microbiology*, 91(6), pp.1023–1029.
- Lévesque, C. et al., 2005. Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. *Applied and environmental microbiology*, 71(7), pp.4057–4068.
- Lopez, S. & Arias, C.F., 2010. How Viruses Hijack Endocytic Machinery. *Nature Education*, 3(9), p.16.
- Lucchini, S., Desiere, F. & Brussow, H., 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J Virol*, 73(10), pp.8647–8656.
- Luesink, E.J. et al., 1998. Transcriptional activation of the glycolytic *las* operon and catabolite repression of the *gal* operon in *Lactococcus lactis* are mediated by the catabolite control protein CcpA. *Molecular Microbiology*, 30(4), pp.789–798.
- Madera, C. et al., 2009. Prophage induction in *Lactococcus lactis* by the bacteriocin Lactococcin 972. *International Journal of Food Microbiology*, 129(1), pp.99–102.
- Maguin, E. et al., 1992. New thermosensitive plasmid for gram-positive bacteria. *Journal of Bacteriology*, 174(17), pp.5633–5638.
- Makarova, K. et al., 2006. Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A*, 103(42), pp.15611–15616.
- Makarova, K.S. & Koonin, E. V, 2007. Evolutionary Genomics of Lactic Acid Bacteria. *Journal of Bacteriology*, 189(4), pp.1199–1208. Available at: <http://jb.asm.org/content/189/4/1199.short>.
- Marino, M., Maifreni, M. & Rondinini, G., 2003. Microbiological characterization of artisanal Montasio cheese: analysis of its indigenous lactic acid bacteria. *FEMS Microbiology Letters*, 229(1), pp.133–140.
- Marth, E.H. & Steele, J., 2001. *Applied dairy microbiology*, CRC Press.

- Martha, R.J. & Clokie, A.M.K., 2008. Bacteriophages: methods and protocols, volume 1: isolation, characterization, and interaction.
- Martin, B. et al., 1997. Effect of milk urea content on characteristics of matured Reblochon cheeses. *Lait*, 77, pp.505–514.
- McDonnell, B. et al., 2017. Global survey and genome exploration of bacteriophages infecting the lactic acid bacterium *Streptococcus thermophilus*. *Frontiers in Microbiology*, 8(SEP).
- McDonnell, B. et al., 2016. Identification and analysis of a novel group of bacteriophages infecting the lactic acid bacterium *Streptococcus thermophilus*. *Applied and Environmental Microbiology*, 82(17), p.AEM.00835-16. Available at: <http://aem.asm.org/lookup/doi/10.1128/AEM.00835-16>.
- McSweeney, P.L.H. et al., 1993. Contribution of the indigenous microflora to the maturation of Cheddar cheese. *International Dairy Journal*, 3(7), pp.613–634.
- Medini, D. et al., 2005. The microbial pan-genome. *Current Opinion in Genetics and Development*, 15(6), pp.589–594.
- Mierau, I. et al., 1996. Multiple-peptidase mutants of *Lactococcus lactis* are severely impaired in their ability to grow in milk. *Journal of bacteriology*, 178(10), pp.2794–2803.
- Mills, S. et al., 2011. A new phage on the “Mozzarella” block: Bacteriophage 5093 shares a low level of homology with other *Streptococcus thermophilus* phages. *International Dairy Journal*, 21(12), pp.963–969.
- Mojica, F.J.M. et al., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, 36(1), pp.244–246.
- Mojica, F.J.M. et al., 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology*, 17(1), pp.85–93.
- Monk, I.R. & Foster, T.J., 2012. Genetic manipulation of Staphylococci—breaking through

the barrier. *Frontiers in Cellular and Infection Microbiology*, 2(April), pp.1–9. Available at: <http://journal.frontiersin.org/article/10.3389/fcimb.2012.00049/abstract>.

Mora, D. et al., 2004. Characterization of urease genes cluster of *Streptococcus thermophilus*. *Journal of Applied Microbiology*, 96(1), pp.209–219.

Mora, D. et al., 2002. Genetic diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy products. *Journal of applied microbiology*, 93(2), pp.278–287.

Müller-Merbach, M. et al., 2007. Diffusion of bacteriophages in an agar gel and in a fermented milk matrix. *Milchwissenschaft*, 62(1), pp.24–27. Available at: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-33846543038&partnerID=40&md5=2fc537970f580deb205bb019b059feca>.

Murphy, J. et al., 2013. Bacteriophage orphan DNA methyltransferases: Insights from their bacterial origin, function, and occurrence. *Applied and Environmental Microbiology*, 79(24), pp.7547–7555.

Murphy, J. et al., 2014. Methyltransferases acquired by lactococcal 936-type phage provide protection against restriction endonuclease activity. *BMC Genomics*, 15(1), pp.1–11.

Neve, H., Laborius, A. & Heller, K., 2003. Testing of the applicability of battery-powered portable microbial air samplers for detection and enumeration of airborne *Lactococcus lactis* dairy bacteriophages. *Kieler Milchwirtschaftliche Forschungsberichte*, 55(4), pp.301–315. Available at: <http://www.mspscorp.com/products-detail.php/aerosol/model-m100110nr-nonrotating-moudi-impactor>.

O’Geen, H., Yu, A.S. & Segal, D.J., 2015. How specific is CRISPR-Cas9 really? *Current Opinion in Chemical Biology*, 29, pp.72–78. Available at: <http://dx.doi.org/10.1016/j.cbpa.2015.10.001>.

O’Sullivan, O. et al., 2009. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiology*, 9(1), p.50. Available at: <http://www.biomedcentral.com/1471-2180/9/50>.

- Ott, A., Germond, J.-E. & Chaintreau, A., 2000. Origin of acetaldehyde during milk fermentation using ¹³C-labeled precursors. *Journal of agricultural and food chemistry*, 48(5), pp.1512–1517.
- Ott, A., Germond, J.E. & Chaintreau, A., 2000. Vicinal diketone formation in yogurt: ¹³C precursors and effect of branched-chain amino acids. *Journal of Agricultural and Food Chemistry*, 48(3), pp.724–731.
- Ozer, B. & Akdemir-Evrendilek, G., 2014. *Dairy Microbiology and Biochemistry: recent developments*, CRC Press.
- Parente, E. & Cogan, T.M., 2004. Starter cultures: General aspects. *Cheese: Chemistry, Physics and Microbiology*, 1(C), pp.123–147.
- Petersen, B.L. et al., 2000. Influence of Capsular and Ropy Exopolysaccharide-Producing *Streptococcus thermophilus* on Mozzarella Cheese and Cheese Whey. *Journal of Dairy Science*, 83(9), pp.1952–1956. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0022030200750715>.
- Van Pijkeren, J.P. & Britton, R.A., 2012. High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Research*, 40(10), pp.1–13.
- Poolman, B., 1993. Energy transduction in lactic acid bacteria. *FEMS microbiology reviews*, 12(1–3), pp.125–147.
- Pourcel, C., Salvignol, G. & Vergnaud, G., 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151(3), pp.653–663.
- Quiberoni, A. et al., 2010. *Streptococcus thermophilus* bacteriophages. *International Dairy Journal*, 20(10), pp.657–664.
- Rasmussen, T.B. et al., 2008. *Streptococcus thermophilus* core genome: Comparative genome hybridization study of 47 strains. In *Applied and Environmental Microbiology*. pp. 4703–4710.
- Rathsam, C., Giffard, P.M. & Jacques, N.A., 1993. The cell-bound fructosyltransferase of *Streptococcus salivarius*: The carboxyl terminus specifies attachment in a

Streptococcus gordonii model system. *Journal of Bacteriology*, 175(14), pp.4520–4527.

Roberts, R.J. et al., 2015. REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, 43(D1), pp.D298–D299.

Rodriguez-Valera, F. et al., 2009. Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology*, 7(11), pp.828–836.

São-José, C. et al., 2006. The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. *Journal of Biological Chemistry*, 281(17), pp.11464–11470.

Schleifer, K.H. et al., 1991. Revival of the species *Streptococcus thermophilus* (ex Orla-Jensen, 1919) nom. rev. *Systematic and applied microbiology*, 14(4), pp.386–388.

Semenova, E. et al., 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences*, 108(25), pp.10098–10103. Available at: <http://www.pnas.org/content/108/25/10098.abstract>.

Siezen, R.J. et al., 2010. Complete genome sequence of *Lactococcus lactis* subsp. *lactis* KF147, a plant-associated lactic acid bacterium. *Journal of Bacteriology*, 192(10), pp.2649–2650.

Slos, P. et al., 1991. Isolation and characterization of chromosomal promoters of *Streptococcus salivarius* subsp. *thermophilus*. *Applied and Environmental Microbiology*, 57(5), pp.1333–1339.

Smid, E.J., Poolman, B. & Konings, W.N., 1991. Casein utilization by lactococci. *Applied and Environmental Microbiology*, 57(9), pp.2447–2452.

Sorek, R., Lawrence, C.M. & Wiedenheft, B., 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry*, 82, pp.237–266.

Stanley, E. et al., 2000. Identification of four loci isolated from two *Streptococcus thermophilus* phage genomes responsible for mediating bacteriophage resistance. *FEMS microbiology letters*, 182(2), pp.271–277.

- Stanley, E. et al., 1997. Sequence analysis and characterization of O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology*, 143(11), pp.3417–3429.
- Stiles, M.E. & Holzapfel, W.H., 1997. Lactic acid bacteria of foods and their current taxonomy. *International journal of food microbiology*, 36(1), pp.1–29.
- Szymczak, P. et al., 2017. Novel variants of *Streptococcus thermophilus* bacteriophages are indicative of genetic recombination among phages from different bacterial species. *Applied and environmental microbiology*, 83(5), pp.e02748-16.
- Titgemeyer, F. & Hillen, W., 2002. Global control of sugar metabolism: A Gram-positive solution. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 82(1–4), pp.59–71.
- Tremblay, D.M. & Moineau, S., 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. *Virology*, 255(1), pp.63–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10049822>.
- Tremblay, D.M. & Moineau, S., 1999. Complete Genomic Sequence of the Lytic Bacteriophage DT1 of *Streptococcus thermophilus*. *Virology*, 255(1), pp.63–76.
- Trun, N. & Trempy, J., 2009. *Fundamental bacterial genetics*, John Wiley & Sons.
- Urshev, Z., 2013. Fast acidifying urease-deficient *Streptococcus thermophilus* isolate shows spontaneous deletion of its complete urease operon. , 19(2), pp.112–116.
- Urshev, Z. et al., 2014. Selection and characterization of naturally occurring high acidification rate *Streptococcus thermophilus* strains. *Biotechnology and Biotechnological Equipment*, 28(5), pp.899–903. Available at: <http://dx.doi.org/10.1080/13102818.2014.966233>.
- Vaillancourt, K. et al., 2004. Characterization of a galactokinase-positive recombinant strain of *Streptococcus thermophilus*. *Applied and Environmental Microbiology*, 70(8), pp.4596–4603.
- Vaillancourt, K. et al., 2002. Galactose and lactose genes from the galactose-positive bacterium *Streptococcus salivarius* and the phylogenetically related galactose-

negative bacterium *Streptococcus thermophilus*: organization, sequence, transcription, and activity of the gal gene product. *Journal of Bacteriology*, 184(3), pp.785–793.

Vasiljevic, T. & Shah, N.P., 2008. Cultured milk and yogurt. *Dairy Processing & Quality Assurance*, pp.219–251.

Vasiljevic, T., Shah, N.P. & Jelen, P., 2005. Growth characteristics of *Lactobacillus delbrueckii* subsp. *bulgaricus* ATCC 11842 as affected by different neutralisers. *Australian Journal of Dairy Technology*, 60(1), pp.3–9.

Vaughan, E.E. et al., 2001. Activation of silent gal genes in the lac-gal regulon of *Streptococcus thermophilus*. *Journal of Bacteriology*, 183(4), pp.1184–1194.

Verreault, D. et al., 2011. Detection of airborne lactococcal bacteriophages in cheese manufacturing plants. *Applied and Environmental Microbiology*, 77(2), pp.491–497.

Welman, A.D. & Maddox, I.S., 2003. Exopolysaccharides from lactic acid bacteria: perspectives and challenges. *Trends in Biotechnology*, 21(6), pp.269–274.

Wigley, D.B., 2013. Bacterial DNA repair: Recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nature Reviews Microbiology*, 11(1), pp.9–13.

Wommack, K.E. & Colwell, R.R., 2000. Virioplankton: Viruses in Aquatic Ecosystems. *Microbiology and Molecular Biology Reviews*, 64(1), pp.69–114. Available at: <http://mmbbr.asm.org/cgi/doi/10.1128/MMBR.64.1.69-114.2000>.

Wren, B.W., 2000. Microbial genome analysis: insights into virulence, host adaptation and evolution. *Nature reviews. Genetics*, 1(1), pp.30–39.

Yasmin, A. et al., 2010. Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. *Journal of Bacteriology*, 192(4), pp.1122–1130.

Zhang, W. et al., 2014. A markerless gene replacement method for *B. amyloliquefaciens* LL3 and its use in genome reduction and improvement of poly- γ -glutamic acid production. *Applied Microbiology and Biotechnology*, 98(21), pp.8963–8973.