



FACULTAD DE CIENCIAS  
DEPARTAMENTO DE BIOLOGÍA  
CENTRO DE CÁLCULO CIENTÍFICO UNIVERSIDAD DE LOS ANDES  
(CECALCULA)

Trabajo Especial de Grado

**Diseño *in silico* de un sistema para la detección de la matriz  
CRISPR-Cas en géneros reportados en sustratos del frente glacial  
del Pico Humboldt**

www.bdigital.ula.ve

**Autor: Br. Layla Paola Chmeit Rangel**  
**Tutor Académico: Dra. Bárbara Huber Baur**  
**Tutor externo: Prof. Ascanio Rojas A.**

**Mérida, julio 2024**

## Resumen

Las bacterias están entre los organismos más abundantes y diversos de la Tierra, habitando casi todos los ambientes conocidos, incluidos aquellos de condiciones extremas. Las bacterias coexisten con virus, que son los entes biológicos más diversos. Algunos de estos virus son bacteriófagos, que infectan bacterias y modelan de diferentes maneras su diversidad genética, alterando en algunos casos el material genético durante la infección, lo que conlleva a modelar la dinámica poblacional del huésped. En consecuencia, al ser las bacterias susceptibles a las infecciones virales, éstas han desarrollado distintos mecanismos para contrarrestar su efecto. Las repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR) y sus proteínas asociadas (Cas) representan un sistema “inmunitario” bacteriano adaptativo, que consiste en identificar y degradar secuencias de ácidos nucleicos exógenos (*i.e.* de los virus), e integrar secuencias cortas de infecciones previas, denominados espaciadores, en el locus CRISPR. Alrededor del 50% de las bacterias conocidas presentan este sistema, que se ha diversificado en diferentes tipos y arreglos de proteínas Cas. Mediante un análisis *in silico* se identificó el sistema CRISPR-Cas de la clase I y subtipo I-E en diferentes especies pertenecientes a los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas* y *Rhodococcus*. Cabe mencionar que para el género *Arthrobacter* se trata del primer reporte del sistema CRISPR-Cas. Por otra parte, se encontró el sistema de clase II subtipo II-C1 en especies del género *Sphingomonas*. Fueron resaltantes las diferencias en la arquitectura genética del sistema entre varios representantes de estos géneros. En este trabajo se diseñaron varios oligonucleótidos para la detección de cada tipo de sistema CRISPR-Cas encontrado en los géneros bajo estudio. Los resultados de las pruebas *in silico* revelaron que los oligos para la identificación del gen *cas7* son los más específicos y eficientes para la detección del sistema tipo I-E en 17 órdenes diferentes, incluidos los géneros de interés. De igual forma, para el sistema tipo II-C1 se pudieron diseñar oligonucleótidos que demostraron ser idóneos (en las pruebas *in silico*), para la identificación del gen *cas9* en 32 especies, incluidos varios representantes del género *Sphingomonas*. Nuestro estudio fue capaz de determinar regiones diana para la identificación del sistema CRISPR-Cas de las clases I y II en géneros reportados en sustratos del frente glaciar del Pico Humboldt y en muchos otros taxones bacterianos.

## Agradecimientos

- Agradezco especialmente a los profesores Ascanio Rojas y Bárbara Huber por creer en mí y apoyarme en el desarrollo del trabajo tanto en los momentos de dificultad como en los momentos de éxito, me hicieron una mejor persona y una mejor investigadora.
- Agradezco a mis padres, a mi hermana y a mis hijas por haber estado allí para mí en todo momento y ayudarme a superar cada uno de los obstáculos que enfrenté en el transcurso de este estudio, que con amor y paciencia me llevaron a superarme a mí misma.
- A mi amiga Andrea que fue un soporte emocional y me guio en muchos momentos para disfrutar el proceso y lograr ser mejor cada día.
- A mi compañera y amiga Valentina que compartió conmigo esta maravillosa carrera de Biología y fue una inspiración para continuar estudiando y convertirme en una investigadora.
- A los profesores Ana Cáceres y Wilfredo Quiñones por haberse tomado el tiempo de leer, corregir y aportarme sus conocimientos para el mejoramiento de este manuscrito, pero, además, por haber impartido las mejores asignaturas que vi en el transcurso de la carrera.
- A mis familiares y amigos, por el apoyo y la motivación para seguir adelante.

www.bdigital.ula.ve

## Índice de Contenido

Abreviaturas .....	10
1. Introducción .....	11
2. Marco Teórico.....	14
2.1. Bacterias asociadas a ambientes glaciales .....	14
2.1.1. <i>Arthrobacter</i> .....	16
2.1.2. <i>Cryobacterium</i> .....	16
2.1.3. <i>Pseudomonas</i> .....	17
2.1.4. <i>Rhodococcus</i> .....	18
2.1.5. <i>Sphingomonas</i> .....	18
2.2. Influencia de los bacteriófagos sobre las comunidades bacterianas.....	19
2.3. Sistema CRISPR-Cas.....	20
2.3.1. Adaptación o adquisición del espaciador .....	22
2.3.2. Biogénesis del ARNcr.....	25
2.3.3. Interferencia .....	25
2.4. Clasificación del sistema CRISPR-Cas .....	26
2.5. Antecedentes de estudios <i>in silico</i> del sistema CRISPR-Cas.....	30
3. Justificación .....	31
4. Hipótesis .....	33
5. Objetivos.....	33
5.1. General.....	33
5.2. Específicos .....	33
6. Metodología.....	34
6.1. Recolección de datos y análisis de la matriz CRISPR-Cas .....	34
6.2. Evaluación de los sitios de interés .....	35
6.3. Diseño y prueba de oligonucleótidos .....	36
6.4. Modelado de proteínas.....	36
7. Resultados.....	37
7.1. Recolección de datos y análisis de la matriz CRISPR-Cas .....	37
7.2. Alineamiento y Evaluación de los sitios de interés.....	41
7.2.1 Genes <i>cas1</i> y <i>cas2</i> .....	41

7.2.2	Gen <i>cas3</i> .....	43
7.2.3	Gen <i>cas5</i> .....	44
7.2.4	Gen <i>cas6</i> .....	51
7.2.5	Gen <i>cas7</i> .....	58
7.2.6	Gen <i>cas8</i> .....	66
7.2.7	Gen <i>cas11</i> .....	72
7.2.8	Gen <i>cas1</i> .....	80
7.2.9	Gen <i>cas9</i> .....	82
7.3	Diseño y prueba de oligonucleótidos .....	84
7.3.1	Gen <i>cas1</i> .....	85
7.3.2	Gen <i>cas5</i> .....	85
7.3.3	Gen <i>cas6</i> .....	86
7.3.4	Gen <i>cas7</i> .....	87
7.3.5	Gen <i>cas8</i> .....	93
7.3.6	Gen <i>cas11</i> .....	94
7.3.7	Gen <i>cas1</i> .....	95
7.2.8.	Gen <i>cas9</i> .....	96
7.1.	Modelado de la proteína Cas7 .....	97
8.	Discusión .....	104
8.1.	Las secuencias nucleotídicas/aminoacídicas de los genes de la matriz CRISPR-Cas en los géneros estudiados son muy diversas, y también la arquitectura genética del sistema puede variar	104
8.2.	Es posible identificar las regiones de interés en la matriz CRISPR-Cas, potenciales para el diseño de un sistema de detección .....	109
8.3.	Los oligonucleótidos diseñados y probados <i>in silico</i> para identificar genes diagnósticos de la clase de sistema CRISPR-Cas resultaron ser eficientes y específicos .....	117
8.4.	El Modelado 3D de las proteínas Cas7 y Cas9 brinda información sobre el papel que juegan las regiones seleccionadas como blanco .....	120
9.	Conclusiones .....	121
10.	Perspectivas .....	122
11.	Bibliografía .....	123
12.	Anexos .....	129

## Índice de figuras

<i>Figura 1.- Fases generales y características generales del sistema CRISPR-Cas .....</i>	<i>23</i>
<i>Figura 2.- Clasificación funcional de las proteínas Cas .....</i>	<i>26</i>
<i>Figura 3.- Arquitectura de los loci genómicos para los subtipos de sistemas CRISPR-Cas de la clase I.....</i>	<i>28</i>
<i>Figura 4.- Arquitectura de los loci genómicos para los subtipos de sistemas CRISPR-Cas de la clase II.....</i>	<i>29</i>
<i>Figura 5.- Terapias basadas en el sistema CRISPR y estado de las fases de pruebas clínicas para diferentes enfermedades.....</i>	<i>32</i>
<i>Figura 6.- Arquitectura genética de la matriz CRISPR-Cas de la clase I y el subtipo tipo I-E en especies representativas de los géneros Arthrobacter, Cryobacterium, Pseudomonas y Rhodococcus .....</i>	<i>39</i>
<i>Figura 7.- Arquitectura genética de la matriz CRISPR-Cas de la clase I y el subtipo I-E en la especie Pseudomonas mendocina.....</i>	<i>40</i>
<i>Figura 8.- Arquitectura genética de la matriz CRISPR-Cas de la clase II y subtipo II-C1 en la especie Sphingomonas horti .....</i>	<i>40</i>
<i>Figura 9.- Análisis de las secuencias del gen cas1 en diferentes especies del género Rhodococcus .....</i>	<i>42</i>
<i>Figura 10.- Análisis de las secuencias del gen cas3 en diferentes especies del género Pseudomonas .....</i>	<i>43</i>
<i>Figura 11.- Análisis de las secuencias del gen cas5 en diferentes especies del género Arthrobacter. ....</i>	<i>46</i>
<i>Figura 12.- Análisis de las secuencias del gen cas5 en diferentes especies del género Cryobacterium .....</i>	<i>47</i>
<i>Figura 13.- Análisis de las secuencias del gen cas5 en diferentes especies del género Pseudomonas .....</i>	<i>49</i>
<i>Figura 14.- Análisis de las secuencias del gen cas5 en diferentes especies del género Rhodococcus. ....</i>	<i>50</i>
<i>Figura 15.- Análisis de las secuencias del gen cas6 en diferentes especies del género Arthrobacter. ....</i>	<i>53</i>
<i>Figura 16.- Análisis de las secuencias del gen cas6 en diferentes especies del género Cryobacterium .....</i>	<i>54</i>
<i>Figura 17.- Análisis de las secuencias del gen cas6 en diferentes especies del género Pseudomonas .....</i>	<i>56</i>

<i>Figura 18.- Análisis de las secuencias del gen cas6 en diferentes especies del género Rhodococcus.</i>	57
<i>Figura 19.- Análisis de las secuencias del gen cas7 en diferentes especies del género Arthrobacter</i>	60
<i>Figura 20.- Análisis de las secuencias del gen cas7 en diferentes especies del género Cyobacterium</i>	61
<i>Figura 21.- Análisis de las secuencias del gen cas7 en diferentes especies del género Pseudomonas</i>	63
<i>Figura 22.- Análisis de las secuencias del gen cas7 en diferentes especies del género Rhodococcus</i>	64
<i>Figura 23.- Alineamiento de la región inicial del gen cas7 (a) y de la proteína Cas7 (b) de diferentes especies de los géneros Arthrobacter, Cryobacterium, Pseudomonas y Rhodococcus.</i>	66
<i>Figura 24.- Análisis de las secuencias del gen cas8 en diferentes especies del género Arthrobacter.</i>	68
<i>Figura 25.- Análisis de las secuencias del gen cas8 en diferentes especies del género Cryobacterium</i>	69
<i>Figura 26.- Análisis de las secuencias del gen cas8 en diferentes especies del género Pseudomonas</i>	71
<i>Figura 27.- Análisis de las secuencias del gen cas8 en diferentes especies del género Rhodococcus.</i>	72
<i>Figura 28.- Análisis de las secuencias del gen cas11 en diferentes especies del género Arthrobacter</i>	75
<i>Figura 29.- Análisis de las secuencias del gen cas11 en diferentes especies del género Cryobacterium</i>	76
<i>Figura 30.- Análisis de las secuencias del gen cas11 en diferentes especies del género Pseudomonas</i>	78
<i>Figura 31.- Análisis de las secuencias del gen cas11 en diferentes especies del género Rhodococcus</i>	79
<i>Figura 32.- Análisis de las secuencias del gen cas1 en diferentes especies del género Sphingomonas</i>	82
<i>Figura 33.- Análisis de las secuencias del gen cas9 en diferentes especies del género Sphingomonas</i>	84
<i>Figura 34.- Resultados de la identificación del gen cas7 en especies del grupo 1 del género Arthrobacter haciendo uso de oligos específicos para el grupo</i>	92

<i>Figura 35.- Resultados de la identificación del gen cas7 en especies del grupo 1 del género Arthrobacter haciendo uso de oligos multigénero .....</i>	<i>93</i>
<i>Figura 36.- Resultados de la identificación del gen cas7 en especies del grupo 2 del género Arthrobacter haciendo uso de oligos multigénero .....</i>	<i>93</i>
<i>Figura 37.- Gráfico de Ramachandran del monómero de la proteína Cas7 de la especie Pseudomonas aeruginosa con un 97,04% favorecido.....</i>	<i>98</i>
<i>Figura 38.- Modelado 3D de la proteína Cas7 para la especie Pseudomonas aeruginosa .....</i>	<i>99</i>
<i>Figura 39.- Gráfico de Ramachandran del hexámero de la proteína Cas7 de la especie Pseudomonas aeruginosa con un 98,06% favorecido.....</i>	<i>99</i>
<i>Figura 40.- Modelado 3D de la proteína Cas7 para la especie Pseudomonas aeruginosa, identificando las regiones donde se ubican las secuencias codificadas por el oligo multigénero. ....</i>	<i>100</i>
<i>Figura 41.- Estructura en criomicroscopía electrónica de la proteína Cas9 en complejo con ADN y ARNsg de la especie Streptococcus pyogenes (Bravo et al., 2022).....</i>	<i>102</i>
<i>Figura 42.- Estructura en criomicroscopía electrónica de la proteína Cas9 en complejo con ADN y ARNsg de la especie Streptococcus pyogenes señalando los dominios funcionales (Bravo et al., 2022) .....</i>	<i>103</i>
<i>Figura 43.- Alineamiento de COBALT de la proteína Cas1 en diferentes especies bacterias</i>	<i>110</i>
<i>Figura 44.- Alineamiento de COBALT de la proteína Cas6 en diferentes especies bacterias</i>	<i>112</i>
<i>Figura A.- Diagrama de flujo de la evaluación e identificación de sitios de interés para el diseño de oligonucleótidos para la identificación de los genes cas del sistema CRISPR-Cas.....</i>	<i>131</i>

### **Índice de tablas**

<i>Tabla 1.- Código IUPAC de nucleótidos para generar bases degeneradas .....</i>	<i>84</i>
<i>Tabla 2.- Oligonucleótidos diseñados para la identificación del gen cas5 en los géneros Cryobacterium, Pseudomonas y Rhodococcus.....</i>	<i>85</i>
<i>Tabla 3.- Oligonucleótidos diseñados para la identificación de la región intermedia entre los genes cas5 y cas6 en el género Pseudomonas .....</i>	<i>87</i>
<i>Tabla 4.- Oligonucleótidos diseñados para la identificación del gen cas7 para los géneros Arthrobacter, Cryobacterium, Pseudomonas, Rhodococcus y oligonucleótidos multigénero .....</i>	<i>88</i>
<i>Tabla 5.- Oligonucleótidos diseñados para la identificación del gen cas8 en el género Pseudomonas .....</i>	<i>94</i>
<i>Tabla 6.- Oligonucleótidos diseñados para la identificación del gen cas11 para los géneros Cryobacterium y Pseudomonas .....</i>	<i>95</i>

<i>Tabla 7.- Oligonucleótidos diseñados para la identificación del gen cas1 en el género Sphingomonas</i> .....	96
<i>Tabla 8.- Oligonucleótidos diseñados para la identificación del gen cas9 en el género Sphingomonas</i> .....	97
<i>Tabla A.- Características de los oligonucleótidos diseñados para la identificación del gen cas1 en el género Rhodococcus</i> .....	132
<i>Tabla B.- Características de los oligonucleótidos diseñados para la identificación del gen cas5 en los géneros Cryobacterium, Pseudomonas y Rhodococcus</i> .....	133
<i>Tabla C.- Características de los oligonucleótidos diseñados para la identificación de la región intermedia entre los genes cas5 y cas6 en el género Pseudomonas</i> .....	134
<i>Tabla D.- Características de los oligonucleótidos diseñados para la identificación del gen cas7 en los géneros Arthrobacter, Cryobacterium, Pseudomonas, Rhodococcus y oligonucleótidos multigénero</i> .....	135
<i>Tabla E.- Características de los oligonucleótidos diseñados para la identificación del gen cas8 en el género Pseudomonas</i> .....	137
<i>Tabla F.- Características de los oligonucleótidos diseñados para la identificación del gen cas11 en los géneros Cryobacterium y Pseudomonas</i> .....	138
<i>Tabla G.- Características de los oligonucleótidos diseñados para la identificación del gen cas1 en el género Sphingomonas</i> .....	139
<i>Tabla H.-Características de los oligonucleótidos diseñados para la identificación del gen cas9 en el género Sphingomonas</i> .....	139

## Abreviaturas

<b><i>ADN</i></b>	Ácido desoxirribonucleico
<b><i>Alfa</i></b>	Tasa de mutaciones sinónimas en un sitio
<b><i>ARN</i></b>	Ácido ribonucleico
<b><i>ARNcr y ARNsg</i></b>	ARN maduro que corresponde al guía del sistema CRISPR-Cas
<b><i>ARNr 16S</i></b>	Marcador molecular para reconstrucción de filogenias, es el componente de la subunidad menor (30S) de los ribosomas procariotas
<b><i>ARNcrtra</i></b>	Fragmento de ARN que interactúa con el sistema de Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas (CRISPR) en los sistemas de la clase II
<b><i>CRISPR</i></b>	<i>Clustered regularly interspaced short palindromic repeats</i> que significa Repeticiones palindrómicas cortas agrupadas y regularmente inter-espaciadas
<b><i>CRISPR-CAS</i></b>	Repeticiones palindrómicas cortas agrupadas y regularmente inter-espaciadas asociadas a proteínas Cas
<b><i>PAM</i></b>	<i>Protospacer adjacent motif</i> , que significa motivo adyacente al protoespaciador
<b><i>pi</i></b>	Estimador de diversidad de nucleótidos, que mide el grado de polimorfismo dentro de una población o entre especies
<b><i>RRM</i></b>	Motivo de reconocimiento de ARN

## 1. Introducción

Las bacterias representan una de las formas de vida más diversas y antiguas de la Tierra, albergando posiblemente más de un millón de especies y colonizando una amplia variedad de ambientes (Sociedad Americana de Microbiología, 1994). Las bacterias habitan en ambientes extraordinariamente diferentes, desde los que ofrecen condiciones ideales para la mayoría de los seres vivos hasta los ambientes extremos donde no se pueden desarrollar la mayoría de las formas de vida. Habitan en los entornos favorables y ricos en nutrientes como suelos, lagos, océanos y otros organismos, pero también se encuentran en entornos extremos como aguas termales a temperaturas cercanas a la ebullición (Brock, 1978), salmueras casi saturadas (Anton et al., 2000), aguas ácidas de minas, a pH cercano a cero (Baker & Banfield, 2003), en las profundidades del hielo antártico y en todos los glaciares (Price, 2000; Christner et al., 2001), y a kilómetros bajo la superficie terrestre (White et al., 1998). De hecho, las bacterias definen los límites medioambientales de la vida.

Las bacterias son abundantes y se estima que el número total de estos microorganismos en la Tierra puede llegar a ser mayor a  $4-6 \times 10^{30}$ , donde la mayor proporción se encuentra en las superficies oceánicas y terrestres:  $3,5 \times 10^{30}$  y  $0,25-2,5 \times 10^{30}$ , respectivamente. Estos organismos contribuyen al flujo global de nutrientes, y su biomasa constituye un reservorio significativo de nutrientes. En la década de los 90 se conocía menos del 1% de todas las especies de bacterias, debido a limitaciones en las técnicas de detección que requerían el crecimiento de los organismos en el laboratorio (Torsvik et al., 1990; Ward et al., 1990; Amann et al., 1995). Sin embargo, gracias al desarrollo de diversas técnicas basadas en biomarcadores y análisis de ADN se ha podido detectar un gran número de especies no cultivables hasta la fecha, que ha permitido obtener un conocimiento más detallado de las comunidades bacterianas.

La diversidad metabólica de las bacterias se puede comparar con su diversidad taxonómica y evolutiva. Se ha demostrado que las bacterias presentan diversos modos de conversión de energía, amplios rangos de uso de sustratos y vías metabólicas únicas. La capacidad de adaptación y las diferentes rutas metabólicas que presentan les permite intervenir en muchos de los procesos medioambientales que sustentan la vida en la Tierra, y su diversidad tiene una gran importancia aplicada en la biorremediación (la degradación biológica de contaminantes) y la bioprospección (la búsqueda de nuevos productos bioquímicos para su uso en medicina e industria). Las comunidades bacterianas desempeñan un papel crucial en los ciclos biogeoquímicos, influyendo en la transformación y movilización de elementos esenciales como el carbono, nitrógeno, fósforo y azufre, los cuales son fundamentales en diferentes ecosistemas. Por ejemplo, en el ciclo del carbono, las bacterias participan en la descomposición de la materia orgánica y la respiración; además, las bacterias metanogénicas producen el gas metano que impacta en el calentamiento global (Yan et al., 2021). Diversas bacterias, entre las que se encuentran representantes del género *Rhizobium*, convierten el nitrógeno atmosférico en amoníaco, y otras bacterias transforman el amoníaco en nitritos y nitratos; estos compuestos pueden ser absorbidos por las plantas, por lo que estas bacterias han representado una herramienta muy importante en el mejoramiento de la

fertilidad del suelo en la agricultura, usadas como bioinoculantes para el aumento del rendimiento agrícola, pero también en la biorremediación (Pessi et al., 2022). Además, muchas bacterias también tienen una influencia significativa en la solubilización y mineralización del fósforo, y en la conversión del sulfuro de hidrógeno en sulfatos que de igual manera pueden ser aprovechados por las plantas (Madsen, 2011). Las bacterias no sólo benefician a las plantas, al permitir la absorción de diferentes nutrientes, sino que también muchas de ellas tienen la capacidad de producir diferentes hormonas (promoviendo el crecimiento vegetal) y sideróforos (permitiendo la protección contra patógenos). En consecuencia, estos microorganismos han sido objetos de estudio en la industria alimentaria (Yarzabal, 2020). Las bacterias también pueden tener impacto en la cría de animales, que pueden ser positivos, al intervenir en el proceso de digestión al descomponer la celulosa y otros polisacáridos; además, las bacterias intestinales de muchos animales sintetizan vitaminas esenciales como la vitamina K y varias del complejo B que son cruciales en el tracto gastrointestinal inferior para asimilar fuentes de carbono (Weiland-Bräuer, 2021). Las bacterias también pueden jugar un papel en la estimulación del sistema inmune, e incluso pueden influir sobre el comportamiento animal. Por otro lado, pueden tener un impacto negativo en diferentes organismos, al ocasionar infecciones que afectan al huésped. Por ejemplo, en humanos, *Bacillus anthracis* es responsable del ántrax, *Borrelia spp.* es causante de la enfermedad de Lyme y *Haemophilus influenza* es responsable de la meningitis bacteriana e infecciones del tracto respiratorio (Weiland-Bräuer, 2021). En el caso de las plantas, se pueden nombrar los patógenos *Xanthomonas spp.* y *Agrobacterium tumefaciens*, responsables de diferentes afecciones (Navarrete et al., 2014). Por esta razón, el estudio de las comunidades bacterianas es de gran importancia y en los últimos años se ha incrementado el interés en éstas, gracias al desarrollo de nuevos enfoques y nuevas tecnologías.

Por su parte, las bacterias también son susceptibles a infecciones, específicamente por virus que son denominados bacteriófagos y que se encuentran en una amplia variedad de ambientes coexistiendo con sus huéspedes en la mayoría de los ecosistemas. Están entre los entes biológicos más abundantes de la biósfera, con un estimado de  $4,8 \times 10^{31}$  partículas virales (Weiland-Bräuer, 2021). Se ha sugerido que en el agua de mar existen alrededor de 10 fagos por cada célula bacteriana y se han mostrado cifras similares para ambientes de agua dulce. En ambientes más extremos la situación no está del todo clara y el número de virus por célula bacteriana puede variar. En suelo seco las bacterias alcanzan abundancias de  $1 \times 10^8$  células/g y los virus de  $5 \times 10^9$  partículas/g. En los océanos, las bacterias alcanzan densidades de  $5 \times 10^5$  células/mL y las partículas virales pueden alcanzar  $1 \times 10^{11}$  virus/mL (Weiland-Bräuer, 2021). Los bacteriófagos ejercen una fuerte influencia sobre la dinámica competitiva entre cepas y especies bacterianas, facilitando además la transferencia horizontal de genes entre bacterias, y controlando la expresión de determinados genes. Además, estos agentes virales afectan directamente la composición de las comunidades al exhibir una gran variedad de estrategias de transmisión que impactan directamente en la célula huésped, asegurando su propia reproducción. Por un lado, los fagos de estilo de vida temperado y los fagos filamentosos establecen asociaciones a largo plazo con las comunidades bacterianas, haciendo uso de mecanismos como la lisogenia y pseudolisogenia (Koskella & Brockhurst, 2014). En contraste, los fagos con un estilo de vida exclusivamente lítico, al inducir la muerte celular, pueden llegar a ocasionar una mortalidad significativa en sus huéspedes, lo que influye directamente sobre la ecología de las poblaciones bacterianas, conllevando a la liberación

de material orgánico al entorno, que puede servir como sustrato para el crecimiento de otras poblaciones microbianas.

Durante la infección viral, algunas bacterias tienen la capacidad de incorporar elementos genéticos en su genoma, como secuencias espaciadoras, en matrices de secuencias palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR, del inglés *Clustered Regularly Interspaced Short Palindromic Repeats*) que pueden incrementarse gradualmente con cada nueva infección. Este mecanismo de inmunidad bacteriana confiere protección contra bacteriófagos, al proporcionar un registro de infecciones anteriores y degradar su material genético gracias a las proteínas Cas.

Por otro lado, los organismos psicrófilos y psicrotolerantes que habitan en ambientes glaciares, caracterizados por restricciones energéticas, tienden a desfavorecer este sistema de defensa, por el alto costo energético que representa mantenerlo. Sin embargo, estudios de metagenómica de flavobacterias en entornos polares han revelado una interacción dinámica y continua entre el huésped y los fagos de las comunidades de hipolitos (comunidades de organismos fotosintéticos y extremófilos) antárticos. Esto ocurre especialmente durante los períodos de fusión del hielo, evidenciando en los genomas la historia de numerosos ataques virales (Heinrichs et al., 2023). Aunque en ambientes glaciares de altas latitudes incluso se han encontrado evidencias de la dinámica bacterias-fagos (Heinrichs et al., 2023), en los ambientes asociados a los glaciares tropicales, el conocimiento sobre la estructura y funcionamiento de las comunidades bacterianas *per se* es escaso.

El glaciar del pico Humboldt, que fue declarado extinto en mayo del 2024, y que está ubicado en el Parque Nacional Sierra Nevada Mérida, Venezuela (8°33'05"N y 70°59'54"O), a unos 4.900 m.s.n.m, constituye en el presente una delgada capa de hielo. Desde 1952 se inició el reporte del retroceso de este glaciar, como consecuencia del cambio climático. Su derretimiento ha desencadenado un proceso de sucesión primaria, dando pie a la colonización de diferentes microorganismos, en un primer momento, seguida de la colonización de líquenes, briófitos, plantas vasculares y polinizadores (Llambí et al., 2021). En el año 2014 se identificaron varios géneros de bacterias psicrófilas y psicrotolerantes presentes en el hielo del glaciar, haciendo uso del marcador *ARNr 16S* (Ball et al., 2014). Recientemente, se aislaron bacterias psicrófilas y psicrotolerantes a partir de biocostras (comunidades complejas de líquenes, musgos, microorganismos, microartrópodos, etc.) ubicadas en el frente glacial, y su identificación molecular determinó su asignación a los géneros *Pseudomonas*, *Arthrobacter*, *Fronidhabitans*, *Rhodococcus*, *Sphingomonas*, *Paenarthrobacter* y *Cryobacterium* (García, 2023). Debido a que la mayoría de especies pertenecientes a estos géneros habitan ecosistemas con condiciones extremas, resulta plausible considerar que estas bacterias coexisten con una gran variedad de bacteriófagos y que son, en consecuencia, susceptibles a infecciones. En este sentido, se puede inferir que algunas poblaciones bacterianas en estos ambientes poli-extremos poseen el sistema CRISPR-Cas como mecanismo de defensa. Además, dadas las condiciones distintivas de estos ambientes, podría

encontrarse un tipo de sistema diferente a los conocidos hasta el momento, con características únicas que podrían ser potenciales para su aplicación en el ámbito de la biotecnología y como herramienta en la edición de genes. El ecosistema del frente glacial corresponde a un ambiente único, con condiciones particulares que están en constante cambio, como lo son la temperatura, radiación y disposición de nutrientes. Muchas de las especies pertenecientes a los géneros identificados en este ecosistema también se han encontrado en una amplia variedad de ambientes con condiciones extremas, por lo que es de gran interés determinar la presencia del sistema CRISPR-Cas en estos géneros tan particulares para tener mayor conocimiento del mecanismo de defensa en ambientes extremos que podrían ser potenciales para usos en el ámbito de la biotecnología.

Siguiendo este razonamiento, nos planteamos las siguientes interrogantes:

¿Se ha reportado el sistema CRISPR-Cas en los géneros bacterianos que han sido previamente identificados en biocostras del frente glacial del Pico Humboldt?

¿Entre los géneros bacterianos que han sido reportados en biocostras del frente glacial del Pico Humboldt se comparte el mismo tipo de sistema y la misma arquitectura genética de la matriz CRISPR-Cas?

¿Es posible diseñar un sistema *in silico* que permita la identificación de la matriz CRISPR-Cas en diferentes géneros de bacterias que han sido reportados en el frente glacial del Pico Humboldt?

## 2. Marco Teórico

### 2.1. Bacterias asociadas a ambientes glaciales

La criósfera terrestre, que incluye glaciares, capas de hielo, hielo marino, hielo lacustre y suelo congelado, es un foco principal de investigación de microorganismos en hábitats fríos. Estos ambientes, con bajas temperaturas que ralentizan los procesos físicos y biológicos, albergan bacterias psicrófilas y psicrotolerantes (Margesin & Collins, 2019). Las bacterias psicrófilas, tolerantes al frío o adaptados al frío, pueden crecer desde los 15 °C o menos para un crecimiento óptimo, 0 °C o menos para un crecimiento mínimo y un máximo de 20 °C (Morita, 1975). Las psicrotolerantes también viven en ambientes fríos, aunque pueden tolerar temperaturas más altas. Estas bacterias enfrentan condiciones adversas como escasez de nutrientes, sequedad, baja actividad del agua, alta radiación ultravioleta y estrés oxidativo, desarrollando diversos mecanismos de supervivencia y adaptación (Morita & Moyer, 2001).

La incorporación de nuevas tecnologías ha permitido una mayor comprensión de la vida microbiana en hábitats fríos. Gracias a la metagenómica se han recolectado alrededor de 1400 conjuntos de datos en ambientes fríos. Dado que en todo el mundo se ha reportado el acelerado retroceso de los glaciares, se han multiplicado los estudios de diversidad bacteriana en glaciares. Un ejemplo es el estudio de la diversidad bacteriana y la estructura comunitaria en el glaciar Midtre Lovénbreen en Svalbard (Noruega), utilizando el marcador del gen *ARNr 16S*, identificando un total de 33 filos, siendo Acidobacteriota, Actinobacteriota, Proteobacteria, Chloroflexi y Verrucomicrobiota los predominantes. El paisaje de Svalbard muestra rápidos cambios debido a la desglaciación. Alrededor del 7% de los glaciares de Svalbard han desaparecido en los últimos 30 años (Venkatachalam et al., 2021). Otro ejemplo, esta vez en glaciares de los Andes, está constituido por una de las primeras descripciones de las comunidades bacterianas de los agujeros de criocanita en glaciares de América del Sur, donde se compararon un glaciar argentino y cuatro chilenos que difieren en sus entornos geográficos y ecológicos. Haciendo uso de la región hipervariable V5-V6 del gen *ARNr 16S*, los autores lograron determinar que los órdenes Betaproteobacteriales, Cytophagales, Chitinophagales, Acetobacterales, Frankiales, Armatimonadales, Sphingobacteriales, Rhizobiales, Bacteroidales, Sphingomonadales y Micrococcales, eran los más abundantes (Pittino et al., 2023).

Para el caso de Venezuela, uno de los estudios pioneros en comunidades bacterianas de los glaciares de los Andes Tropicales corresponde al estudio de bacterias psicrófilas asociadas al hielo del glaciar del Pico Humboldt, donde reportaron el aislamiento, caracterización e identificación de las mismas, haciendo uso del marcador molecular *ARNr 16S*. Los resultados mostraron la presencia de cinco filos/clases diferentes: Alfa-, Beta y Gammaproteobacterias, Actinobacterias y Flavobacterias (Ball et al., 2014). Posteriormente, Balcázar y colaboradores (2015) focalizaron su investigación en varias especies del género *Pseudomonas*, estudiando la potencialidad de estas bacterias para promover el crecimiento vegetal, y demostrando que los microorganismos del hielo de los glaciares tropicales pueden usarse para desarrollar bioinoculantes que operen a bajas temperaturas (Yarzabal, 2020).

En 2023, en el Laboratorio de Microbiología Molecular y Biotecnología de la Universidad de los Andes de Mérida, Venezuela, se aislaron bacterias psicrófilas y psicotolerantes de biocostras presentes en el frente del glaciar del Pico Humboldt. Mediante el marcador del gen *ARNr 16S* se identificaron los géneros bacterianos *Pseudomonas*, *Rhodococcus*, *Arthrobacter*, *Cryobacterium*, *Frondehabitans*, *Paenarthrobacter* y *Sphingomonas*, lo que aportó información sobre algunas de las cepas bacterianas constituyentes de las comunidades bacterianas asociadas a este glaciar tropical (García, 2023). El ecosistema incipiente del frente glacial corresponde a un ambiente muy único con condiciones particulares que están en constante cambio, como la temperatura, radiación y la disponibilidad de nutrientes. Muchas de las especies pertenecientes a los géneros identificados en este ecosistema también se han encontrado en diversos ambientes con condiciones extremas. Por lo tanto, es de gran interés conocer algunos de sus atributos adaptativos, como la presencia del sistema CRISPR-Cas, para darles posibles usos en el ámbito de la biotecnología.

## **Características generales de algunos géneros asociados al frente glacial del Pico Humboldt: *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y *Sphigomonas***

### *2.1.1. Arthrobacter*

Género bacteriano perteneciente al filo Actinomycetota, la clase Actinomycetes, al orden Micrococcales y a la familia Micrococcaceae (Reimer et al., 2021). Es un género de bacterias principalmente del suelo, caracterizadas por tener un ciclo de crecimiento de bastones y cocos. Los bastones irregulares en cultivos jóvenes son reemplazados por células cocoides en fase estacionaria, que cuando se transfieren a un medio fresco, producen excrecencias que dan nuevamente lugar a bastones irregulares. Tanto la forma bastoncillo como la cocoide son Gram- positivas. Las células no forman endosporas; son inmóviles o móviles por un flagelo subpolar o unos pocos laterales. Se trata de aerobios obligados y catalasa positivos. Su metabolismo es respiratorio y no es fermentativo. El contenido de GC se ubica en el rango de 59-66% mol (rama de actinomiceto) y el peptidoglicano de la pared celular contiene lisina como diaminoácido. Un total de 15 especies de *Arthrobacter* (en sentido estricto) fueron reportadas en el Manual de Bacteriología Sistemática de Bergey hasta el año 2000, mientras que ahora se reconocen alrededor de 70 especies (Euzéby: lista de nombres procarióticos con posición en la nomenclatura Género *Arthrobacter*). La versatilidad nutricional es característica: carbohidratos, ácidos orgánicos, aminoácidos, compuestos aromáticos y ácidos nucleicos se utilizan como fuentes de carbono y energía. Con excepción de la biotina, no requieren vitaminas u otros factores de crecimiento orgánico. Utilizan principalmente nitrógeno inorgánico. La temperatura óptima para el crecimiento es de 25 a 30 °C y la mayoría crecen en el rango de aproximadamente 10 a 35 °C. Muchas cepas crecen a 5 °C y a 37 °C (Gobbetti & Rizzello, 2014).

Las comunidades de este género son numéricamente importantes en la biota autóctona de suelos y rizósferas. La versatilidad nutricional, la extrema resistencia a la desecación e inanición aseguran su predominio en suelos de diferentes ubicaciones geográficas. La acidez del suelo disminuye su viabilidad celular. Las cepas psicrotóficas del género son abundantes en los ambientes del subsuelo terrestre y se encuentran en el Ártico y la Antártida, limos de glaciares, crioconitas de glaciares alpinos y cuevas de hielo (Gobbetti & Rizzello, 2014).

Hasta el momento, aún no hay datos concluyentes sobre la presencia del sistema CRISPR-Cas en especies del género *Arthrobacter*.

### *2.1.2. Cryobacterium*

Este género pertenece al filo Actinomycetota, la clase Actinomycetes, al orden Micrococcales y a la familia Microbacteriaceae. Representa un grupo específico de bacterias que habitan ambientes de baja temperatura. Fue propuesto por Suzuki et al. (1997) y contiene bacterias Gram-positivas y aeróbicas. La mayoría de las cepas de *Cryobacterium* son tolerantes al frío y algunas de ellas son psicrófilas que crecen temperaturas de  $\leq 20$  °C (Liu et al., 2020). Hasta el 2020 se habían registrado 14 especies con nombres válidamente publicados (Parte et al., 2020) de las cuales 12 especies fueron aisladas de la criósfera.

Dastager et al. (2008) aislaron diferentes especies del género *Cryobacterium* y las especies analizadas mostraron un crecimiento óptimo de 0 a 25 °C, a excepción de la cepa *C. mesophilum* que creció entre 20 y 28 °C, y tuvo un crecimiento óptimo entre 25 y 28 °C. El análisis fenotípico mostró que el género *Cryobacterium* se diferenciaba de sus contrapartes Microbacteriaceae en la presencia de una cantidad significativa de ácido 12-metilpentadecenoico (es decir, a-15:1). La presencia de a-15:1 es inusual pero razonable para las bacterias psicrófilas Gram-positivas con el fin de mantener la fluidez de la membrana a bajas temperaturas (Suzuki et al., 1997).

Se ha reportado la presencia del sistema CRISPR-Cas en un análisis genómico de muestras del género *Cryobacterium* aisladas de núcleos de hielo, revelando adaptaciones a nivel genómico en respuesta a la vida en el hielo. El estudio se basó en la secuenciación de 14 muestras y los resultados los analizaron mediante CRISPR-finder para lograr identificar la presencia del sistema en cuatro cepas (Liu et al., 2020).

### 2.1.3. *Pseudomonas*

Género bacteriano perteneciente al filo Pseudomonadota, la clase Gammaproteobacteria, al orden Pseudomonadales y a la familia Pseudomonadaceae (Reimer et al., 2021). Es uno de los géneros bacterianos más complejos y actualmente contiene el mayor número de especies entre las bacterias Gram-negativas. El número actual de especies reconocidas y válidamente publicadas es 180, incluidas 10 subespecies (Gomila et al., 2015). Las especies de *Pseudomonas* son bacilos aerobios que miden de 0,5-0,8 µm por 1,5-3,0 µm. Poseen un solo flagelo polar. Son bacterias quimiorganotróficas y algunas son quimiolitotróficas que utilizan hidrógeno o dióxido de carbono como fuente de energía. No requieren factores de crecimiento orgánicos y acumulan reservas de carbono de poli-β-hidroxibutirato. Son bacterias catalasa positivas y normalmente oxidasas positivas. No son muy tolerantes a los ácidos y rara vez crecen por debajo de un pH de 5,0 a 6,0. La diversidad de especies dentro del género y los problemas para crear un buen esquema de clasificación para las especies de *Pseudomonas* se traduce en que la identificación de los aislados a nivel de especie sea difícil (Cousin, 1999). *P. aeruginosa* crece bien a 37°C y también puede sobrevivir en un amplio rango de temperaturas de 4°C a 42°C (LaBauve & Wargo, 2012).

*Pseudomonas* es uno de los géneros más dominantes registrados en muchos glaciares. Sus cepas también están presentes en gran variedad de ambientes y pueden mantener una íntima relación con plantas y animales. Se considera que es uno de los grupos de gamma-proteobacterias más versátiles y ecológicamente importantes, desempeñando un papel fundamental en los ciclos del nitrógeno y el carbono (Mingma, et al., 2019).

Un estudio evidenció la presencia del sistema CRISPR en el 19,6% de especies del género *Pseudomonas* analizadas. Entre estas, el 81,2% pertenecía a la especie *P. aeruginosa*, mientras que el 18,8% restante correspondía a otras especies como *Pseudomonas putida*, *P. stutzeri*, *P. chlororaphis*, *P. mendocina*, *P. pseudoalcaligenes*, *P. alcaliphila*, *P. balearica*, *P. parafulva* y *P. citronellolis*. Se identificaron un total de 113 conjuntos CRISPR típicos en los 48 genomas analizados, con un promedio de dos loci CRISPR por genoma que correspondían a los tipos IF, IE,

IC y IV de Clase I (Parra-Sánchez et al., 2023). Además, otros estudios realizados en *P. aeruginosa* determinaron que los principales tipos CRISPR observados son IF e IE es esta especie. Los genomas de *P. aeruginosa* son grandes (normalmente de 6 a 7 Mbp) y se ha predicho que aproximadamente el 50% de los genomas secuenciados poseen un sistema CRISPR-Cas activo (Wheatley & MacLean, 2021).

#### 2.1.4. *Rhodococcus*

Género bacteriano perteneciente al filo Actinomycetota, la clase Actinomycetes, al orden Mycobacteriales y a la familia Nocardiaceae (Reimer et al., 2021). Con alrededor de 15 especies, son aeróbicas obligadas, Gram-positivas, parcialmente acidorresistentes, catalasa positivas, inmóviles y sin endosporas. No poseen flagelo, aunque a veces pueden existir algunos pili en este género. Tienen forma de cocobacilos filamentosos y ramificados. La pared celular contiene micolato, con una longitud de 34 a 64 átomos de carbono. Además, los miembros de este género producen pigmento rojo, lo que les confiere su nombre. Las reacciones metabólicas son oxidativas y la mayoría de las cepas de este género crecen bien en medios convencionales, y la tiamina (vitamina B<sub>1</sub>) es esencial para el crecimiento de algunas especies. Las especies son sensibles a la lisozima (50 g/ml) y a la mitomicina C (5 µg/ml). El contenido de GC (% molar) en su ADN está entre 67 y 73; sin embargo, puede ser menor en algunas especies (Majidzadeh & Fatahi-Bafghi, 2018).

Este género se distribuye en una gran variedad de ambientes, incluidos suelos, océanos y aguas dulces, así como en los intestinos de insectos o en asociación con esponjas marinas. Algunas especies son patógenas, como el *R. hoagii*, que causa infecciones zoonóticas en animales de pastoreo y *R. fascians*, el agente causante de la enfermedad de las agallas de las hojas en las plantas. Múltiples especies de *Rhodococcus* degradan diversos compuestos orgánicos, incluidos bifenilos policlorados (PCB), hidrocarburos aromáticos policíclicos (PAH) e hidrocarburos alifáticos, otorgando a este género una potencialidad muy prometedora para fines de biorremediación. Probablemente el hecho de tener grandes genomas y la presencia de múltiples elementos extracromosómicos que agregan nuevos rasgos funcionales al contenido general, les permite establecerse en un diverso rango de nichos y tener amplio potencial catabólico (Garrido-Sanz et al., 2020).

En la actualidad hay pocos registros de la presencia del sistema CRISPR-Cas en especies del género *Rhodococcus*. Entre ellos, se encuentra el desarrollado por Napolitano et al. (2022), donde estudiaron la coevolución virus-huésped de *Rhodococcus sp.* cepa ADH y un profago que la parasita.

#### 2.1.5. *Sphingomonas*

Género bacteriano perteneciente al filo Pseudomonadota, la clase Alphaproteobacteria, al orden Sphingomonadales y a la familia Sphingomonadaceae (Reimer et al., 2021). Es un género de bacterias Gram-negativas compuesto por más de 103 especies. Morfológicamente, se pueden

identificar como un bacilo pigmentado de color amarillo, inmóvil y no fermentativo. Sin embargo, la forma mótil tiene un único flagelo polar. La membrana externa de todas las especies de este género exhibe glicoesfingolípidos en lugar de lipopolisacáridos (LPS). Las especies de *Sphingomonas* desempeñan un papel vital en la tolerancia al estrés abiótico de las plantas, la biorremediación y la biodegradación de contaminantes ambientales (Asaf et al., 2020).

Se han aislado cepas de *Sphingomonas* a partir de diferentes fuentes, ya sean contaminadas o no contaminadas, como agua de mar, agua dulce, agua subterránea, aguas residuales, endófitos, hábitats terrestres, sedimentos (de ríos y subsuelo), suelo rizosférico y suelo terrestre. Otras fuentes de *Sphingomonas* incluyen la rizósfera y las superficies de las plantas, agua potable, tuberías de cobre corroídas, sedimentos profundos del subsuelo, infecciones de peritonitis, bacteriemia, pacientes con meningitis y septicemia, equipos de diálisis hospitalarios, heridas y sangre. Existen muy pocas cepas en hábitats naturales extremos, es decir, ambientes higrófilos, hipersalinos, tundra y pelágicos, por lo que es de gran importancia comprender su distribución y su posible rol en dichos ecosistemas (Asaf et al., 2020).

En un estudio del genoma de la cepa Kira de la especie *S. paucimobilis* se encontró una secuencia CRISPR, pero no se determinó el tipo de sistema (Nishimura et al., 2021). En otro estudio del genoma de la cepa *S. sp.* Cra20, se predijeron las secuencias repetidas directamente (DR) y los espaciadores de las secuencias completas del genoma mediante la herramienta *CRISPR recognition tool* (CRT) (Luo et al., 2020).

## 2.2. Influencia de los bacteriófagos sobre las comunidades bacterianas

Los fagos desempeñan un papel crítico en hábitats extremadamente fríos. Sin embargo, aún es muy limitado el conocimiento de la diversidad de fagos polares y el papel que desempeñan en la fisiología y ecología microbianas, debido a la dificultad de obtener muestras ambientales de las regiones polares y debido a la presencia de una gran cantidad de huéspedes de fagos no cultivables. Análisis de metagenómica han determinado que los fagos están muy extendidos en diversos entornos, incluido el océano, hielo marino, suelo, lagos y agujeros de crioonita (Heinrichs et al., 2023). No obstante, sólo se han aislado unos pocos sistemas fago-bacterias de las regiones polares. En un estudio reciente se aislaron seis sistemas cultivados de fago-huésped a partir de muestras de suelo recolectadas en islas antárticas, con cinco fagos que podrían infectar cepas de *Pseudomonas* aisladas de las mismas muestras de suelo. Sin embargo, aún no se tiene clara la interacción de los mismos o la dinámica que tienen en estos ecosistemas (Liu et al., 2023).

Los fagos pueden modular las comunidades bacterianas al tener distintas estrategias de infección, mediante el ciclo lítico, pseudolisogenia y la lisogenia, que pueden afectar de distintas maneras a estas comunidades. En ambientes con alta disponibilidad de recursos se cree que el ciclo lítico se ve favorecido, mientras que en ambientes donde las condiciones son adversas o desfavorables, se sugiere que la lisogenia o pseudolisogenia son los mecanismos más frecuentes, debido principalmente a bajas densidades poblacionales y aislamiento tanto de los huéspedes como de los fagos, dificultando así los eventos de encuentros (Heinrichs et al., 2023). En consecuencia, es de esperarse que esta última estrategia prevalezca en ambientes polares, lo que fue respaldado

en un estudio de los fagos tanto en el Ártico como en la Antártida, donde las condiciones ambientales son poli-extremas, con fuertes cambios estacionales. En dicho estudio se observó un cambio de un modo de replicación predominantemente lisogénico a un modo de replicación lítico cuando la disponibilidad de recursos incrementaba (Angly et al., 2006).

Sin embargo, los mecanismos ambientales y las señales que lleven al balance lisis-lisogenia siguen siendo discutibles, y la tendencia no siempre es tan clara en ecosistemas cambiantes y diversos. Esto se ve reflejado en un estudio en un ambiente pelágico ártico que investigó la dinámica de bacterias y fagos a lo largo de un año, y donde se observó que la lisis mediada por fagos era más importante en invierno (Heinrichst et al., 2023). La liberación de materia orgánica producto de la lisis proporciona nutrientes a los ecosistemas polares y apoya la productividad de los ecosistemas, que es comúnmente limitada por la escasez de nitrógeno y fósforo. En un estudio de la dinámica del bucle viral estacional en lagos árticos de agua dulce, se estimó que la lisis de bacterias mediada por fagos puede aportar más del 60% del carbono al depósito de carbono orgánico disuelto, especialmente en los meses de invierno (Sävström et al., 2007). Los nutrientes liberados por efecto de la acción viral proporcionan recursos al medio, que quedan a disposición de organismos circundantes, lo que nuevamente influye en las estructuras de la comunidad microbiana.

### 2.3. Sistema CRISPR-Cas

En 1987 se reportan por primera vez las repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR) en las secuencias de ADN de la bacteria *Escherichia coli*, descritas en la Universidad de Osaka (Japón) por Ishino y colaboradores (Ishino et al., 1987). Sin embargo, en su investigación, los autores no comprendieron el origen y el significado del sistema en la célula bacteriana. Aunque no se sabía la función de sistema CRISPR, los científicos de la época ya habían propuesto una forma de utilizar la información codificada en los loci CRISPR en la investigación médica, para genotipar varias cepas de bacterias, como *Mycobacterium tuberculosis* y *Streptococcus pyogenes*, encontrando que los loci CRISPR tenían un alto grado de polimorfismo en diferentes cepas de la misma especie de bacterias patógenas, lo que permitió la identificación de cepas bacterianas en condiciones clínicas (Gostimskaya, 2022).

En 1995 en la Universidad de Alicante (España) se llevó a cabo una investigación que condujo a la comprensión de la función biológica de los loci CRISPR, encontrando estructuras genéticas similares en el genoma de la arquea *Haloferax mediterranei*. Posteriormente se estableció la relación de estos elementos descritos en arqueas, con repeticiones de ADN previamente encontradas en genomas bacterianos, planteando por primera vez la hipótesis de que estos inusuales loci incluyen fragmentos de ADN foráneo, lo que sugirió la existencia de un sistema de defensa en bacterias y arqueas (Mojica et al., 1993).

En 2002, el aislamiento y la secuenciación de pequeñas especies de ARN no codificante de *Archaeoglobus fulgidus* reveló que los loci CRISPR se transcriben en pequeños ARN, y los genes *cas* fueron identificados como una familia de genes asociados al locus CRISPR. La homología de

secuencia indicó que muchas de las proteínas Cas participan en reacciones químicas que involucran ácidos nucleicos (Tang et al., 2002).

Los fragmentos de ADN viral “espaciadores” tienen de 17 a 84 bases de largo, están separados por repeticiones palindrómicas cortas de 23 a 50 bases y están agrupados en regiones intergénicas, que conforman una biblioteca de información genética de previas infecciones. En un principio se pensó que el sistema funcionaba como el mecanismo de interferencia del ARN, pero el estudio de Marraffini & Sontheimer (2008) demostró por primera vez que el objetivo del sistema inmunológico de los procariontes era el ADN extraño, y no el ARNm, por lo que el uso de dicho sistema en el laboratorio podría representar una herramienta potencial para la edición genómica (Marraffini & Sontheimer, 2008).

Gracias a tres estudios fundamentales se descubrieron las características básicas de CRISPR-Cas, la adaptación, el papel de las guías ARNcr y la orientación del ácido nucleico invasor. El primer estudio demostró la presunta función de CRISPR-Cas en la prevención de la infección por fagos en *S. thermophilus*. El trabajo demostró un aspecto fundamental de la inmunidad CRISPR-Cas que no estaba previsto en los modelos: que la inmunidad es adaptativa. Brouns y colaboradores (2008) demostraron la necesidad de los ARNcr para la inmunidad, propuesta en los primeros modelos. También, descubrieron un complejo ribonucleoproteico Cas que alberga ARNasa responsable de la generación de ARNcr pequeños (maduros) en *E. coli* (Brouns et al., 2008). Marraffini (2015) demostró que un espaciador presente en el locus CRISPR de *Staphylococcus epidermidis*, que coincidía con una región del gen de la *nickasa* de los plásmidos estafilocócicos, impedía la transferencia conjugativa de estos plásmidos. El estudio también reveló que, a diferencia del mecanismo propuesto similar a la interferencia del ARN, los sistemas CRISPR-Cas proporcionan inmunidad al apuntar al ADN, y no al ARN. También sugirieron la existencia de ADN nucleasas Cas programables y guiadas por ARNcr, lo que abría la posibilidad para el desarrollo de herramientas de manipulación de ADN basadas en CRISPR (Marraffini, 2015).

Se ha encontrado una gran diversidad de genes asociados a CRISPR desde su descubrimiento en procariontes, incluidos representantes de los genes codificantes para las familias de helicasas, nucleasas, polimerasas y otras. Las proteínas asociadas a este sistema pueden asignarse al módulo adaptativo (si participa en la adquisición de inmunidad, representado principalmente por las proteínas Cas1 y Cas2) o al módulo efector, que está directamente involucrado en el procesamiento de elementos genéticos móviles mediante su reconocimiento y escisión, con algunas proteínas adicionales y reguladoras que también se encuentran asociadas al sistema (Hille & Charpentier, 2016).

En la actualidad el sistema CRISPR-Cas se divide principalmente en dos clases, una clasificación basada en el módulo efector. El de la Clase I está representado por un complejo de varias proteínas, mientras que en la Clase II, el sistema sólo consiste en una única proteína multidominio (Cas9, Cas12 o Cas13). Luego cada clase se divide en tres tipos que, a su vez, también se dividen en numerosos subtipos. No obstante, esta clasificación no es concluyente

debido a que se siguen encontrando variaciones de cada uno los niveles de clasificación (Hille & Charpentier, 2016).

El mecanismo de CRISPR-Cas se puede dividir en tres etapas: la adaptación o adquisición del espaciador, la biogénesis de ARNcr y la interferencia del objetivo.

### 2.3.1. Adaptación o adquisición del espaciador

La primera etapa consta de la incorporación de una secuencia distinta de elementos genéticos móviles del invasor (llamado protoespaciador) a la matriz CRISPR, lo que genera un nuevo espaciador (*Spacer* en la figura 1a). De esta manera el organismo huésped puede registrar el material genético del bacteriófago y esto demuestra la naturaleza adaptativa de este sistema de defensa. Las proteínas Cas1 y Cas2, juegan un papel crucial en este proceso de adquisición de espaciadores, y se encuentran en casi todos los tipos CRISPR-Cas. Las excepciones son los sistemas CRISPR-Cas de tipo III-C, III-D y IV, que no albergan proteínas homólogas. Además, el tipo VC muestra una composición mínima, ya que comprende sólo una supuesta proteína efectora denominada C2C3 y un homólogo de Cas1. Diferentes estudios han revelado la bioquímica del proceso de integración del espaciador, donde demuestran que Cas1 y Cas2 del sistema tipo IE de *Escherichia coli* forman un complejo que promueve la integración de nuevos espaciadores (Figura 1b). Tanto Cas1 como Cas2 son nucleasas, pero el sitio catalíticamente activo de Cas2 es prescindible para la adquisición del espaciador. Por lo general, se incorpora un nuevo espaciador en el límite de repetición líder de la matriz CRISPR, mientras que la primera repetición de la matriz se duplica (Hille & Charpentier, 2016).

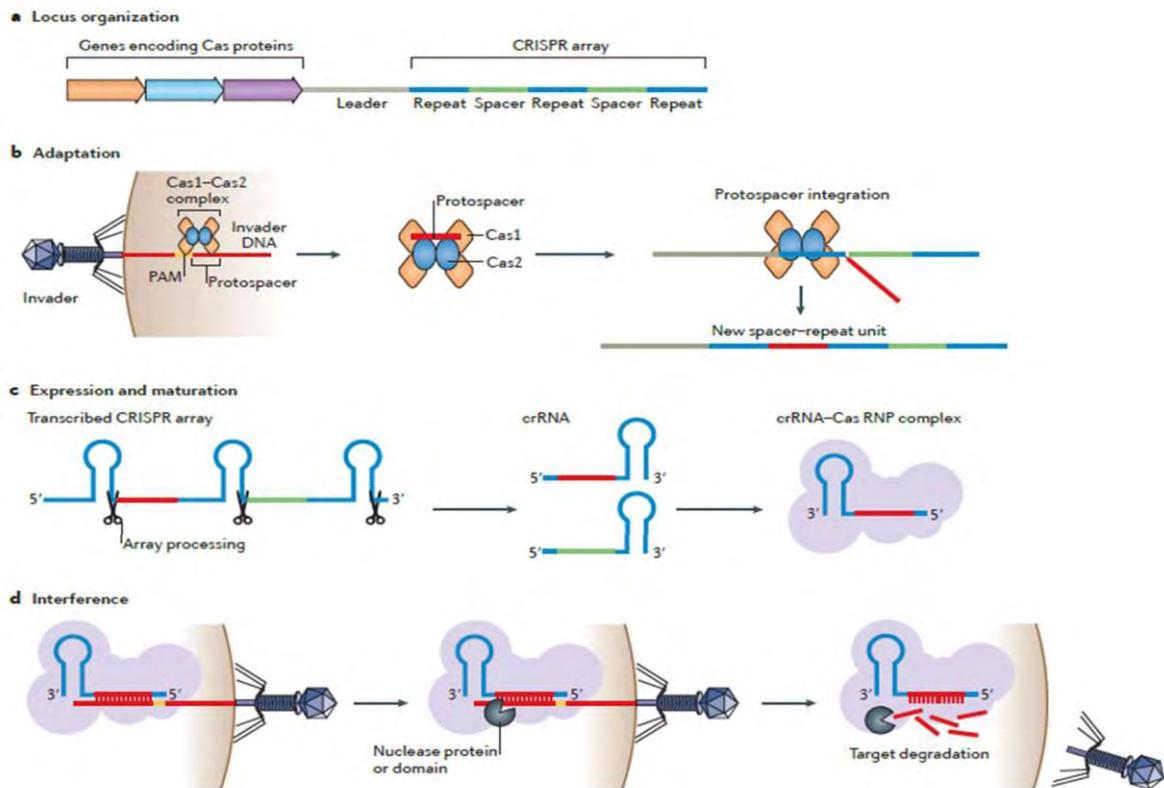


Figura 1.- Fases generales y características generales del sistema CRISPR-Cas. **a.** Organización de un locus CRISPR-Cas típico en un genoma bacteriano o arqueal. **b.** En la etapa de adaptación, el complejo Cas1-Cas2, que comprende dos dímeros Cas1 y un solo dímero Cas2, adquiere un protoespaciador del ADN invasor y lo integra como un nuevo espaciador en la matriz CRISPR. **c.** En la etapa de expresión y maduración, la matriz CRISPR se transcribe y luego se procesa en ARN CRISPR maduros (ARNcr), que forman complejos de ribonucleoproteína (RNP) con proteínas Cas. **d.** En la etapa de interferencia, el complejo ARNcr-Cas RNP identifica el ADN objetivo mediante emparejamiento de bases complementarias en presencia de un motivo protoespaciador adyacente (PAM) y la secuencia diana luego es degradada por proteínas o dominios de nucleasa. Figura tomada de Amitai & Sorek (2016).

Los mecanismos de los diferentes tipos de CRISPR-Cas pueden conservarse sólo hasta cierto punto. Varios estudios han mostrado variaciones en cuanto a los requisitos y objetivos de la maquinaria de adaptación. En el sistema del tipo IB además de las proteínas Cas1 y Cas2, se requiere la proteína Cas4 para la adaptación. El sistema CRISPR-Cas tipo IF de *Pseudomonas aeruginosa* requiere además maquinaria de interferencia para promover la incorporación de nuevos espaciadores. De manera similar, los sistemas de tipo II-A requieren Csn2, Cas9 y del ARN CRISPR transactivador (ARNcrtra) para la adquisición. En el tipo III-B se encontró otro modo de adaptación, hasta ahora único, donde la proteína Cas1 está fusionada a una transcriptasa inversa, proporcionando la adquisición tanto de ADN como de ARN (Hille & Charpentier, 2016).

En los sistemas CRISPR-Cas de tipo I, II y V, se ha demostrado que una secuencia corta es crucial para la adquisición y la interferencia, y es denominada motivo adyacente al protoespaciador (PAM), encontrándose directamente al lado del protoespaciador. En los sistemas CRISPR-Cas tipo II-A, el dominio de reconocimiento PAM de Cas9 es responsable de la selección del protoespaciador. Se cree que después de la selección del protoespaciador, Cas9 recluta Cas1, Cas2

y posiblemente Csn2 para la integración del nuevo espaciador en la matriz CRISPR. Para el tipo IE, el complejo Cas1-Cas2 es suficiente para la selección e integración del espaciador, aunque se ha informado que la presencia del complejo de interferencia aumenta la frecuencia de los espaciadores integrados que son adyacentes a un PAM adecuado. Además, en un proceso llamado cebado, la maquinaria de interferencia de varios sistemas CRISPR-Cas tipo I puede estimular la mayor absorción de nuevos espaciadores tras la unión guiada por ARNcr a un protoespaciador que se seleccionó en una primera infección (Hille & Charpentier, 2016).

Un elemento central del mecanismo de selección de nuevas secuencias espaciadoras es la prevención de la autoinmunidad, es decir, la capacidad de la maquinaria de adquisición para distinguir el ADN propio (cromosómico) del no propio (invasor). En un estudio de una cepa de *E. coli* se sobre-expresaron las proteínas Cas1-Cas2, y se eliminó la maquinaria de inmunidad Cascada (en inglés *Cascade*, complejo efector) y Cas3, determinando que existe una fuerte preferencia por la integración de plásmidos sobre secuencias espaciadoras cromosómicas (Levy et al., 2015). Se ha demostrado que los sitios *Chi* (Secuencias cortas de ADN en el genoma de una bacteria donde es más probable que se produzca una recombinación homóloga que en todo el genoma en promedio) motivos de 8 nucleótidos presentes aproximadamente una vez cada 5 kb en el genoma de *E. coli*, limitan la adquisición de secuencias cromosómicas. En *E. coli*, la sobre-expresión de Cas1-Cas2 en ausencia de inmunidad conduce a la adquisición de secuencias espaciadoras derivadas de regiones limitadas por sitios Chi en un lado y por sitios Ter (secuencias terminadoras que están asociadas con la terminación de la transcripción en el ADN) en el otro. Hay estudios donde los datos de adquisición cromosómica obtenidos en *E. coli*, sugieren que el procesamiento nucleolítico de las roturas del ADNds generadas durante la replicación es la fuente de nuevas secuencias espaciadoras para la maquinaria de adquisición CRISPR-Cas (Levy et al., 2015). Este trabajo predice que la degradación del ADN que sigue a la formación de roturas espontáneas durante la replicación viral o del plásmido produce fragmentos de ADN que son capturados por la maquinaria de adaptación de Cas para su incorporación como espaciadores no propios en la matriz CRISPR. No obstante, hasta el momento no existen pruebas directas (Marraffini, 2015).

Por otro lado, los espaciadores se integran en la matriz CRISPR en una reacción que se asemeja a la integración retroviral. En *E. coli* se demostró que las proteínas Cas1 y Cas2 realizan la reacción de integración del espaciador. Mediante un sistema sobreexpresión Cas1-Cas2 evidenciaron la presencia de intermediarios de integración *in vivo* y utilizando sondas específicas para ensayos de transferencia *Southern* demostraron que cada cadena de la primera secuencia repetida se separa y se liga al extremo 3' de la nueva secuencia espaciadora. Esta reacción es catalizada por el complejo Cas1-Cas2, que está compuesto por dos unidades Cas1-Cas2. El complejo se une a secuencias CRISPR de *E. coli* (*in vitro*) y cuando se incuba con un sustrato de ADNds de 33 nucleótidos (que imita una secuencia espaciadora capturada), media la adición covalente de este ADNds a los plásmidos que albergan la matriz CRISPR. Esta reacción requiere extremos 3' OH libres en el sustrato espaciador de 33 nucleótidos, que presumiblemente realizan un ataque nucleofílico directo al enlace fosfodiéster entre la primera repetición y las secuencias espaciadoras. Los resultados de secuenciación profunda de los productos de reacción revelaron un fuerte sesgo para la integración de espaciadores con un nucleótido de citosina en el extremo 3' (es decir, la base complementaria a la primera guanina (extremo 5') de la secuencia espaciadora)

(Yosef et al., 2012). En base a los estudios se puede argumentar que el complejo Cas1-Cas2 proporciona la especificidad de orientación de la reacción de integración (Marraffini, 2015).

### 2.3.2. Biogénesis del ARNcr

La segunda etapa del mecanismo de CRISPR-Cas es la biogénesis, donde la matriz CRISPR se transcribe en un ARNcr precursor largo (pre-ARNcr) que se procesa posteriormente en ARNcr guías maduros compuestos por las secuencias memorizadas de los invasores (Figura 1c). Los miembros de la familia Cas6 en los sistemas de tipo I y III, realizan el paso de procesamiento produciendo especies intermedias de ARNcr que están flanqueadas por una etiqueta corta en el extremo 5'; en los sistemas de tipo IC, la proteína Cas5d procesa el pre-ARNcr. Puede producirse un recorte adicional del extremo 3' del ARNcr intermedio mediante una nucleasa desconocida y se obtienen especies de ARNcr maduras compuestas por una porción espaciadora completa (extremo 5') y una porción repetida (extremo 3'), que generalmente muestra una estructura en horquilla en la mayoría de los sistemas de tipo I. Por otro lado, la maduración de los ARNcr en los sistemas CRISPR-Cas de clase II, requiere de ARNcrtra para el procesamiento del pre-ARNcr. La secuencia anti-repetición de este ARN permite la formación de un dúplex de ARN con cada una de las repeticiones del pre-ARNcr, que es estabilizado por Cas9. Posteriormente, el dúplex es reconocido y procesado por la ARNasa III del huésped, lo que produce una forma intermedia de ARNcr que sufre una mayor maduración mediante un mecanismo aún desconocido, para conducir al ARN guía pequeño maduro (Deltcheva et al., 2011). Existe un mecanismo independiente de la ARNasa III en el sistema CRISPR-Cas tipo II-C de *Neisseria meningitidis*; en un estudio se identificaron secuencias promotoras que se encuentran dentro de cada repetición y algunas pudieron iniciar la transcripción que conduce a especies de ARNcr intermedias (Hille & Charpentier, 2016).

### 2.3.3. Interferencia

La última etapa del mecanismo de CRISPR-Cas es la interferencia, en la que los ARNcr maduros corresponden a guías para interferir específicamente con los ácidos nucleicos invasores. Los sistemas de clase I emplean complejos Cascade compuesto por proteínas CasABCDE para lograr la degradación del objetivo, mientras que en los sistemas de clase II, una sola proteína efectora es suficiente para la interferencia del objetivo. Una forma de evitar la autofocalización en los sistemas de tipo I, II y V es el reconocimiento específico de la secuencia PAM que se encuentra aguas arriba (tipos I y V) o aguas abajo (tipo II) del protoespaciador (Hille & Charpentier, 2016). La presencia de un PAM promueve la unión del complejo Cascade a su objetivo y la formación de un bucle R entre la secuencia espaciadora de ARNcr y el ADNds. Los primeros 8 pb en el extremo 5' del dúplex ADN-ARNcr son críticos para la inmunidad y para definir una secuencia 'semilla' dentro del objetivo (Marraffini, 2015). En los sistemas de tipo III, la discriminación entre lo propio y lo no propio se logra a través de la etiqueta 5' del ARNcr maduro, que no debe emparejarse con el objetivo para permitir la degradación por parte del complejo (Hille & Charpentier, 2016).

En los sistemas de tipo I, el ARNcr unido al complejo Cascade reconoce el ADN que contiene protoespaciadores que coinciden con el espaciador de ARNcr. La presencia de PAM aumenta la fuerza de la interacción entre 10 y 50 veces; así, permite el reclutamiento de la nucleasa

Cas3 que induce una mella en el ADN extraño y posteriormente degrada el ADN objetivo. En los sistemas CRISPR-Cas de tipo II, el dúplex ARNcrtra-ARNcr guía a la proteína efectora Cas9 para introducir una rotura de doble hebra en el ADN objetivo. La maquinaria de interferencia de los sistemas tipo III comprende complejos Cas10-Csm (tipos III-A y III-D) y Cas10-Cmr (tipos III-B y III-C), que pueden apuntar tanto al ADN como al ARN (Figura 1d) (Hille & Charpentier, 2016).

## 2.4. Clasificación del sistema CRISPR-Cas

La rápida evolución de la mayoría de los genes *cas* y la notable variabilidad en la arquitectura genómica del loci CRISPR-Cas plantea un desafío importante para la correcta anotación de las proteínas Cas y para la clasificación de estos sistemas. Sin embargo, un esquema de clasificación consistente es esencial para una caracterización oportuna y sólida del loci CRISPR-Cas en nuevos genomas y, por lo tanto, es importante para seguir avanzando en la investigación de CRISPR. Debido a la complejidad de la composición genética y la arquitectura genómica de los sistemas CRISPR-Cas, se ha optado por una clasificación basada en evidencias combinadas a partir de análisis filogenéticos, genómicos comparativos y estructurales (Figura 2) (Makarova et al., 2019).

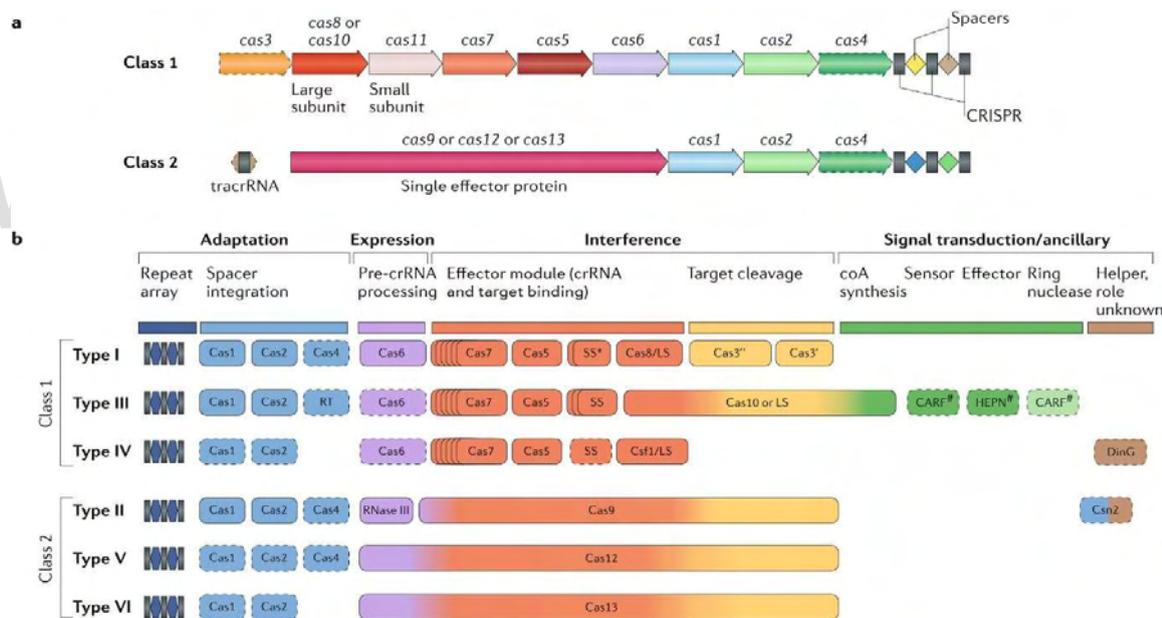


Figura 2.- Clasificación funcional de las proteínas Cas. **a.** Locus representativo de la clase I y la clase II. **b.** Clasificación del sistema CRISPR-Cas identificando las variaciones de las proteínas cas en los tipos de cada una de las clases. Figura tomada de Makarova et al. (2019).

Los sistemas CRISPR-Cas de clase I se definen por la presencia de un complejo ARNcr-efector de múltiples subunidades. Estos incluyen sistemas del tipo I y tipo III, así como el supuesto nuevo tipo IV. Todos los loci de tipo I contienen el gen característico *cas3*, que codifica una helicasa de la superfamilia 2 estimulada por ADN monocatenario (ADNss) con una capacidad demostrada para desenrollar ADN bicatenario (ADNds) y dúplex ARN-ADN. A menudo, el

dominio helicasa se fusiona con un dominio endonucleasa de la familia HD que participa en la escisión del ADN diana. El dominio HD comúnmente se encuentra en el extremo amino de las proteínas Cas3 (Makarova et al., 2019).

Los sistemas de tipo I se dividen actualmente en siete subtipos, del IA al IF adicionando IU. En el caso del subtipo IU, U significa no caracterizado porque el mecanismo de escisión del pre-ARNcr y la arquitectura del complejo efector de este sistema siguen siendo desconocidos. Los sistemas CRISPR-Cas de tipo IC, ID, IE e IF generalmente están codificados por un único operón que abarca los genes *cas1*, *cas2* y *cas3* junto con los genes de las subunidades del complejo Cas. Por el contrario, muchos loci de tipo IA y IB parecen tener una organización diferente en la que los genes *cas* están agrupados en dos o más operones (predichos). En la mayoría de los loci de tipo I, cada una de las familias de genes *cas* está representada por un único gen (Figura 3) (Makarova et al., 2019).

Los sistemas CRISPR-Cas de clase II se definen por la presencia de un módulo efector de ARNcr de una sola subunidad. Esta clase incluye sistemas CRISPR-Cas de tipo II, así como una supuesta nueva clasificación, el tipo V. El gen característico del tipo II es el *cas9*, que codifica una proteína multidominio que combina las funciones del complejo ARNcr-efector con la escisión del ADN objetivo y también contribuye a la adaptación. Además de *cas9*, todos los loci CRISPR-Cas de tipo II identificados contienen *cas1* y *cas2*, y la mayoría de los loci de tipo II también codifican un ARNcrtra, que es parcialmente complementario a las repeticiones dentro de la matriz CRISPR respectiva (Figura 4) (Jinek et al., 2012).

El núcleo de Cas9, que incluye dominios de nucleasa y un grupo característico rico en arginina (Arg), muy probablemente evolucionó a partir de genes de elementos transponibles que no están asociados con CRISPR. Gracias a la similitud de secuencias entre Cas9 y sus homólogos que no están relacionados con el sistema, Cas9 no puede utilizarse como única firma para la identificación de sistemas de tipo II. Sin embargo, la presencia de *cas9* en las proximidades de los genes *cas1* y *cas2* es un sello distintivo de los loci de tipo II. Los sistemas CRISPR-Cas de tipo II se clasifican actualmente en tres subtipos, que se introdujeron en la clasificación anterior (II-A y II-B) o se propusieron posteriormente sobre la base de una organización de locus distinta (II-C). Los sistemas del subtipo II-A incluyen un gen adicional, *cas2*, que se considera un gen característico de este subtipo. Las variantes largas y cortas de *Csn2* forman grupos compactos cuando se superponen a la filogenia de Cas9 y parecen corresponder a dos variantes distintas del subtipo II-A. Sin embargo, al igual que con el subtipo IB, se mantienen en un solo subtipo (Makarova et al., 2019).

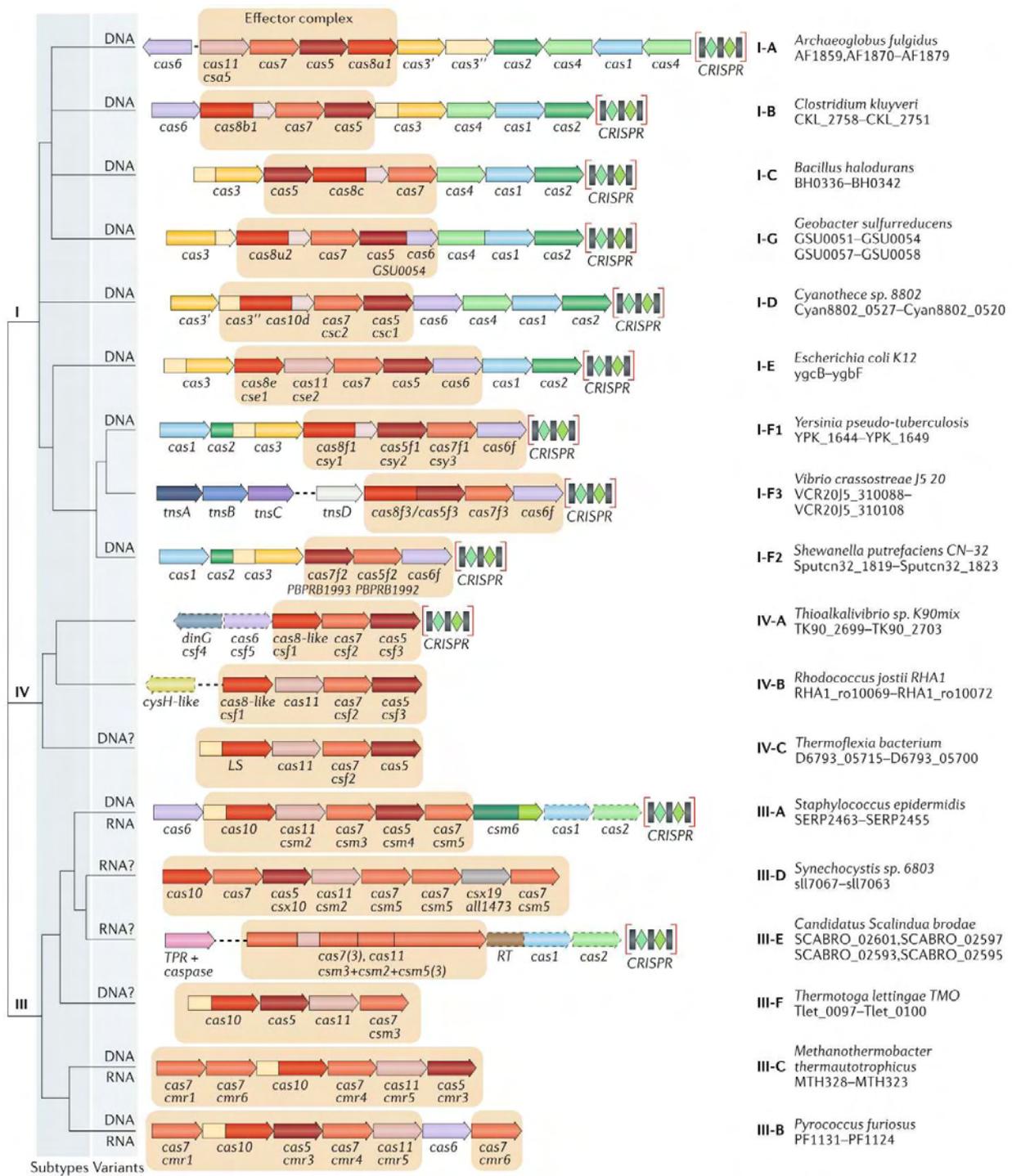


Figura 3.- Arquitectura de los loci genómicos para los subtipos de sistemas CRISPR-Cas de la clase I. Se muestra la organización típica de operones para cada subtipo de sistema CRISPR-Cas. Para cada genoma representativo, se indican los respectivos nombres de las etiquetas del locus genético para cada subunidad. Figura tomada de Makarova et al. (2019).

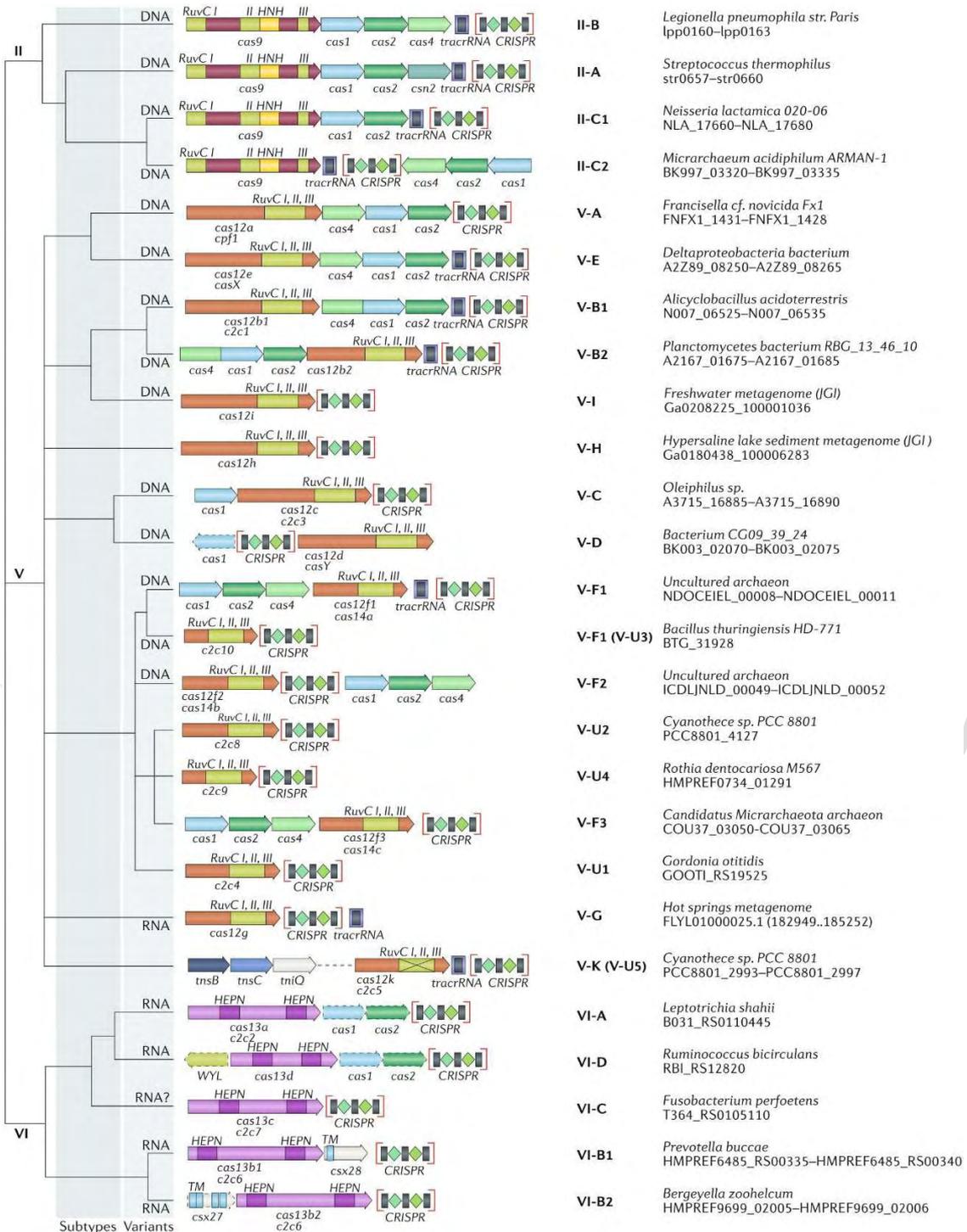


Figura 4.- Arquitectura de los loci genómicos para los subtipos de sistemas CRISPR-Cas de la clase II. Se muestra la organización típica de operones para cada subtipo de sistema CRISPR-Cas. Para cada genoma representativo, se indican los respectivos nombres de las etiquetas del locus genético para cada subunidad. Figura tomada de Makarova et al. (2019).

## 2.5. Antecedentes de estudios *in silico* del sistema CRISPR-Cas

En los últimos años, las investigaciones *in silico* en el campo de la genómica bacteriana han incrementado significativamente, impulsadas por los avances en la tecnología de secuenciación y la capacidad computacional. Utilizando bases de datos génicas, algoritmos bioinformáticos y modelos de predicción se pueden analizar las secuencias de ADN bacteriano de manera precisa, permitiendo la identificación de genes, la predicción de funciones génicas, el estudio de la evolución bacteriana y de la arquitectura genética, entre otros (Beyleveld et al., 2013).

Siendo el sistema CRISPR-Cas una herramienta novedosa y de gran interés para el desarrollo de nuevas biotecnologías, se han realizado diversos estudios mediante el uso de la bioinformática (análisis *in silico*). Entre ellos se encuentra un estudio realizado en Egipto sobre la detección de sistemas CRISPR-Cas en *Pseudomonas aeruginosa* y su correlación con la resistencia a antibióticos. Para ello, se diseñaron diferentes pares de oligos para varios sistemas CRISPR-Cas mediante la obtención de 25 secuencias provenientes de la base de datos NCBI; posteriormente se alinearon las secuencias y se identificaron los sitios conservados, con el fin de diseñar oligonucleótidos para cada gen. Finalmente, los oligos fueron puestos a prueba en 122 cepas de *Pseudomonas aeruginosa* aisladas de muestras clínicas. Los resultados de la PCR arrojaron que 32 aislados albergaban sistemas CRISPR-Cas, de los cuales 10 tenían el subtipo I-C, mientras 9 de los aislados contenían el subtipo I-E y 12 de los aislados albergaban el subtipo I-F (Soliman et al., 2022).

En un estudio del Instituto Aggeu Magalhães de Pernambuco, Brasil, se diseñaron varios pares de oligos para la identificación del sistema CRISPR-Cas subtipo I-F y I-E en 130 cepas clínicas de *Pseudomonas aeruginosa*. Los resultados de PCR indicaron que 73 muestras no presentaban ningún tipo de sistema, 50 aislados fueron positivos para el sistema tipo I-F, 5 aislados fueron positivos para el tipo I-E y 2 aislados dieron positivo para ambos tipos de sistema (De Oliveira et al., 2019). Estos oligos también fueron utilizados en un estudio de la Universidad Federal de Pernambuco para la identificación del sistema I-F en 13 genomas secuenciados de cepas de *Pseudomonas aeruginosa* provenientes de muestras clínicas y 161 genomas anotados de NCBI de muestras de *Pseudomonas aeruginosa* originarias de Brasil. Además, se diseñaron pares de oligonucleótidos para los genes *cas1* y *cas3*, para la identificación del sistema CRISPR-Cas subtipo I-E. Mediante PCR se identificó el sistema en los 13 aislados; en adición, se identificó el subtipo I-F en 56 aislados, ambos subtipos en 6 aislados, el subtipo I-C en 7 aislados y ningún tipo de sistema en 90 aislados (De Oliveira et al., 2021).

### 3. Justificación

Debido a la abundancia de fagos en todos los ambientes y su potencial impacto en las poblaciones bacterianas, se ha incrementado el interés por comprender los procesos coevolutivos que determinan las interacciones entre bacterias y fagos (S. Zhang et al., 2019). Aunque se desconoce con precisión el papel que podrían desempeñar las poblaciones de bacteriófagos en la configuración de poblaciones y comunidades bacterianas, se sabe que éstos cumplen un rol central en este proceso, mediando la transferencia horizontal de genes, impulsando la evolución bacteriana, modulando la competencia entre especies y moldeando la estabilidad de los ecosistemas (Díaz-Muñoz & Koskella, 2014).

El sistema CRISPR-Cas constituye un mecanismo de defensa contra los fagos, y en la actualidad se han desarrollado diferentes aplicaciones de este sistema para su uso en la biotecnología (Doudna & Charpentier, 2014). En un principio, la aplicación de estos sistemas CRISPR-Cas estaba dirigida a la edición de genes, haciéndose cada vez más precisa, estable y versátil, lo que ha permitido la edición en un solo punto, en múltiples sitios y a escala del genoma. Sin embargo, la aplicación del sistema ha escalado a nivel de diversas industrias, como por ejemplo en la agricultura, donde se ha aplicado la edición de genes mediante el sistema CRISPR-Cas de la clase I y la clase II a diversas plantas y se han obtenido mutantes genéticos estables. Por ejemplo, se ha utilizado para generar mutaciones específicas en el alelo *TaMLO-A1* del locus de resistencia al moho en plantas transgénicas (S. Zhang et al., 2019), para generar líneas de trigo bajas en gluten que muestran una reducción de  $\alpha$ -gliadinas y para generar mutaciones en los genes implicados en la arquitectura de la planta que son específicos para cada especie. Este último es un rasgo muy importante para el desarrollo de cultivos de alto rendimiento, por lo que se ha buscado producir mutaciones en estos genes para obtener mejores resultados en los cultivos como el arroz (*Oryza sativa*), trigo (*Triticum aestivum*), maíz (*Zea mays*), cebada (*Hordeum vulgare*), algodón (*Gossypium hirsutum*), tomate (*Solanum lycopersicum*) y soja (*Glycine max*) (Bao et al., 2019). A nivel general se han realizado diversas pruebas para el desarrollo de plantas resistentes a enfermedades, plagas y condiciones climáticas adversas, además de perseguir el aumento de la producción y el valor nutricional de los cultivos (Shan et al., 2013).

En el ámbito de la medicina, una de las aplicaciones más resaltantes que se ha desarrollado en los últimos años es la terapia génica. En 2023, la Administración de Alimentos y Medicamentos de los Estados Unidos (FDA, por sus siglas en inglés) aprobó las primeras terapias génicas Casgevy y Lyfgenia para tratar a pacientes con anemia de células falciformes (SCD) y otro trastorno genético que afecta la hemoglobina en los glóbulos rojos, la beta talasemia dependiente de transfusiones (TDT). La fase tres del ensayo arrojó datos muy exitosos donde 25 de 27 personas con TDT ya no dependían de transfusiones después del tratamiento, algunos durante más de tres años y los otros dos pacientes tuvieron reducciones significativas en la frecuencia de transfusiones (80%, 96%). Por otro lado, 16 de 17 pacientes con SCD no presentaron las crisis vasooclusivas que caracterizan la enfermedad después del tratamiento. El otro paciente se ha mantenido libre de hospitalizaciones relacionadas con crisis vasooclusivas (Liu et al., 2021; Henderson, 2024). Actualmente se están realizando diferentes ensayos clínicos usando diversos tipos de sistemas CRISPR-Cas para el desarrollo de tratamientos para enfermedades como infecciones del tracto

urinario, amiloidosis hereditaria por transtiretina, angioedema hereditario y cáncer (Henderson, 2024).

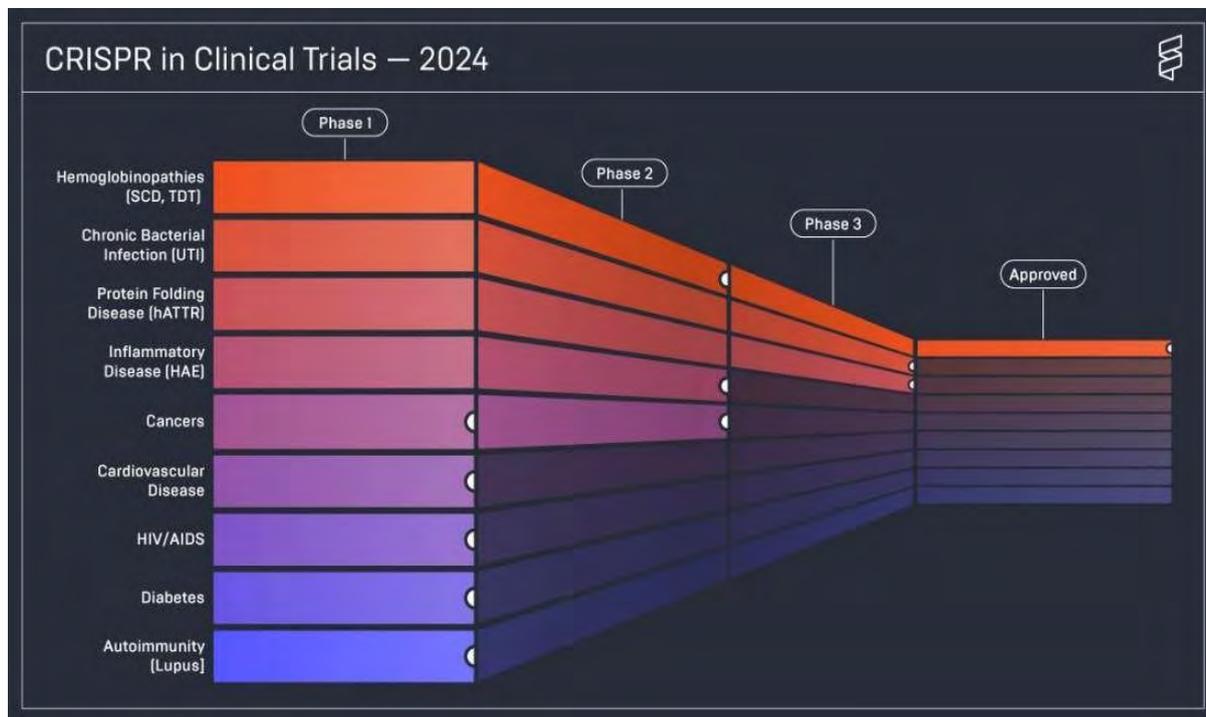


Figura 5.- Terapias basadas en el sistema CRISPR y estado de las fases de pruebas clínicas para diferentes enfermedades. Figura tomada de Henderson (2024).

El éxito de estos tratamientos representa un avance muy importante en el uso de la terapia génica mediante el sistema CRISPR-Cas para la cura de enfermedades y es una oportunidad para seguir investigando nuevos sistemas que permitan el desarrollo de nuevos y mejores tratamientos para diferentes enfermedades (Figura 5). También se ha utilizado este sistema como herramienta para el diagnóstico. Un ejemplo es su aplicación terapéutica en COVID-19, mediante el uso de PAC-MAN (CRISPR antiviral profiláctico en células humanas) basado en CRISPR-Cas13, que utiliza la actividad de la endonucleasa de ARN guiada por ARN de Cas13d en células humanas para eliminar el virus SARS-CoV-2. Este estudio indicó que el sistema CRISPR-Cas13d se puede utilizar para apuntar y escindir eficazmente las secuencias de ARN de fragmentos de SARS-CoV-2 con ARNs guía adecuadamente diseñados en células epiteliales de pulmón (Liu et al., 2021). Este estudio también representa una ventana al desarrollo de nuevos sistemas diagnósticos rápidos y eficientes mediante el sistema CRISPR-Cas. Por otro lado, un enfoque ampliamente explorado es el uso del sistema CRISPR-Cas como herramienta antimicrobiana y de detección para la eliminación de patógenos (Bhattacharjee et al., 2022). También se ha usado para la eliminación específica de algunos plásmidos portadores del gen de resistencia antimicrobiana (AMR), llevando a que las bacterias sean susceptibles a los antibióticos, y aportando así grandes avances en el área de la salud (Bhattacharjee et al., 2022).

La alta diversidad de tipos de sistemas CRISPR-Cas ha conducido a un rápido desarrollo de aplicaciones de edición del genoma, con capacidades como eliminación de genes, edición de bases, edición primaria, inserción de genes, modulación epigenética, modulación transcripcional y edición de ARN. Se ha convertido en una herramienta muy valiosa para su aplicación en una diversidad de áreas como la biología molecular y la biotecnología (Koonin et al., 2023).

Se estima que el 50% de las bacterias poseen el sistema CRISPR-Cas, lo que plantea un desafío para su identificación en distintas cepas. Actualmente, el enfoque predominante para esta identificación es la secuenciación de genoma completo. Aunque es muy efectiva, esta técnica puede resultar costosa, especialmente cuando se desea analizar una amplia variedad de bacterias. Por lo tanto, el desarrollo de un sistema de detección rápido y de bajo costo constituye una herramienta valiosa para detectar la presencia del sistema CRISPR-Cas en diversas bacterias. Esto permitiría seleccionar las muestras de mayor interés para una verificación posterior mediante secuenciación de genoma completo. En consecuencia, diseñar un sistema para la detección de genes diagnósticos de la matriz CRISPR-Cas de la clase I, subtipo I-E, y de la clase II en los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y *Sphingomonas* podría proporcionar una herramienta eficaz para identificar este sistema en representantes pertenecientes a estos géneros bacterianos caracterizados por colonizar ambientes diversos y extremos, como los diferentes sustratos asociados al glaciar del Pico Humboldt, que al tener características únicas como una tolerancia elevada a condiciones extremas podrían contener sistemas únicos con potenciales para el uso en la biotecnología y en particular para el mejoramiento de los cultivos a bajas temperaturas.

## 4. Hipótesis

Mediante la aproximación *in silico* es posible la identificación y clasificación del sistema CRISPR-Cas en géneros de bacterias reportadas en sustratos del frente glacial del Pico Humboldt.

## 5. Objetivos

### 5.1. General

Diseñar un método para la detección *in silico* del sistema CRISPR-Cas en géneros bacterianos reportados en sustratos del frente glacial del Pico Humboldt.

### 5.2. Específicos

1. Analizar la matriz CRISPR-Cas en diferentes especies bacterianas pertenecientes a los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y *Sphingomonas*.

2. Seleccionar las regiones de interés en la matriz CRISPR-Cas, potenciales para el diseño de un sistema de detección basado en PCR en los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus*, y *Sphingomonas*.
3. Definir y probar *in silico* oligonucleótidos que permitan estudiar la arquitectura genética de la matriz CRISPR-Cas en los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus*, y *Sphingomonas*.

## 6. Metodología

### 6.1. Recolección de datos y análisis de la matriz CRISPR-Cas

Se obtuvieron secuencias disponibles, tanto de nucleótidos como de aminoácidos, correspondientes a los genes *cas* en los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y *Sphingomonas*, a partir de bases de datos de acceso libre, como NCBI (NCBI Resource Coordinators, 2018) y Uniprot (The UniProt Consortium, 2023) y bases de datos específicas del sistema CRISPR-Cas como CRISPRfinder (Couvin et al., 2018) y CRISPRone (Zhang & Ye, 2017). Para identificar los loci CRISPR-Cas se realizaron búsquedas mediante la herramienta Blast (Altschul et al., 1990) para cada uno de los géneros, en distintas bases de datos que contengan información de genomas de bacterias, seleccionando aquellos resultados que arrojaran mejores valores de identidad y cobertura, prestando atención a aquellos resultados con el valor de E cercano a 0. Se clasificaron los resultados para cada género, por tipo de sistema, procedencia y de estar disponible, por estilo de vida. Así mismo, se identificó la dirección y el orden de los genes de la matriz. Todas las secuencias fueron procesadas para filtrar errores, como codones de paradas, eliminación de bases erradas, información repetida y posibles fallos presentes que pudieran afectar los resultados. Una vez procesadas las secuencias, se procedió a verificar que estuvieran en el marco abierto de lectura correcto mediante la herramienta MEGA (Tamura et al., 2021). Para las secuencias donde no estaban anotados los genes *cas* en el genoma, se procedió a reconstruir el gen mediante las herramientas Genewise (Birney et al., 2004) y ORFfinder, que permiten buscar marcos de lectura abiertos (ORF) en la secuencia de ADN y predicen el rango de cada ORF, junto con su traducción a proteína; además, verifica la proteína predicha utilizando SMART BLAST o el BLASTP normal.

Luego de este procesamiento, las secuencias nucleotídicas obtenidas fueron alineadas mediante el programa MAFFT (Katoh et al., 2002). Por otro lado, se utilizó la herramienta COBALT de NCBI (Papadopoulos & Agarwala, 2007) para alinear las secuencias aminoacídicas.

## 6.2. Evaluación de los sitios de interés

Con el fin de determinar los sitios potencialmente útiles para el diseño de los oligonucleótidos, en primer lugar, se identificaron las regiones conservadas en la proteína mediante el alineamiento de la secuencia aminoacídica en la herramienta COBALT, que utiliza la base de datos y dominios conocidos para realizar el alineamiento, es decir que efectúa el alineamiento en base a los dominios conservados y no a la secuencia de aminoácidos. Las regiones señaladas con alto nivel de similitud, fueron verificadas al buscar la proteína Cas en *Conserved Domains and Protein Classification* (CDD) (Wang et al., 2023) en NCBI, realizando un alineamiento de la proteína de diferentes especies de bacterias e identifica motivos y dominios. Una vez que se hubiera verificado que la región señalada en el alineamiento de COBALT fuera un dominio conservado, se buscó la región correspondiente en la secuencia nucleotídica para finalmente en los casos en los que tenía poca variabilidad nucleotídica determinar si representan sitios óptimos para el diseño de oligonucleótidos.

Simultáneamente se llevaron a cabo otros dos enfoques para la selección de sitios de interés para cada gen que conforma la matriz CRISPR-Cas. Para ello se estimó la diversidad nucleotídica ( $\pi$ ), que se define como el número de diferencias por sitio nucleotídico entre dos secuencias elegidas al azar de una población (Nei & Li, 1979). Esta medida es la proporción de nucleótidos diferentes entre dos secuencias aleatorias. Cuando se dispone de más de dos secuencias, la diversidad media de nucleótidos se calcula como la media aritmética de todas las comparaciones por pares, que es el número promedio de diferencias de nucleótidos por pares entre secuencias en una muestra. El valor de  $\pi$  depende tanto del número de sitios polimórficos como de su frecuencia. Este análisis se realizó haciendo uso del programa DnaSP (Rozas & Rozas, 1995) que permite hacer análisis de polimorfismos de ADN utilizando datos de un único locus o de varios loci (como los formatos generados por algunos programas ensambladores de data RAD-seq). Esta herramienta permite estimar varias medidas de variación de las secuencias de ADN intra e inter poblaciones (en este caso entre especies) en sitios no codificantes, sinónimos o no sinónimos, o en diferentes posiciones de codones. Los valores de  $\pi$  utilizados como referencia abarcaron valores de alrededor de 0.001 como diversidad nucleotídica baja, mientras que valores de 0.005 - 0.01 se consideraron como diversidad nucleotídica alta, que son valores obtenidos en estudios de diversidad nucleotídica en especies bacterianas (Ochman & Wilson, 1987 y Jaureguy et al., 2008).

Del mismo modo, se realizaron pruebas de neutralidad estimando dos estadísticos, el test de Tajima ( $D$ ) (Tajima, 1989), que compara el número de sitios segregantes por sitio, contra la diversidad de nucleótidos. Un sitio se considera segregante sí, en una comparación de  $n$  secuencias, hay dos o más nucleótidos en ese sitio. La diversidad de nucleótidos se define como el número promedio de diferencias de nucleótidos por sitio entre dos secuencias. El test de Tajima mide la diferencia entre el número de sitios de segregación y el número promedio de diferencias por pares en un conjunto de secuencias de ADN. Un valor positivo de Tajima  $D$  sugiere un exceso de variantes de frecuencia intermedia, lo que puede indicar una selección equilibrada o cuellos de botella en la población. Un valor negativo de Tajima  $D$  sugiere un exceso de variantes de baja frecuencia, lo que puede indicar una selección positiva o una expansión de la población. Un valor

de Tajima  $D$  cercano a cero sugiere un modelo evolutivo neutral. La interpretación de los valores del  $D$  de Tajima depende del conjunto de datos y del contexto de los datos de estudio. La segunda prueba estadística aplicada fue la estimación de los sitios bajo selección negativa *Alfa*, que incluye el test de MEME (Modelo de evolución de efectos mixtos) (Murrell et al., 2012), disponible en el servidor Datamonkey (Weaver et al., 2018). Esta permite estimar una tasa de sitio sinónimo *Alfa* ( $\alpha$ ) y una mezcla de dos categorías de tasas de recambio no sinónimo (beta ( $\beta$ ), con proporción  $\rho^-$ , y  $\beta^+$  con proporción  $1-\rho^-$ ), y utiliza una prueba de razón de verosimilitud para determinar si hay mayor cantidad de mutaciones no sinónimas ( $\beta^+$ ) que mutaciones sinónimas ( $\alpha$ ) sobre un sitio. En este estudio la tasa de mutaciones sinónimas *Alfa* fue la utilizada, ya que se requiere saber cuáles sitios están bajo selección negativa. En la Figura A se esquematiza el proceso implementado para la selección de los sitios de interés para el diseño de oligonucleótidos.

### 6.3. Diseño y prueba de oligonucleótidos

Se utilizaron los principales algoritmos para el diseño y prueba de oligonucleótidos, como Primer3 (Untergasser et al., 2012), oligoexplorer (Untergasser et al., 2012) y Primer-BLAST (Ye et al., 2012) usando como blanco las regiones definidas con las pruebas de neutralidad y diversidad nucleotídica. Se seleccionaron aquellos oligos que cumplieron con los parámetros adecuados de longitud, temperatura de fusión, porcentaje de guaninas y citosinas, formación de dímeros de oligos y bucles. Los oligos escogidos fueron puestos a prueba haciendo uso de la herramienta Primer-BLAST (Ye et al., 2012), que utiliza Primer3, BLAST y el algoritmo de alineación global para comparar los cebadores con diferentes bases de datos, con el fin de evitar pares de cebadores (todas las combinaciones, incluido el par de cebadores *forward-reverse*, así como pares inverso-reverso) que puedan causar amplificaciones inespecíficas. Además, gracias a estas pruebas *in silico* se evaluó la especificidad de los oligos diseñados en diferentes bases de datos.

### 6.4. Modelado de proteínas

Debido a que la proteína Cas7 dentro de los géneros estudiados estaba muy conservada, y que la región complementaria a los oligos era compartida en todos los grupos estudiados, se escogió a la especie *Pseudomonas aeruginosa* para realizar el modelado en 3D, dado que esta especie tenía los datos disponibles mejor representados y curados.

Para el modelado de la proteína Cas7 inicialmente se identificaron las estructuras 3D de la proteína disponibles en diferentes bacterias. El estudio de Jackson et al. (2014) muestra la estructura cristalina del complejo efector *Cascade* y especifica la conformación de la proteína Cas7 en el complejo, destacando que es un hexámero en forma de un caballito de mar y que corresponde al esqueleto principal de este, donde interactúa con el ARNcr y el ADN foráneo. Haciendo uso del servidor AlphaFold 3 (Abramson et al., 2024), se realizó el modelado del monómero de la proteína Cas7 de la especie *Pseudomonas aeruginosa*, con la finalidad de comparar este modelo con lo propuesto en el trabajo de Jackson et al. (2014). Una vez corroborada la estructura correcta del monómero se procedió al modelado, nuevamente en AlphaFold 3, del hexámero (que forma la estructura de

caballito de mar) junto a una secuencia de ADN y ARN del bacteriófago *Lambdavirus lambda LC730321* con la finalidad de visualizar las interacciones entre ellos.

Para verificar la calidad del modelado se utilizó la plataforma de *SWISS-MODEL* (Waterhouse et al., 2018), haciendo uso de la herramienta *structure assesment* (análisis de estructura), donde este servicio integra diferentes aproximaciones estadísticas que analizan el modelado para explorar su calidad, destacando el gráfico de Ramachandran y el QMEANDisCo (una estimación estadística de la calidad del modelado).

Finalmente, se identificaron las regiones en la proteína donde fueron diseñados los oligonucleótidos para corroborar que su ubicación correspondía a regiones de importancia dentro de la proteína. Tanto la estructura cristalina de la proteína Cas7, como el modelo obtenido, fueron visualizados en el programa Chimera (Pettersen et al., 2004).

En el caso de Cas9 se realizó una búsqueda de estructuras cristalizadas, crio-microscopía electrónica y modelados 3D de la proteína junto al ADN y ARN de un bacteriófago en otras bacterias (Berman et al., 2000), todo esto para identificar y así poder confirmar que las regiones donde fueron diseñados los oligos tienen un rol importante para el funcionamiento de la proteína.

## 7. Resultados

### 7.1. Recolección de datos y análisis de la matriz CRISPR-Cas

La mayoría de las secuencias de los genes y proteínas que conforman la matriz CRISPR para los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y *Sphingomonas* se obtuvieron de las bases de datos NCBI y Uniprot, respectivamente. Cabe destacar que las bases de datos especializadas para el sistema CRISPR (como CRISPRCasFinder y CRISPRone) carecían de información completa para los géneros estudiados. Además, en aquellos casos donde estaban reportados algunos de los genes *cas* para estas especies, al corroborar el marco abierto de lectura y hacer su correspondiente verificación por Blast, los resultados obtenidos mostraban poca identidad con respecto a las secuencias reportadas en otras bases de datos. Por lo tanto, los genes que mostraron este comportamiento fueron excluidos del análisis.

No todas las especies tenían anotados correctamente los genes de la matriz CRISPR-Cas en el genoma, por lo que se procedió a hacer uso de las herramientas bioinformáticas como Blastn y Blastp, seleccionando los resultados de mayor porcentaje de cobertura y de identidad cercana al 100%, y un valor de E cercano a 0. Estas secuencias fueron analizadas mediante la herramienta ORFfinder para determinar el marco abierto de lectura, lo que nos permitió reconstruir la secuencia codificante (CDS). De esta manera se obtuvieron 16 secuencias de especies diferentes del género *Arthrobacter*, 5 secuencias y 4 especies para el género *Cryobacterium*, 5 secuencias de especies diferentes del género *Rhodococcus*, 50 secuencias y 29 especies para el género *Pseudomonas* y 28 secuencias de especies diferentes para el género *Sphingomonas*.

Una vez ordenados los genes y reconstruida la arquitectura genética para cada uno de los géneros, se encontró que en *Arthrobacter*, *Cryobacterium*, *Pseudomonas* y *Rhodococcus* el sistema prevalente era el CRISPR-Cas de la clase I y el tipo I-E, y para el género *Sphingomonas* se encontró que la mayoría de las especies poseían el sistema de la clase II, como se señala en la Figura 6.

No obstante, en el interior de cada género se observó una diversidad notoria en la arquitectura genética de la matriz CRISPR-Cas. Esa variabilidad fue importante en el género *Arthrobacter*, quizás debido a la baja calidad de los datos obtenidos. Además, en algunas especies (e.g. *Arthrobacter oryzae*, *Arthrobacter sp. isolate S7A\_St\_54*, entre otras) se apreciaron variaciones en el orden de los genes, siguiendo la secuencia *cas3*, *cas6*, *cas5*, *cas7*, *cas1* y *cas8* (resultados no mostrados). Cabe destacar que diversas especies pertenecientes a los géneros *Arthrobacter* y *Cryobacterium* carecían de los genes *cas1* y *cas2* (e.g. *Arthrobacter sp. ES1 BJQ89\_9*, *Cryobacterium sinapicolor strain TMT1-23-1*, entre otras), aunque en algunas especies sólo el gen *cas2* estaba ausente (e.g. *Arthrobacter sp. HMSC06H05* y *Cryobacterium sp. TMS1-13-1*). También se encontró que la orientación de la matriz podía ser variable en algunas especies pertenecientes a los géneros estudiados.

De manera general, el orden de los genes de la matriz CRISPR-Cas en especies de los géneros *Cryobacterium*, *Pseudomonas* y *Rhodococcus* coincidía con el reportado previamente para el sistema tipo I-E (Makarova & Koonin, 2015). Sin embargo, en varias especies del género *Pseudomonas* procedentes de distintas regiones geográficas (e.g. *Pseudomonas dryadis* y *Pseudomonas indoloxydans*), se observaron varios genes que no correspondían a genes *cas* dentro de la matriz. Por ejemplo, entre el gen *cas3* y *cas8* se observaron los genes que codifican para las *type II toxin-antitoxin system RelE/ParE family toxin*, *type II toxin-antitoxin system Phd/YefM family antitoxin* y *type II toxin-antitoxin system VapC family toxin*. Además, entre los genes *cas6* y *cas1* se encontró un gen que codifica para la *KilA-N domain-containing protein* (Figura 7).

Con respecto al género *Sphingomonas*, todas las cepas encontradas presentan la clase II y el tipo II-C1 del sistema CRISPR-Cas (Figura 8), aunque también se observó que la orientación de los genes variaba en algunas especies. Este hallazgo nos permitió explorar arquitecturas genéticas de la matriz CRISPR-Cas fuertemente contrastantes (i.e. clase I versus clase II) entre los distintos géneros estudiados.

Los alineamientos de cada uno de los genes que conforman la matriz CRISPR-Cas en los géneros estudiados, nos permitieron en un primer momento filtrar los datos, eliminando aquellas secuencias que tuvieran inserciones o *gaps* excesivos que pudieran afectar el análisis. En el caso del género *Arthrobacter*, se evidenciaron grandes diferencias entre las secuencias disponibles en las bases de datos, lo que nos llevó a eliminar de los alineamientos varias secuencias por la baja calidad de las mismas. Esto trajo como consecuencia que los alineamientos se dividieran en dos grupos. El grupo 1 contiene a las especies *Arthrobacter oryzae strain TNB02*, *Arthrobacter sp. ES1 BJQ89\_9*, *Arthrobacter sp. isolate S7A\_St\_54*, *Arthrobacter sp. isolate ERR9968748*, *Arthrobacter sp. efr-133-R2A-120*, *Arthrobacter globiformis* y *Arthrobacter sp. ISL-28 ISL28\_10*, que provienen de diferentes ambientes y tienen diversos estilos de vida, siendo en su mayoría de

vida libre. A pesar de la alta variabilidad entre las secuencias de este grupo, se observaron regiones conservadas en diversos genes de la matriz. El grupo 2 está compuesto por las especies *Arthrobacter cumminsii*, *Arthrobacter albus*, *Arthrobacter sp. HMSC08H08*, *Arthrobacter sp. HMSC06H05* y *Arthrobacter sp. isolate CPoEBzCyWp\_bin.3*, que se encuentran principalmente relacionadas al humano y presentan alta similitud en la secuencia de los distintos genes de la matriz CRISPR-Cas.

Por otro lado, la arquitectura genómica observada en todas las especies del género *Sphingomonas* fue la reportada para la clase II y el tipo II-C1 del sistema CRISPR-Cas, también se observó que la orientación de los genes variaba en algunas especies. Las secuencias de los genes *cas1* y *cas9* fueron muy variables. Por otro lado, el análisis del gen *cas2* fue descartado debido a que tiene un tamaño pequeño.

Una vez depurada y organizada la data genética, se realizaron nuevos alineamientos para proceder a identificar los sitios de interés. Con base en esto, se realizó un análisis de polimorfismo para todos los genes, graficando el valor de *pi* (medida de diversidad nucleotídica) a lo largo del gen. Además, se evaluaron los genes bajo el modelo de MEME, que permite graficar los eventos de selección negativa a lo largo de la proteína. Haciendo uso de estas herramientas se obtuvieron las regiones más resaltantes que podrían ser blancos para el diseño de oligonucleótidos.

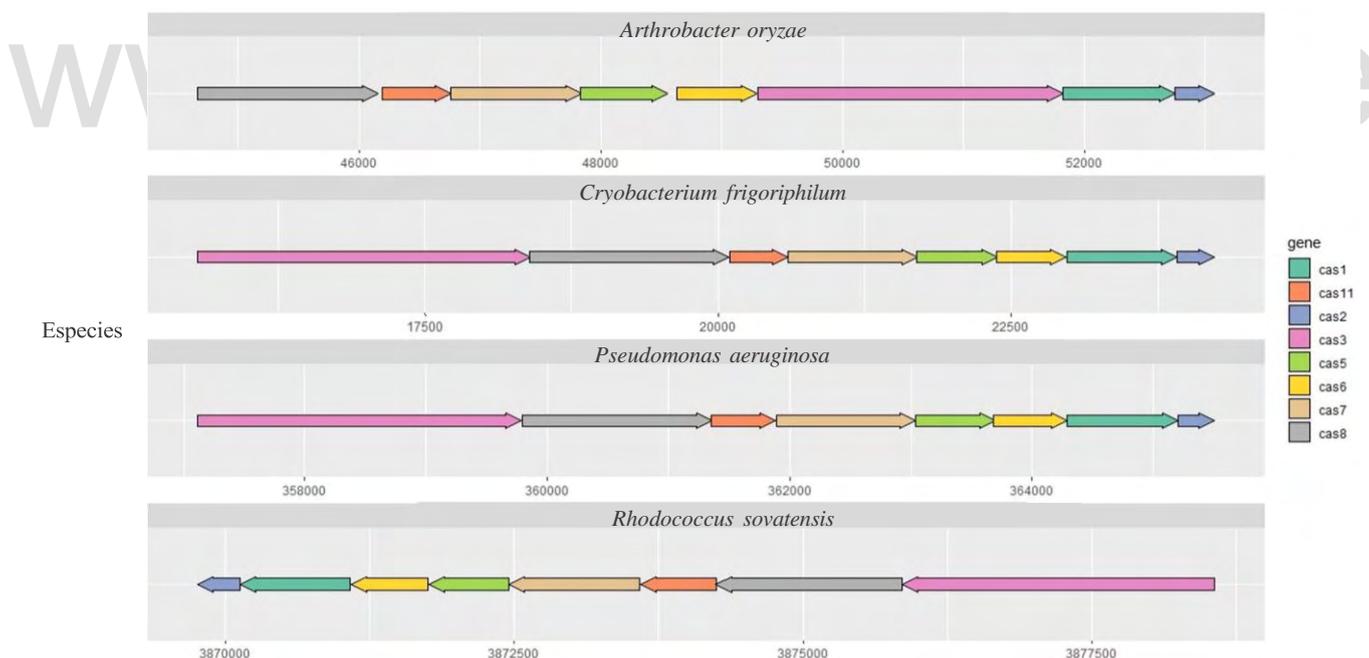


Figura 6.- Arquitectura genética de la matriz CRISPR-Cas de la clase I y el subtipo tipo I-E en especies representativas de los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas* y *Rhodococcus*.

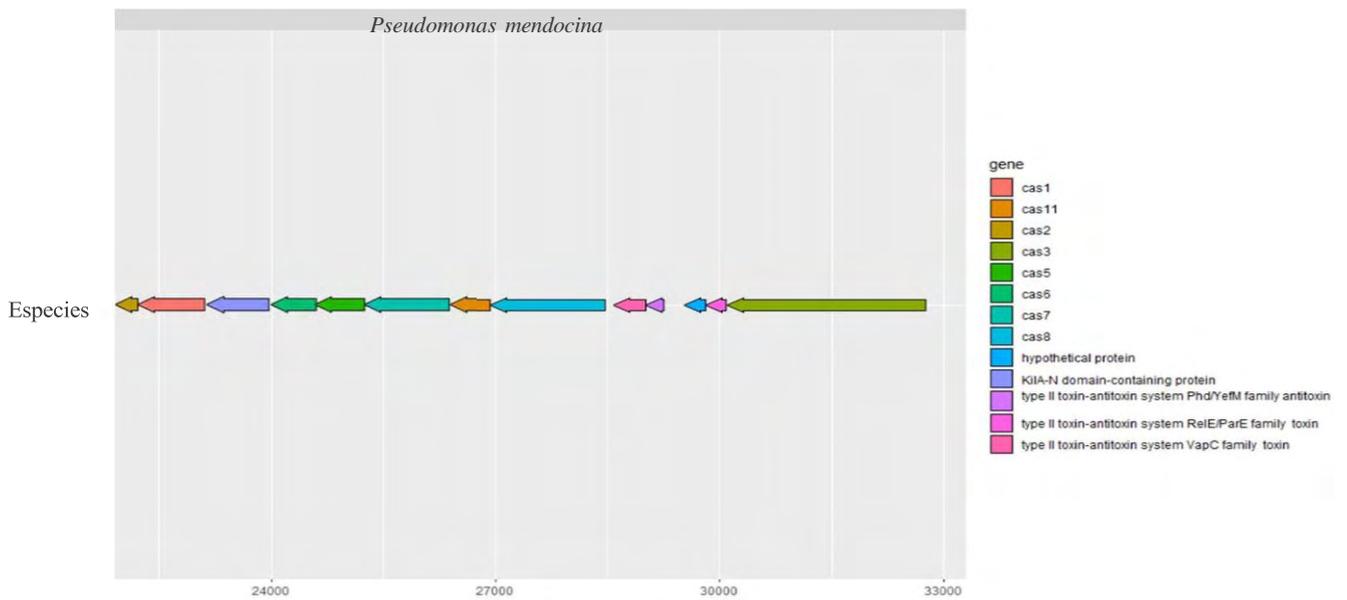


Figura 7.- Arquitectura genética de la matriz CRISPR-Cas de la clase I y el subtipo I-E en la especie *Pseudomonas mendocina*.

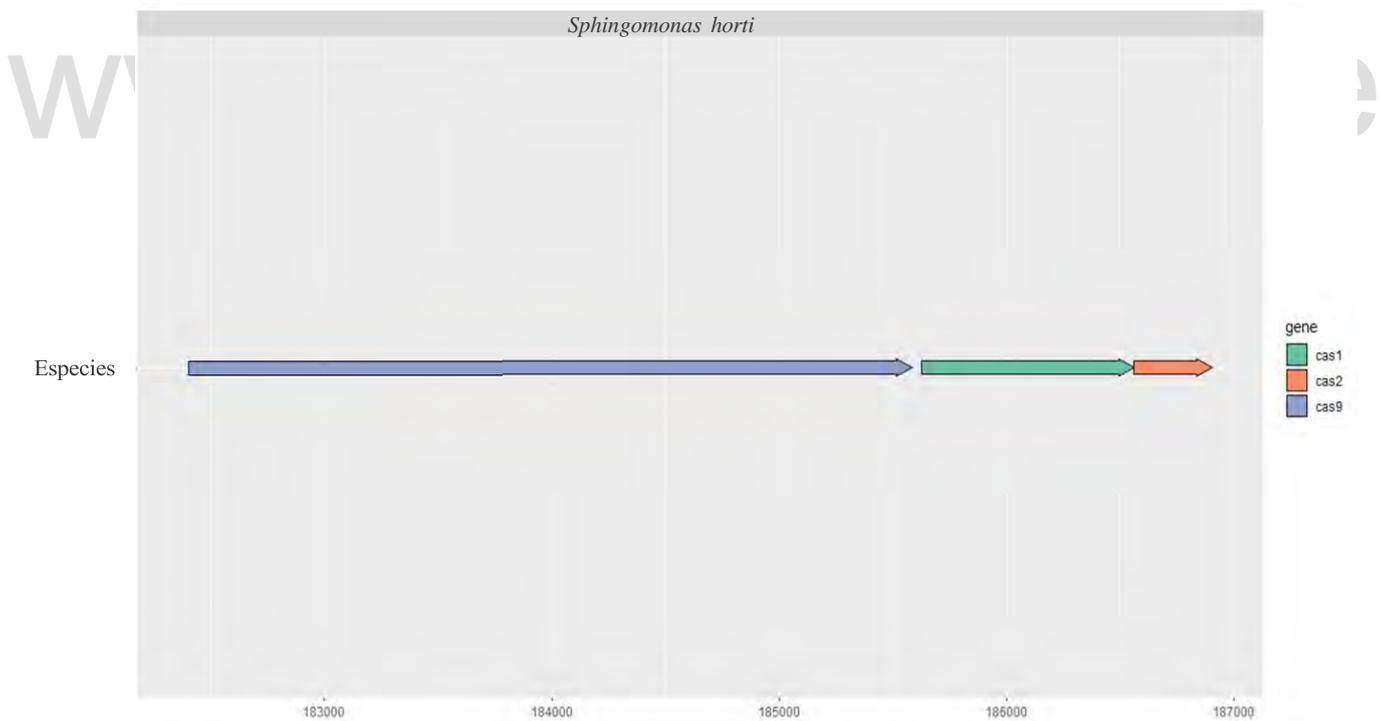


Figura 8.- Arquitectura genética de la matriz CRISPR-Cas de la clase II y subtipo II-C1 en la especie *Sphingomonas horti*.

## 7.2 Alineamiento y Evaluación de los sitios de interés

### **Sistema CRISPR-Cas clase I, subtipo I-E**

#### 7.2.1 *Genes cas1 y cas2*

Los genes *cas1* y *cas2* codifican para proteínas que cumplen funciones cruciales en la integración del material genético externo en la matriz CRISPR. Sin embargo, estos genes no fueron considerados para el análisis y el diseño de oligonucleótidos debido a que en varias especies pertenecientes a los géneros *Artrhobacter* y *Cryobacterium* estaban ausentes. Solamente se tomó en consideración el análisis del gen *cas1* en el género *Rhodococcus*, donde se determinaron dos regiones conservadas a pesar de la alta variabilidad de las secuencias (Figura 9a). La selección de los sitios de interés se basó en los gráficos de diversidad nucleotídica ( $\pi$ ) y de *Alfa* (Figura 9c), que determinaron las regiones conservadas a nivel de proteína, que a su vez coincidían con las regiones que habían sido señaladas como poco variables en el alineamiento de COBALT (Figura 9d). Se seleccionaron dos regiones, una ubicada entre los aminoácidos 56 y 71 (Figura 9b) y otra entre los aminoácidos 140 y 158, que fueron candidatas para el diseño de oligonucleótidos.

www.bdigital.ula.ve

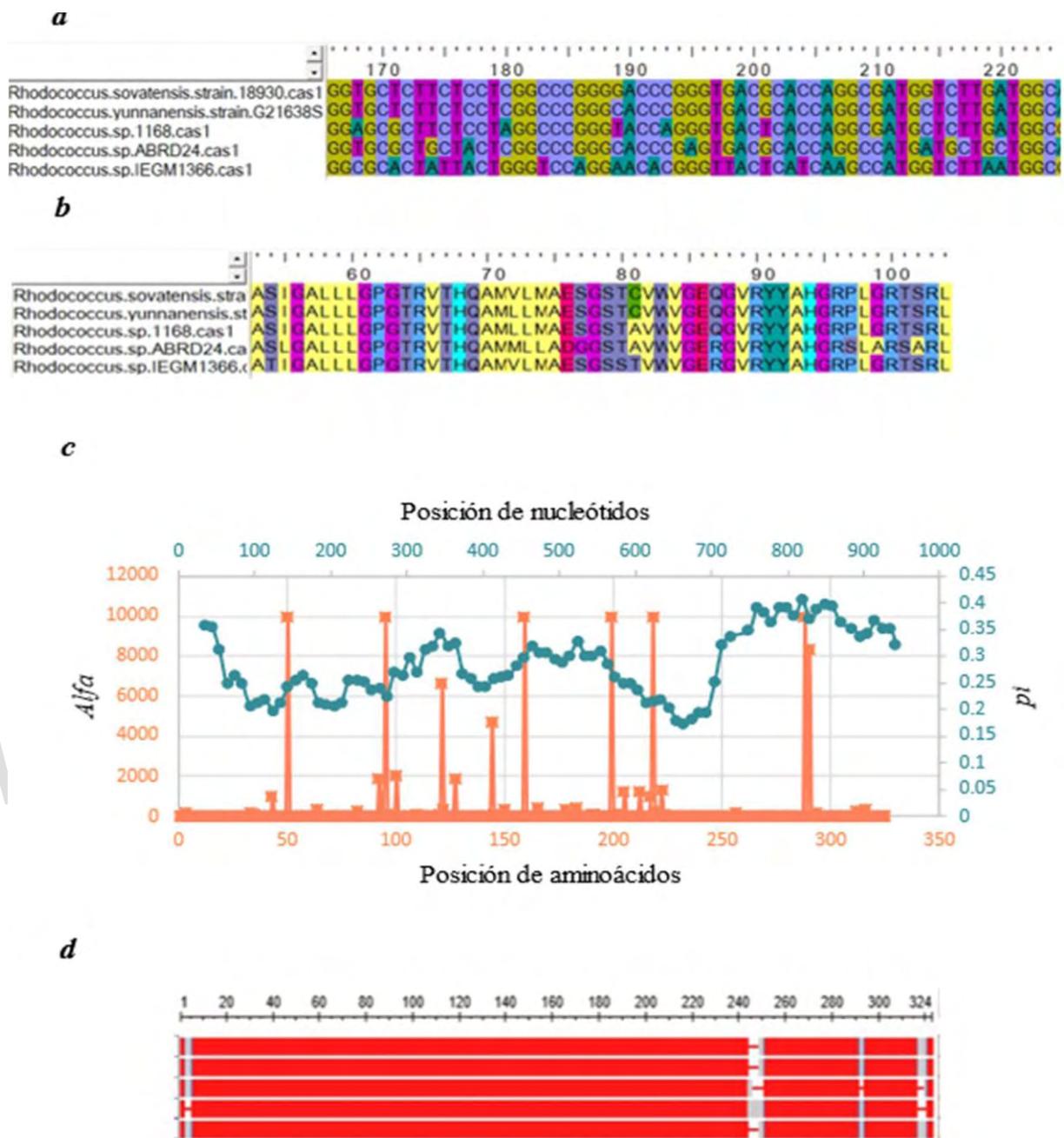


Figura 9.- Análisis de las secuencias del gen *cas1* en diferentes especies del género *Rhodococcus*. **a.** Alineamiento del gen desde la posición 166 a la 220. **b.** Alineamiento de la proteína *Cas1* desde la posición 53 a la 100. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas1* en diferentes especies del género *Rhodococcus*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas1* resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

### 7.2.2 Gen cas3

Al analizar los alineamientos del gen *cas3* para los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas* y *Rhodococcus*, se observó que tanto las secuencias nucleotídicas como las aminoacídicas tenían alta similitud (Figura 10a), lo que además fue evidenciado en los análisis de polimorfismo y de Alfa que mostraron diversas regiones conservadas (Figura 10b). Sin embargo, al evaluar estas regiones por separado y al realizar un Blastn, los resultados mostraron que efectivamente se identificaba al gen *cas3*, pero también se detectaban otros genes que codifican para proteínas con funciones similares como *DEAD/DEAH box helicase*, *MFS transporter*, *ATP-dependent helicase*, *PAAR domain-containing protein* y *dipeptide ABC transporter ATP-binding protein*. Por este motivo, el gen *cas3* no fue considerado para el diseño de oligonucleótidos.

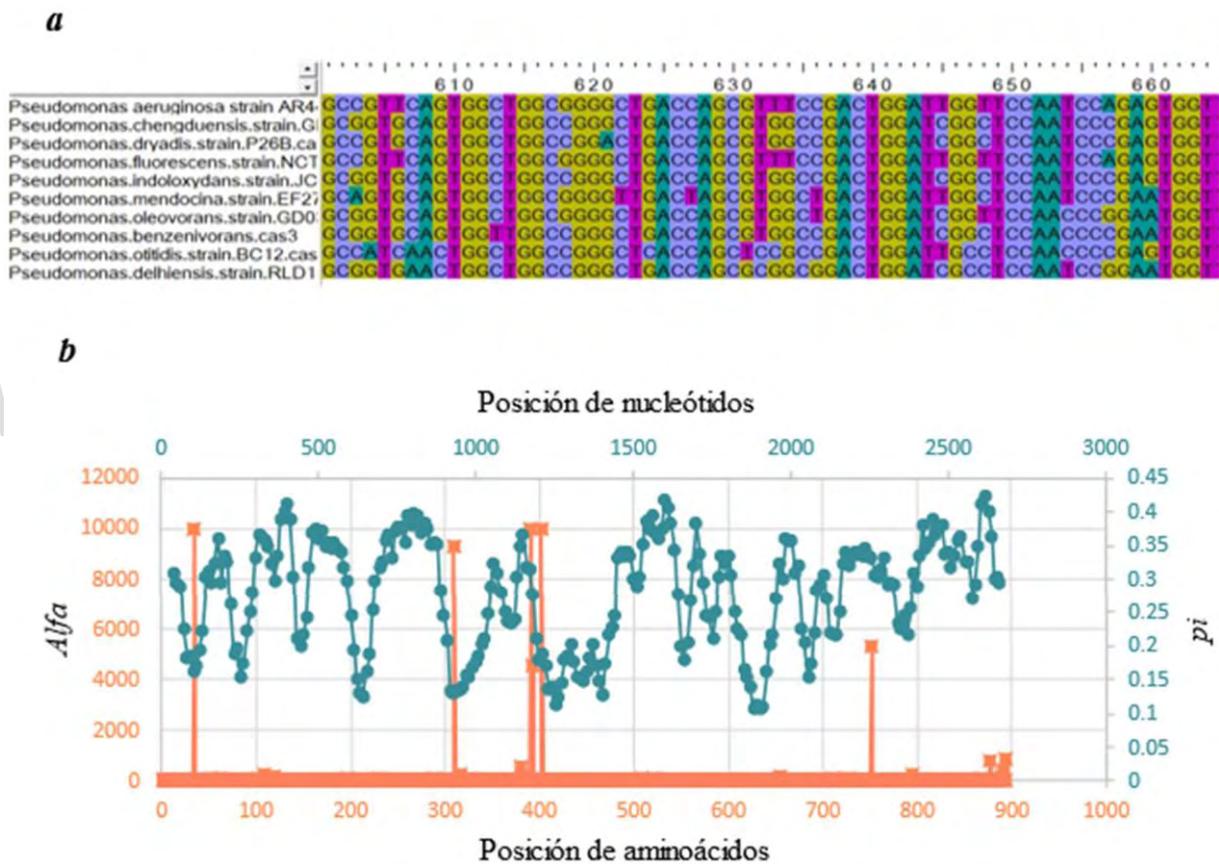
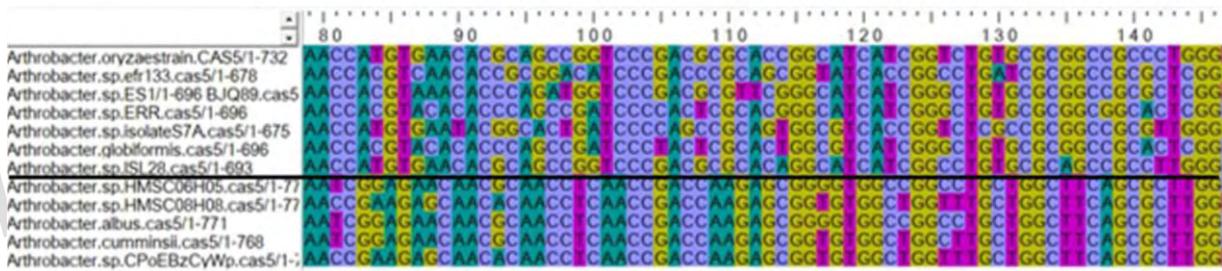


Figura 10.- Análisis de las secuencias del gen *cas3* en diferentes especies del género *Pseudomonas*. **a.** Alineamiento del gen *cas3*, desde la posición 601 a la 660. **a.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas3* en diferentes especies del género *Pseudomonas*, donde *pi* está representado por ● y Alfa está representado por ■.

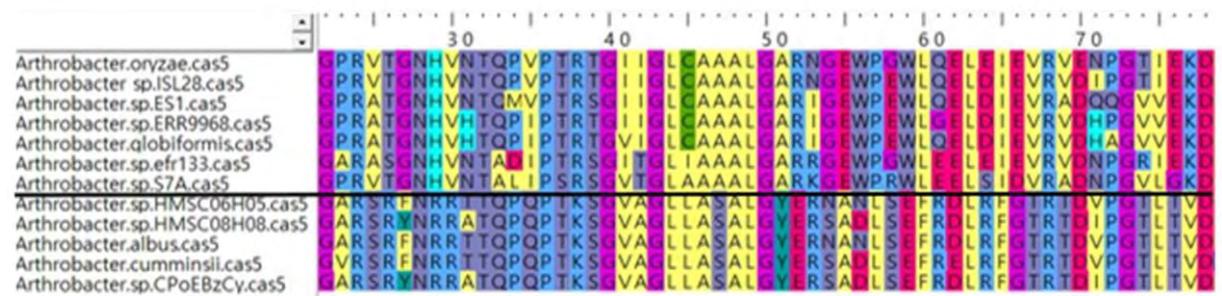
### 7.2.3 Gen cas5

El gen *cas5* es muy variable entre los géneros. En el caso del género *Arthrobacter* en particular, se observó una mayor variabilidad en las secuencias nucleotídicas/aminoacídicas en el grupo 1 que en el grupo 2. Se identificó una región con poca diversidad nucleotídica entre ambos grupos, comprendida entre las bases 79 y 145 (Figura 11a), correspondiente a la secuencia de aminoácidos en la Figura 11b. El gráfico de polimorfismo arrojó mucha variabilidad en el grupo 1, mientras que el gráfico de *Alfa* mostró cuatro sitios de selección negativa correspondientes a los aminoácidos 8, 71, 155 y 207 (líneas color naranja en la Figura 11c.1). Estos sitios y los indicados en el alineamiento de COBALT (Figura 11d.1) fueron seleccionados para el posible diseño de oligonucleótidos en el grupo 1. Por otro lado, el gráfico de polimorfismo para el grupo 2 indicó una región de poca variabilidad en la parte inicial del gen, y el gráfico de *Alfa* solo señaló un sitio de selección negativa correspondiente al aminoácido 2 (Figuras 11c.2). En este caso se tomaron en cuenta las regiones observadas en el gráfico de polimorfismo y el alineamiento de COBALT para el diseño de oligonucleótidos.

**a**



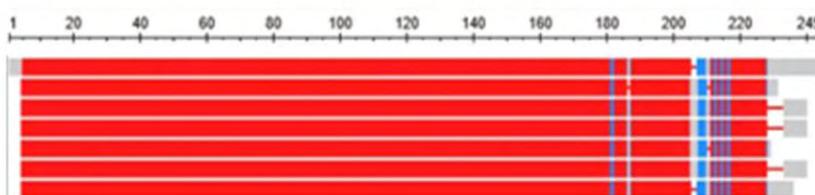
**b**



c.1



d.1



c.2



d.2

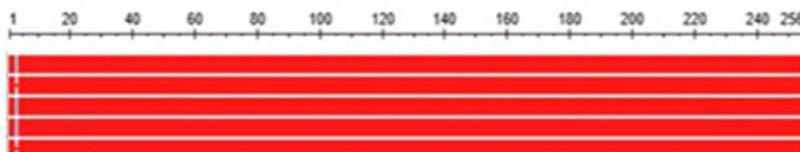
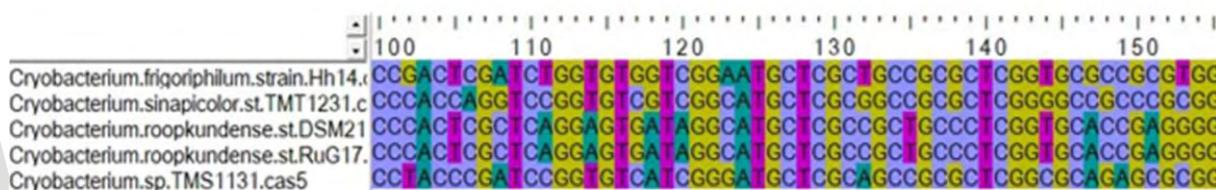


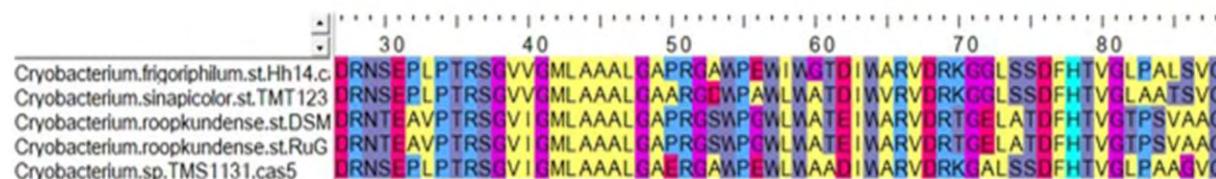
Figura 11.- Análisis de las secuencias del gen *cas5* en diferentes especies del género *Arthrobacter*. **a.** Alineamiento del gen *cas5*, desde la posición 79 a la 145. Las primeras siete especies corresponden al grupo 1, mientras las últimas cinco especies corresponden al grupo 2. **b.** Alineamiento de la proteína *Cas5*, desde la posición 21 a la 73. Las primeras siete especies corresponden al grupo 1, mientras las últimas cinco especies corresponden al grupo 2. **c.1** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas5* en el grupo 1 de *Arthrobacter*, donde pi está representado por ● y Alfa está representado por ■. **c.2** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas5* en el grupo 2 de *Arthrobacter*, donde pi está representado por ● y Alfa está representado por ■. **d.1** Alineamiento general de la proteína *Cas5* de diferentes especies del grupo 1 resaltando las regiones conservadas de color rojo y las regiones diferentes en gris y azul, realizado en COBALT. **d.2.** Alineamiento general de la proteína *Cas5*, del grupo 2, resaltando las regiones conservadas en rojo y las regiones diferentes en gris y azul, realizado en COBALT.

El análisis de la secuencia del gen *cas5* en especies del género *Cryobacterium* mostró una región conservada en la zona intermedia del gen mientras que a lo largo de la secuencia proteica se evidenció un alto nivel de similitud (Figura 12c). El gráfico de polimorfismo destacó una alta diversidad nucleotídica a lo largo del gen, en contraste con el gráfico de Alfa que identificó cinco sitios bajo selección negativa en la proteína (Figura 12c). Dado que los sitios observados en el gráfico de Alfa coincidieron con las regiones señaladas en el alineamiento de COBALT (Figura 12d), estas regiones se seleccionaron para el diseño de oligonucleótidos (Figuras 12a y 12b).

**a**



**b**



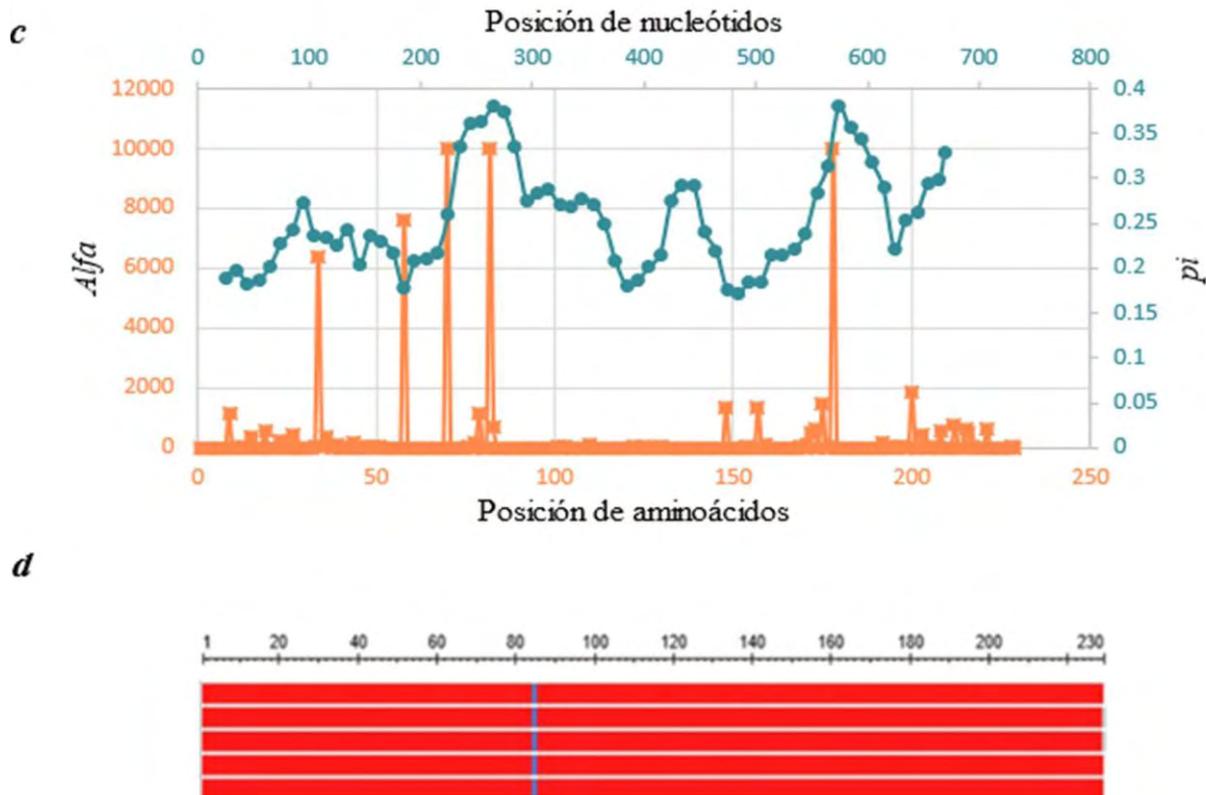
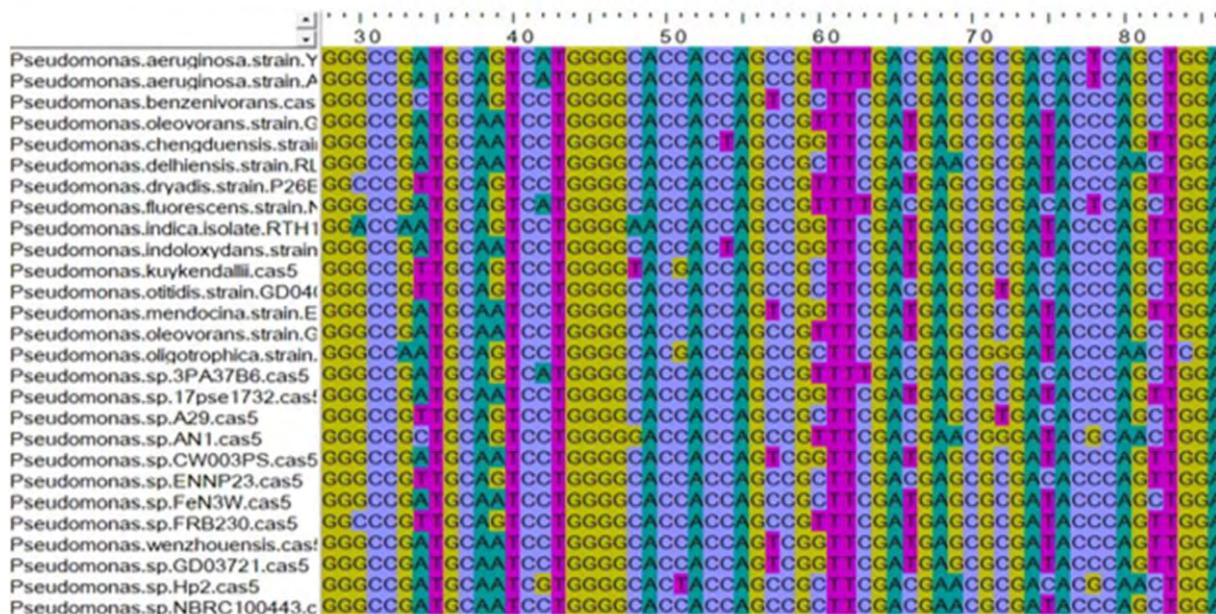


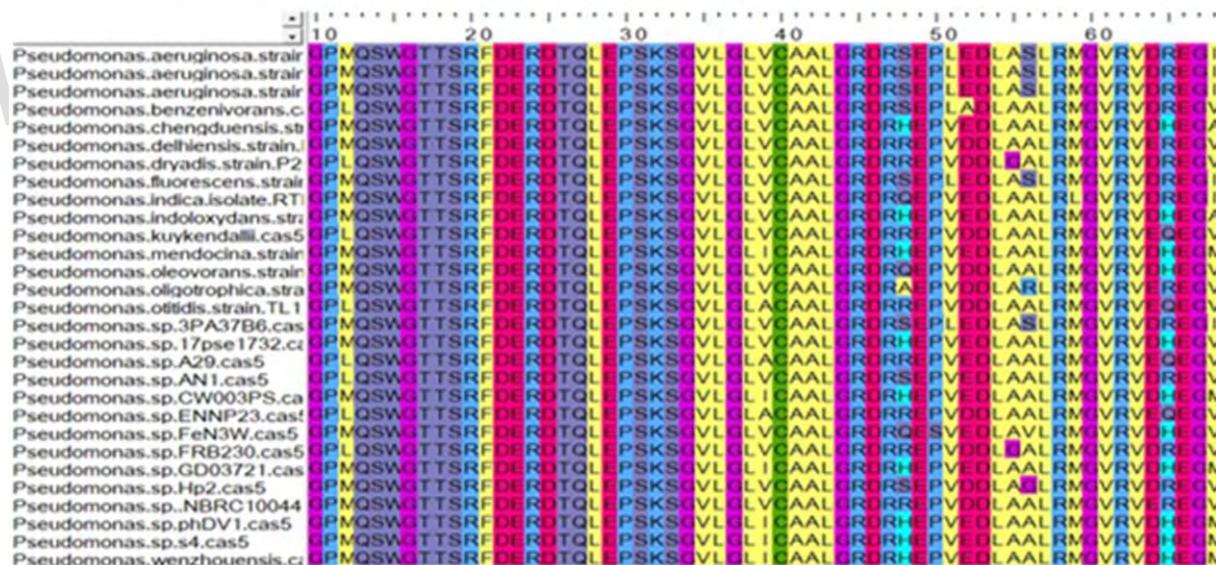
Figura 12.- Análisis de las secuencias del gen *cas5* en diferentes especies del género *Cryobacterium* **a.** Alineamiento del gen *cas5* desde la posición 100 a las 155. **b.** Alineamiento de la proteína Cas5 desde la posición 26 a las 80. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas5*, donde  $\pi$  está representado por  $\bullet$  y Alfa está representado por  $\blacksquare$ . **d.** Alineamiento general de la proteína Cas5, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

En el género *Pseudomonas* se observó una región muy conservada al realizar el alineamiento del gen *cas5*, ubicada en la región inicial desde el nucleótido 28 al 77 (Figuras 13a) y reflejado en la proteína en los aminoácidos 10 y 65 (Figura 13b y 13d). El gráfico de polimorfismo mostró mayor variabilidad en la región final del gen, mientras que el gráfico de Alfa identificó diversos puntos de selección negativa desde el aminoácido 19 al 59, y un sitio al final de la secuencia, en la posición 213 (Figura 13c). Por lo tanto, estas regiones se evaluaron para el diseño de los oligonucleótidos.

**a**



**b**



c



d

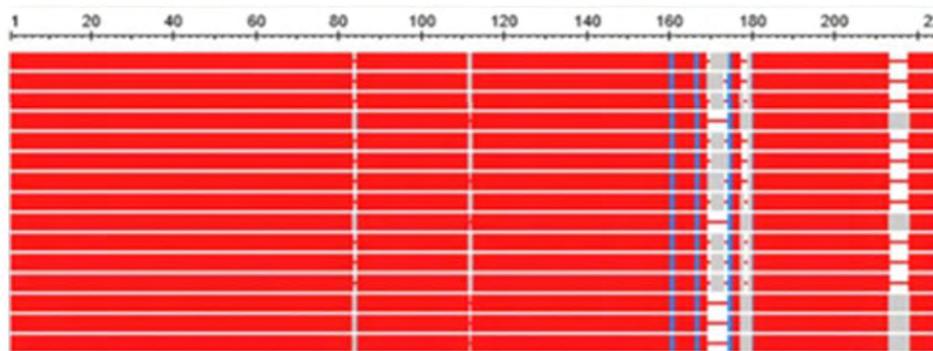
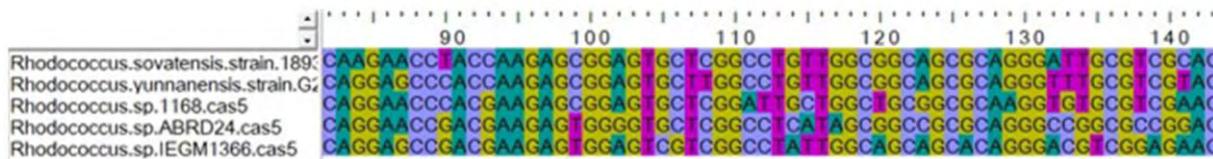


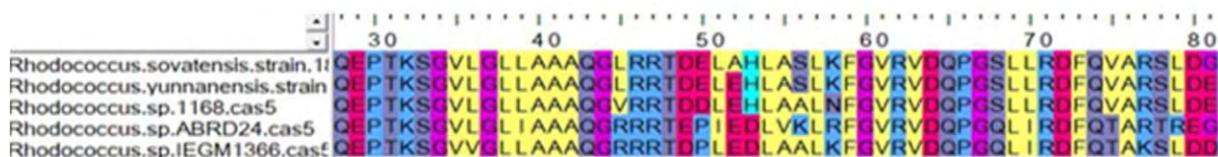
Figura 13.- Análisis de las secuencias del gen *cas5* en diferentes especies del género *Pseudomonas*. **a.** Alineamiento del gen *cas5* desde la posición 28 hasta la 79. **b.** Alineamiento de la proteína Cas5 desde la posición 10 a la 62. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas5*, donde  $\pi$  está representado por  $\bullet$  y Alfa está representado por  $\blacksquare$ . **d.** Alineamiento general de la proteína Cas5, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

Al igual que en el género *Pseudomonas*, el alineamiento del gen *cas5* para especies del género *Rhodococcus* reflejó una región conservada en la zona inicial desde el nucleótido 28 al 74 (Figuras 14a-14d). El gráfico de polimorfismo mostró una alta diversidad nucleotídica a lo largo del gen, mientras que el gráfico Alfa mostró varias regiones bajo selección negativa, la mayoría de ellas ubicadas en la región central de la proteína. También se observaron dos sitios en el inicio, en las posiciones 20 y 93, y en la región final, en las posiciones 213 y 226 (Figura 14c). Debido a que había muchas regiones potenciales para el diseño de oligonucleótidos, se seleccionaron aquellos sitios que coincidían con los identificados en el alineamiento de COBALT.

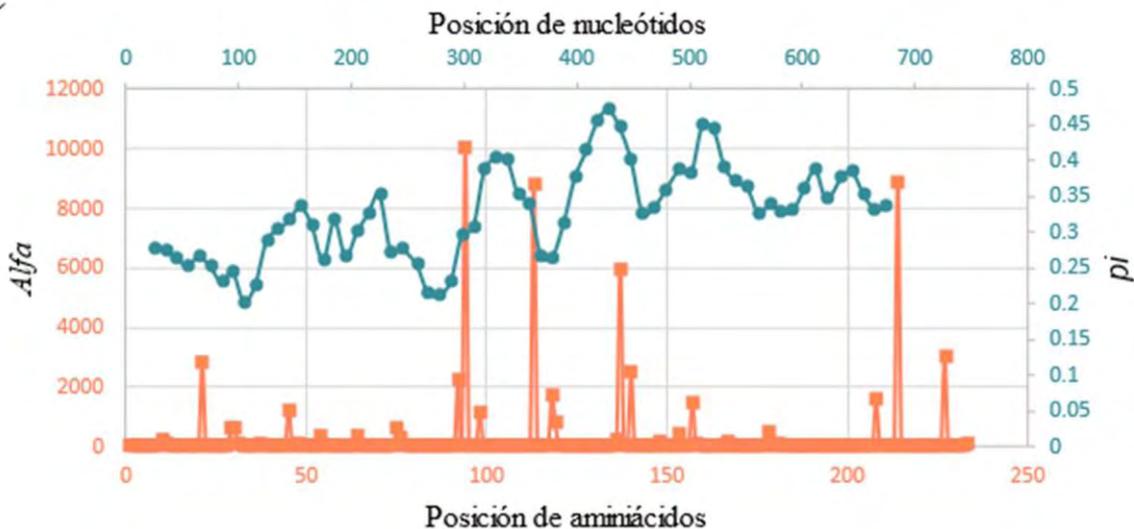
**a**



**b**



**c**



**d**

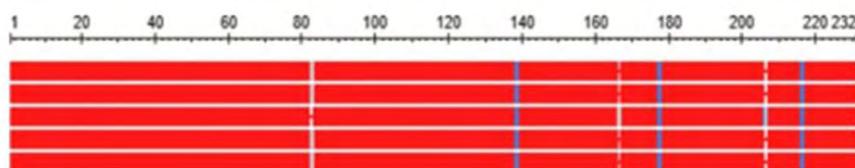


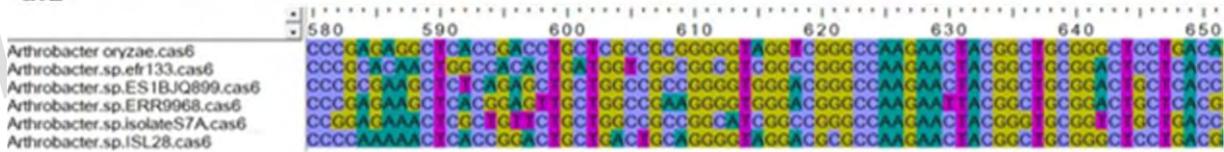
Figura 14.- Análisis de las secuencias del gen *cas5* en diferentes especies del género *Rhodococcus*. **a.** Alineamiento del gen *cas5*, desde la posición 83 hasta la 128. **b.** Alineamiento de la proteína Cas5, desde la posición 28 a la 80. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas5*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína Cas5, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

### 7.2.4 Gen *cas6*

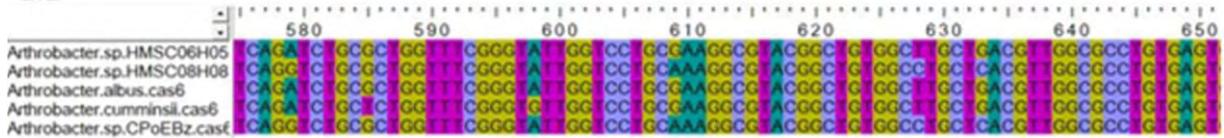
La proteína codificada por el gen *cas6* cumple funciones muy importantes en el procesamiento del pre-ARNcr y en promedio tiene una longitud de 600 pb (Makarova et al., 2019). Se encontraron muchas variaciones en el alineamiento de las secuencias de este gen para los géneros estudiados.

El grupo 2 del género *Arthrobacter* mostró una mayor similitud entre las secuencias del gen *cas6* que el grupo 1, que presentó una alta variabilidad. Cabe destacar que la especie *Arthrobacter globiformis* fue descartada del análisis debido a que no estaba anotado el gen completo en la base de datos. Las regiones de interés se identificaron en el grupo 1 mediante el alineamiento de COBALT (Figura 15d.1) y el gráfico de *Alfa*, que indicó seis sitios bajo selección negativa en la proteína, en las posiciones 23, 49, 84, 143, 199 y 203 (Figura 15c.1). En el análisis del grupo 2, varias regiones se seleccionaron para el diseño de oligonucleótidos, con base en el gráfico de polimorfismo; dicho gráfico señaló menor variación en la zona inicial de la secuencia (desde el nucleótido 65 al 95), en la zona central (en la posición 315) y en la zona terminal (en la posición 545) (Figura 15c.2). Por otro lado, el gráfico de *Alfa* indicó un sitio bajo selección negativa en el aminoácido 157 (Figura 15c.2). Así, para el grupo 2 se seleccionaron diferentes zonas potenciales para el diseño de oligonucleótidos.

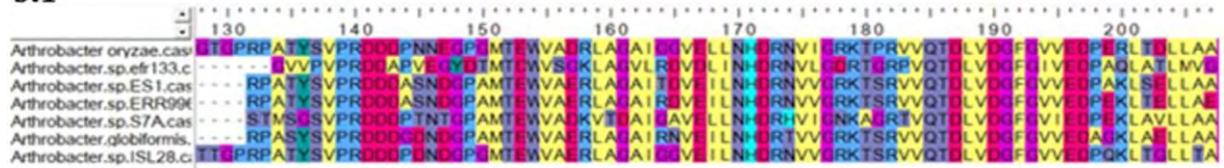
**a.1**



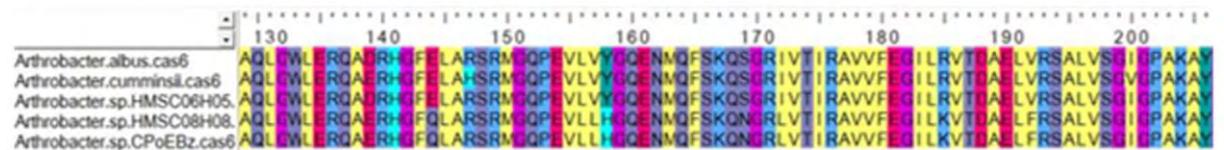
**a.2**



**b.1**



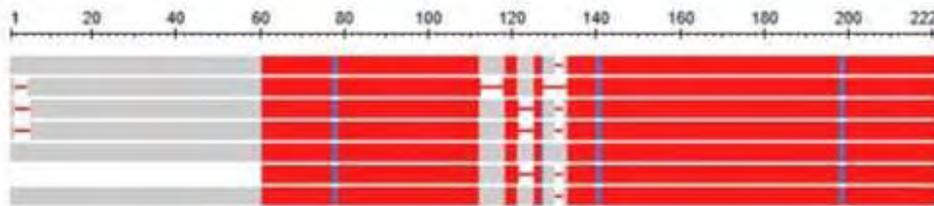
**b.2**



c.1



d.1



c.2



d.2

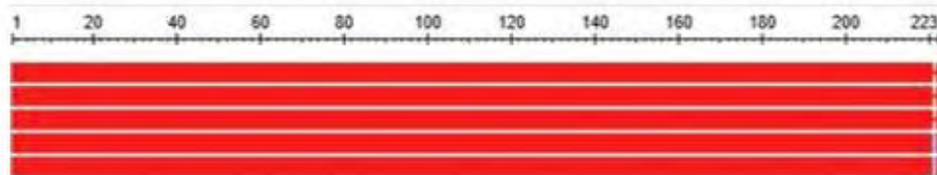


Figura 15.- Análisis de las secuencias del gen *cas6* en diferentes especies del género *Arthrobacter*. **a.1.** Alineamiento del gen *cas6* en el grupo 1 del género *Arthrobacter*, desde la posición 604 a la 649. **a.2.** Alineamiento de la proteína *Cas6* en el grupo 1 del género *Arthrobacter*, desde la posición 162 a la 217. **b.1.** Alineamiento del gen *cas6* en el grupo 2, desde la posición 580 a la 649. **b.2.** Alineamiento de la proteína *Cas6* en el grupo 2, desde la posición 130 a la 212. **c.1** y **c.2.** Gráficos de polimorfismo y Alfa para el gen *cas6* en diferentes especies de los grupos 1 y 2 del género *Arthrobacter*, respectivamente. La diversidad nucleotídica ( $\pi$ ) está representada por ● y Alfa está representado por ■. **d.1.** Alineamiento general de la proteína *Cas6* en el grupo 1, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT. **d.2.** Alineamiento general de la proteína *Cas6* en el grupo 2, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

En el género *Cryobacterium*, el alineamiento del gen *cas6* mostró dos regiones muy conservadas al inicio y al final del gen. En el alineamiento de COBALT se observó que la secuencia de la proteína estaba muy conservada especialmente en el extremo de la proteína (Figura 16d). El gráfico de polimorfismo señaló una alta diversidad a lo largo del gen, excepto hacia la región terminal, donde el valor de  $\pi$  fue más bajo (Figura 16c). Se identificaron dos sitios potenciales para el diseño de oligonucleótidos (en base al gráfico de Alfa) ubicados en las posiciones 89 y 157 (Figura 16c). Por lo tanto, para el diseño de oligonucleótidos se delimitaron estas regiones, junto con las observadas en el alineamiento de la proteína en COBALT.

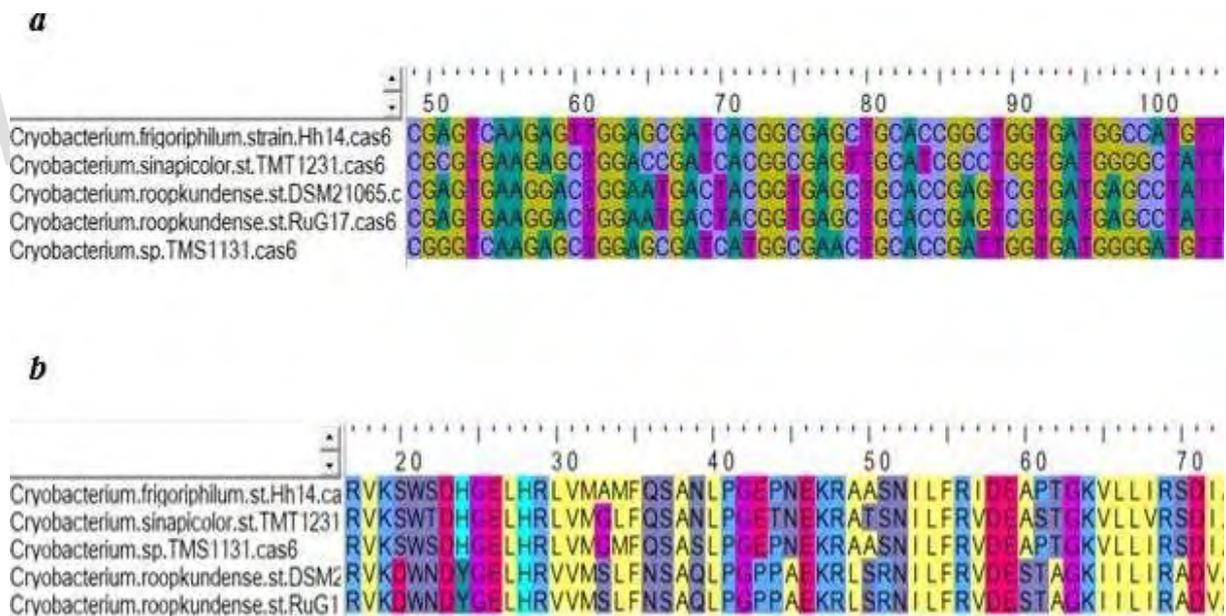
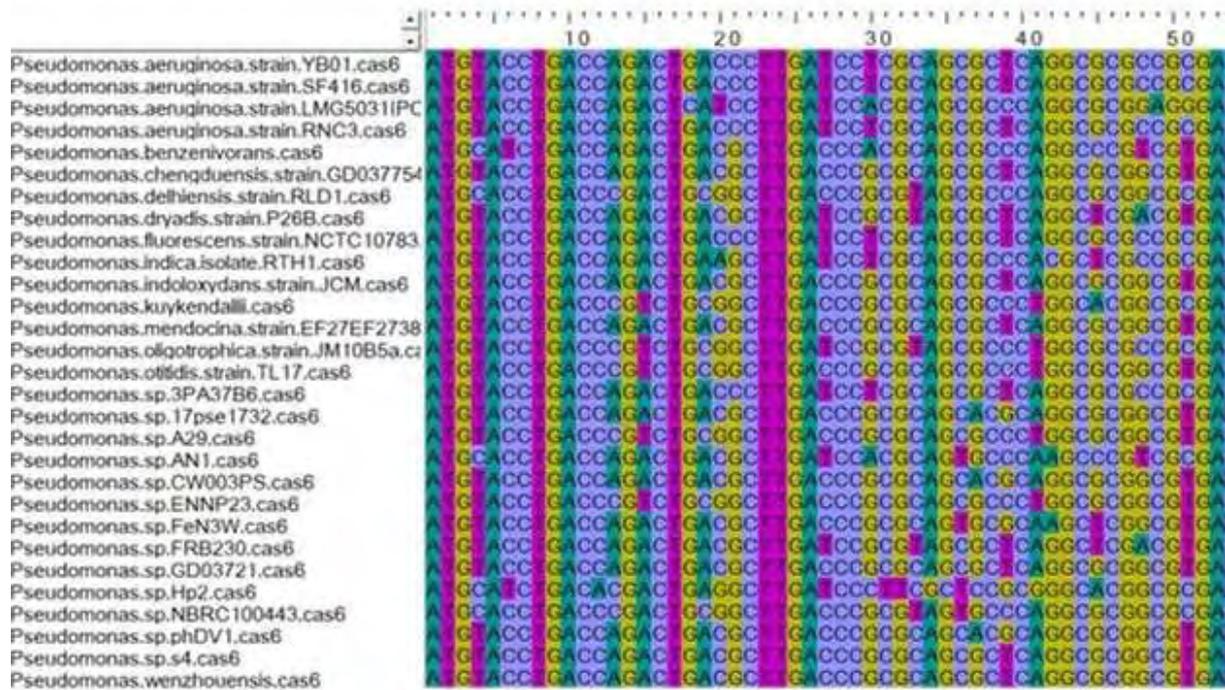




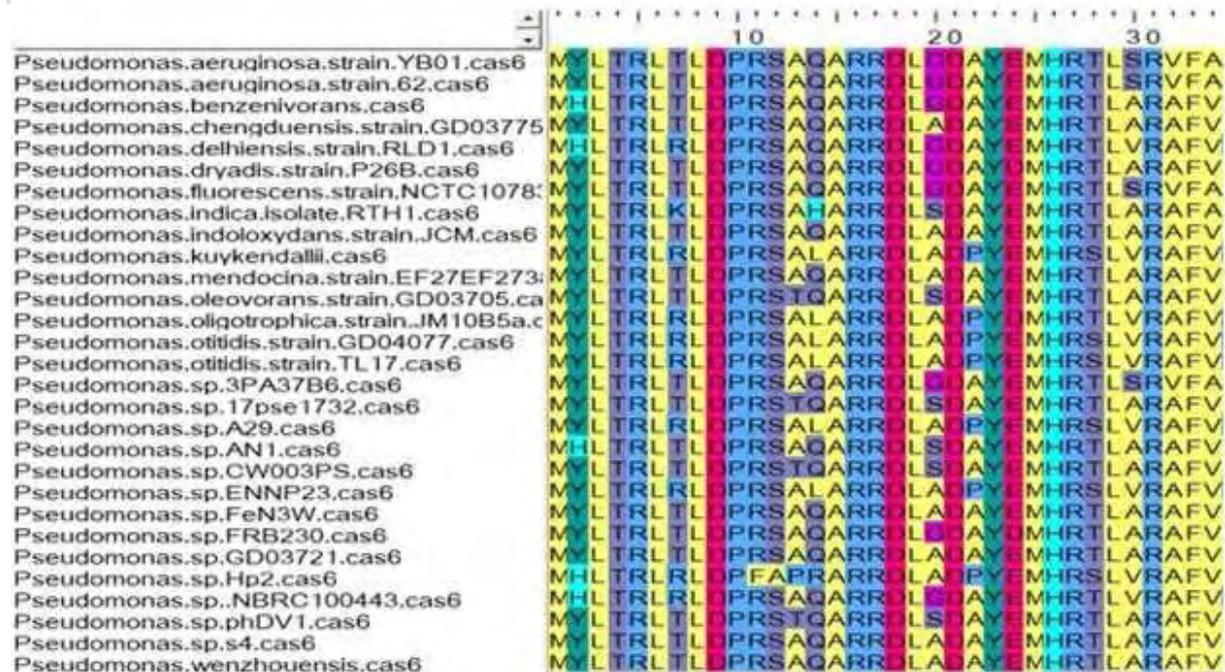
Figura 16.- Análisis de las secuencias del gen *cas6* en diferentes especies del género *Cryobacterium*. **a.** Alineamiento del gen *cas6* desde la posición 49 a la 101. **b.** Alineamiento de la proteína *Cas6* desde la posición 17 a la 69. **c. d.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas6* en diferentes especies del género *Cryobacterium*, donde  $\pi$  está representado por  $\bullet$  y Alfa está representado por  $\blacksquare$ . **d.** Alineamiento general de la proteína *Cas6*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

Al realizar el alineamiento de las secuencias nucleotídicas y aminoacídicas del gen *cas6* en el género *Pseudomonas* se observaron dos regiones muy conservadas, al inicio, entre los nucleótidos en posición 1 a 100 (Figura 17a), y en la región intermedia, alrededor de la posición 300 de los nucleótidos. La Figura 17c muestra dos regiones de baja diversidad nucleotídica en las posiciones 25, 341 y 591. En el gráfico de Alfa también se muestra un sitio de selección negativa en la posición 63 (Figura 17c). Para llevar a cabo el diseño de oligonucleótidos, se evaluaron todas las regiones identificadas mediante los diferentes análisis.

**a**



**b**



**e**

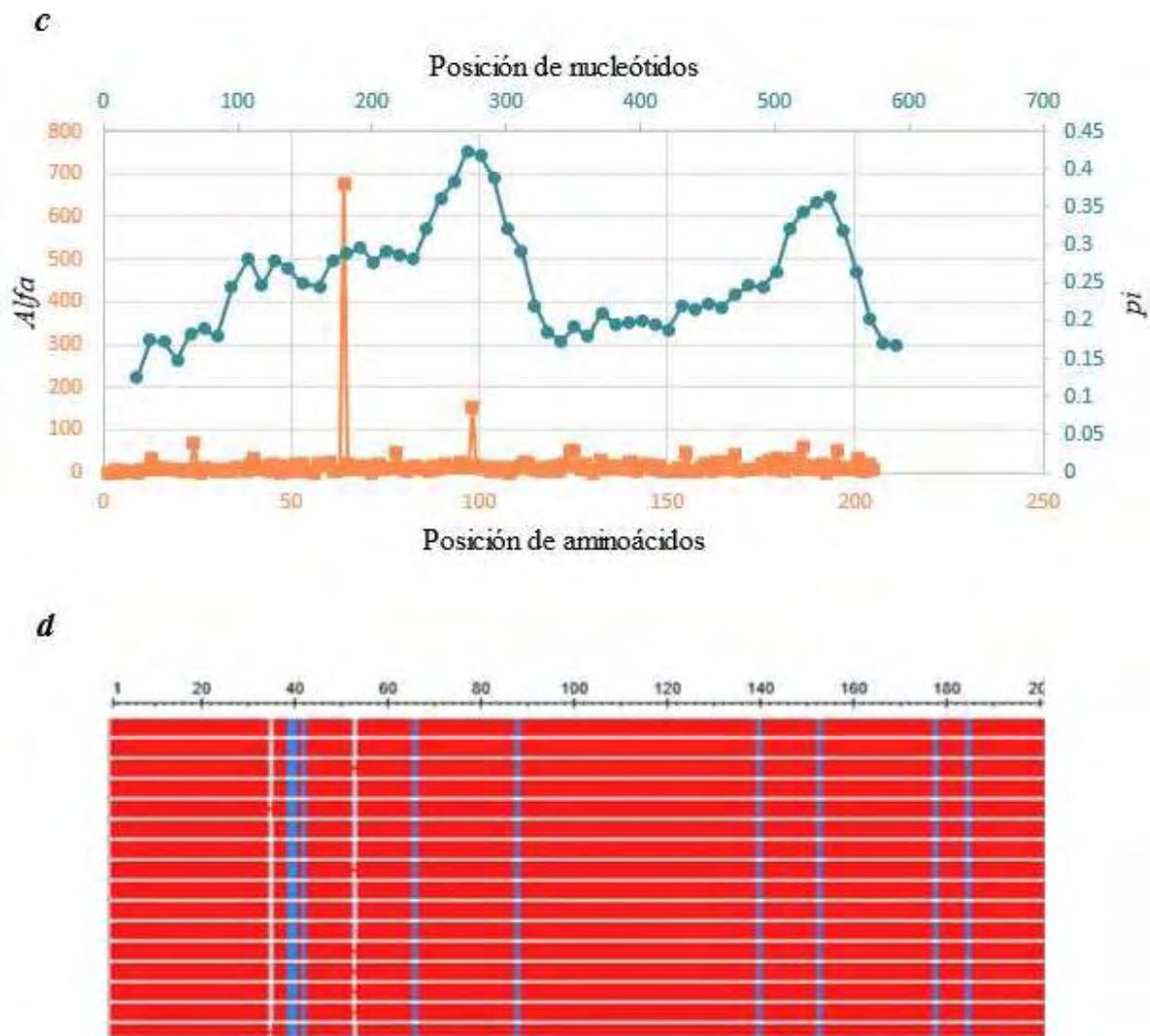
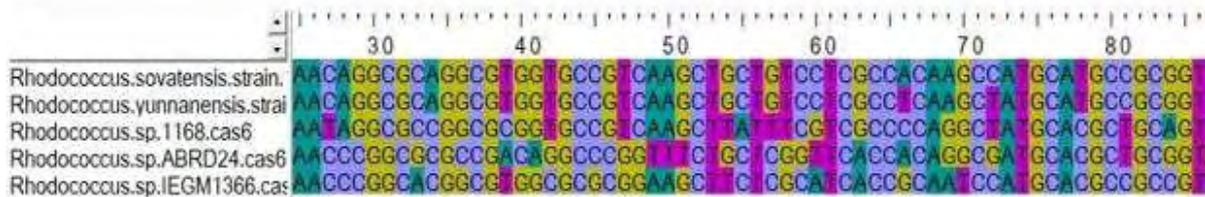


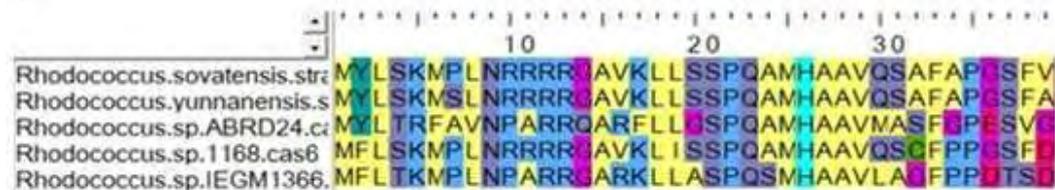
Figura 17.- Análisis de las secuencias del gen *cas6* en diferentes especies del género *Pseudomonas*. **a.** Alineamiento del gen *cas6*, desde la posición 1 hasta la 47. **b.** Alineamiento de la proteína Cas6, desde la posición 1 hasta la 34. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas6*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína Cas6, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT

Se identificaron tres regiones conservadas en las secuencias nucleotídicas del gen *cas6* para el género *Rhodococcus*. La primera está ubicada entre las posiciones 25 y 80, la segunda está alrededor de la posición 290 y la tercera, cerca de la posición 610. Tanto en la secuencia nucleotídica como en la aminoacídica de la especie *Rhodococcus sp. IEGM* se observaron varias diferencias con respecto al resto de especies. En la Figura 18c se observa una alta diversidad nucleotídica a lo largo del gen, en contraste con el gráfico de Alfa, donde se indican varios sitios bajo selección negativa (aminoácidos 51, 110, 179 y 219). Por lo tanto, se evaluaron las tres regiones señaladas en el alineamiento y los sitios mostrados en el gráfico de Alfa para el diseño de oligonucleótidos.

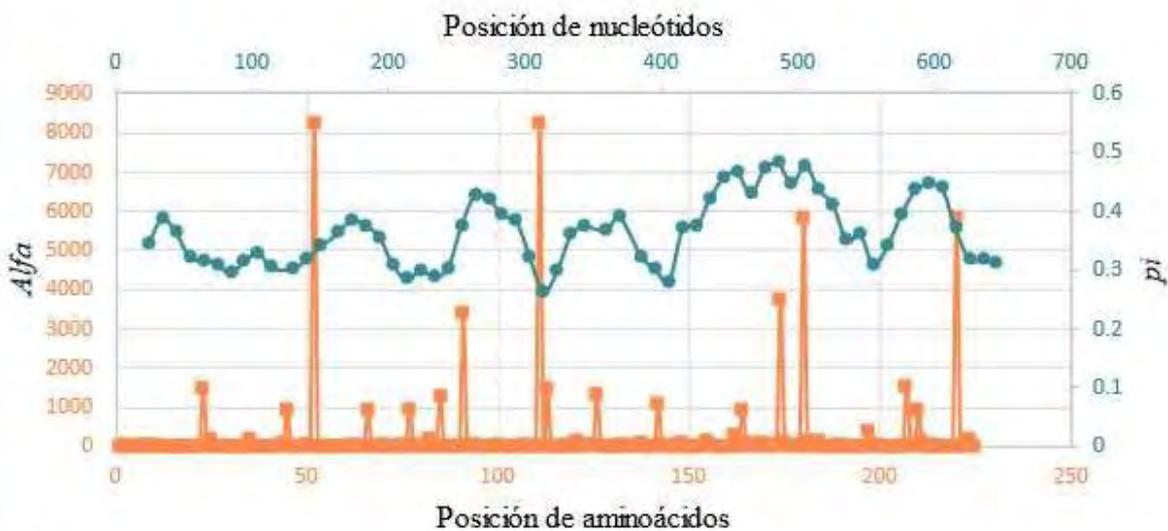
**a**



**b**



**c**



**d**

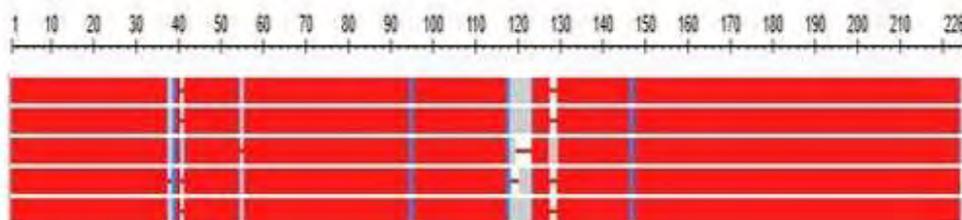


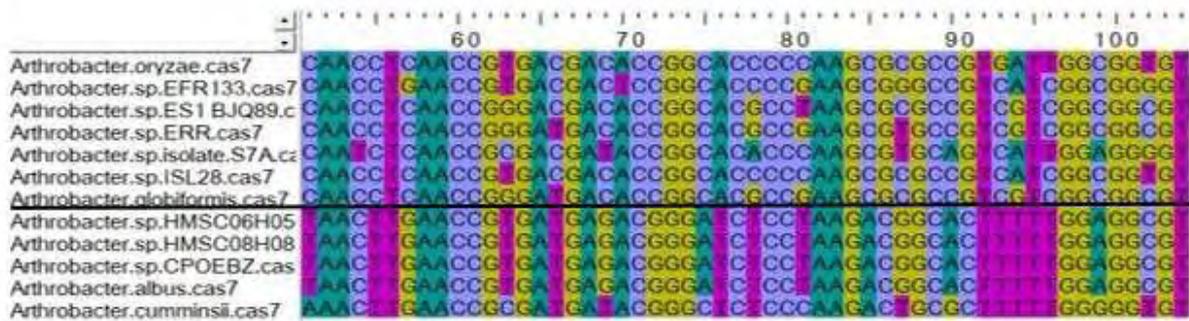
Figura 18.- Análisis de las secuencias del gen *cas6* en diferentes especies del género *Rhodococcus*. **a.** Alineamiento del gen *cas6* desde la posición 25 a la 77. **b.** Alineamiento de la proteína Cas6 desde la posición 1 a la 38. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas6*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína Cas6, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

### 7.2.5 Gen *cas7*

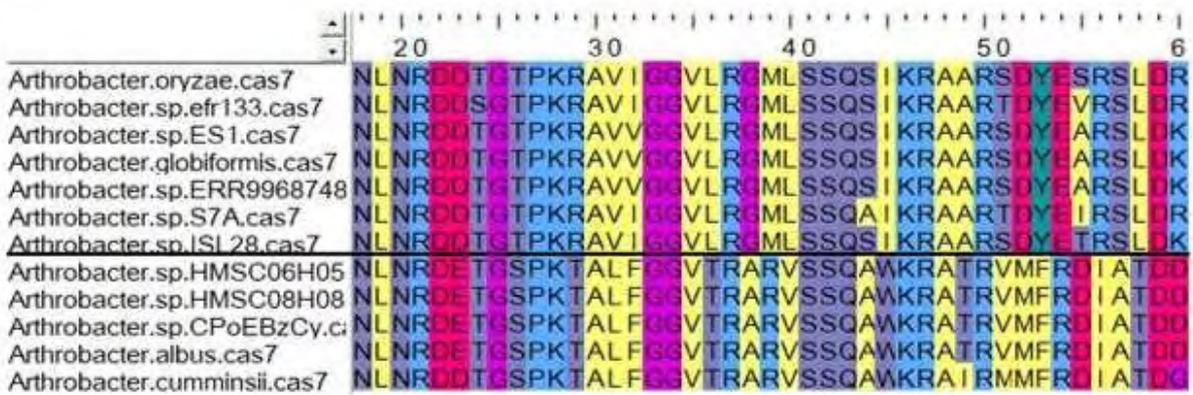
El análisis de las secuencias nucleotídicas y aminoacídicas del gen *cas7* en los géneros estudiados, mostró que este es un gen muy conservado, independientemente del género, la especie y la procedencia (Figura 23). De hecho, se identificaron dos regiones muy conservadas entre todos los géneros analizados: la primera ubicada entre los nucleótidos 50 y 100, y la segunda, entre los nucleótidos 500 y 640 (resultados no mostrados). Esto brinda un indicio de la alta importancia que tiene la proteína Cas7 y su funcionalidad en el complejo efector del sistema CRISPR-Cas del tipo I-E.

En el género *Arthrobacter* se evidenció que tanto el gen como la proteína estaban muy conservados en ambos grupos, con ligeras diferencias de algunos aminoácidos. En algunos casos sí cambiaba la naturaleza del mismo, lo que podría influir en la conformación de la proteína y en la interacción con las otras proteínas que conforman el complejo *Cascade*. Los gráficos de polimorfismo fueron muy similares entre los dos grupos del género *Arthrobacter*, mostrando una menor variabilidad en la región inicial del gen (Figura 19c). En los géneros *Cryobacterium*, *Pseudomonas* y *Rhodococcus* se observó que había menor variación en la región central. Por otra parte, el gráfico de *Alfa* arrojó en el género *Arthrobacter* grupo 1, varios sitios bajo selección negativa concentrados en la región central de la proteína, mientras que en el grupo 2 se apreció que toda la proteína tenía muchos sitios bajo selección negativa. En el género *Cryobacterium* se identificaron varios sitios de selección negativa en el gráfico de *Alfa*, en la posición 89, 139 y 202 (figura 19), por el contrario, en el género *Pseudomonas* los sitios bajo selección negativa fueron en la posición 71 y la 349 (figura 20). Finalmente, en el género *Rhodococcus* el gráfico de *Alfa* mostró que los sitios bajo selección negativa estaban concentrados en la región inicial de la proteína (figura 21). A pesar de las diferencias observadas en los gráficos de polimorfismo y de *Alfa* entre los géneros, en todas las especies, las regiones de interés se concentraron al inicio y al final tanto en los alineamientos del gen como de la proteína (figura 22), en consecuencia, se procedió a realizar un alineamiento de estas regiones de todos los géneros, mostrando que tanto a nivel de nucleótidos como de aminoácidos son regiones muy conservadas y son idóneas para el diseño de oligonucleótidos para todos los géneros.

**a**

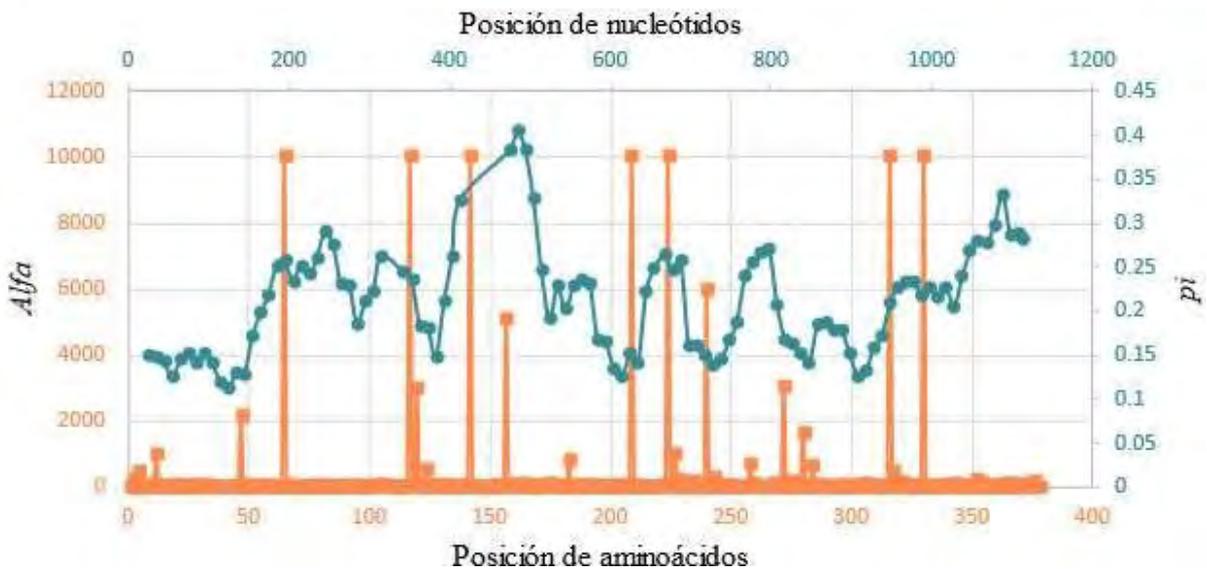


**b**

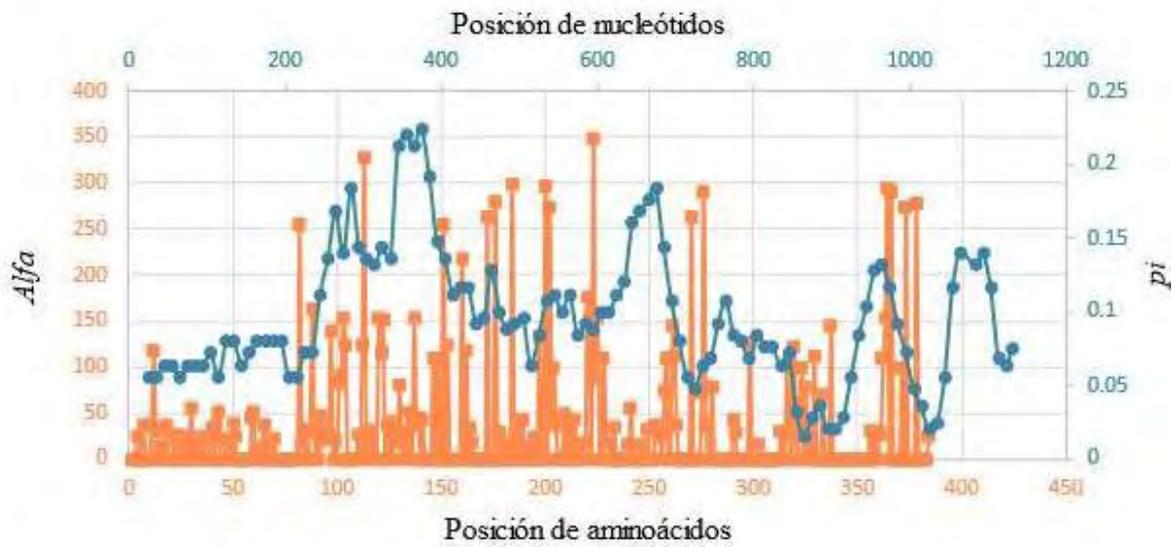


W  
c.1

e



c.2



d

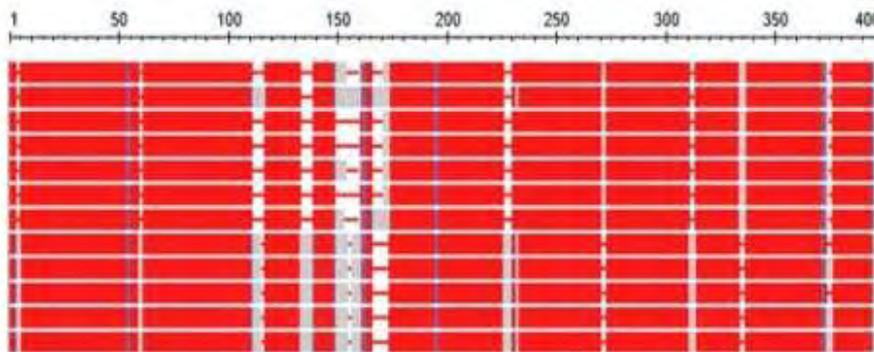
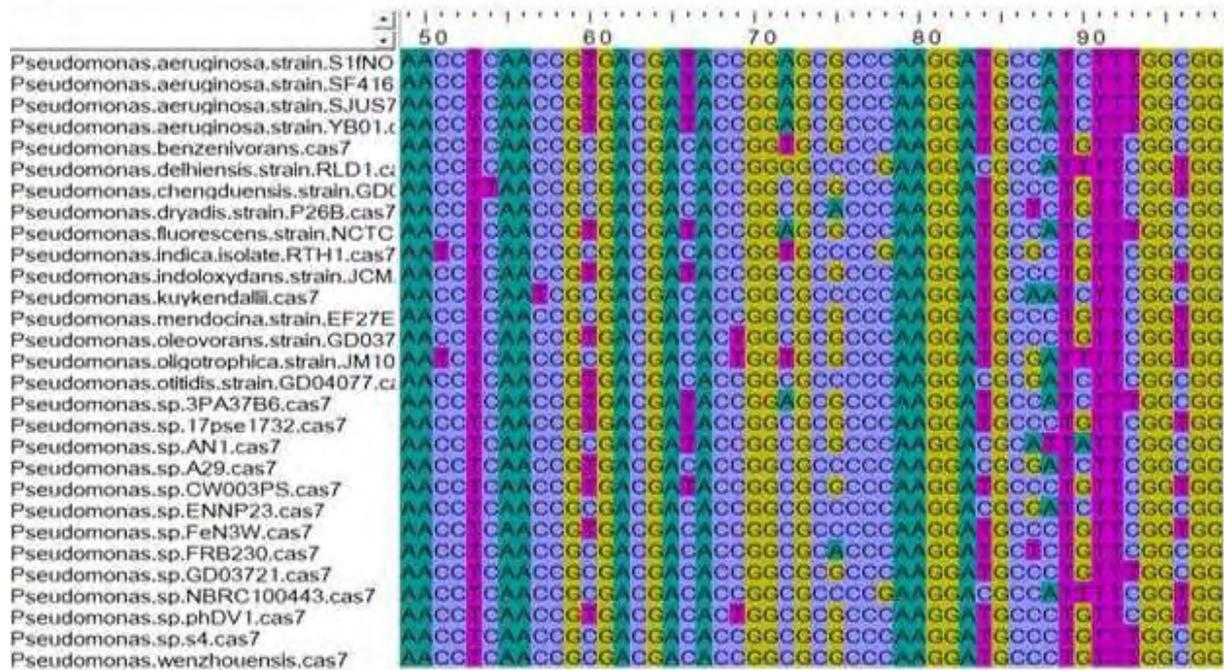


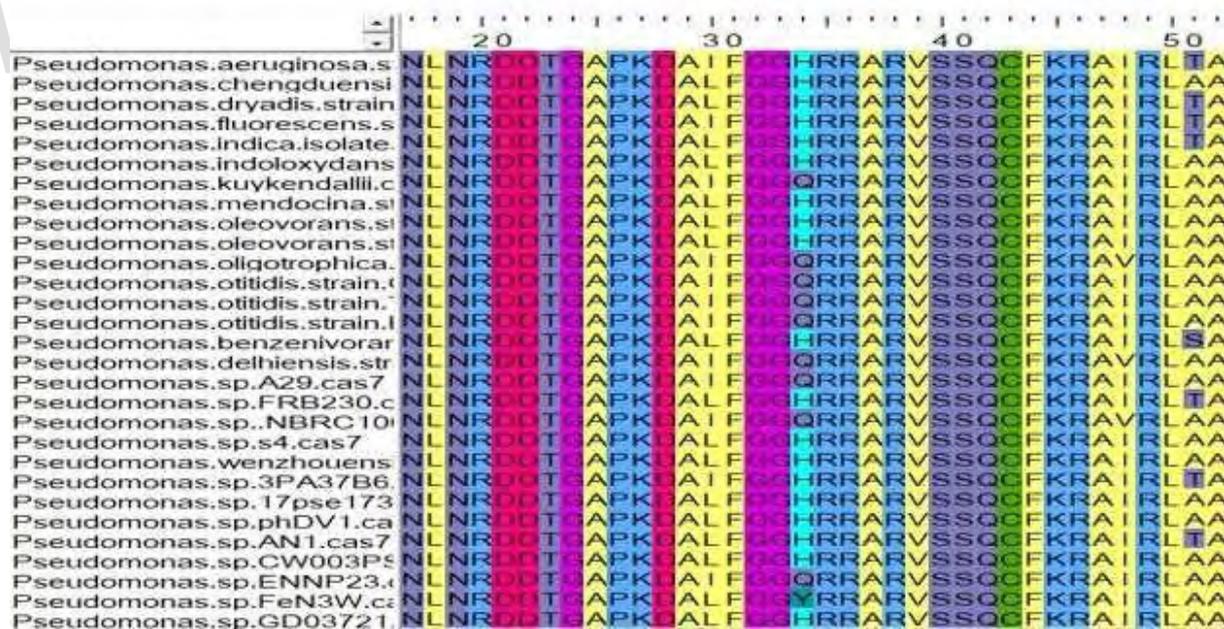
Figura 19.- Análisis de las secuencias del gen *cas7* en diferentes especies del género *Arthrobacter* **a.** Alineamiento del gen *cas7* de diferentes especies del género *Arthrobacter*, desde la posición 51 a la 98. Las primeras siete especies corresponden al grupo 1, mientras las últimas cinco especies corresponden al grupo 2. **b.** Alineamiento de la proteína Cas7 desde la posición 18 a la 50. Las primeras siete especies corresponden al grupo 1, mientras las últimas cinco especies corresponden al grupo 2, se aprecia gran similitud entre los aminoácidos de ambos grupos con ligeras diferencias. **c.1.** y **c.2.** Gráficos de polimorfismo y de Alfa para el gen *cas7* en diferentes especies del grupo 1 y 2 del género *Arthrobacter* respectivamente. La diversidad nucleotídica ( $\pi$ ) está representado por  $\bullet$  y Alfa está representado por  $\blacksquare$ . **d.** Alineamiento general de la proteína Cas7 incluidos ambos grupos, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT. Las primeras siete especies conforman el grupo 1 y las últimas cinco corresponden a especies del grupo 2.



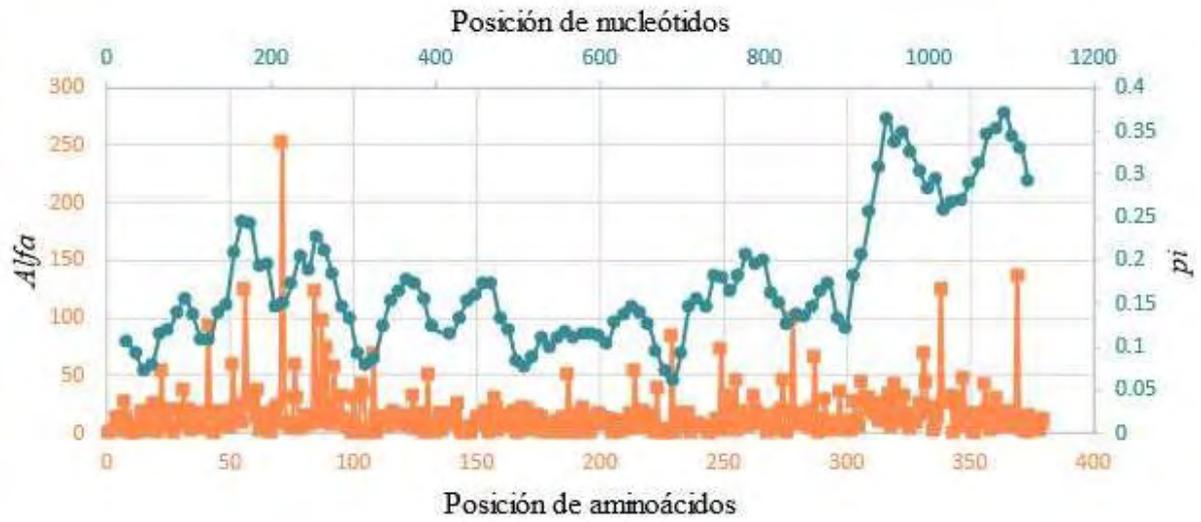
**a**



**b**



**c**



**d**



Figura 21.- Análisis de las secuencias del gen *cas7* en diferentes especies del género *Pseudomonas* **a.** Alineamiento del gen *cas7*, desde la posición 49 a la 94. **b.** Alineamiento de la proteína Cas7, desde la posición 17 a la 49. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas7* en diferentes especies del género *Pseudomonas*, donde la diversidad nucleotídica ( $\pi$ ) está representado por  $\bullet$  y Alfa está representado por  $\blacksquare$ . **d.** Alineamiento general de la proteína Cas7, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT

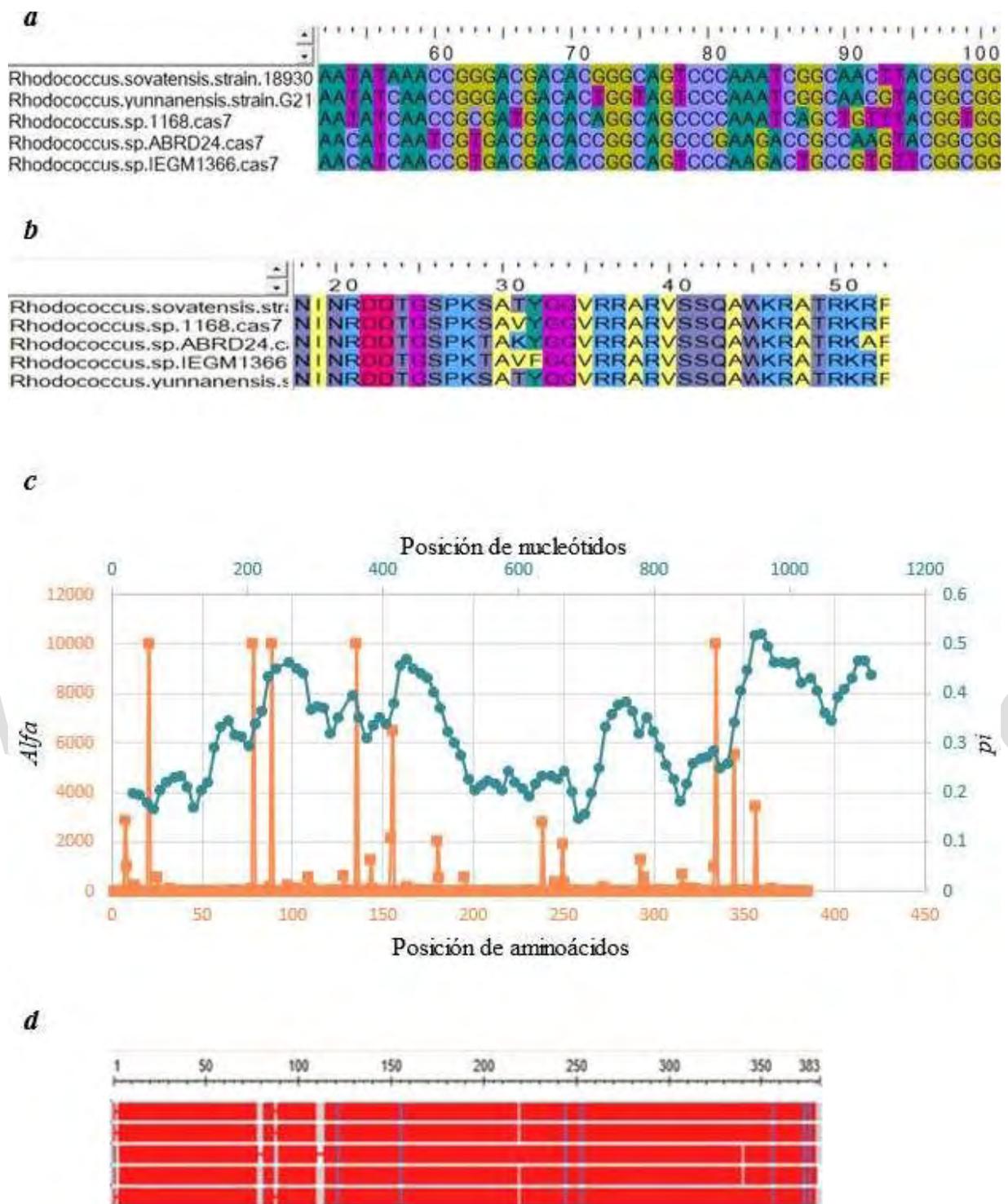
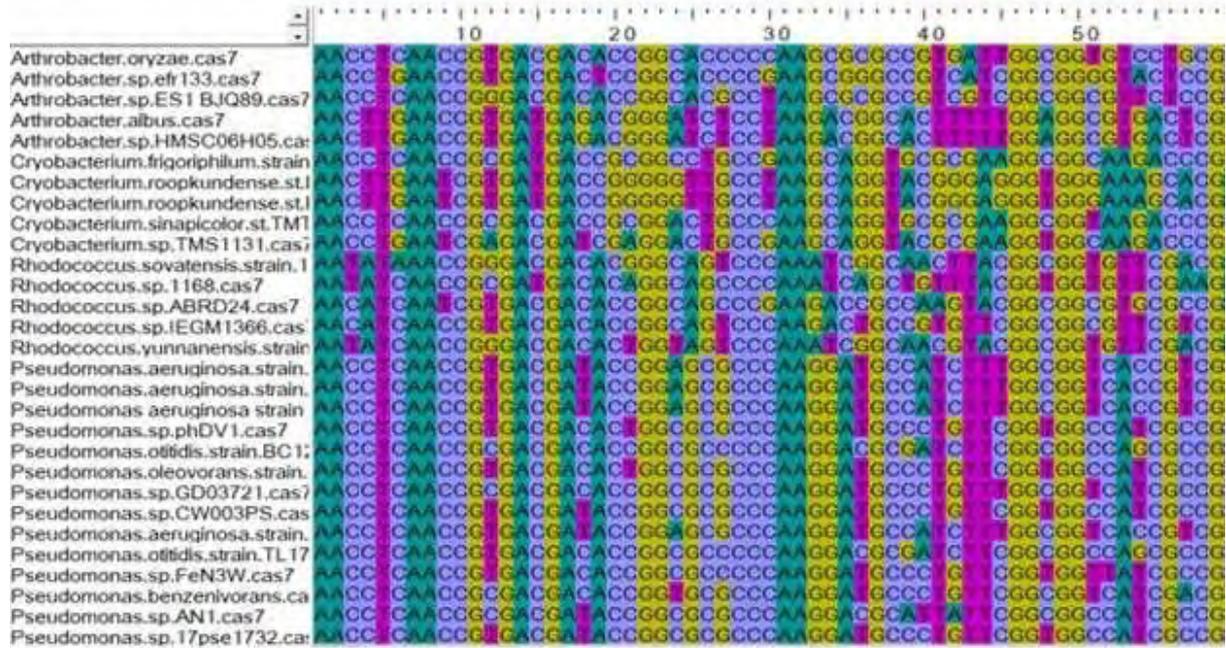


Figura 22.- Análisis de las secuencias del gen *cas7* en diferentes especies del género *Rhodococcus*. **a.** Alineamiento del gen *cas7*, desde la posición 52 a la 93. **b.** Alineamiento de la proteína Cas7, desde la posición 18 a la 50. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas7* en diferentes especies del género *Rhodococcus*, donde la diversidad nucleotídica ( $\pi$ ) está

representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína Cas7, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

**a**



**b**

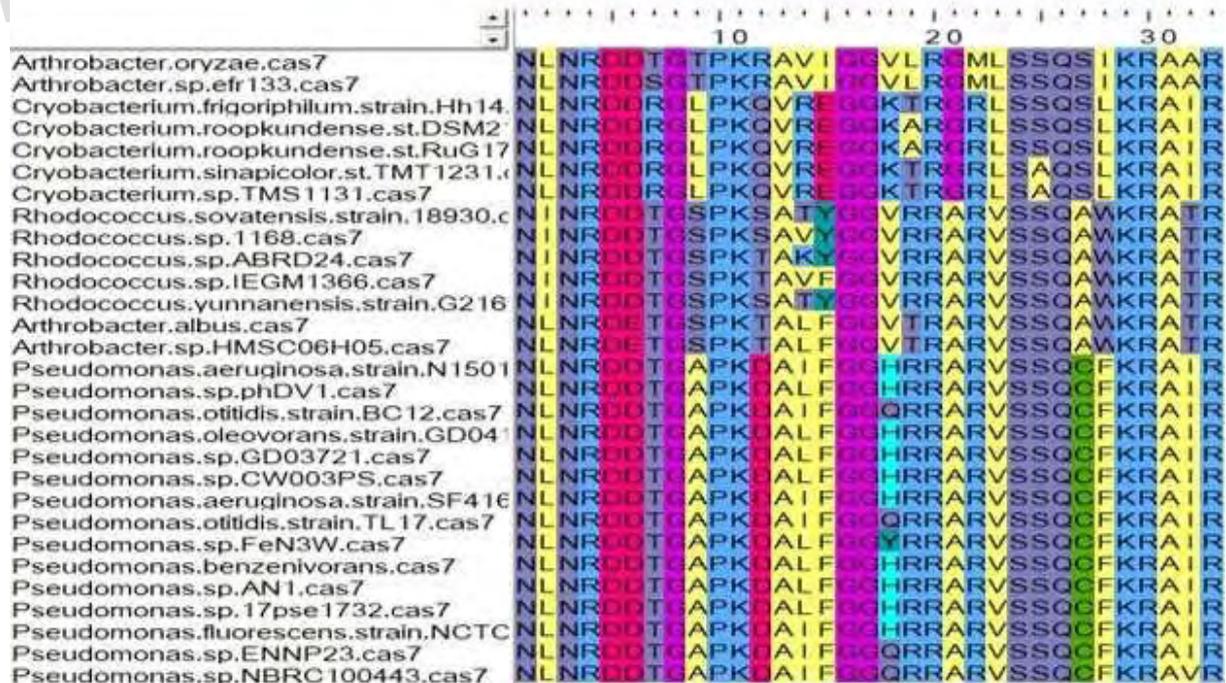
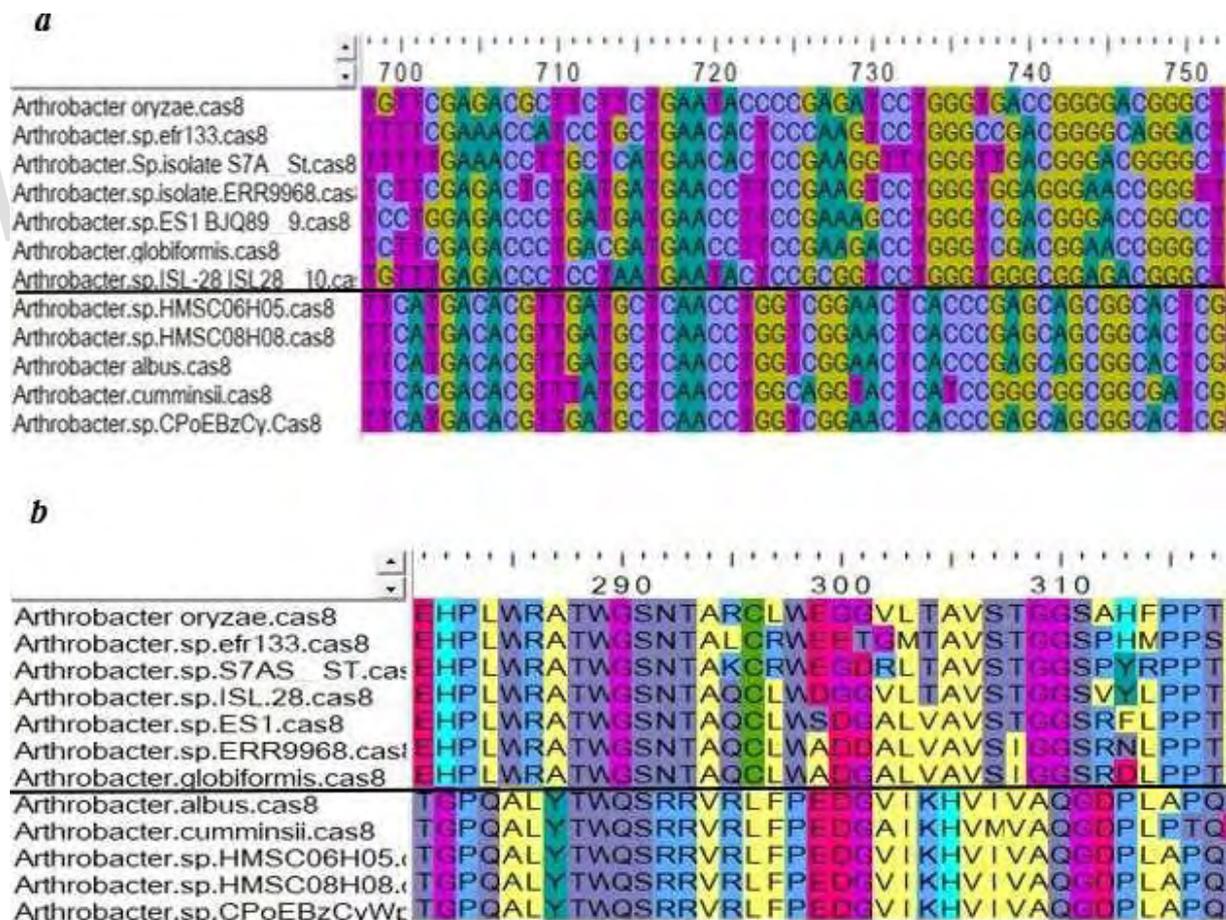


Figura 23.- Alineamiento de la región inicial del gen *cas7* (a) y de la proteína Cas7 (b) de diferentes especies de los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas* y *Rhodococcus*.

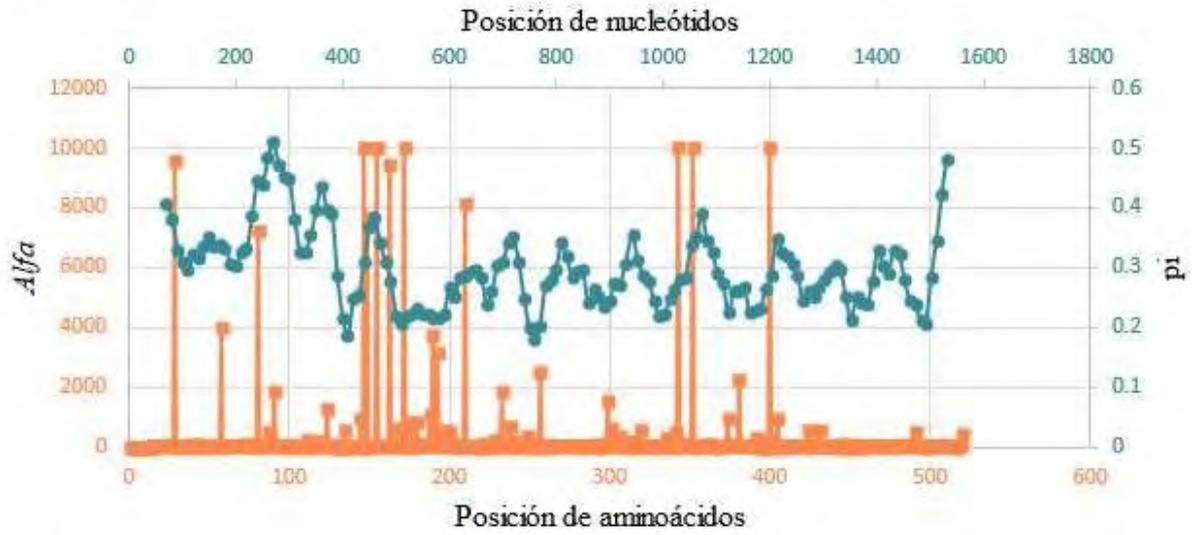
### 7.2.6 Gen *cas8*

La proteína Cas8 es parte del complejo *Cascade* y ayuda a mantener la estabilidad del mismo, además de reclutar a la proteína Cas3 y posicionar el complejo sobre el ADN diana, lo que permite que la Cas3 lo degrade (Xue & Sashital, 2019). Por lo tanto, es una proteína muy importante dentro del sistema CRISPR-Cas. El análisis de las secuencias nucleotídicas y aminoácidas del gen *cas8* mostraron diversas regiones conservadas en cada uno de los géneros, pero con grandes diferencias entre los mismos.

En el género *Arthrobacter* se identificó una región conservada tanto en el grupo 1 como en el grupo 2, ubicada entre los nucleótidos 700 al 750 (Figura 24a). El alineamiento COBALT señaló varias regiones conservadas entre ambos grupos (Figura 24d). Además, estas regiones concordaron con varios sitios bajo selección negativa observados en el gráfico *Alfa* del grupo 1 (Figura 24c.1), por lo que se seleccionaron estas regiones para el diseño de oligonucleótidos.



c.1



c.2



d

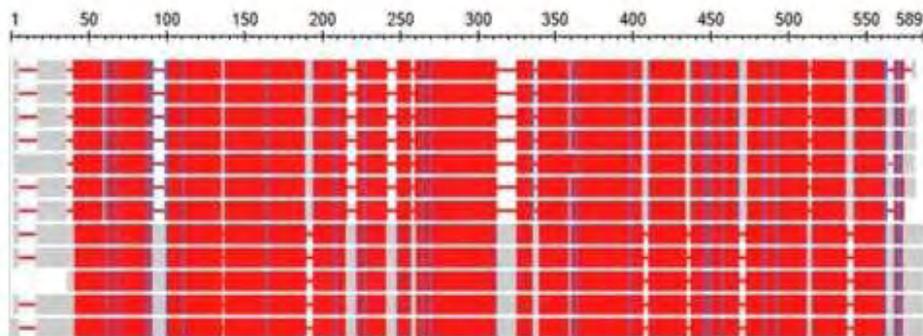
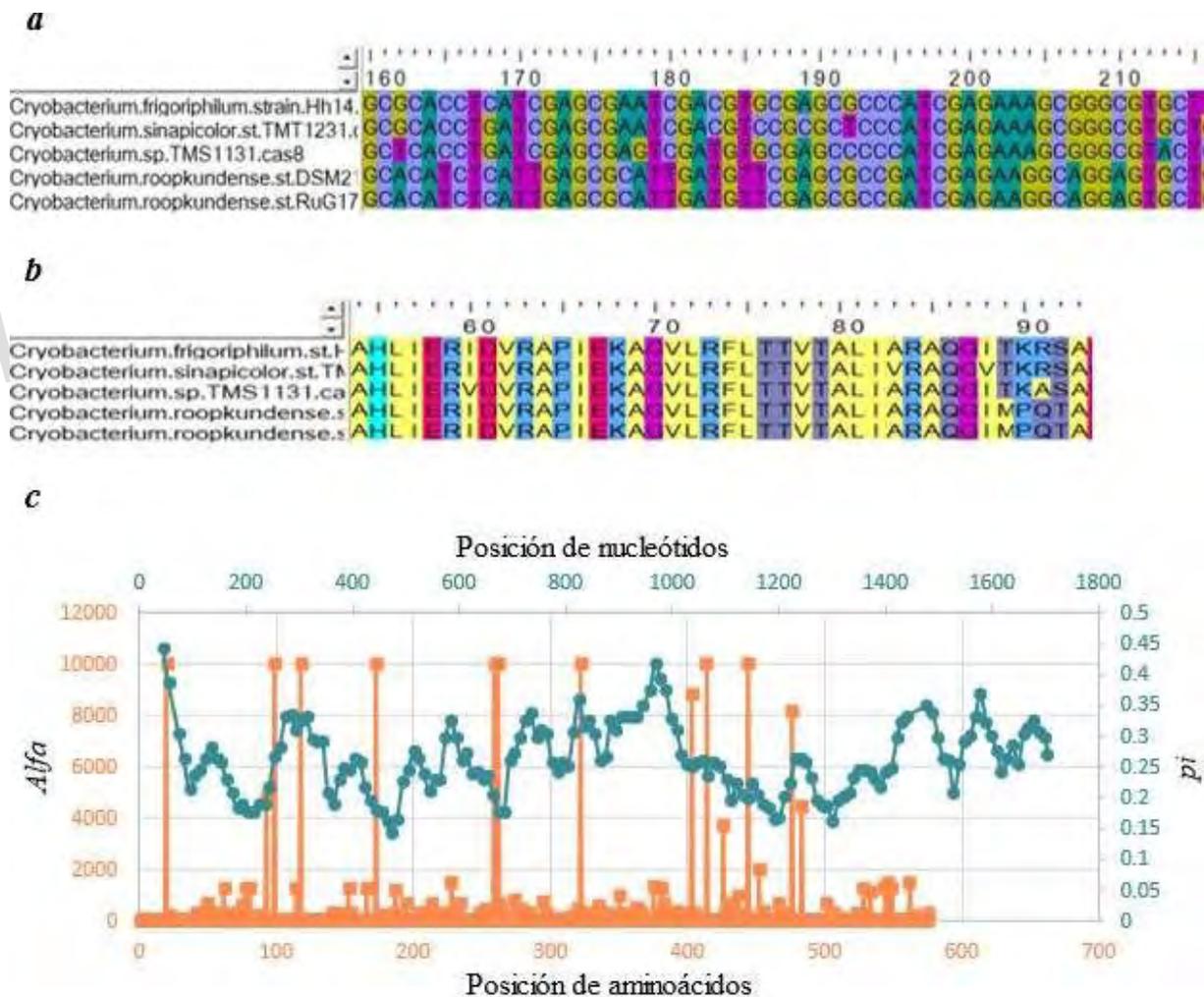


Figura 24.- Análisis de las secuencias del gen *cas8* en diferentes especies del género *Arthrobacter*. **a.** Alineamiento del gen *cas8* desde la posición 698 hasta la 735. **b.** Alineamiento de la proteína *Cas8* de la posición 281 a la 310. **c.1.** y **c.2.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas8* de los grupos 1 y 2 del género *Arthrobacter*, respectivamente, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas8*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT. Las primeras 7 secuencias corresponden al grupo 1 y las 5 secuencias restantes corresponden al grupo 2.

El alineamiento de las secuencias del gen *cas8* en especies del género *Cryobacterium* reveló un contraste entre algunas regiones muy conservadas a nivel de la secuencia de la proteína (e.g. Figura 25b), pero con alta diversidad a nivel de la secuencia nucleotídica (Figura 25a). En la Figura 25c se observan diferentes regiones con valores bajos de *pi*, que se vieron reflejados en el gráfico de Alfa (aminoácidos 19, 98, 117, 172, 258 y 413 bajo selección negativa). Estos sitios, junto con las regiones registradas en los alineamientos fueron evaluados posteriormente para el diseño de oligonucleótidos.



**d**

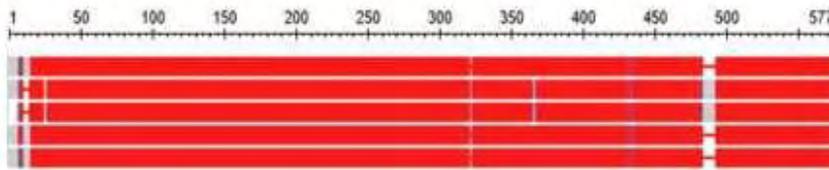
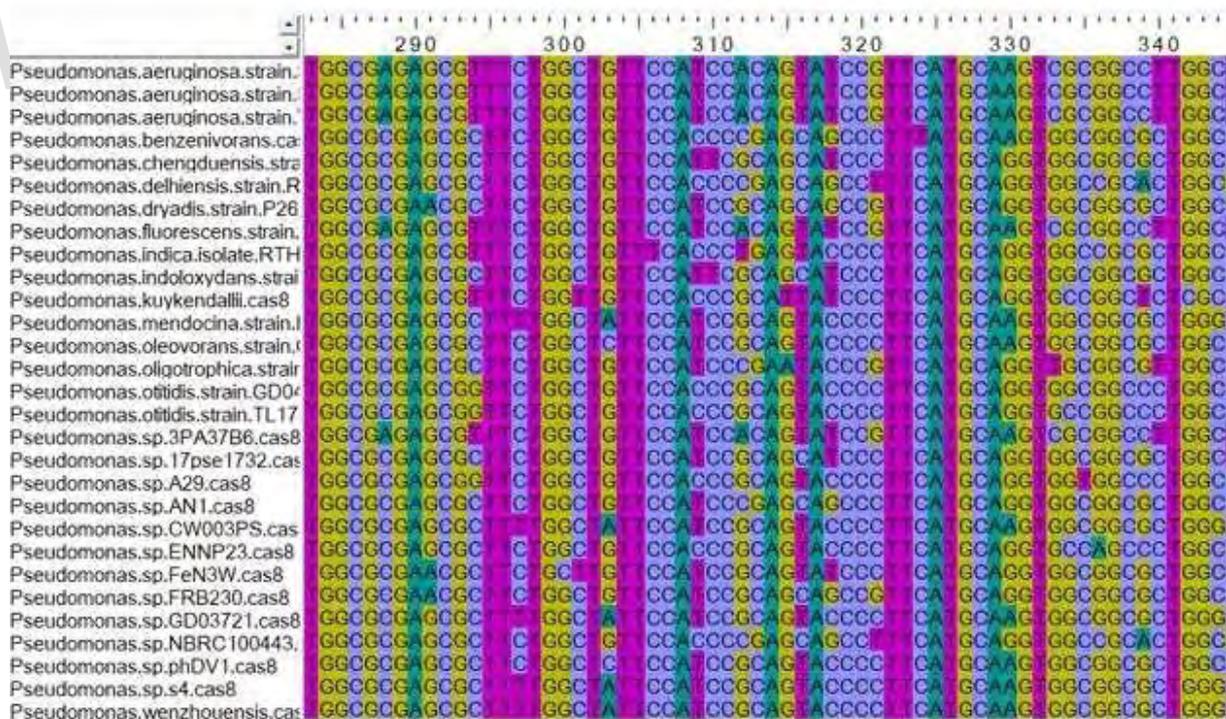


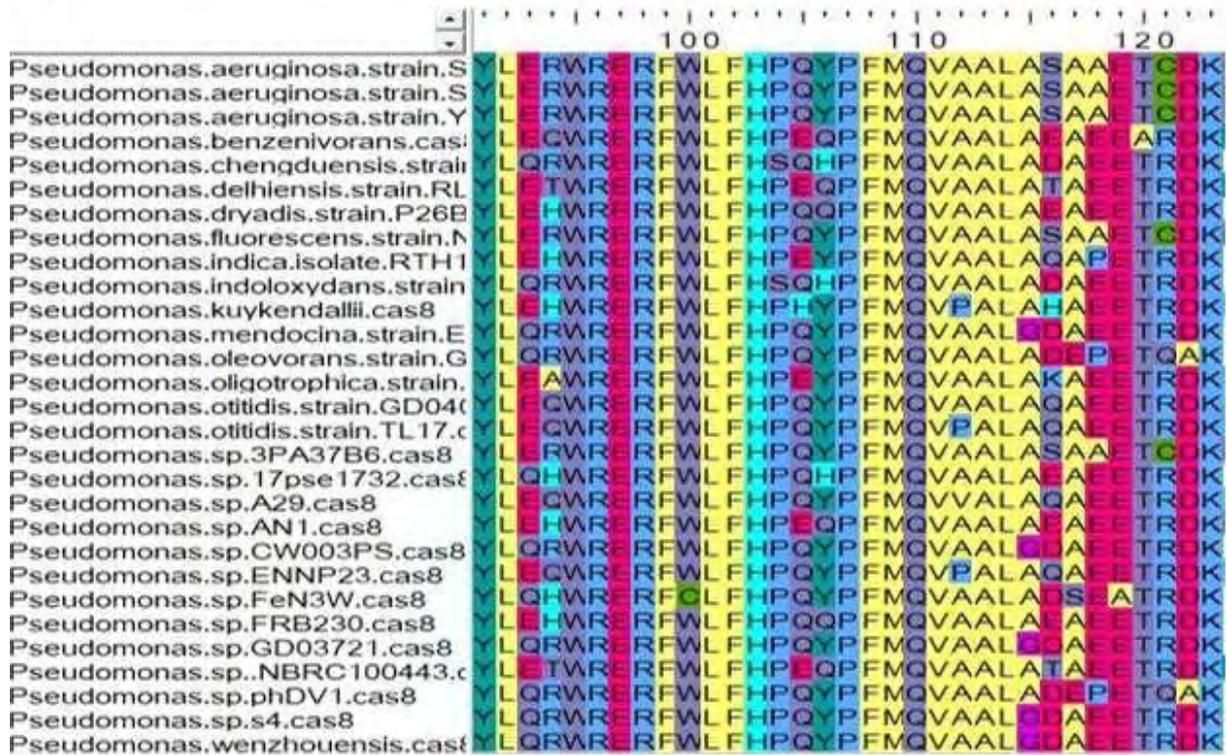
Figura 25.- Análisis de las secuencias del gen *cas8* en diferentes especies del género *Cryobacterium*. **a.** Alineamiento del gen *cas8* desde la posición 160 a la 220. **b.** Alineamiento de la proteína *Cas8* desde la posición 54 a la 90. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas8*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas8*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

Para la selección de los sitios de interés del gen *cas8* en el género *Pseudomonas*, se identificaron las regiones con valores bajos de *pi* (Figura 26c) y se observaron las regiones conservadas en el alineamiento de COBALT (Figura 26d). Por su parte, el alineamiento de las secuencias nucleotídicas puso en evidencia una alta variación y el gráfico de Alfa mostró sólo un sitio bajo selección negativa ubicado en el aminoácido 276 (Figura 26c).

**a**

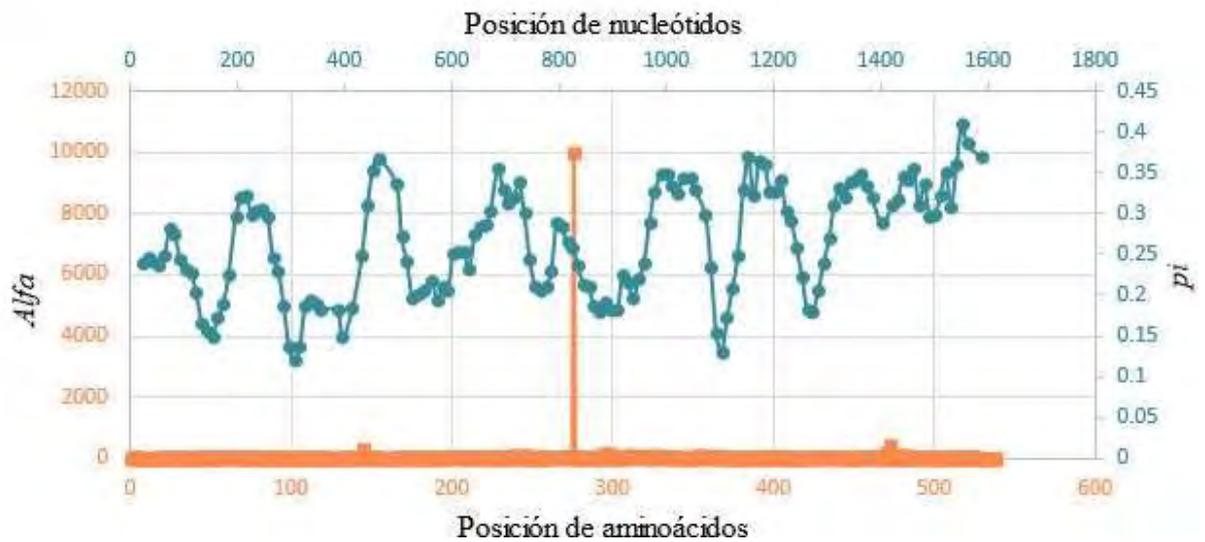


b



W c

e



**d**

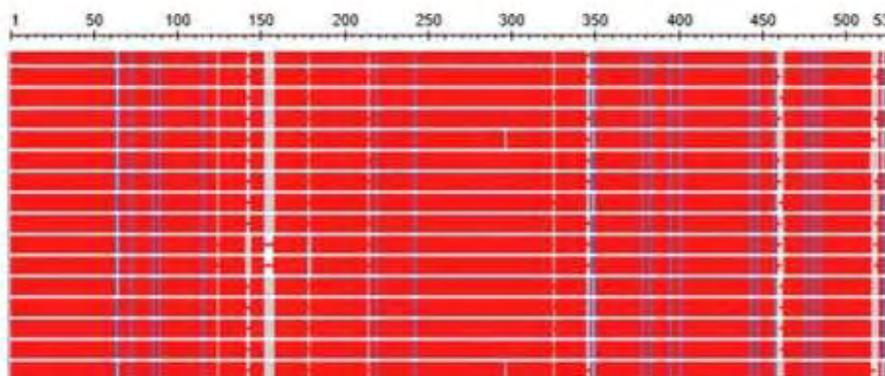
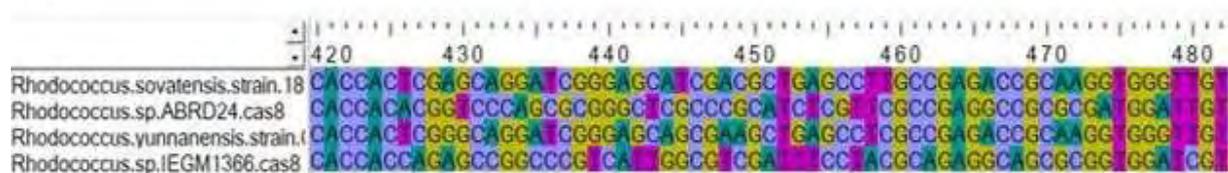


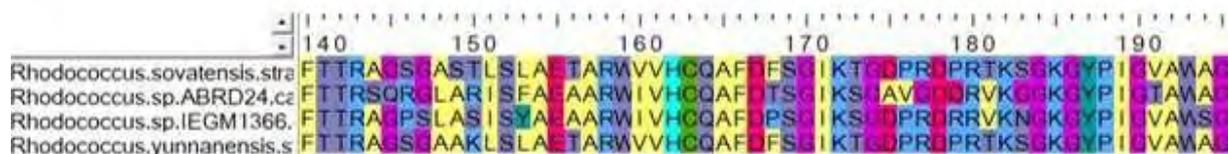
Figura 26.- Análisis de las secuencias del gen *cas8* en diferentes especies del género *Pseudomonas*. **a.** Alineamiento del gen *cas8* desde la posición 283 a la 342. **b.** Alineamiento de la proteína *Cas8* desde la posición 91 a la 119. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas8*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas8*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

Para el análisis del gen *cas8* en el género *Rhodococcus* se descartó la secuencia de la especie *Rhodococcus sp.1168*, debido a que tenía *gaps* excesivos y que el gen no estaba completo en la base de datos. Para identificar los sitios potenciales para el diseño de oligonucleótidos se buscaron las regiones de menor valor de *pi* (Figura 27c). Sin embargo, esta figura mostró mucha variación y sólo se encontró una posición potencial ubicada en el nucleótido 497. Se observaron dos sitios bajo selección negativa, en el aminoácido 7 y en el 461 (ver valores de Alfa en la Figura 27c). Los tres sitios mencionados y las regiones conservadas dadas por el alineamiento de COBALT (Figura 27d) fueron las candidatas para el diseño de oligonucleótidos.

**a**



**b**



c



d

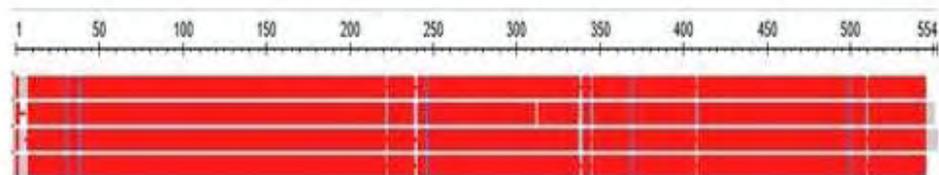


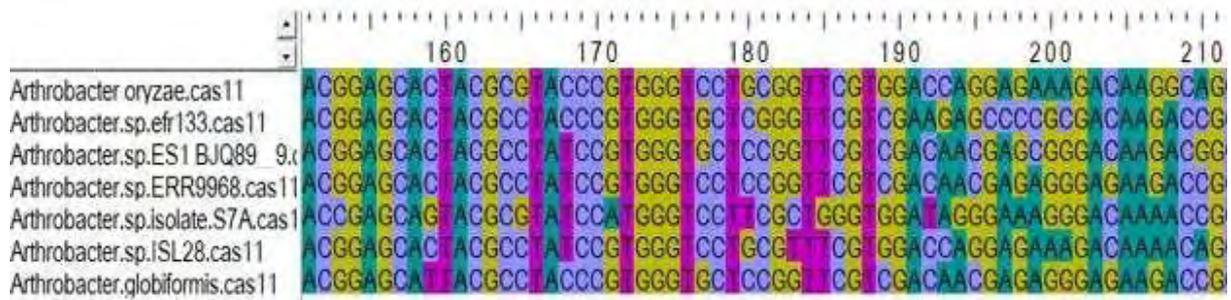
Figura 27.- Análisis de las secuencias del gen *cas8* en diferentes especies del género *Rhodococcus*. a. Alineamiento del gen *cas8* desde la posición 420 a la 488. b. Alineamiento de la proteína Cas8 desde la posición 140 a la 190. c. Gráfico de polimorfismo y gráfico de Alfa para el gen *cas8*, donde pi está representado por ● y Alfa está representado por ■. d. Alineamiento general de la proteína Cas8, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

### 7.2.7 Gen *cas11*

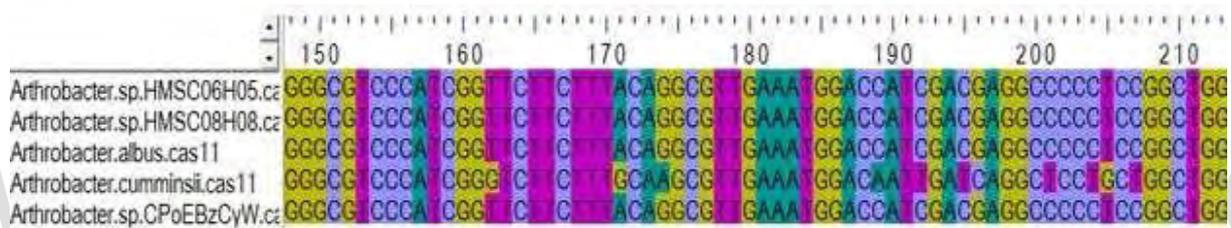
La proteína Cas11 es importante para la formación y estabilidad del complejo *Cascade*, es decir que cumple una función estructural en el sistema CRISPR-Cas (Amitai & Sorek, 2016). Al analizar las secuencias del gen *cas11* se observaron diferentes variaciones entre los géneros. Para el género *Arthrobacter* se encontraron muchas diferencias entre los grupos 1 y 2. Para el grupo 1, se identificaron varias regiones de interés proporcionadas por el alineamiento de COBALT (Figura 28d) y por el gráfico de Alfa que destacó los aminoácidos 92, 129, 137, 151 y 170 como sitios bajo selección negativa (Figura 28c.1). Por otro lado, el alineamiento del gen *cas11* en el grupo 2 indicó mucha similitud entre las secuencias (Figuras 28a.2 y 28b.2). El gráfico de polimorfismo reveló tres regiones de menor variabilidad alrededor de los nucleótidos 165, 305 y 465 (Figura 28c.2), mientras que los sitios bajo selección negativa se encontraron en la región inicial de la proteína

(Figura 28c.2). En este caso se seleccionaron aquellos sitios que coincidieron con regiones conservadas en el alineamiento de COBALT (Figura 28d).

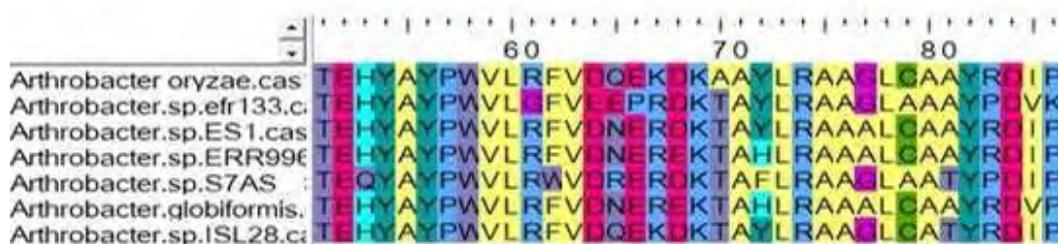
**a.1**



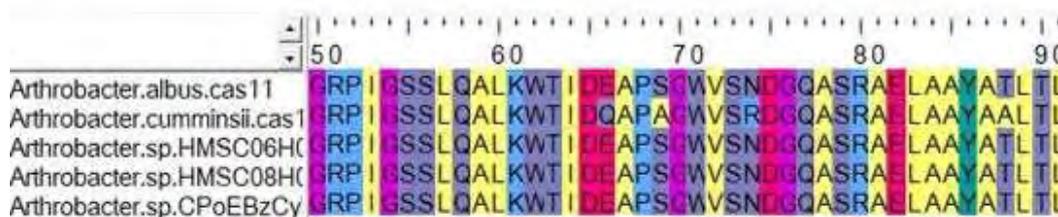
**a.2**



**b.1**



**b.2**



c.1



c.2



d

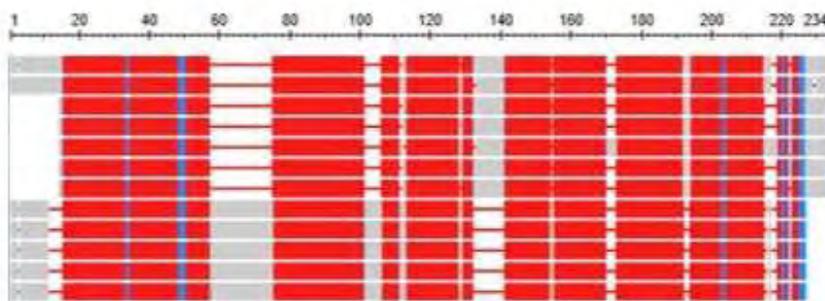
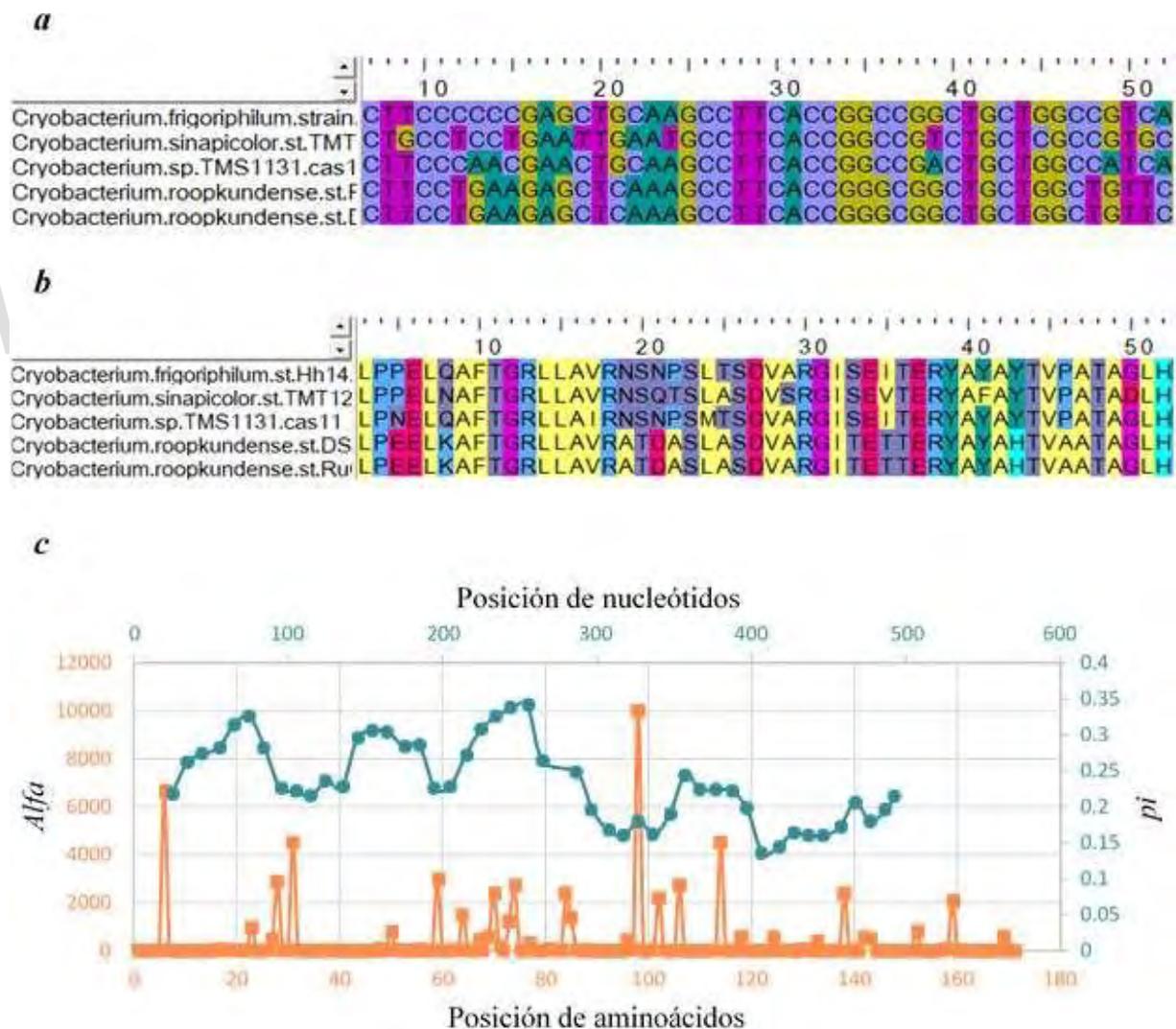


Figura 28.- Análisis de las secuencias del gen *cas11* en diferentes especies del género *Arthrobacter*. **a.1.** Alineamiento del gen *cas11* en el grupo 1 del género *Arthrobacter* desde la posición 151 a la 205. **a.2.** Alineamiento de la proteína *Cas11* de diferentes especies del grupo 2, desde la posición 51 a la 85. **b.1.** Alineamiento del gen *cas11* de diferentes especies del grupo 2 desde la posición 148 a la 210. **b.2.** Alineamiento de la proteína *Cas11* en el grupo 2, desde la posición 50 a la 100. **c.1.** y **c.2.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas11* en los grupos 1 y 2 del género *Arthrobacter*, respectivamente, donde  $\pi$  está representado por  $\bullet$  y Alfa está representado por  $\blacksquare$ . **d.** Alineamiento general de la proteína *Cas11* de diferentes especies del género *Arthrobacter*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT. Las primeras 7 secuencias corresponden al grupo 1 y las 5 secuencias restantes corresponden al grupo 2.

En el análisis de las secuencias del gen *cas11* para el género *Cryobacterium* se observaron dos regiones muy conservadas tanto en la secuencia nucleotídica como aminoacídica. La primera región ubicada entre los aminoácidos 3 y 18 (Figura 29a), y la segunda entre los aminoácidos 132 y 147 (no mostrada). Estas regiones coincidieron con sitios bajo selección negativa (Figura 29c) y se seleccionaron para el diseño de oligonucleótidos.



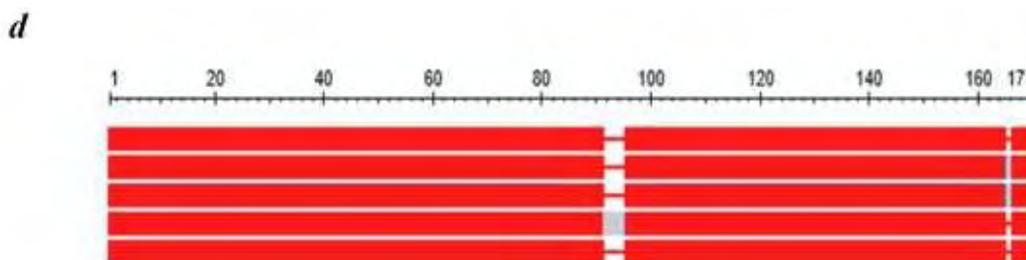


Figura 29.- Análisis de las secuencias del gen *cas11* en diferentes especies del género *Cryobacterium*. **a.** Alineamiento del gen *cas11* desde la posición 7 a la 50. **b.** Alineamiento de la proteína *Cas11* desde la posición 3 a la 50. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas11*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas11*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

Para el género *Pseudomonas* se detectó una alta variabilidad entre las secuencias del gen *cas11*, que fue evidenciado en altos valores de *pi* (Figura 30c). El gráfico de COBALT indicó varias regiones conservadas en la proteína, entre las cuales resaltan la región de aminoácidos 19 a 47 al inicio, y la región de aminoácidos 93 a 122, en el medio de la secuencia (Figura 30d). Varios de los sitios identificados en el gráfico de Alfa (Figura 30c) se encontraban entre las regiones determinadas por el alineamiento de COBALT (Figura 30d), por lo que los oligonucleótidos fueron diseñados sobre estas regiones.

www.bdigital.ula.ve



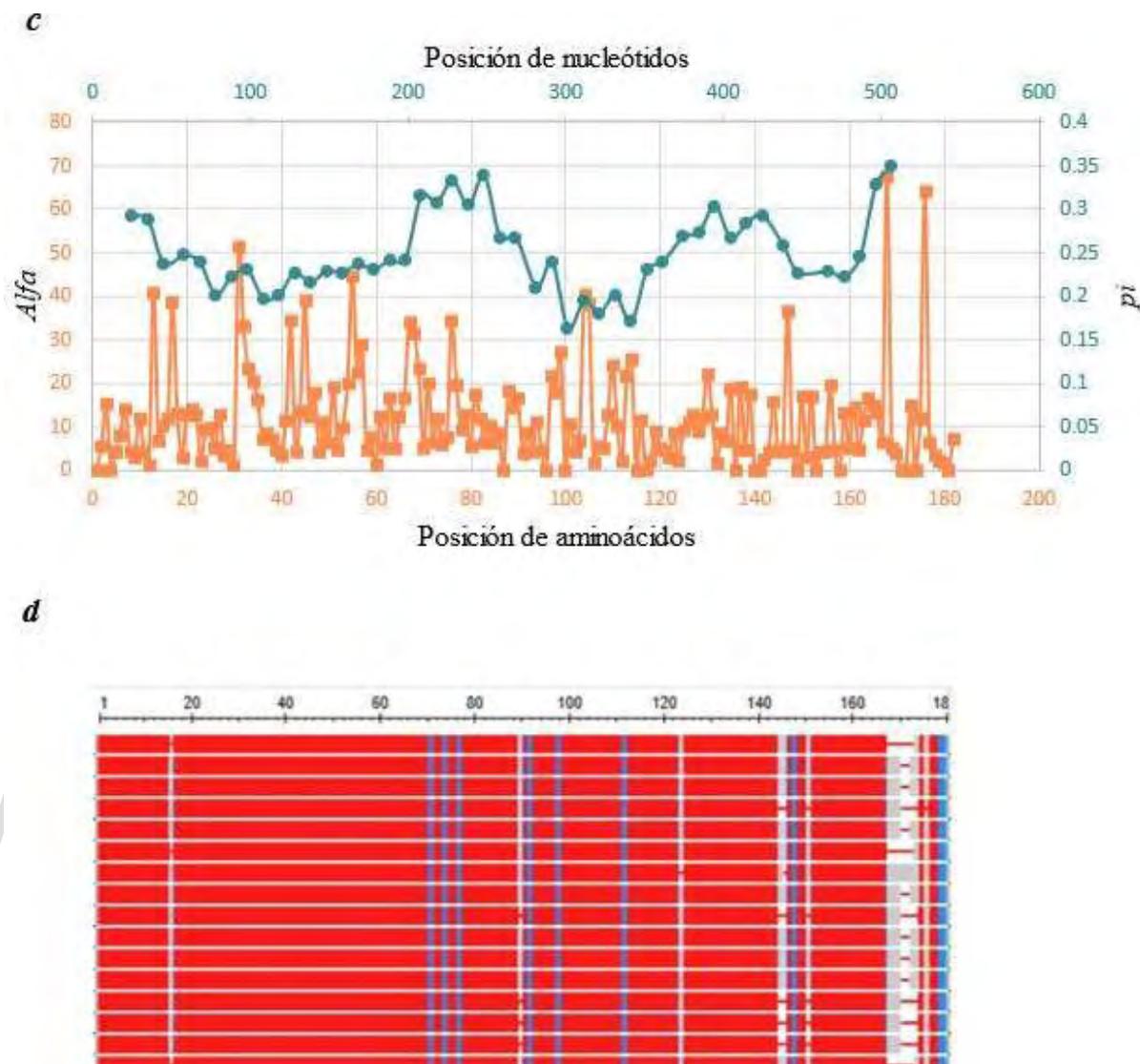


Figura 30.- Análisis de las secuencias del gen *cas11* en diferentes especies del género *Pseudomonas*. **a.** Alineamiento del gen *cas11* desde la posición 277 hasta la 351. **b.** Alineamiento de la proteína *Cas11* de la posición 88 a la 121. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas11*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas11*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

Los sitios de interés en el gen *cas11* para el género *Rhodococcus* fueron seleccionados en base a las regiones concordantes entre los sitios conservados de acuerdo al alineamiento de COBALT (Figura 31d) y varios sitios mostrados en el gráfico de Alfa que, a pesar de no tener un elevado valor de Alfa (Figura 31c), fueron regiones muy conservadas y convenientes para el diseño de oligonucleótidos.

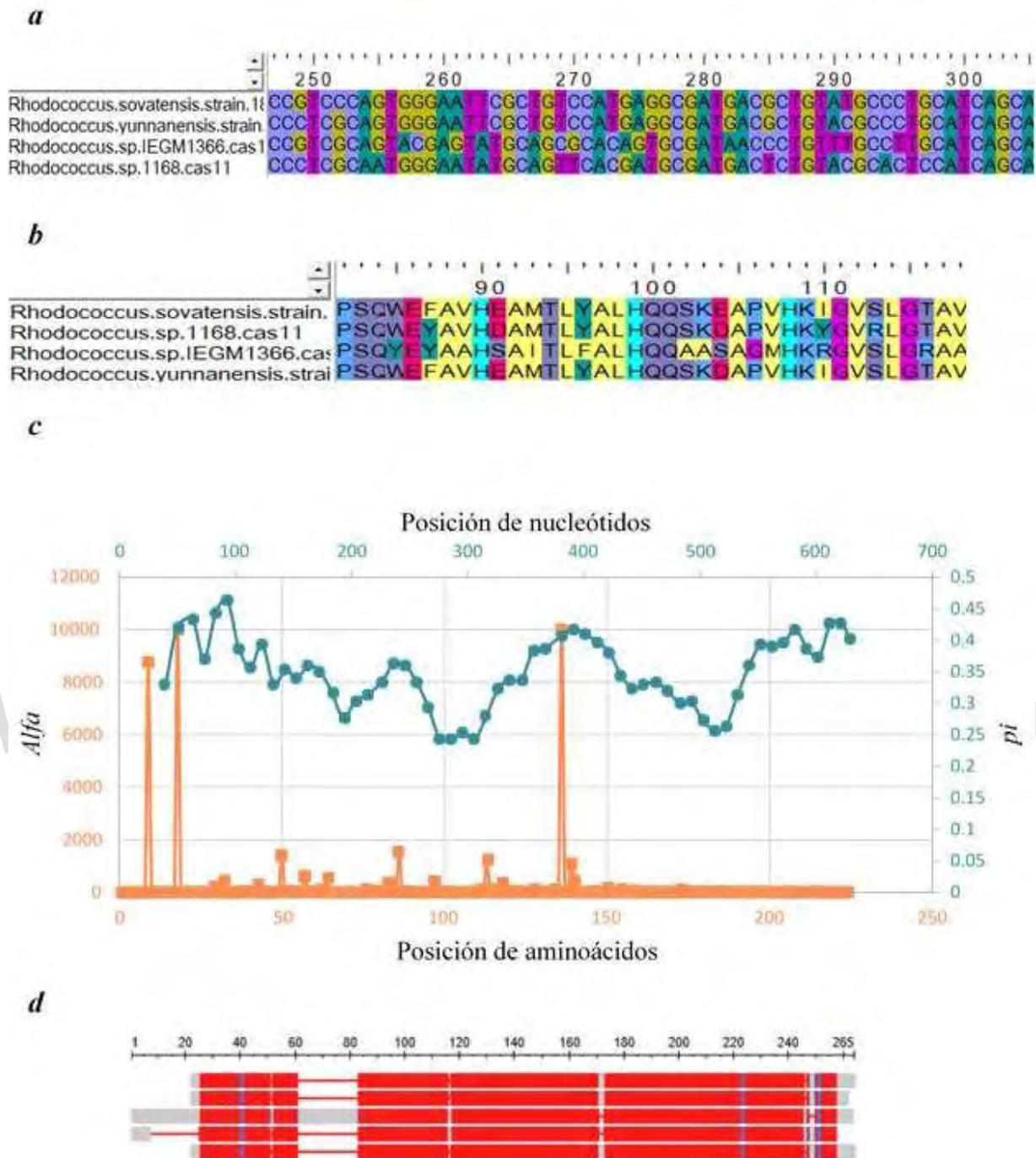


Figura 31.- Análisis de las secuencias del gen *cas11* en diferentes especies del género *Rhodococcus*. **a.** Alineamiento del gen *cas11* desde la posición 448 a la 305. **b.** Alineamiento de la proteína *Cas11* desde la posición 82 a la 118. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas11*, donde *pi* está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas11*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

## Sistema CRISPR-Cas clase II, subtipo II-1C

### 7.2.8 Gen *cas1*

De igual forma que en el sistema de la clase I, en la clase II, la proteína Cas1 está implicada en la incorporación del espaciador a la matriz CRISPR (Amitai & Sorek, 2016). Las secuencias nucleotídicas y aminoácidas del gen *cas1* en el género *Sphingomonas* son muy diversas, particularmente la secuencia nucleotídica, lo que fue reflejado en los altos valores de *pi* (Figura 32c). En cuanto a la proteína, se identificaron diversas regiones conservadas (Figura 32d), mientras que sólo se identificaron dos sitios bajo selección negativa, correspondientes a los aminoácidos 198 y 236 (Figura 32c). Para el diseño de oligonucleótidos se tomaron en cuenta las regiones señaladas en el alineamiento de COBALT, siendo las más destacadas desde los aminoácidos 40 al 60 y del 200 al 230.

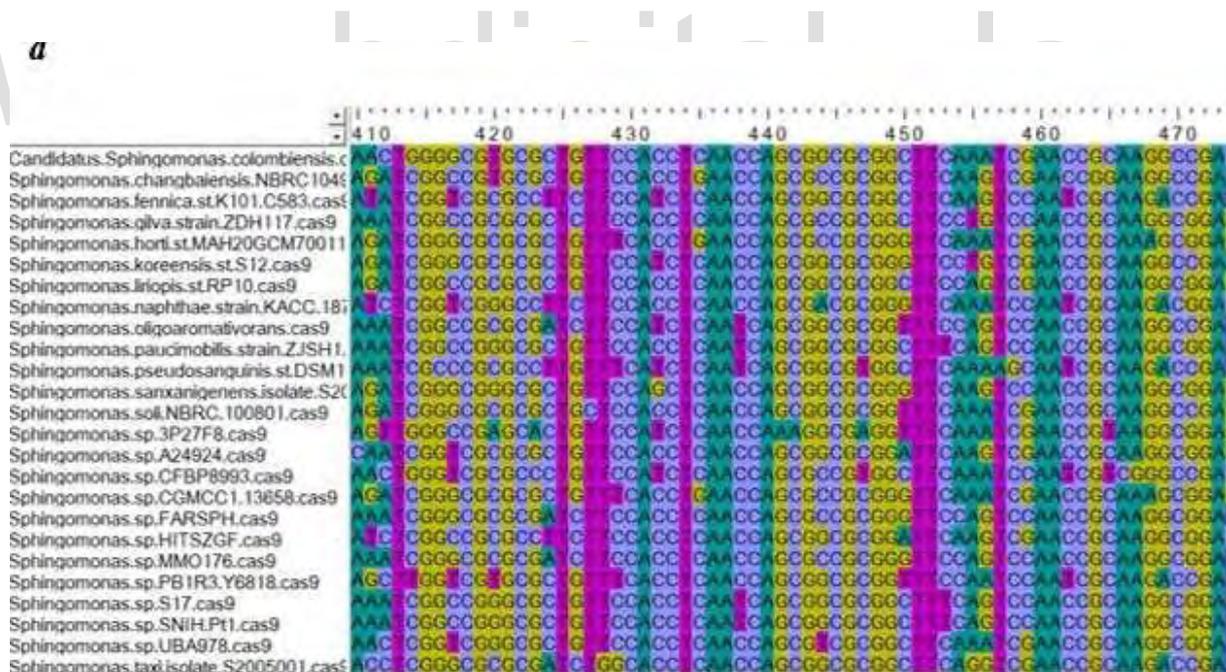
	630	640	650	660	670	680
<i>Sphingomonas</i> .sp.CGMCC.1.13658.cas1	GCGGCAACGCC	CGCGC	CGCCGACGACC	GGCGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .sp.FARSPH.cas1	GGGGCAA	CCC	CGCGC	CGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.MMO176.cas1	GCGGCAA	CCC	CGCGC	CGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .fennica.strain.K101C583	GCGGCAACCCC	CGCGC	CGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .horti.strain.MAH20.cas1	GCGGCAACGCC	CGCGC	CGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .sp.SRS2.cas1	GCGGCAACGCC	CGCGC	CGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .koreensis.strain.S12.cas1	GCGGCAA	GCGC	GGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sanxanigenens.isolate.ca	GGCTGAA	GCC	GCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .soli.NBRC.100801.cas1	GCGGCAA	GCC	CGCA	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .bacterium.cas1	GCGGCAACCGCC	GGCGC	CGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .sp.isolate.SMAGU3283.c	GCGGCAA	CGGC	GGCGC	CGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .changbaiensis.NBRC.ca	GCGGCAA	GCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .gliva.strain.ZDH117.cas1	GCGGCAACGCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .naphthae.strain.KACC18	GCGGCAACGCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .liriopsis.strain.RP10.cas1	GCGGCAACGCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .oligoparomativorans.cas1	GCGGCAA	CCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
Candidatus. <i>Sphingomonas</i> .colombiensis	GGACCAA	GCC	GCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.3P27F8.cas1	GGCTTAA	CGCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.A24924.cas1	GCGGCAACGCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG
<i>Sphingomonas</i> .sp.CFBP8993.cas1	GAGCAA	CCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .paucimobilis.strain.ZJSH	GCGGCAA	CCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.HITSZGF.cas1	GCGGCAA	GCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .pseudosanguinis.strain.D	GCGGCAA	CCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.PB1R3.Y6818.cas1	GAGCAA	CCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.S17.cas1	GCGGCAA	GCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.SNIH.Pt1.0416.cas1	GCGGCAA	CCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .sp.UBA978.cas1	GAGCAA	CCC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC
<i>Sphingomonas</i> .taxi.isolate.S2.cas1	GCGGCAACCGC	CGCGC	GGCCGACGACC	GAGGAGCCG	CGGCCCGC	GG



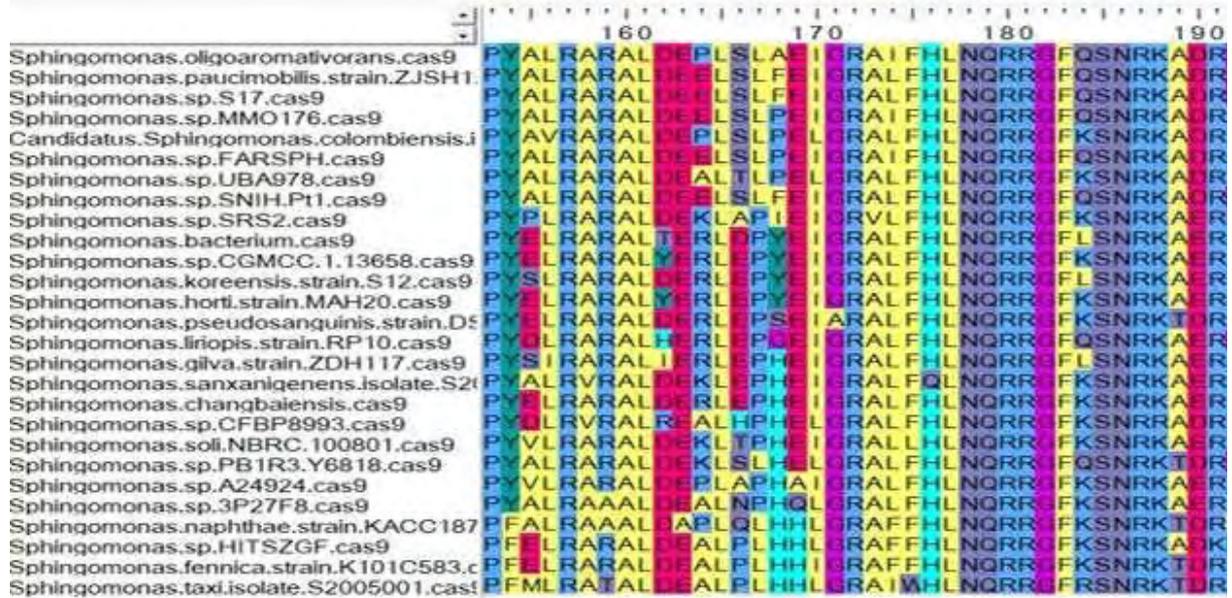
Figura 32.- Análisis de las secuencias del gen *cas1* en diferentes especies del género *Sphingomonas*. **a.** Alineamiento del gen *cas1* desde la posición 630 a la 680. **b.** Alineamiento de la proteína *Cas1* desde la posición 181 a la 230. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas1*, donde  $\pi$  está representado por  $\bullet$  y Alfa está representado por  $\blacksquare$ . **d.** Alineamiento general de la proteína *Cas1*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

### 7.2.9 Gen *cas9*

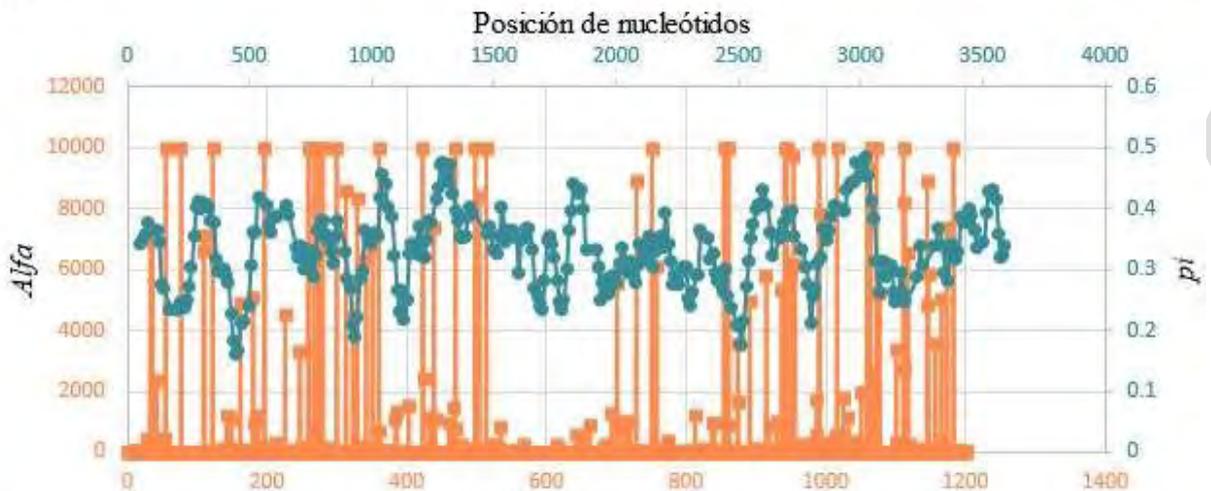
La proteína *Cas9* está implicada en diferentes funciones dentro del mecanismo de defensa CRISPR-Cas de la clase II, siendo la proteína efectora, por lo que es de esperarse que entre las especies de un género existan regiones muy conservadas. Sin embargo, este no fue el caso para el género *Sphingomonas*, donde se observó una alta variabilidad de las secuencias nucleotídicas y aminoacídicas, con la presencia de muchos *gaps*. A pesar de la alta diversidad, se lograron identificar varias regiones cortas, muy conservadas, que fueron candidatas para el diseño de oligonucleótidos. El gráfico de polimorfismo mostró dos regiones de valores bajos de  $\pi$  en los nucleótidos 500 y 2511 (Figura 33c). El alineamiento de COBALT mostró una baja similitud entre las secuencias, pero señaló varias regiones conservadas, idóneas para el diseño de oligonucleótidos (Figura 33d). Finalmente, el gráfico de Alfa mostró múltiples sitios de selección negativa (Figura 33c) algunos de los cuales también fueron considerados para el diseño de oligos para ese gen.



**b**



**c**



**d**

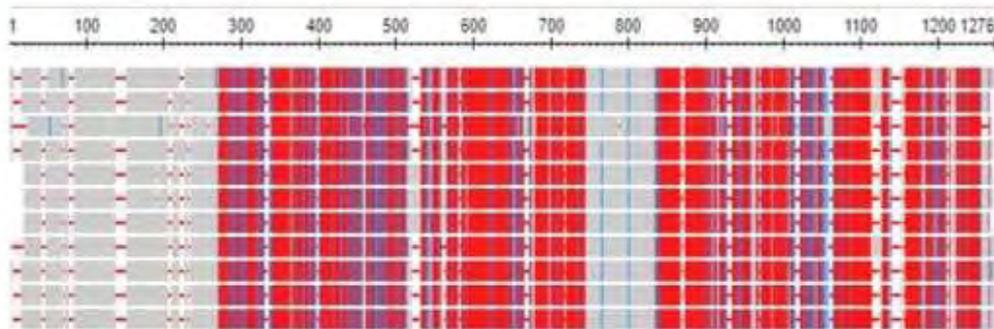


Figura 33.- Análisis de las secuencias del gen *cas9* en diferentes especies del género *Sphingomonas*. **a.** Alineamiento del gen *cas9* desde la posición 410 a la 470. **b.** Alineamiento de la proteína *Cas9* desde la posición 153 a la 191. **c.** Gráfico de polimorfismo y gráfico de Alfa para el gen *cas9*, donde pi está representado por ● y Alfa está representado por ■. **d.** Alineamiento general de la proteína *Cas9*, resaltando las regiones conservadas en color rojo y las regiones diferentes en gris y azul, realizado en COBALT.

### 7.3 Diseño y prueba de oligonucleótidos

Las regiones seleccionadas en los análisis previos se sometieron a evaluación en el programa Oligoexplorer para el diseño de los oligonucleótidos, teniendo en cuenta que la temperatura de fusión (*melting temperature*) debía ser menor a 65 °C y similar entre ambos oligos con una máximo de 3 °C de diferencia, el porcentaje de guaninas y citocinas (%GC) debía ser de 68% máximo, la probabilidad de alineamiento complementario debía ser minimizada y los oligos debían tener una longitud entre 18 y 24 pares de bases. Las regiones que tuvieran los valores adecuados de los parámetros se sometieron a diferentes pruebas en el programa de Primer-BLAST, utilizando las bases de datos nr (*no redundant*) y RefSeq (Genomas representativos optimizados para secuencias altamente similares), con el fin de obtener la mayor cantidad de resultados posibles. En aquellos casos donde los oligos variaban en una o dos bases se agregaban bases degeneradas para poder identificar el mayor número de especies (Tabla 1).

Tabla 1.- Código IUPAC de nucleótidos para generar bases degeneradas.

Símbolo	Bases
G	Guanina
A	Adenina
T	Timina
C	Citosina
R	G / A
Y	T / C
M	A / C
K	G / T
S	G / C
W	A / T
H	A / C / T
B	G / T / C
V	G / C / A
D	G / A / T
N	G / A / T / C

## Sistema CRISPR-Cas clase I, subtipo I-E

### 7.3.1 *Gen cas1*

Se diseñaron oligonucleótidos para la identificación del gen *cas1* en especies del género *Rhodococcus* mediante el programa Oligoexplorer. Los siguientes oligos, sentido o *forward* GTGCTCTTCTCCTCGGCC y TGCTCCCGATAGCTCTTCCG como antisentido o *reverse*, fueron los que mostraron mejores valores de los parámetros evaluados con 63,8 °C para el *forward* y 62,5 °C para el *reverse*. Al probar los oligonucleótidos en Primer-BLAST se detectó el gen *cas1* únicamente en la especie *Rhodococcus sovatusensis* con un amplicón de 307 pares de bases. Los valores de cada uno de los parámetros se encuentran en la Tabla A de Anexos.

### 7.3.2 *Gen cas5*

Diversos oligonucleótidos fueron diseñados para la identificación del gen *cas5* en los géneros bacterianos *Cryobacterium*, *Pseudomonas* y *Rhodococcus* (Tabla 2), ya que para el género *Arthrobacter* en el grupo 1 no se logró el diseño de oligonucleótidos que cumplieran con los requerimientos necesarios y para el grupo 2, se realizaron diferentes pruebas en Primer-BLAST que detectaban una sola especie. Para los demás géneros, se procedió a diseñar oligos degenerados que permitan la identificación del gen en el mayor número posible de especies. Todos los oligos diseñados tuvieron valores adecuados de temperatura, formación de dímeros y porcentaje de GC (mostrados en Tabla B de los Anexos).

Tabla 2.- Oligonucleótidos diseñados para la identificación del gen *cas5* en los géneros *Cryobacterium*, *Pseudomonas* y *Rhodococcus*.

<b>Género <i>Cryobacterium</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
TCTGGGCSAGAGTCGACC	TGCGACCCARRTACGTCATG	309
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
60,7	58,4	- <i>C.roopkundense</i> - <i>C.frigoriphilum</i> - <i>C.sinapicolor</i>
<b>Género <i>Pseudomonas</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
CARTCCTGGGGCACCACC	GTGGCGGTCTGGTARTCGC	194
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>

60,7	63,8	<ul style="list-style-type: none"> <li>- <i>P. sp. FeN3W</i></li> <li>- <i>P. oligotrophica</i></li> <li>- <i>P. sp. phDV1</i></li> <li>- <i>P. sp. CW003PS</i></li> <li>- <i>P. oleovorans</i></li> <li>- <i>P. sp. GD03919</i></li> <li>- <i>P. sp. GD03721</i></li> <li>- <i>P. pseudoalcaligenes</i></li> <li>- <i>P. dryadis</i></li> <li>- <i>P. otitidis</i></li> <li>- <i>P. delhiensis</i></li> <li>- <i>P. indoloxydans</i></li> <li>- <i>Aeromonas media strain T5-1</i></li> <li>- <i>Aeromonas caviae KAM339</i></li> <li>- <i>Azotobacter sp. NL3</i></li> <li>- <i>Azotobacter vinelandii DJ</i></li> <li>- <i>Chromobacterium sp. Beijing</i></li> <li>- <i>Ideonella dechloratans</i></li> <li>- <i>Chromobacterium rhizoryzae</i></li> <li>- <i>Aromatoleum diolicum</i></li> <li>- <i>Aeromonas diversa</i></li> </ul>
<b>Género <i>Rhodococcus</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
ACGAAGAGCGGVGTGCTCG	CCBAGGTAGAGCGGGAAGKT	287
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
61,6	62,5	<ul style="list-style-type: none"> <li>- <i>Rhodococcus sovatisensis</i></li> <li>- <i>Saccharopolyspora rectivirgula</i></li> <li>- <i>Nocardioides soli</i></li> <li>- <i>Propionibacterium cyclohexanicum</i></li> </ul>

### 7.3.3 Gen *cas6*

Para el gen *cas6* no fue posible diseñar oligos específicos que cumplieran con las condiciones idóneas de parámetros en todos los géneros (ver primer párrafo de la sección 7.3). En consecuencia, se diseñaron oligos que cumplieron con los valores adecuados de los parámetros (Tabla C en Anexos) para identificar una región intermedia entre los genes *cas5* y *cas6* para el género *Pseudomonas* (Tabla 3), lo que permitiría evaluar no sólo la presencia de dos genes al mismo tiempo, sino también el conservatismo de la sintenia de estos dos genes. Los análisis de Primer-BLAST lograron identificar la región intermedia *cas5-cas6* en 17 especies incluso en otros taxones diferentes de *Pseudomonas*.

Tabla 3.- Oligonucleótidos diseñados para la identificación de la región intermedia entre los genes *cas5* y *cas6* en el género *Pseudomonas*.

<b>Género <i>Pseudomonas</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
ACCTGACCAGACTGACSCCTTGAC	CTCGCCCACCAGVCCATAG	365
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
66,4	63,8	<ul style="list-style-type: none"> <li>- <i>P. aeruginosa</i></li> <li>- <i>P. fluorescens</i></li> <li>- <i>P. sp. AK6U</i></li> <li>- <i>P. oleovorans</i></li> <li>- <i>P. dryadis</i></li> <li>- <i>P. indoloxydans</i></li> <li>- <i>P. mendocina</i></li> <li>- <i>P. sp. GD03919</i></li> <li>- <i>P. sp. GD03721</i></li> <li>- <i>P. sp. PhDV1</i></li> <li>- <i>P. sp. FeN3W</i></li> <li>- <i>P. sp. B111</i></li> <li>- <i>P. sp. CW003PS</i></li> <li>- <i>Azotobacter spp.</i></li> <li>- <i>Microbulbifer celer</i></li> </ul>

#### 7.3.4 *Gen cas7*

El análisis de las diferentes regiones conservadas en el gen *cas7* permitió el diseño de oligonucleótidos tanto generales como específicos para diversos géneros bacterianos, además todos cumplieron los parámetros requeridos para unos oligos efectivos (Tabla D en Anexos), facilitando la identificación del gen *cas7* en distintos taxones. Para el género *Arthrobacter*, los oligonucleótidos se diseñaron basándose en las regiones conservadas del grupo 1. Sin embargo, las pruebas realizadas con Primer-BLAST no lograron identificar el gen *cas7* en especies de *Arthrobacter*, aunque permitieron su detección en otras especies, como se muestra en la Tabla 4. Por consiguiente, se realizaron pruebas para evaluar la especificidad de estos oligonucleótidos haciendo uso de la herramienta *CAP contig assemblt program* de Bioedit, con el fin de ensamblar las secuencias junto con los oligos para corroborar que amplifique en las secuencias estudiadas. Para ello se probaron los oligonucleótidos diseñados específicamente para el grupo 1, que no pudieron detectar las especies mediante Primer-BLAST, obteniendo la amplificación de todas las especies con diferencias en algunas bases (Figura 34).

Tabla 4.- Oligonucleótidos diseñados para la identificación del gen *cas7* para los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y oligonucleótidos multigénero.

<b>Género <i>Arthrobacter</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
CCTCAACCGKGACGACACCR	GCGGTGGMGTTCTTCTTGS	822
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
63,34	62,6	- <i>Actinomadura spp.</i> - <i>Thermomonospora spp.</i> - <i>Streptomyces spp.</i> - <i>Catenulispora pinistramenti</i>
<b>Género <i>Cryobacterium</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
GCCCTGGCACAACCTSAAYC	GAGCGGTAGTAGACGCCYGACG	639
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
62,5	67,7	- <i>C.roopkundense</i> - <i>C.frigoriphilum</i> - <i>C.sinapicolor</i>
<b>Género <i>Pseudomonas</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
SCTCAAGCTCAAGGACGAYG	CGGARTTGAACTCCACCTGG	404
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>

60,5	60,5	<ul style="list-style-type: none"> <li>- <i>P. aeruginosa</i></li> <li>- <i>P. sp. B111</i></li> <li>- <i>P. sp. AK6U</i></li> <li>- <i>P. benzenivorans</i></li> <li>- <i>P. sp. FeN3W</i></li> <li>- <i>P. sp. phDV1</i></li> <li>- <i>P. sp. CW003PS</i></li> <li>- <i>P. sp. GD03919</i></li> <li>- <i>P. sp. GD03721</i></li> <li>- <i>P. otitidis</i></li> <li>- <i>P. oleovorans</i></li> <li>- <i>P. dryadis</i></li> <li>- <i>P. oligotrophica</i></li> <li>- <i>P. indoloxydans</i></li> <li>- <i>P. delhiensis</i></li> <li>- <i>Ideonella spp.</i></li> <li>- <i>Azotobacter spp.</i></li> <li>- <i>Stutzerimonas spp.</i></li> <li>- <i>Chromobacterium spp.</i></li> <li>- <i>Aromatoleum anaerobium</i></li> <li>- <i>Allochromatium palmeri</i></li> </ul>
<b>Género <i>Rhodococcus</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
GAGCAATMTCAACCGTGACG	CAGGGTGGAKGAGTTGAACT	655
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>

58,4	60,5	<ul style="list-style-type: none"> <li>- <i>R. sovatisensis</i></li> <li>- <i>R. sp. ABRD24</i></li> <li>- <i>Crossiella spp.</i></li> <li>- <i>Actinokineospora spp.</i></li> <li>- <i>Georgenia sp. TF02-10</i></li> <li>- <i>Amycolatopsis spp.</i></li> <li>- <i>Nocardia spp</i></li> <li>- <i>Actinobacteria bacterium</i></li> <li>- <i>Tessaracoccus flavescens</i></li> <li>- <i>Actinoalloteichus spp.</i></li> <li>- <i>Actinoplanes spp.</i></li> <li>- <i>Bifidobacterium spp.</i></li> <li>- <i>Compostimonas suwonensis</i></li> <li>- <i>Austwickia chelonae</i></li> <li>- <i>Phytoactinopolyspora halophila</i></li> <li>- <i>Specibacter cremeus</i></li> <li>- <i>Demequina lutea</i></li> <li>- <i>Haloactinomyces albus</i></li> <li>- <i>Myceligenans pegani</i></li> <li>- <i>Saccharothrix spp.</i></li> <li>- <i>Umezawaea beigongshangensis</i></li> <li>- <i>Streptomonospora spp.</i></li> <li>- <i>Micromonospora spp.</i></li> <li>- <i>Xylanimonas allomyrinae</i></li> <li>- <i>Prauserella marina</i></li> <li>- <i>Nakamurella multipartita</i></li> <li>- <i>Nocardioides acrostichi</i></li> <li>- <i>Nocardiopsis ansamitocini</i></li> <li>- <i>Streptomyces hainanensis</i></li> <li>- <i>Actinomadura hallensis</i></li> </ul>
<b>Género Multigénero</b>		
<b>Forward</b>	<b>Reverse</b>	<b>Tamaño del amplificado</b>
ACCTCAACCGNGACGACACC	TCGACGGCGGTGWAGTAGTC	586
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>

62,5	62,5	<ul style="list-style-type: none"> <li>- <i>Pseudomonadales</i></li> <li>- <i>Micrococcales</i></li> <li>- <i>Micromonosporales</i></li> <li>- <i>Kitasatosporales</i></li> <li>- <i>Pseudonocardiales</i></li> <li>- <i>Catenulisporales</i></li> <li>- <i>Streptosporangiales</i></li> <li>- <i>Rhodocyclales</i></li> <li>- <i>Nakamurellales</i></li> <li>- <i>Propionibacteriales</i></li> <li>- <i>Aeromonadales</i></li> <li>- <i>Alteromonadales</i></li> <li>- <i>Actinomycetales</i></li> <li>- <i>Rhodospirillales</i></li> <li>- <i>Neisseriales</i></li> <li>- <i>Hyphomicrobiales</i></li> <li>- <i>Burkholderiales</i></li> </ul>
------	------	--

Para el género *Cryobacterium*, se diseñaron oligonucleótidos específicos que, al ser probados con Primer-BLAST, demostraron alta especificidad para este género (Tabla 4). En el caso de *Pseudomonas*, se diseñaron tres pares de oligos, seleccionando finalmente aquel que permitió la identificación del gen en el mayor número de especies, como se detalla en la Tabla 3. Así mismo, para el género *Rhodococcus* se generaron diversos oligonucleótidos y se seleccionó el más efectivo en la detección del gen en el mayor número de especies, también mostrado en la Tabla 4.

El par de oligonucleótidos diseñado en base a todos los géneros estudiados permitió la identificación del gen *cas7* en diversas especies pertenecientes a 17 órdenes bacterianos, demostrando así la versatilidad y eficiencia de estos oligos en la detección de este gen en una amplia gama de bacterias y determinando su uso potencial en la caracterización del sistema CRISPR-Cas en muestras ambientales, mediante el uso de la técnica de *metabarcoding*. Para verificar que los oligos también logran identificar las especies estudiadas en el género *Arthrobacter* se procedió a analizar su eficiencia mediante *CAP contig assemblt program* de Bioedit, tomando en cuenta las bases degeneradas específicas para cada grupo del género *Arthrobacter* y obteniendo la amplificación *in silico* en todas las especies de ambos grupos (Figuras 35 y 36).

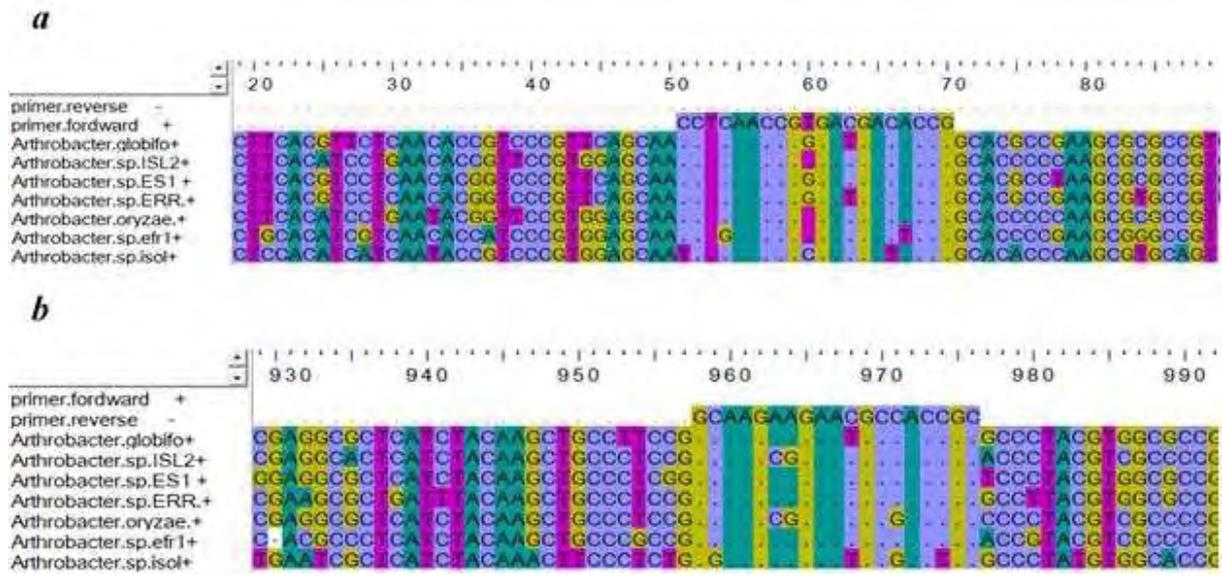


Figura 34.- Resultados de la identificación del gen *cas7* en especies del grupo 1 del género *Arthrobacter* haciendo uso de oligos específicos para el grupo. **a.** Unión del oligo forward a las secuencias de diferentes especies de *Arthrobacter* del grupo 1. **b.** Unión del oligo reverse a las secuencias de diferentes especies de *Arthrobacter* del grupo 1.

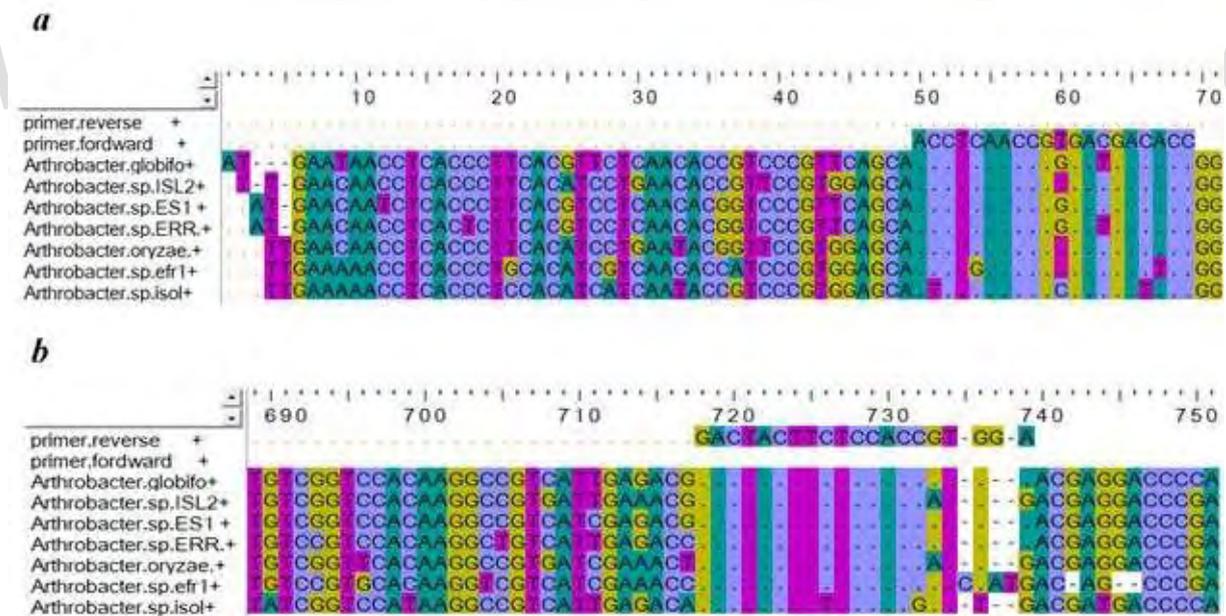


Figura 35.- Resultados de la identificación del gen *cas7* en especies del grupo 1 del género *Arthrobacter* haciendo uso de oligos multigénero. **a.** Unión del oligo forward a las secuencias de diferentes especies de *Arthrobacter* del grupo 1. **b.** Unión del oligo reverse a las secuencias de diferentes especies de *Arthrobacter* del grupo 1.

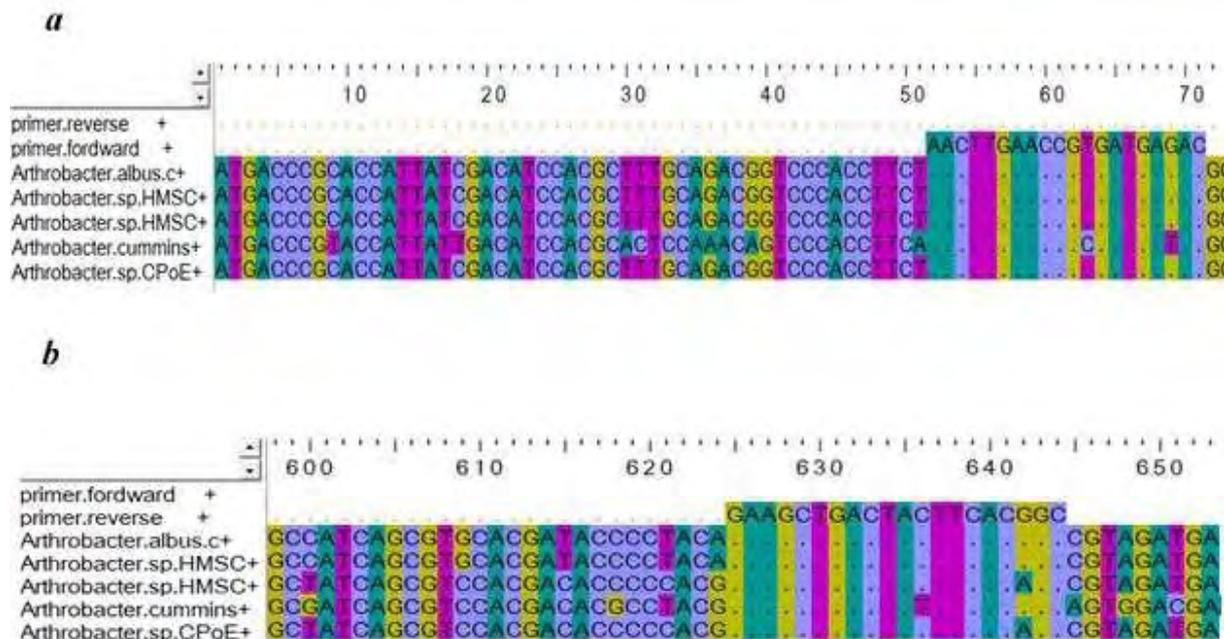


Figura 36.- Resultados de la identificación del gen *cas7* en especies del grupo 2 del género *Arthrobacter* haciendo uso de oligos multigénero. **a.** Unión del oligo forward a las secuencias de diferentes especies de *Arthrobacter* del grupo 2. **b.** Unión del oligo reverse a las secuencias de diferentes especies de *Arthrobacter* del grupo 2.

### 7.3.5 Gen *cas8*

Al analizar la proteína codificada por el gen *cas8* se obtuvieron regiones candidatas para el diseño de oligonucleótidos. Sin embargo, en tal análisis se evidenció una alta diversidad nucleotídica aparentemente debida a una alta tasa de mutaciones sinónimas, lo que probablemente impidió el diseño de oligos para los géneros *Arthrobacter*, *Cryobacterium* y *Rhodococcus*. No obstante, para el género *Pseudomonas* sí fue posible el diseño de oligonucleótidos, que tenían valores adecuados de los parámetros requeridos (Tabla E en Anexos) y que logró detectar *in silico* el gen *cas8* en 21 especies, tanto del género *Pseudomonas* como de otros taxones (Tabla 5).

Tabla 5.- Oligonucleótidos diseñados para la identificación del gen *cas8* en el género *Pseudomonas*.

<b>Género <i>Pseudomonas</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
CGSTTCTGGCTGTTCCATCC	CAGCAGCTTGGCCTGRTCRCCT	774
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
62,5	63,3	<ul style="list-style-type: none"> <li>- <i>P. sp. phDVI</i></li> <li>- <i>P. oleovorans</i></li> <li>- <i>P. benzenivorans</i></li> <li>- <i>P. mendocina</i></li> <li>- <i>P. otitidis</i></li> <li>- <i>P. sp. GD03919</i></li> <li>- <i>P. sp. GD03721</i></li> <li>- <i>P. sp. CW003PS</i></li> <li>- <i>P. indoloxydans</i></li> <li>- <i>P. delhiensis</i></li> <li>- <i>P. dryadis</i></li> <li>- <i>Azotobacter sp. NL3</i></li> <li>- <i>Azotobacter vinelandii DJ</i></li> <li>- <i>Azotobacter chroococcum</i></li> <li>- <i>Azotobacter salinestris</i></li> <li>- <i>Aquabacterium lacunae</i></li> <li>- <i>Azotobacter beijerinckii</i></li> <li>- <i>Aromatoleum anaerobium</i></li> <li>- <i>Stutzerimonas stutzeri</i></li> <li>- <i>Thiocapsa bogorovii</i></li> <li>- <i>Chitiniphilus shinanonensis</i></li> </ul>

### 7.3.6 Gen *cas11*

Se diseñaron oligonucleótidos para la identificación del gen *cas11* sólo para los géneros *Cryobacterium* y *Pseudomonas* (Tabla 6), que cumplieron los parámetros adecuados (Tabla F en Anexos) mientras que para los otros géneros éstos no se cumplieron, especialmente la temperatura de fusión, que en la mayoría de los casos superaba los 67°C. Además, sólo amplificaban una sola especie, por lo que éstos fueron descartados del estudio

Tabla 6.- Oligonucleótidos diseñados para la identificación del gen *cas11* para los géneros *Cryobacterium* y *Pseudomonas*.

<b>Género <i>Cryobacterium</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
CCGAGCTGCAAGCCTTCAC	CCGGTGTCCCAGAACGTGAG	400
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
64,6	64,6	- <i>C.roopkundense</i> - <i>C.frigoriphilum</i>
<b>Género <i>Pseudomonas</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
CCGGCYTATCCSTATGTGGA	ACACCGMCAGGTTTTCCG	226
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
60,54	60,59	- <i>P. aeruginosa</i> - <i>P. sp. B111</i> - <i>P. fluorescens</i> - <i>P. sp. AK6U</i> - <i>P. sp. CW003PS</i> - <i>P. sp. GD03919</i> - <i>P. sp. GD03721</i> - <i>P. sp. phDV1</i> - <i>P. oleovorans</i> - <i>P. sp. FeN3W</i>

### Sistema CRISPR-Cas clase II, subtipo II-1C

#### 7.3.7 Gen *cas1*

Se realizaron diferentes pares de oligonucleótidos basados principalmente en las regiones señaladas por COBALT, para la identificación del gen *cas1* en especies del género *Sphingomonas*. Se seleccionaron aquellos que cumplieron todos los parámetros necesarios para garantizar el funcionamiento de las PCR (Tabla G en Anexos). Los resultados de Primer-BLAST arrojaron una alta especificidad y lograron la identificación del gen en diferentes especies (Tabla 7).

Tabla 7.- Oligonucleótidos diseñados para la identificación del gen *cas1* en el género *Sphingomonas*.

<b>Género <i>Sphingomonas</i></b>		
<b><i>Forward</i></b>	<b><i>Reverse</i></b>	<b>Tamaño del amplificado</b>
CSCACGGCRTCACCTGGTC	SGCGGAACGGYTCGATCAVA	534
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
66	64,6	<ul style="list-style-type: none"> <li>- <i>S. sp. FARSPH</i></li> <li>- <i>S. naphthae</i></li> <li>- <i>S. paucimobilis</i></li> <li>- <i>S. fennica</i></li> <li>- <i>S. pseudosanguinis</i></li> <li>- <i>S. oligoaromativorans</i></li> <li>- <i>Sphingobium limneticum s</i></li> <li>- <i>Microvirga flavescens</i></li> <li>- <i>Sphingobium jiangsuense</i></li> <li>- <i>Sphingobium psychrophilum</i></li> </ul>

#### 7.2.8. *Gen cas9*

Se diseñaron dos pares de oligos para la identificación del gen *cas9* en especies del género *Sphingomonas*, utilizando dos regiones destacadas en el alineamiento de COBALT. Ambos oligos cumplieron con los parámetros adecuados (Tabla H en Anexos) y fueron puestos a prueba en Primer-Blast. Aunque la eficiencia potencial de la PCR fue buena para ambos pares de oligos, se seleccionó el par que logró identificar este gen en el mayor número de especies. El par de oligonucleótidos seleccionado logró amplificar el gen en 32 especies diferentes pertenecientes al género *Sphingomonas* y otros taxones (Tabla 8).

Tabla 8.- Oligonucleótidos diseñados para la identificación del gen *cas9* en el género *Sphingomonas*.

<b>Género <i>Sphingomonas</i></b>		
<b>Forward</b>	<b>Reverse</b>	<b>Tamaño del amplificado</b>
TBTTCCACCTCAAMCAGCGS	RAABAGCGGATGCGCCTTS	482
<b>Temperatura de fusión (°C)</b>	<b>Temperatura de fusión (°C)</b>	<b>Taxones identificados</b>
62,5	61,6	<ul style="list-style-type: none"> <li>- <i>S. colombiensis</i></li> <li>- <i>S. paucimobilis</i></li> <li>- <i>S. sanxanigenens NX02</i></li> <li>- <i>S. sp. FARSPH</i></li> <li>- <i>S. soli</i></li> <li>- <i>S. liriopsis</i></li> <li>- <i>S. changbaiensis</i></li> <li>- <i>S. gilva</i></li> <li>- <i>S. horti</i></li> <li>- <i>S. naphthae</i></li> <li>- <i>Sphingobium sp. CAP-1</i></li> <li>- <i>Mesorhizobium sp. 8</i></li> <li>- <i>Erythrobacteraceae bacterium</i></li> <li>- <i>Sphingopyxis sp. FD7 DNA</i></li> <li>- <i>Sphingobium baderi</i></li> <li>- <i>Marinovum algicola</i></li> <li>- <i>Pseudaminobacter arsenicus</i></li> <li>- <i>Ancylobacter lacus</i></li> <li>- <i>Chenggangzhangella methanolivorans</i></li> <li>- <i>Sphingobium jiangsuense</i></li> <li>- <i>Sphingobium psychrophilum s</i></li> <li>- <i>Sphingobium limneticum</i></li> <li>- <i>Sphingobium wexiniae</i></li> <li>- <i>Pseudaminobacter manganicus</i></li> <li>- <i>Sphingobium baderi</i></li> <li>- <i>Shinella yambaruensis</i></li> <li>- <i>Nitratireductor alexandrii</i></li> <li>- <i>Methylopila jiangsuensis</i></li> <li>- <i>Hephaestia caeni</i></li> <li>- <i>Paracoccus thiocyanatus</i></li> <li>- <i>Frigidibacter mobilis</i></li> <li>- <i>Xanthobacter agilis</i></li> </ul>

### 7.1. Modelado de la proteína Cas7

El modelo 3D del monómero de la proteína Cas7 de la especie *Pseudomonas aeruginosa*, luego de un paso de refinamiento en el programa *Galaxy refine complex* del portal *Galaxy web* (Heo et al., 2016), mostró un gráfico de Ramachandran con un 97,04% favorecido y un QMEANDisCo global de 0,60 +/- 0,05 (Figura 37). Este modelo (Figura 38a) concuerda con la estructura para *Escherchia coli* propuesta por Jackson et al. (2014). Sobre este modelo se identificaron las regiones donde se diseñaron los oligos multigénero como se muestra en la figura Figura 38b; se señala el *forward* en color amarillo conformado por una hélice alfa, seguido de una lámina beta y una estructura desordenada; el *reverse* se indica en color rojo y comprende una región desordenada entre dos láminas beta.

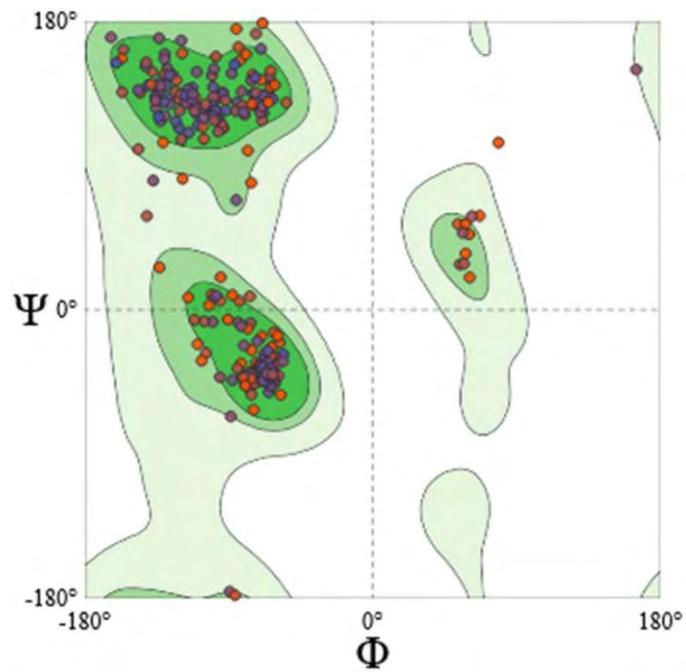
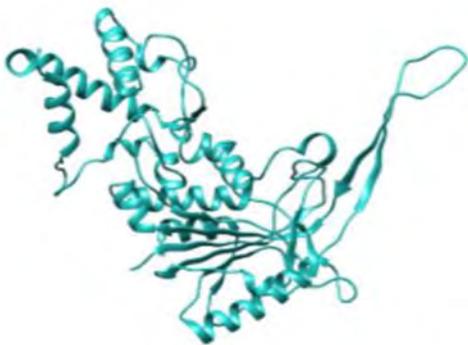
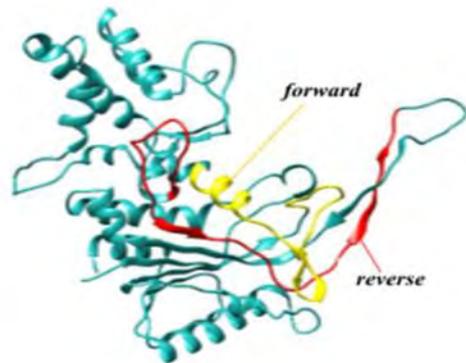


Figura 37.- Gráfico de Ramachandran del monómero de la proteína Cas7 de la especie *Pseudomonas aeruginosa* con un 97,04% favorecido.

**a**



**b**



**c**

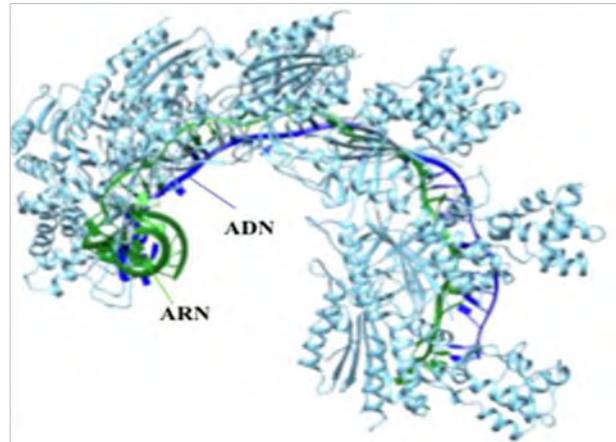
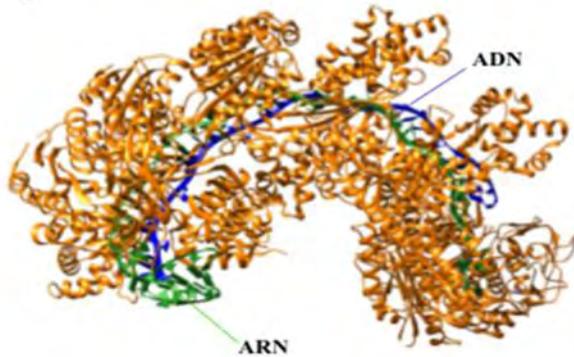


Figura 38.- Modelado 3D de la proteína Cas7 para la especie *Pseudomonas aeruginosa*. **a.** Monómero de la proteína Cas7. **b.** Monómero de la proteína Cas7, indicando en color amarillo la secuencia del forward y en color rojo la secuencia del reverse. **c.** Comparación del modelado 3D del hexámero de la proteína Cas7 de las especies *Escherichia coli* (en color naranja) y *Pseudomonas aeruginosa* (en azul claro), y su interacción con una molécula de ADN (en color azul oscuro) y de ARN (en color verde) del bacteriófago *Lambdavirus lambda LC730321*.

Por otra parte, el modelo del hexámero de la proteína Cas7 junto con ADN y ARN del bacteriófago *Lambdavirus lambda LC730321* presentó unos buenos valores para el gráfico de Ramachandran, con 98,06% favorecido y QMEANDisCo global de 0,61 +/- 0,05 (Figura 39). Con el fin de definir si la estructura obtenida del hexámero era correcta, se comparó con la de la proteína cristalizada a partir de *Escherichia coli* por parte de Mulepati et al. (2014), número de acceso PDB 4QYZ. En la Figura 38c se muestra que las estructuras cuaternarias de ambos modelos adoptan una forma de caballito de mar. Se observa que tanto en el modelo de *Escherichia coli* como en el de *Pseudomonas aeruginosa*, el hexámero envuelve al ADN y al ARN en la región central, y que cada una de las unidades interactúa en algunos sitios con tales moléculas, para estabilizarlas. Entre las diferencias observadas entre los modelos de ambas especies, destaca la disposición de la unidad de la proteína Cas7 ubicada en el extremo “cola” de la estructura; esta unidad es más laxa en *P. aeruginosa*, en comparación con la estructura cristalizada de la Cas7 de *E. coli*, que es más compacta.

En la Figura 40b se puede observar la ubicación de los oligos sobre los sitios correspondientes en el hexámero. En color rojo se muestra el *reverse* donde la lámina beta mantiene una interacción con el ARN (hélice de color verde), y el *forward* (en color amarillo) ubicado en el extremo de “la cola”, también interactúa con el ARN. En la Figura 40c se observa en mayor detalle cómo los átomos del *forward* interactúan con la cadena de ARN, mientras que los átomos del *reverse* también lo hacen, pero en menor medida (Figura 40d).

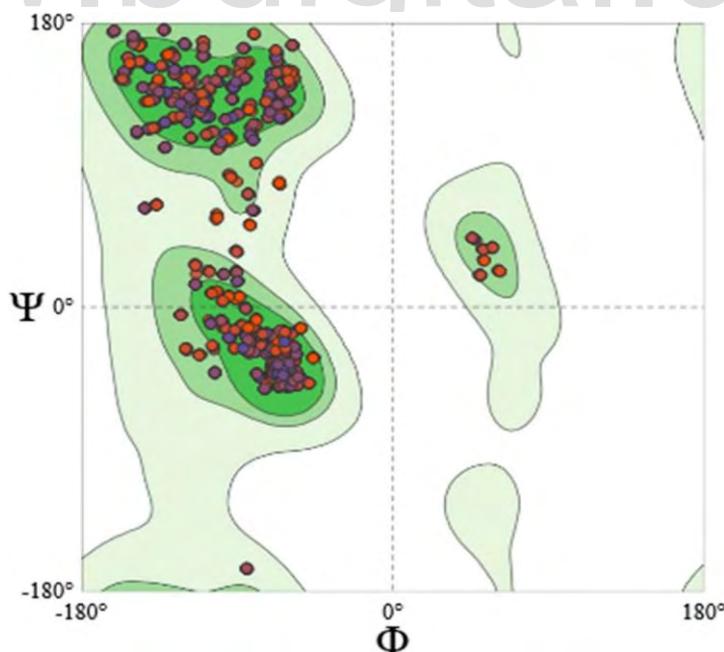


Figura 39.- Gráfico de Ramachandran del hexámero de la proteína Cas7 de la especie *Pseudomonas aeruginosa* con un 98,06% favorecido.

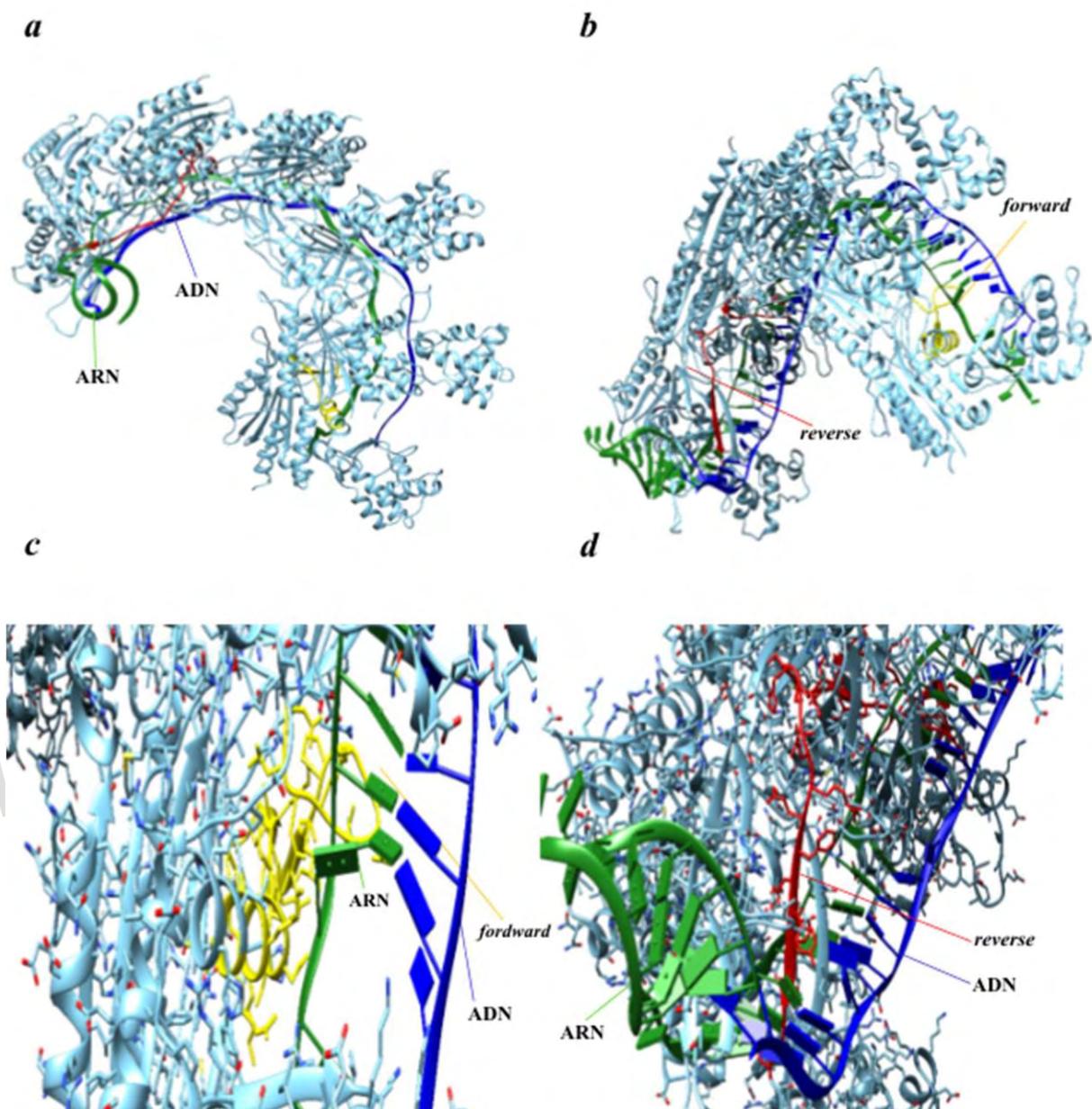


Figura 40.- Modelado 3D de la proteína Cas7 para la especie *Pseudomonas aeruginosa*, identificando las regiones donde se ubican las secuencias codificadas por el oligo multigénero. **a y b.** Distintas vistas del modelado 3D del hexámero de la proteína Cas7 e identificación de los sitios de unión de la proteína codificados por la secuencia complementaria del oligo multigénero en el hexámero, donde el forward se encuentra en una unidad, en color amarillo y el reverse en otra unidad, en color rojo. **c.** Interacción de la banda de ARN (en color verde) con la región de la proteína codificada por la secuencia complementaria del forward (en color amarillo). **d.** Interacción de la banda de ARN (en color verde) con la región de la proteína codificada por la secuencia complementaria al reverse (en color rojo).

Para la proteína Cas9 también se identificaron las regiones donde se ubican los oligos diseñados para su amplificación. Para ello, se realizó una búsqueda de estructuras 3D de la proteína Cas9, escogiéndose una en complejo con ARNsg (ARN guía) y ADN determinada por Bravo et al. (2022), mediante criomicroscopía electrónica guiada por cinética. La estructura obtenida pertenece a la especie *Streptococcus pyogenes* y se encuentra en su conformación activa (Figura 41a). Para poder visualizar los oligos en especies distantes a nivel filogenético, se realizó un alineamiento por COBALT de las secuencias aminoacídicas de especies del género *Sphingomonas* junto con la secuencia de la especie *Streptococcus pyogenes*, observando que tanto el *forward* como el *reverse* corresponden a los motivos conservados **AXXHXXXXXRG** y **TXXXXXXLXXXXXXN**, respectivamente (Figura 41b).

La estructura de la proteína Cas9 de la especie *Streptococcus pyogenes* mostró que ambos oligos se encuentra en el mismo lóbulo de la proteína, que se encuentra uno frente al otro y que mantienen una cercanía con el ARNsg (Figura 41c). La proteína Cas9 contiene varios dominios funcionales caracterizados, dos de los cuales son el lóbulo de reconocimiento/REC y el lóbulo de nucleasa/NUC; este último está compuesto por dos dominios, el dominio HNH y el RuvC, los cuales se señalan en la Figura 42a y se identificaron en la estructura proteica, con la finalidad de visualizar en cuál dominio se localizaban los oligos. Lo que se encontró es que el *forward* se ubica en el lóbulo REC mientras el *reverse* se ubica en el lóbulo NUC, específicamente en el dominio HNH, como se muestra en la Figura 42b.

www.bdigital.ula.ve

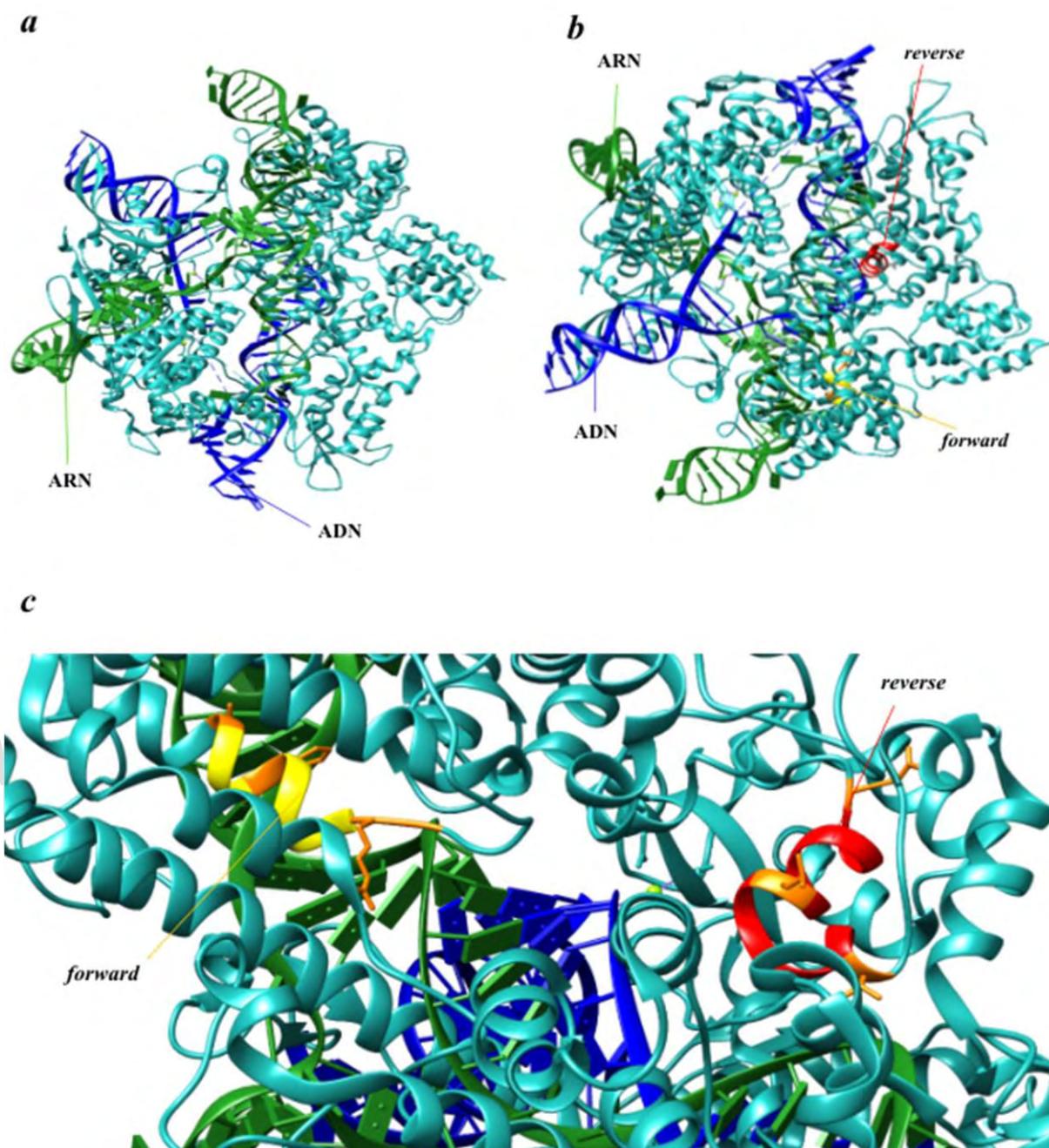
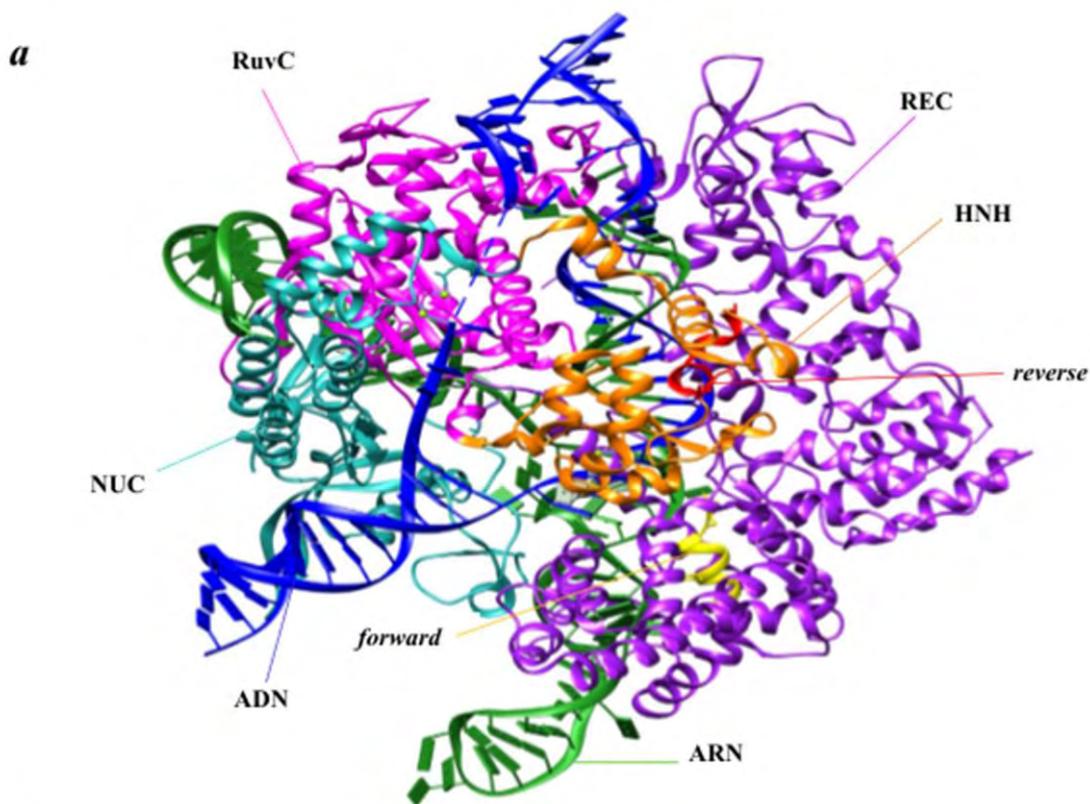
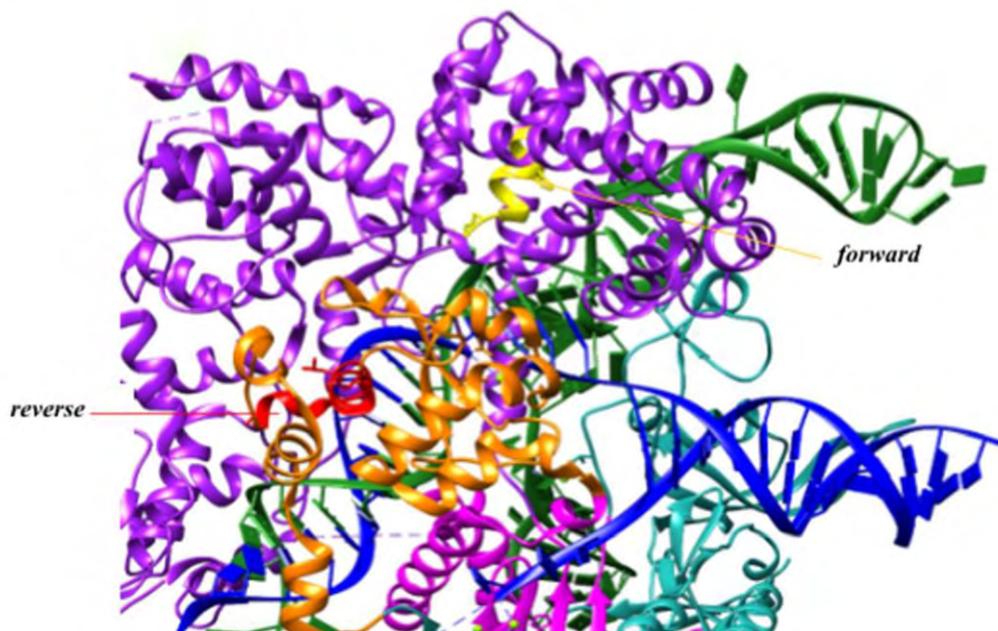


Figura 41.- Estructura en criomicroscopía electrónica de la proteína Cas9 en complejo con ADN y ARNsg de la especie *Streptococcus pyogenes* (Bravo et al., 2022). **a.** Estructura general de la proteína Cas9 en la conformación activa resaltando el ADN en color azul y el ARNsg en color verde. **b.** Estructura de la proteína Cas9 destacando los oligos diseñados, donde el forward se señala en color amarillo y el reverse, en color rojo. **c.** Acercamiento de la estructura de la proteína, indicando la ubicación de los oligos, donde se muestra la cercanía de ambos oligos con el ARNsg; además, se resaltan los aminoácidos conservados (A, H, R y G) en color naranja para el forward, y los aminoácidos conservados (T, L y N) en color naranja para el reverse.



W **b**



ve

Figura 42.- Estructura en criomicroscopía electrónica de la proteína Cas9 en complejo con ADN y ARNs<sub>g</sub> de la especie *Streptococcus pyogenes* señalando los dominios funcionales (Bravo et al., 2022). **a.** Estructura general señalando el lóbulo de reconocimiento (REC) en color púrpura y el lóbulo nucleasa (NUC) en color azul celeste, y señalando específicamente los dominios de NUC, HNH (en color naranja) y RuvC (en color magenta). Además, se identifican los sitios de unión del forward en color

amarillo y el reverse, en color rojo. **b.** Acercamiento de la estructura general, identificando la ubicación exacta de los oligonucleótidos diseñados.

## 8. Discusión

- 8.1. Las secuencias nucleotídicas/aminoacídicas de los genes de la matriz CRISPR-Cas en los géneros estudiados son muy diversas, y también la arquitectura genética del sistema puede variar

### Sistema CRISPR-Cas de la clase I

En este estudio se analizó la diversidad de las secuencias de nucleótidos y aminoácidos de diferentes genes constituyentes de la matriz CRISPR-Cas en cinco géneros diferentes, *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y *Sphingomonas*, además de la arquitectura genética de las distintas clases del sistema CRISPR-Cas en dichos géneros. Estos géneros contienen cepas psicrófilas y psicrotolerantes que fueron aisladas a partir de los sustratos asociados al último glaciar de Venezuela, el del Pico Humboldt. El sistema CRISPR-Cas cumple un rol importante como “sistema inmune bacteriano” en la defensa contra los virus, y participa de esta manera en la modulación potencial de la estructura y función de las comunidades microbianas en los ambientes poli-extremos, como lo son las regiones glaciales y periglaciales (Heinrichs et al., 2023). Por lo tanto, resulta muy interesante analizar las variaciones del sistema CRISPR-Cas en otras especies de los géneros mencionados, gracias a la data genómica disponible en las bases de datos. En particular, resulta de gran utilidad el diseño de oligonucleótidos que pudieran potencialmente detectar genes diagnósticos de los diferentes sistemas CRISPR-Cas en cepas cultivables únicas, aplicando un enfoque basado en la PCR.

El análisis de las secuencias nucleotídicas/aminoacídicas de la mayoría de los genes *cas* demostró una enorme variabilidad genética entre los diferentes géneros. En el interior de cada género, también se observaron diferencias importantes entre las especies/cepas disponibles en las bases de datos, aunque las regiones conservadas siempre estaban presentes. Este resultado era de esperarse, ya que el sistema CRISPR-Cas es muy diverso entre las bacterias que lo contienen, debido a la constante coevolución con los bacteriófagos y otros elementos genéticos móviles. Existen, de hecho, múltiples clases de este sistema, con varios subtipos por clase (ver sección 2.4 en el Marco Teórico).

Cada clase de sistema tiene sus propias características que les permite cumplir su función. Se ha registrado que el sistema de la clase I es el más frecuente en la naturaleza y representa el 90% de los loci CRISPR-Cas reportados, en comparación con el sistema de la clase II que es menos común (Xue & Sashital, 2019). En consistencia con esta observación, cuatro de los cinco géneros bacterianos estudiados (*Arthrobacter*, *Cryobacterium*, *Pseudomonas* y *Rhodococcus*) presentaron el sistema CRISPR-Cas de la clase I y el subtipo I-E, a pesar de no se encuentran estrechamente relacionados filogenéticamente. Además, para *Arthrobacter*, este representa uno de los primeros

estudios en la identificación de este sistema en especies representantes de este género. Por otra parte, sólo uno de los cinco géneros estudiados, *Sphingomonas*, presentó la clase II del sistema.

En cuanto a las variaciones de la arquitectura genética en la matriz CRISPR-Cas de la clase I, subtipo I-E, cabe destacar que éstas fueron notorias, principalmente en el género *Arthrobacter*, pero también en *Pseudomonas* y *Cryobacterium*. Un resultado muy llamativo fue que algunas especies del género *Arthrobacter* y *Cryobacterium* carecían de los genes *cas1* y *cas2*, que codifican proteínas involucradas en la adquisición de espaciadores del ADN foráneo involucrado en el mecanismo de “memoria” del sistema CRISPR-Cas. Cas1 es la proteína más conservada presente en la mayoría de los sistemas CRISPR-Cas, lo que supone que evoluciona más lentamente que otras proteínas Cas (Nuñez et al., 2014). Por lo tanto, Cas1 ha sido utilizada en filogenias para la clasificación del sistema CRISPR-Cas (Chylinski et al., 2014).

Estudios bioquímicos argumentan que las proteínas Cas1 y Cas2 son nucleasas dependientes de metales y forman un complejo heterohexámero que es necesario y suficiente para la integración del espaciador (Nuñez et al., 2014). Sin embargo, se determinó que para la funcionalidad de este complejo en *Escherichia coli* sólo se requiere la actividad enzimática de Cas1, para la integración del espaciador mediante el complejo Cas1-Cas2, mientras que la actividad de Cas2 es prescindible, lo que indica que esta última proteína tiene un papel estructural en la adquisición del espaciador (Yosef et al., 2012).

Se han caracterizado experimentalmente la proteína Cas1 de tres especies, *Escherichia coli*, *Pseudomonas aeruginosa* y *Archaeoglobus fulgidus*, y se han resuelto sus estructuras, demostrando que la proteína Cas1 forma un homodímero y es una nucleasa que escinde el ADNsb y el ADNdb (Makarova & Koonin, 2015). El monómero Cas1 consta de dos dominios, con el dominio catalítico de hélice  $\alpha$  C-terminal y el dominio N-terminal en su mayoría de cadena  $\beta$ , que probablemente esté involucrado en la dimerización y la interacción con otras proteínas, en particular con Cas2 (Jackson et al., 2014). A pesar de que la proteína Cas1 se ha considerado la firma de la presencia de sistemas CRISPR-Cas en un genoma, se ha descubierto que muchos genomas carecen de un gen *cas1* y poseen loci *cas* que codifican complejos efectores aparentemente activos y, por lo tanto, podrían funcionar de manera independiente de Cas1.

Los ejemplos de sistemas que carecen de *cas1* incluyen los sistemas de tipo IV, el subtipo III-B y una variante del subtipo I-F. La identificación de complejos efectores muy diversos denominados “independientes”, es decir, que no están asociados con el par de genes *cas1-cas2* se ha hecho cada vez más común. Estos complejos efectores "solos" a menudo están presentes en plásmidos o asociados con genes relacionados con transposones, como *TniQ/TnsD*, una proteína de unión al ADN necesaria para la transposición. Muchos de estos casos son derivados del subtipo I-F y algunos otros, con genes que codifican Cas6, Cas8, Cas5, Cas7, son derivados del subtipo I-C (Makarova & Koonin, 2015). Nuestros resultados muestran, por otra parte, que el gen *cas2* estaba ausente en varias especies estudiadas de los géneros *Arthrobacter* y *Cryobacterium*; este es un gen muy pequeño y se ha reportado que puede estar fusionado con el gen *cas3* en el sistema de subtipo I-F. Por esa razón se realizó una búsqueda en BLASTn para identificar si se encontraba fusionado en algunas especies, pero los resultados obtenidos descartaron la presencia de los genes

cas2 y cas3 fusionados en las especies estudiadas. En consecuencia, se puede inferir que otras proteínas nucleasas constitutivas lleven a cabo la función de las nucleasas Cas1 y Cas2 en la integración de espaciadores en la matriz CRISPR, principalmente aquellas que comparten dominios funcionales similares, como las proteínas de la superfamilia *Adenylation DNA ligase like, una SsRNA endonuclease* o una *Single-stranded DNA endonuclease*, reportadas en otras bacterias.

En el género *Arthrobacter* no sólo se observaron caracteres especiales de la arquitectura genética, como la ausencia de los genes *cas1* y *cas2*, sino que también se detectó una alta variabilidad de las secuencias nucleotídicas y aminoacídicas entre las especies analizadas. Al observar los alineamientos de cada uno de los genes *cas*, se evidenció un patrón entre las especies donde se distinguieron tres grupos, por lo que se evaluaron los alineamientos de cada gen en base a cada grupo, obteniendo así una mayor similitud entre las secuencias de cada grupo. Es importante mencionar que la filogenia del género *Arthrobacter* no ha sido del todo resuelta, y ha sufrido diferentes rearrreglos a lo largo del tiempo. De hecho, un análisis filogenético que empleó el algoritmo *neighbor-joining* y que abarcó la gran mayoría de especies del género *Arthrobacter*, indicó una estructura interna inestable del género (Busse, 2016).

El estudio de Busse (2016) confirmó la dificultad de establecer relaciones filogenéticas estables entre las llamadas especies de *Arthrobacter*, ya que es un grupo muy diverso de bacterias con elevadas diferencias morfológicas y metabólicas. Los análisis filogenéticos revelaron una heterogeneidad significativa, lo que llevó a reevaluar la taxonomía del género en base a las similitudes en la secuencia del gen *16S ARNr*, la homogeneidad en los tipos de peptidoglicanos, los sistemas de quinonas y los perfiles lipídicos polares. Los resultados arrojaron que *Arthrobacter oryzae* y *Arthrobacter humicola* pueden considerarse los parientes más cercanos de *A. pascens* y *A. globiformis* (la especie tipo del género), con las que también comparten un porcentaje mayor al 98 % de similitud en la secuencia del gen *16S ARNr*. Por otro lado, el árbol filogenético separó una sublínea profundamente ramificada conformada por las especies *Arthrobacter albus* y *Arthrobacter cumminsii*, que se encuentra distantemente relacionada con la especie tipo del género y, aparentemente, más estrechamente relacionada con otros miembros de la familia *Micrococcaceae*. También, las especies *A. woluwensis* y *A. sanguinis* se encontraron muy separadas de las demás especies dentro del árbol.

El mismo autor reconstruyó la filogenia de este género con 12 especies establecidas, en base al gen *recA*, demostrando que las especies de referencia se separaron en al menos seis grupos. Busse (2016) igualmente indicó que las filogenias del gen *16S ARNr* y *recA* estaban de acuerdo entre sí. En conclusión, este estudio determinó que el género *Arthrobacter sensu stricto* debería restringirse a las especies *A. globiformis*, *A. pascens*, *A. oryzae* y *A. humicola*, que comparten altas similitudes en la secuencia del gen *16S ARNr*, un sistema de quinona con MK-9(H<sub>2</sub>) predominante y un peptidoglicano con Lys-Ala 2-3 (A3 $\alpha$ ; A11.5, A11.6). Sin embargo, esta propuesta excluye a muchas especies del género *Arthrobacter* sin la opción de reclasificarlas en otros géneros, ya que aún no se dispone de rasgos diagnósticos estables. Por eso es necesaria la colecta de datos que permitan la asignación a otros géneros de especies que difieren de la descripción del género de *Arthrobacter sensu stricto*. No obstante, diversas especies sí fueron reclasificadas a otros cinco

géneros, *Paenarthrobacter*, *Pseudarthrobacter*, *Glutamicibacter*, *Paeniglutamicibacter* y *Pseudoglutamicibacter*, que comprenden dos de las especies analizadas en este estudio, *A. cumminsii* y *A. albus* (Busse, 2016). Por lo tanto, es evidente que en el presente estudio se están incluyendo bajo el nombre de *Arthrobacter* a especies/cepas que están cercanas filogenéticamente a este género, pero que no pertenecen al género *per se*.

Debido a que 11 de las 16 especies estudiadas acá eran no cultivables, es posible que su clasificación deba ser reevaluada motivado a las diferencias entre ellas. El primer grupo (grupo 1) incluye a las especies *A. oryzae* y *A. globiformis* (la especie tipo) (Figura 15a.1). Por otro lado, en el segundo grupo (grupo 2) se encuentran tres especies no cultivables que son del género *Arthrobacter*, junto con *A. cumminsii* y *A. albus*, que han sido reclasificadas al género *Pseudoglutamicibacter* (Busse, 2016). Debido a la elevada similitud de secuencias dentro de este grupo, es posible que todas correspondan al género *Pseudoglutamicibacter*. Finalmente, se evaluó un tercer grupo compuesto por cuatro especies, de las cuales tres eran no cultivables y una era *A. woluwensis*. Esta última, al estar muy alejada de la especie tipo en el árbol filogenético, podría no ser un representante del género *Arthrobacter*. Además, las especies de este grupo eran muy disímiles y presentaban una gran cantidad de *gaps* en su alineamiento, por lo que se decidió descartar este grupo del análisis.

Reforzando los argumentos que sostienen la alta diversidad en la arquitectura genética en el género *Arthrobacter*, es notable que en el grupo 1 se observó un contraste en el ordenamiento de los genes. En particular, ocho especies presentan el orden *normal*: *cas3-cas8-cas11-cas7-cas5-cas6-cas1-cas2* y otras especies muestran diferente orden de los genes: *cas1-cas3-cas6-cas5-cas7-cas11-cas8*. En las especies pertenecientes al grupo 2 de este género y en los otros estudiados, la arquitectura genética observada fue la reportada específicamente para el sistema CRISPR-Cas subtipo I-E (orden canónico), que podría suponerse que se trata de la arquitectura ancestral, por ser la más común.

Los sistemas CRISPR-Cas de tipo I-C, I-D, I-E y I-F normalmente están codificados por un único operón canónico que abarca los genes *cas1*, *cas2* y *cas3* junto con los genes para las subunidades del complejo *Cascade*. En contraste, muchos sistemas de tipo I-A y I-B parecen tener una organización diferente en la que los genes *cas* están agrupados en dos o más operones canónicos. En la mayoría de las matrices de clase I, cada una de las familias de genes *cas* está representada por un único gen. Cada subtipo de los sistemas tipo I tiene una combinación de genes característica y distintiva de la organización del operón. En particular, *cas4* está ausente en los sistemas I-E e I-F. Los subtipos I-E e I-F son monofiléticos en los árboles filogenéticos de *cas1* y *cas3*, y cada uno tiene uno o más genes característicos distintos. En el caso del subtipo I-E, que es el que presentan los géneros *Cryobacterium*, *Pseudomonas*, *Arthrobacter* y *Rhodococcus* estudiados aquí, el gen único es *cas8* (Makarova & Koonin, 2015). Para el subtipo I-F se han identificado diferentes combinaciones en el orden de los genes. No obstante, no hay evidencias de este comportamiento en el subtipo I-E, por lo que es posible que la inversión de los genes observada en el grupo 1 de *Arthrobacter* sea una variación reciente o adquirida vía transferencia horizontal de genes o transposones.

Finalmente, en el caso del género *Pseudomonas*, resultó muy interesante notar que el operón contenía diferentes genes que no codifican a proteínas Cas. Lo resaltante es que todas las especies que mostraron esta arquitectura genética, incluyendo especies cultivables y no cultivables, y de diferentes orígenes geográficos, compartían en algunos casos los mismos genes y el mismo orden, por lo que parece tratarse de una variación mantenida y generalizada de la arquitectura del sistema CRISPR-Cas en el género *Pseudomonas*.

Para diferentes especies del género *Sphingomonas* se identificó la presencia del sistema CRISPR-Cas de la clase II. Con el fin de determinar el subtipo de sistema, se observó que el orden de los genes dentro de los genomas anotados era *cas9-cas1-cas2*-transactivador-espaciadores, que corresponde al sistema del subtipo II-C1 (ver Figura 4). Los sistemas CRISPR/Cas de tipo II según la clasificación actual incluyen tres subtipos, II-A, II-B y II-C, el último de los cuales tiene dos variantes (II-C1, II-C2) (Makarova et al., 2019). El CRISPR-Cas tipo II es considerado el más simple de los sistemas CRISPR-Cas, ya que incluye la matriz de repeticiones-espaciadores CRISPR y sólo cuatro, aunque normalmente son tres, genes *cas*. Además, incluye otros factores bacterianos, en particular el ARNcr transactivador (tracrRNA) y la RNasa III, que contribuyen a la función de este sistema (Chylinski et al., 2014). Los sistemas CRISPR-Cas del subtipo II-C son frecuentes en patógenos y comensales, y son los subtipos más comunes entre los sistemas de clase II; no obstante, estos subtipos son los menos estudiados (Chylinski et al., 2014).

El sistema del tipo II se basa en el uso de una única proteína Cas9 multidominio de gran tamaño, como complejo efector. Además, todos los loci CRISPR/Cas de tipo II contienen genes *cas1* y *cas2*, esenciales para la fase de adaptación, que en este caso parecen estar presentes en todos los géneros estudiados, y ARNtracr, un ARN no codificante, necesario para madurar el pre-ARNcr largo y para las interacciones con la proteína Cas9. El complejo proteico ARNcr-ARNcrtra- Cas9 es capaz de reconocer y escindir las secuencias de ADN diana (Hryhorowicz et al., 2023).

Dentro de los diferentes tipos y subtipos que conforman la clase II existen diferencias sustanciales en cuanto a los mecanismos de procesamiento del pre-ARNcr. En los sistemas de tipo VI y subtipo VA, la proteína efectora también cumple la actividad de RNasa de procesamiento de pre-ARNcr, mientras que en el tipo II y en varios subtipos del tipo V, este procesamiento generalmente es llevado a cabo por una enzima no Cas, la RNasa III, donde el módulo efector incluye una molécula de ARN adicional, el ARN CRISPR transactivador (tracr), que forma dúplex estables con la repetición directa parcialmente complementaria del pre-ARNcr. Después de la escisión del dúplex de ARN por la RNasa III, el ARN guía maduro, es decir, el complejo ARNcr-ARNtracr, permanece unido de forma estable a los efectores, lo que permite una interferencia específica del ADN (Makarova et al., 2019).

En un estudio se determinó que el sistema CRISPR-Cas de tipo II es relativamente raro en bacterias (Chylinski et al., 2014). Sin embargo, los resultados de un análisis estadístico de genomas, tanto de bacterias como de arqueas, indicaron la presencia del tipo II en bacterias, pero no en arqueas. La presencia exclusiva de sistemas CRISPR-Cas de tipo II en bacterias es compatible con la participación de la RNasa III, una enzima principalmente bacteriana y que media el procesamiento del pre-ARNcr en este tipo de sistemas CRISPR-Cas (Wolf et al., 2012).

## 8.2. Es posible identificar las regiones de interés en la matriz CRISPR-Cas, potenciales para el diseño de un sistema de detección

El gen *cas1* resultó altamente variable entre las especies de cada uno de los géneros estudiados, tanto a nivel de las secuencias nucleotídicas como aminoacídicas. Como se explicó previamente, varias especies del género *Arthrobacter* y *Cryobacterium* carecían de este gen. A pesar de ser una de las proteínas más importantes y conservadas en el contexto de los sistemas CRISPR-Cas, este gen no fue considerado para el diseño de oligonucleótidos en todos géneros, a excepción del género *Rhodococcus*, ya que al evaluar las secuencias correspondientes del gen y la proteína se identificaron regiones muy conservadas que fueron candidatas para el diseño de oligonucleótidos. En el estudio de la estructura cristalina del complejo de Cas1 y Cas2 de *Escherichia coli* se determinó que la interfaz entre Cas1 y Cas2 consta de enlaces de hidrógeno, interacciones electrostáticas e hidrofóbicas (Nuñez et al., 2014). Se observaron contactos electrostáticos extensos entre tres residuos de arginina (R245, R252, R256) en  $\alpha 8$  de Cas1 con dos residuos ácidos (E65 y D84) de Cas2. El residuo R252 se coloca entre E65 y D84 y puede formar puentes salinos entre los dos residuos ácidos, aunque se observó una densidad continua entre R252 y E65 en la interfaz Cas1-Cas2. Las regiones conservadas identificadas en especies de *Rhodococcus* no se encontraron dentro del rango de estos residuos planteados para *Escherichia coli* (Nuñez et al., 2014). Sin embargo, en otro estudio de una forma cristalina alternativa de *Escherichia coli* cultivada en presencia de  $Mg^{2+}$ , se revelaron interacciones secundarias Cas1-ADN que brindan información adicional sobre el mecanismo de unión del ADN genómico Cas1-Cas2 y la posterior integración (Nuñez et al., 2015). Un primer plano del sitio activo mostró una densidad continua de  $Mg^{2+}$  con E141, H208, D221. En los sitios activos, la base terminal 3' está involucrada en una interacción de apilamiento con Y217 que posiciona los extremos nucleofílicos 3'-OH del protoespaciador cerca de los residuos conservados de unión a metales, E141, H208 y D221 (Nuñez et al., 2015). Se ha demostrado previamente que estos tres residuos coordinan un ion  $Mn^{2+}$  en el sitio activo de Cas1 de *P. aeruginosa* (Wiedenheft et al., 2009a).

Una de las regiones conservadas que se identificaron en la proteína Cas1 se encuentra entre los aminoácidos 140 y 158 en especies del género *Rhodococcus*. Con el fin de evaluar si la región observada correspondía a un dominio conservado entre varias especies y a una región funcional, se realizó una comparación mediante COBALT de la proteína Cas1 de *Rhodococcus sovatusensis* y *Escherichia coli*, encontrándose un dominio muy conservado que contiene el aminoácido E141 implicado en la unión a metales (Nuñez et al., 2015). Así mismo, al evaluar si este dominio estaba presente en otras bacterias, se logró identificar el motivo LRGXEGXRV en distintas especies de bacterias no filogenéticamente relacionadas, haciendo uso de la Clasificación de Proteínas y Dominios Conservados, del inglés *Conserved Domains and Protein Classification (CDD)*, del NCBI (Figura 43). En consecuencia, esta región fue seleccionada para el diseño de oligonucleótidos para el género *Rhodococcus*.

<i>Rhodococcus sovatus</i> . strain.18930	EDTS-ELTMQQ <b>LRGREGAR</b> VRKSYREHSERTGVEWVRR	149
<i>Escherichia coli</i> .strain.K12	KRAPKEEDLAL <b>LRGIEGSR</b> VRQTYALLAKQYGKWNGR	158
<i>Rhodospirillum rubrum</i>	DDPPARRSVDQ <b>LRGMEGVR</b> REIYRLLAKKYAVDWNAR	146
<i>Methanospirillum hungatei</i>	EDFSDAYSVEQ <b>LRGLEGVR</b> REGYRKIARDTGVIWNGR	145
<i>Rhodoferax ferrireducens</i>	DHMPPSFTIEK <b>LRGLEGSR</b> VKEIYVNLADKLGWVWQGR	147

Figura 43.- Alineamiento de COBALT de la proteína Cas1 en diferentes especies bacterias, identificando en color rojo los aminoácidos conservados y el dominio LRGXEGXRV, y las regiones variables en color azul y gris.

Como se mencionó previamente, el gen *cas2* no fue tomado en cuenta para su evaluación debido a su tamaño y a la ausencia del mismo en algunas especies.

En el caso del gen *cas3* se identificaron múltiples regiones de interés en las secuencias nucleotídicas de todos los géneros analizados. El gen *cas3* codifica para una proteína "distintiva" de los sistemas CRISPR de tipo I, y funciona con un complejo de ribonucleoproteína llamado "*Cascade*" que empareja las bases del ARN CRISPR (ARNcr) con un ADN objetivo formando un bucle R. El reclutamiento de la proteína Cas3 a un bucle R del complejo *Cascade* posiciona el dominio de la actividad de la ADN nucleasa-translocasa de Cas3 en disposición de degradar el ADN blanco. Esta es la base general de la interferencia de CRISPR en los sistemas CRISPR de tipo I (Makarova & Koonin, 2015). Esta proteína comprende un dominio de nucleasa HD fusionado a dos dominios similares a RecA característicos de las helicasas de la superfamilia 2. Estos dominios cooperan para ofrecer translocación y degradación del ADN simple banda (ADNss) dependiente de ATP. En común con muchas helicasas, Cas3 tiene un dominio accesorio ubicado en el extremo C-terminal de la proteína, aunque sus funciones exactas no están claras. La función de la nucleasa HD puede ser apoyada por una variedad de iones metálicos. Las estructuras cristalinas de Cas3 muestran sitios activos de HD unidos a hierro, manganeso y calcio. La bioquímica de Cas3 *in vitro* indica actividad de nucleasa estimulada por manganeso y cobalto, e inhibida por hierro; el cobalto es necesario para observar una interacción estable entre Cas3 y *Cascade* al menos en un sistema. Las proteínas de esta familia tienen un dominio de nucleasa N-terminal y un dominio C-terminal helicase/ATPase. En algunos sistemas CRISPR/Cas los dominios se intercambian, en otros se codifican por separado (He et al., 2020). Esta proteína incluso ha servido de base para desarrollar filogenias dentro de los subtipos, al ser una proteína tan importante (Makarova et al., 2013), por lo que se consideró que podría ser un buen marcador para la detección del sistema. Sin embargo, no se consiguieron sitios adecuados para el diseño de oligos ideales, diagnósticos de este tipo de sistema.

Para el gen *cas5* se observó una alta variabilidad en las secuencias nucleotídicas, específicamente en el grupo 1 del género *Arthrobacter*. Por otro lado, en los géneros *Cryobacterium*, *Pseudomonas* y *Rhodococcus* se observaron diversas regiones conservadas, concentradas en las regiones iniciales y finales del gen para el caso de *Cryobacterium* y *Pseudomonas*, mientras que para el género *Rhodococcus* se concentraron en la región central del gen. La proteína Cas5 forma parte del complejo *Cascade* que junto con la proteína Cas6e (las proteínas denominadas con la letra (e) corresponden a el sistema del subtipo I-E) realiza

interacciones específicas con porciones de la secuencia de repetición CRISPR. La estructura de *Cascade* revela una característica común similar a un pulgar en las proteínas Cas7e y Cas5e, que es fundamental para el ensamblaje oligomérico de la estructura helicoidal (Jackson et al., 2014). La proteína Cas5e también exhibe una forma similar a la de la mano derecha, pero carece del dominio "dedos", lo que permite que Cas5 cubra el filamento de Cas7 en el extremo 5' del ARNcr al realizar la misma interacción de Cas7 entre las subunidades de Cas7. Además, la proteína Cas5e también contiene los primeros seis nucleótidos de la porción de repetición 5' del ARNcr. Una subunidad de Cas11 que está próxima al extremo 5' del ARNcr, junto con Cas5e, contacta a Cas8 en la parte inferior del complejo (Jackson et al., 2014). Se ha identificado que la Cas5 de sistemas CRISPR distantes relacionados contienen una hélice  $\alpha 1$  rica en glicina y un bolsillo de unión cargado positivamente que puede desempeñar un papel similar en el reconocimiento de nucleótidos en el extremo 5' del ARNcr (Garside et al., 2012). Debido a su importancia estructural en el complejo efector, las regiones conservadas y cercanas a éstas, fueron evaluadas para el diseño de oligonucleótidos para la posible detección del gen *cas5*.

Del mismo modo, para el gen *cas5*, para el gen *cas6* se observó una alta variabilidad en las secuencias, pero se lograron identificar unas pocas regiones conservadas. Para los géneros *Rhodococcus* y *Arthrobacter* (grupo 1) se observaron diversos sitios encontrados a lo largo del gen, mientras que para *Cryobacterium*, *Pseudomonas* y el grupo 2 de *Arthrobacter*, sólo se identificaron regiones en el inicio y el final del gen. El procesamiento de ARN CRISPR por la proteína Cas6 es esencial para la protección guiada por ARN contra el ADN invasor. Se ha demostrado que las proteínas de la familia Cas6 son filogenéticamente diversas, pero todas las proteínas Cas6 son endorribonucleasas independientes de metales que unen y escinden selectivamente transcripciones largas de ARN CRISPR (Brouns et al., 2008). En un estudio de la estructura cristalina del complejo cascada del sistema CRISPR-Cas del subtipo I-E se reveló que la proteína Cas6e de *E. coli* consta de motivos de reconocimiento de ARN (RRM) en tándem conectados por un enlace de ocho residuos. Cada RRM, también llamado pliegue tipo ferredoxina, consta de una disposición  $\beta 1 \alpha 1 \beta 2 \beta 3 \alpha 2 \beta 4$  conservada, en la que las hebras  $\beta$  están dispuestas en una lámina  $\beta$  antiparalela de cuatro hebras y las dos hélices se agrupan en un lado de la lámina (Jackson et al., 2014). Las láminas  $\beta$  en cada uno de los dos RRM se enfrentan entre sí, creando una hendidura en forma de V a lo largo de una cara de la proteína. Una superficie cargada positivamente en la cara opuesta de la proteína establece contactos electrostáticos significativos con la cadena 3' del tallo-bucle del ARNcr. Además, un bucle de ranura cargado positivamente (residuos 90-119) en el dominio RRM C-terminal hace contactos electrostáticos extensos con el surco principal del bucle de tallo de ARNcr. En el estudio realizado por Jackson (2014) se determinaron los aminoácidos esenciales de interacción; sin embargo, no fue posible encontrar un sitio idóneo en base a estas regiones conservadas en el gen *cas6*, dada la alta variabilidad de la secuencia nucleotídica, quizás debido a una alta tasa de mutaciones sinónimas en estas regiones.

Dentro de los ocho residuos identificados en la proteína Cas6 de *E. coli* por Jackson (2014), resaltaron dos aminoácidos en el N-ter (Tyr17 y His20) que corresponden al sitio activo. Una de las regiones conservadas en especies del género *Pseudomonas* se encuentra entre los aminoácidos 1 y 27. Para evaluar si esta región contenía los aminoácidos esenciales del sitio activo de la proteína se realizó el mismo procedimiento utilizado para la proteína Cas1 en el género *Rhodococcus*. El

resultado del alineamiento de varias especies bacterianas, incluido un representante del género *Pseudomonas*, mostró un motivo conservado (YXXH) que contiene los aminoácidos tirosina e histidina implicados en la actividad endorribonucleasa (Figura 44). Por esta razón, se seleccionó esta región para el diseño de oligonucleótidos para la identificación del gen *cas6* para especies del género *Pseudomonas*.

<i>Pseudomonas aeruginosa</i> . strain. AR442	GDAYEMHRTLSRVFAD	69
<i>Escherichia coli</i> . strain. K12	RGEYVMHQWLWDLFPG	68
<i>Chlorobaculum tepidum</i>	TDSYSLHRVIYSLFED	73
<i>Salmonella</i> . sp	RGEYVMHQWLWELFPG	68

Figura 44.- Alineamiento de COBALT de la proteína Cas6 en diferentes especies bacterias, identificando en color rojo los aminoácidos conservados y el dominio YXXH, y las regiones variables en color azul y gris.

En el caso del gen *cas7* se identificaron diversas regiones conservadas en cada uno de los géneros. Lo que resulta muy interesante es que se detectaron dos regiones muy conservadas que fueron compartidas entre las especies de todos los géneros estudiados. La proteína Cas7 es una proteína esencial del sistema, ya que constituye la base principal de andamiaje para el complejo *Cascade*, compuesta por seis proteínas Cas7 que oligomerizan a lo largo del ARNcr formando una arquitectura entrelazada que presenta la secuencia guía del ARNcr en seis segmentos discretos (Mulepati et al., 2014). La estructura de la proteína dilucidada en *E. coli* mostró que cada segmento consta de un nucleótido incrustado seguido de cinco bases accesibles al solvente que están ordenadas en una configuración de forma pseudo A mediante interacciones con tres subunidades proteicas diferentes (Jackson et al., 2014). La proteína Cas7 se pliega formando una estructura con forma de mano derecha. Esta forma es creada por un RRM modificado que forma la palma, un dominio helicoidal que se asemeja a los dedos (residuos 59-181), un bucle de 30 aminoácidos que toma la forma de un pulgar (residuos 193-223) y dos bucles más pequeños insertados en el RRM que forman una red entre el pulgar y los dedos (Jackson et al., 2014). A diferencia de la mayoría de los RRM, que se unen al ARN utilizando residuos conservados ubicados en la cara de la lámina  $\beta$  antiparalela, la estructura revela una serie de interacciones con la cadena principal de fosfato que se limitan principalmente a la primera hélice  $\alpha$  ( $\alpha 1$ ) del RRM, la membrana y el pulgar. La primera hélice  $\alpha$  de la mayoría de los RRM se ubica en la parte posterior de la lámina  $\beta$  y no entra en contacto directo con el ARN. Sin embargo, en Cas7 la hélice  $\alpha 1$  se ubica perpendicular y alejada de un lado de la lámina  $\beta$  central. Los residuos conservados en la hélice  $\alpha 1$  interactúan con tres fosfatos consecutivos de una manera que se introducen dos giros consecutivos de  $\sim 90^\circ$  (en inglés *chicanes*) en la estructura principal del ARNcr. Estos giros en el ARNcr ocurren con una periodicidad regular de seis nucleótidos definida por la distancia entre las hélices  $\alpha 1$  en las subunidades Cas7 adyacentes. Cada uno de estos giros está separado por cinco bases que mantienen contacto con la parte exterior en forma pseudo A, mientras que la sexta base se voltea fuera de la presentación helicoidal y se cubre con el pulgar de una molécula Cas7 adyacente. La posición de cada pulgar se estabiliza mediante interacciones electrostáticas con la hélice  $\alpha 1$  en la

palma de una molécula adyacente, y se ha demostrado que las mutaciones en el pulgar de las proteínas Cas7 homólogas reducen significativamente las afinidades de unión del ARN (Jackson et al., 2014). En el estudio mencionado se identificaron los aminoácidos que son necesarios para la interacción con el ARNcr y que son fundamentales para la estabilización del complejo. Resulta imprescindible observar el modelado en 3D de la proteína Cas7 en especies de los géneros estudiados para evaluar si las regiones donde fueron diseñados los oligonucleótidos corresponden a las regiones de interacción.

La proteína Cas8, también llamada Cse1, es una proteína grande que tiene dos funciones importantes dentro del mecanismo del sistema CRISPR-Cas: el reconocimiento de la secuencia PAM y el reclutamiento de la proteína Cas3. La Cas8 se ha identificado como una proteína característica del subtipo I-E. El análisis de las regiones de interés mostró que en el grupo 1 del género *Arthrobacter* y en *Pseudomonas*, las regiones conservadas se concentraron en el centro del gen. Para el caso de *Cryobacterium*, tales regiones se repartieron a lo largo del gen. En el género *Rhodococcus* las regiones de menor variabilidad se encontraron al inicio y al final del gen, y en el grupo 2 del género *Arthrobacter*, las regiones poco variables se concentraron únicamente al final. Este resultado podría estar indicando que la evolución del gen *cas8* no parece haber determinado una única manera posible de garantizar la función de la proteína Cas8 codificada.

Un estudio de la estructura de la proteína Cas8 en *Escherichia coli* determinó que la proteína está conformada por dos dominios que adoptan un pliegue globular único que contiene un ion metálico coordinado por cuatro cisteínas y un haz de cuatro hélices. El motivo de unión del ion metálico crea una perilla en el extremo de un bucle que puede estar involucrado en el posicionamiento de la hélice L1 (un bucle de los residuos 130-143 que está desordenado en la estructura cristalina de la proteína Cse1) para el acoplamiento. Además de la interacción de acoplamiento por L1, el dominio globular de Cse1 también hace contactos con el RRM modificado de Cas5, y el haz de cuatro hélices en Cse1 se extiende desde la parte superior del dominio globular, haciendo contactos con el dominio C-terminal de Cse2 (Cas 11). Esta interacción completa el puente estructural que conecta el haz de cuatro hélices de la cola de Cse1 a la cabeza de Cas6 para la conformación del complejo *Cascade* (Jackson et al., 2014).

Por otro lado, la proteína Cas8 es responsable del reconocimiento de PAM en *Cascade* donde múltiples motivos peptídicos dentro de Cas8 leen el surco menor del PAM. Las interacciones específicas entre las secuencias Cas8 y PAM detienen el complejo de vigilancia en los sitios PAM e inician el desenrollado del ADN. Dos motivos adicionales en el complejo *Cascade* de *E. coli* han sido implicados en interacciones no específicas y específicas de PAM. Una horquilla P rica en lisina ayuda a posicionar el ADNdb para la lectura de PAM. Además, dos bucles ricos en lisina en las dos subunidades Cas7 adyacentes de la cadena principal de *Cascade* forman una prensa cargada positivamente que entra en contacto con la cadena principal de ADN mediante de interacciones electrostáticas no específicas. Este motivo es necesario para el muestreo de ADN no específico y para la unión a un objetivo específico, y se cree que mantiene el ADN en su lugar durante el reconocimiento de PAM y el desenrollado del ADN (Xue & Sashital, 2019).

Cuando *Cascade* localiza un sitio PAM, se cree que las interacciones favorables entre Cas8 y el PAM permiten la flexión del ADN doble banda (ADNdb), desestabilizando el ADNdb en la región proximal al PAM. El desenrollado posterior del ADNdb doblado se ve facilitado por la inserción de motivos en forma de cuña en Cas8. Esta desestabilización puede impulsar la fusión local del ADNdb, lo que permite la invasión de la cadena y la formación de un heterodúplex ARNcr-ADN para iniciar la formación del bucle R. Por lo tanto, la pérdida de energía durante la flexión y el desenrollado del ADN se compensa con la formación de pares de bases entre el ARNcr y el ADN diana. Cabe destacar que cada sexto nucleótido del dúplex ARNcr-ADN permanece sin aparear para acomodar el dúplex híbrido enrollado en el canal de unión del ácido nucleico. Una estructura reciente de criomicroscopía electrónica (crio-EM) del complejo *Cascade* en *Thermobifida fusca* capturó un bucle R parcial intermedio que contiene dos tramos de ARNcr-ADN de 5 pb. Con base en esta estructura, se planteó la hipótesis de que el ADNdb puede fundirse en pasos de 6 pb, formando los dúplex más cortos de 5 pb en ruta a la formación completa del bucle R (Xue & Sashital, 2019).

Los cambios conformacionales que ocurren dentro de la subunidad Cas8 tras el bloqueo del bucle R son necesarios para el reclutamiento de Cas3 por *Cascade*. En algunos sistemas de tipo I, Cas3 y Cas8 se fusionan en una sola proteína, lo que indica la importancia de Cas8 en el reclutamiento de Cas3. Una estructura de microscopía electrónica de tinción negativa de la estructura *Cascade*-Cas3 de *E. coli* confirmó la interacción directa entre Cas8 y Cas3 (Hochstrasser et al., 2014). La estructura crio-EM de alta resolución de la estructura *Cascade*-Cas3 de *T. fusca* reveló que la rotación del CTD de Cas8e durante el bloqueo del bucle R es responsable del reclutamiento de Cas3, por medio de la exposición de una región Cas8 que interactúa directamente con Cas3 (Xiao et al., 2018).

Debido a las diversas funciones que desempeña la proteína Cas8 en el mecanismo CRISPR-Cas es de esperar que se observen muchas regiones conservadas. En el análisis de las secuencias se identificaron diversas regiones conservadas en la proteína. No obstante, al evaluar estos sitios en la secuencia nucleotídica se observó una alta variabilidad y en consecuencia las regiones de interés para la identificación del gen *cas8* fueron muy reducidas.

Finalmente, el análisis del gen *cas11* mostró resultados muy variables entre los géneros estudiados. Para los géneros *Cryobacterium* y *Pseudomonas*, se identificaron diferentes regiones conservadas a lo largo del gen y la proteína. Para el género *Rhodococcus* se identificaron pocas regiones conservadas localizadas en el inicio y al final del gen. Finalmente, para el género *Arthrobacter* se observaron grandes diferencias entre los dos grupos: en el grupo 1 se detectaron regiones conservadas en la región inicial del gen, mientras que en el grupo 2 estas regiones se localizaron en la región intermedia del gen.

La proteína Cas11, también denominada Cse2, forma un dímero que interactúa con subunidades de Cas7 mediante puentes salinos para formar el vientre que ayuda a estabilizar el ARNcr y el ADN objetivo. Aunque las subunidades de Cse2 no hacen contacto directo con el ARNcr, los cálculos electrostáticos muestran que ambas caras del dímero de Cse2 están cargadas

positivamente, lo que indica un posible papel de Cse2 en la estabilización de las cadenas unidas y desplazadas del objetivo de ADN (Nam et al., 2012).

En *Cascade*, la cadena principal Cas7 del complejo permanece rígida, mientras que las subunidades Cas11 y Cas8 experimentan cambios para ensanchar el canal de unión al ADN. El dímero Cas11 se desliza hacia la cola del complejo, empujando hacia abajo el dominio C-terminal alfa-helicoidal (CTD) de Cas8, que se mueve hacia el dominio N-terminal que interactúa con PAM (NTD) (van Erp et al., 2015). Debido a la alta variabilidad entre las secuencias y las pocas regiones identificadas para el diseño de oligonucleótidos, es posible que este gen no sea un buen marcador para la detección del sistema CRISPR-Cas.

Hasta ahora se han interpretado los resultados concernientes a los genes de la clase I. Con respecto al sistema CRISPR-Cas de la clase II, la evaluación del gen *cas1* en especies del género *Sphingomonas* mostró grandes diferencias con respecto a la clase I, debido a que se identificaron alrededor de seis regiones en la proteína mediante el alineamiento por COBALT, que fueron potenciales para el diseño de oligonucleótidos, diferentes a las señaladas en el gráfico de polimorfismo y de Alfa (Figura 32c). Dentro de las regiones identificadas, destaca la ubicada entre los aminoácidos 190 y 230, de gran interés debido a que en un estudio de la estructura cristalina de Cas1 en *Streptococcus pyogenes* (SpCas1), que presenta el sistema de clase II, se determinó que en SpCas1 y otros homólogos de Cas1 tipo II, los residuos E149, H205 y E220 (en SpCas1) representan los aminoácidos del sitio potencial de unión al metal catalítico y se conservan en secuencia (Ka et al., 2016). Además, en UniProt se identificó en la especie *Francisella tularensis subsp. novicida* (strain U112) los aminoácidos 157, 222 y 237 como sitios de unión a metales, específicamente a manganeso. Al ser especies tan distantes y dada la alta diversidad del sistema, es posible que la región potencial estudiada para el diseño de oligonucleótidos corresponda a la que codifique para los aminoácidos esenciales para la unión a metales de la Cas1, caracterizada por ser una nucleasa dependiente de metales (Wiedenheft et al., 2009b).

En un estudio sobre la adquisición de espaciadores del sistema CRISPR-Cas de tipo II-C de *Riemerella anatipestifer* (RA-CH-2) (He et al., 2018) se demostró mediante la introducción de un plásmido que tanto Cas1 como Cas2 son necesarios para la adquisición del espaciador del sistema CRISPR-Cas RA-CH-2, en consonancia con el papel informado de Cas1 y Cas2 en los sistemas de tipo I-E y II-A, ya que la eliminación de *cas1* o *cas 2* de RA-CH-2 anuló la adquisición del espaciador.

Del mismo modo que para el gen *cas2* de la clase I, el diseño de oligonucleótidos para este gen en la clase II no fue posible debido a que es un gen muy pequeño y no presentaba sitios óptimos para el diseño de oligos.

El análisis de la proteína Cas9 realizado en COBALT de diferentes especies del género *Sphingomonas* mostró una alta variabilidad entre las secuencias. La Cas9 es una proteína de aproximadamente 1100 aminoácidos, que al ser la única proteína efectora del sistema CRISPR-Cas de la clase II consta de varios dominios.

Un estudio de la estructura de resolución atómica de la proteína Cas9 en la especie *Streptococcus pyogenes* reveló que ésta adopta una arquitectura bilobulada que comprende un lóbulo  $\alpha$ -helicoidal (REC) que media la unión del ácido nucleico (residuos 56–718), y un lóbulo de nucleasa (NUC) (residuos 1–55 y 719–1368) (Nishimasu et al., 2015). El primero contiene los dominios REC-I, REC-II y REC-III que son responsables del reconocimiento de nucleótidos. Una hélice rica en arginina une los dos lóbulos, lo que constituye un ancla para la unión del ARN guía, mientras que los dominios C-terminal (C-term) y de interacción con PAM (PI) de la proteína participan en los procesos de reconocimiento y unión del ADN (Palermo et al., 2016). Por otro lado, el lóbulo de nucleasa tiene dos dominios de endonucleasa, el dominio HNH (residuos 766–909, rico en residuos de histidina y asparagina) y el dominio RuvC (residuos 1–55, 719–765 y 910–1099). El dominio HNH escinde la cadena de ADN objetivo (ADNt), mientras que el dominio RuvC escinde la cadena de ADN no objetivo (ADNnt). Además, el dominio HNH contiene dos regiones en forma de bisagra cerca de sus extremos N-ter y C-ter, los enlazadores L1 y L2, que crean una comunicación cruzada entre los dominios RuvC y HNH (Nishimasu et al., 2015).

El dominio de interacción PAM (PI, residuos 1100-1368) confiere especificidad PAM y, por lo tanto, es responsable de iniciar la unión al ADN. Una vez unido el ADN, los residuos cargados positivamente presentes en la interfaz entre los lóbulos REC y NUC, particularmente en la hélice puente, estabilizan el híbrido ARN guía único (ARNsg) ARNsg-ADN cargado negativamente. Sin embargo, los residuos cargados positivamente presentes en la región de enlace (L1 y L2) entre los dominios RuvC y HNH ayudan a estabilizar el ADNnt desplazado (Nishimasu et al., 2015).

El reconocimiento y la escisión de ADNdb requiere estrictamente la presencia de un motivo adyacente al protoespaciador (PAM) en la cadena de ADN no objetivo y depende de la complementariedad de pares de bases de la cadena de ADN objetivo con la plantilla guía de ARN. Como se mencionó previamente, Cas9 adopta una arquitectura bilobulada general, en la que el heterodúplex ARNsg-ADNt reside dentro de un canal central entre los lóbulos de reconocimiento  $\alpha$ -helicoidal (REC) y nucleasa (NUC), mientras que el ADNnt desplazado se enrosca en un canal lateral dentro del lóbulo NUC. El lóbulo NUC comprende dos dominios de nucleasa dependientes de iones metálicos (denominados HNH) responsables de escindir el ADNt mediante un mecanismo de un ion metálico, y RuvC responsable de escindir el ADNnt por medio de un mecanismo de dos iones metálicos (Zuo & Liu, 2017).

La Cas9, al ser una proteína tan grande, debe tener una conformación adecuada para que los dominios funcionales tengan la posición correcta para unirse a los sustratos. En consecuencia, las regiones no funcionales pueden ser muy variables mientras su ordenamiento permita la ubicación de los aminoácidos implicados en las diferentes funciones de la Cas9. Es quizás por esta razón que en el alineamiento de COBALT se observaron regiones muy variables, pero también regiones pequeñas muy conservadas. Entre las más conservadas resalta una región entre los aminoácidos 130 al 660, que se encontrarían ubicados dentro del lóbulo  $\alpha$ -helicoidal (REC) reportado para la especie *Streptococcus pyogenes* y que se encarga de la unión del ácido nucleico, como se mencionó previamente (Nishimasu et al., 2014). Por lo tanto, podría tratarse de una región muy importante dentro de la proteína Cas9. Sin embargo, se hace necesaria la visualización

del modelo 3D de la proteína Cas9 y el alineamiento de las secuencias aminoacídicas de especies pertenecientes al género *Sphingomonas* junto con la especie *Streptococcus pyogenes*, para confirmar si los sitios donde fueron diseñados los oligonucleótidos se encuentran conservados entre especies filogenéticamente distantes, y para poder conocer su ubicación en la estructura proteica, evaluando así si se trata de sitios de importancia funcional dentro de la proteína.

### 8.3. Los oligonucleótidos diseñados y probados *in silico* para identificar genes diagnósticos de la clase de sistema CRISPR-Cas resultaron ser eficientes y específicos

El sistema de detección del sistema CRISPR-Cas de la clase I mediante un enfoque basado en PCR podría funcionar correctamente, según su evaluación *in silico*, para algunos de los genes en los distintos géneros estudiados. Para el gen *cas1* sólo se diseñaron oligonucleótidos para el género *Rhodococcus*, que cumplieron con los parámetros necesarios para su selección (Tabla A en Anexos). Sin embargo, en las pruebas de Primer-BLAST, estos oligos sólo amplificaron la especie *Rhodococcus sovatisensis*. Teniendo en cuenta que la matriz sólo fue hallada en cinco especies en la base de datos, y dos de ellas son especies no cultivables, es posible que los oligos diseñados para esta proteína no puedan detectar este gen en otras especies.

Como se mencionó anteriormente, la proteína Cas2 no es indispensable para la adquisición de espaciadores (Nuñez et al., 2014), y no está presente en las todas las especies analizadas. Además, tratándose de un gen pequeño (de 300pb) no se tomó en cuenta para el diseño de oligonucleótidos.

Todos los oligos generados para el gen *cas3* en los géneros evaluados, que cumplieran los parámetros requeridos, fueron puestos a prueba mediante Primer-BLAST. Los resultados mostraron la amplificación *in silico* del gen *cas3*, pero no únicamente. Como se describió previamente, el gen *cas3* codifica para una proteína que contiene un dominio nucleasa N-terminal y un dominio C-terminal helicasa/ATPasa, regiones que deben ser muy conservadas para mantener la funcionalidad. A pesar de que los oligos fueron diseñados en base a estos sitios, no se trataba de regiones óptimas para la identificación del gen *cas3* y del sistema CRISPR-Cas, ya que existe una alta probabilidad de que amplifiquen otro tipo de genes que codifiquen para proteínas que compartan las mismas funciones. Por lo tanto, se prescindió del diseño de oligonucleótidos para este gen.

El gen *cas5* resultó ser un buen marcador para la detección del sistema CRISPR-Cas de clase I, ya que algunos oligos permitirían incluso la identificación de especies de géneros diferentes a los estudiados, y no sólo de especies pertenecientes a los géneros *Cryobacterium*, *Rhodococcus* y *Pseudomonas* (Tabla 2). Estos oligos cumplieron los parámetros adecuados para realizar la respectiva prueba de la PCR (Tabla B en Anexos). En el caso del género *Arthrobacter* los oligos diseñados para cada uno de los grupos no cumplieron los parámetros adecuados de temperatura de fusión. A nivel general se ha determinado que el contenido de guaninas y citosinas del ADN de especies *Arthrobacter* examinadas del género varía entre aproximadamente 60 y 72 moles % (Tm)

lo que afecta directamente la temperatura de fusión (Busse, 2016). En consecuencia, este factor afectó el diseño de oligonucleótidos para todos los genes en este género. Cabe destacar que la proteína Cas5 cumple un papel fundamentalmente estructural en el mecanismo CRISPR-Cas y está presente en otros subtipos de la clase I. Por lo tanto, el uso de los oligos diseñados permitiría identificar este gen en bacterias que presenten distintos subtipos de la clase I, y determinar, naturalmente, la presencia del sistema CRISPR-Cas.

Los diferentes oligonucleótidos diseñados para el gen *cas6* no fueron adecuados en los géneros *Arthrobacter*, *Cryobacterium* y *Rhodococcus*, debido a los parámetros de temperatura y formación de dímeros en algunos casos, y la inespecificidad para la identificación *in silico* de este gen. La proteína Cas6 está implicada en el procesamiento de la cadena larga del pre-ARNcr para formar los ARNcr maduros mediante su actividad endorribonucleasa. Por tal motivo, al realizar las pruebas de algunos oligos en Primer-BLAST se logró la amplificación del gen *cas6*, pero también la de otros genes que codifican para proteínas con actividad endorribonucleasa. Al no ser específicos los oligos diseñados, éstos fueron descartados.

En el género *Pseudomonas*, los oligonucleótidos seleccionados se adecuaron a los parámetros requeridos para la identificación del gen mediante la PCR (Tabla C en Anexos) y fueron efectivos ya que se diseñaron en una región intergénica entre los genes *cas5* y *cas6*, y permitieron la identificación de ambos genes en 18 especies incluidas las del género *Pseudomonas*, pero también algunas de otros géneros distintos, todos ellos de la clase de las Gamma-proteobacteria.

Un estudio de la actividad funcional de la proteína Cas6 del subtipo I-E en *Escherichia coli* determinó que la interferencia de CRISPR y la adaptación cebada pueden proceder en ausencia de la actividad endonucleolítica de Cas6e o en ausencia de ésta, cuando el ARNcr maduro se produce a partir de unidades de transcripción diseñadas por medio de la terminación de la transcripción independiente del factor. La proteína Cas6 también parece cumplir un papel en la estabilización de la interacción de ARNcr con el complejo *Cascade*, probablemente mediante la participación en interacciones simultáneas con el mango de repetición 3' de ARNcr y subunidades de *Cascade*. Si bien contribuye a la robustez general de la interferencia CRISPR, esta función es secundaria y, tanto la interferencia como la adaptación pueden ocurrir en ausencia de Cas6 (Semenova et al., 2015). Por lo tanto, el gen *cas6* no representa un marcador diagnóstico ideal para la detección del sistema CRISPR-Cas.

A diferencia de los otros genes que codifican para las proteínas Cas que conforman el complejo multiprotéico, la proteína Cas7 es fundamental para formar la estructura del complejo efector debido a que es necesaria la formación de un hexámero que conforma el esqueleto central (Jackson et al., 2014). Además, esta proteína no sólo interactúa con otras proteínas Cas, sino que también interactúa con el ADN y el ARNcr guía (Xue & Sashital, 2019). Por consiguiente, era de esperarse que muchas regiones estuvieran muy conservadas para garantizar la función de la proteína, y fueran óptimas para el diseño de oligonucleótidos. Lo interesante es que, entre todos los genes estudiados, éste es el único que presentaba regiones conservadas entre los diferentes géneros analizados. Esto podría indicar que se trata de un gen cuya evolución ha sido restringida y sus posibilidades de variación se hayan visto limitadas porque existen sólo unas pocas

constituciones (*i. e.* estructuras terciarias) en las que la proteína resulta funcional. A pesar de que existían sitios altamente conservados a nivel general, se observaron pequeñas diferencias específicas para cada género. Por esta razón, se diseñaron oligos para cada género separadamente y oligos generales (multiespecie), en base a la secuencia consenso de todos los géneros. Todos los oligonucleótidos que cumplieron con los parámetros (Tabla D en Anexos) fueron puestos a prueba *in silico* y se obtuvieron buenos resultados de identificación en una amplia variedad de especies, incluso para especies representantes de diferentes órdenes. También se logró una alta especificidad. Así, el gen *cas7* representa un excelente marcador diagnóstico para la identificación del sistema CRISPR-Cas de la clase I, aunque no es específico del subtipo I-E (Figura 3). No obstante, es necesario conocer si las regiones donde fueron diseñados los oligos corresponden a regiones funcionales en la proteína, para confirmar si existe selección negativa en esos sitios. Para ello, es necesario realizar un modelado molecular en el que se detecten las diferentes regiones.

Para el gen *cas8* se observó algo similar que para el gen *cas6*. En ambos genes se identificaron diversas regiones idóneas para el diseño de oligonucleótidos, pero también se hallaron altas tasas de sustituciones sinónimas. Por lo tanto, no fue posible el diseño de oligonucleótidos en todos los géneros estudiados, pero sí en el género *Pseudomonas*. Estos últimos tuvieron buenos valores de los parámetros requeridos (Tabla E en Anexos), además, resultaron ser altamente específicos y también efectivos, identificando el gen *cas8* en 21 especies diferentes tanto del género *Pseudomonas* como de otros taxones. En consecuencia, este gen representa un buen marcador para la identificación del sistema CRISPR-Cas de clase I (y de distintos subtipos) en las especies mostradas en la Tabla 5.

Finalmente, el gen *cas11* no representó un buen marcador para la identificación del sistema CRISPR-Cas del subtipo I-E, ya que sólo los oligonucleótidos diseñados para el género *Pseudomonas* cumplieron todos los parámetros requeridos para la prueba de la PCR (Tabla E en Anexos) e identificaron pocas especies (Tabla 6). Esta proteína cumple un papel importante al ayudar a estabilizar el ARNcr y el ADN objetivo, pero su rol no tan es fundamental como el de la proteína Cas7.

Los oligonucleótidos diseñados para la identificación del sistema CRISPR-Cas de clase II para el género *Sphingomonas*, fueron muy efectivos. Los oligos diseñados para el gen *cas1* se basaron en una de las regiones funcionales más importantes de la proteína, lo que permitió la identificación de 10 especies bacterianas incluyendo representantes del género *Sphingomonas*. Por lo tanto, parece tratarse de un marcador efectivo para la identificación del sistema. Por otra parte, los oligos diseñados para el gen *cas9*, realizados sobre la región correspondiente al lóbulo  $\alpha$ -helicoidal (REC), resultaron aún mejor identificando el sistema CRISPR-Cas de clase II, ya que permitieron la identificación de este gen en al menos 32 especies diferentes de varios taxones, incluido el género *Sphingomonas*. Ambos pares de oligos mostraron buenos valores de todos los parámetros necesarios para las pruebas de PCR (Tablas G y H en Anexos).

Debido a que los oligos diseñados para la identificación de algunos genes en los géneros estudiados presentan tamaños de amplificadas con una diferencia considerable de pares de bases nucleotídicas y Tm equiparables, es posible plantearse realizar PCRs múltiples, o *PCR multiplex*,

que permitan la identificación de varios genes de la matriz CRISPR-Cas simultáneamente. Por ejemplo, para el género *Cryobacterium* se podrían identificar los genes *cas5* y *cas7* en simultáneo, con una diferencia de ~300 pb en el tamaño de los amplicones, que pueden ser fácilmente diferenciados en geles de agarosa. Sin embargo, la diferencia de temperatura de fusión entre los oligos es superior a 5 °C. Para este género también se podría realizar la identificación por *multiplexing* de los genes *cas11* y *cas7*, pues las temperaturas de fusión de ambos pares de oligos rondan los 64 °C y la diferencia del tamaño de los amplificadores es de ~200 pb. También es posible la identificación de varios genes en el género *Pseudomonas*. Por ejemplo, los genes *cas5* y *cas7*, con diferencias de temperaturas de fusión de ~3 °C entre oligos y con amplicones de más de 100 pb de diferencia, podrían amplificarse simultáneamente. Para el género *Sphingomonas* no es posible la identificación de los genes *cas1* y *cas2* simultáneamente en una PCR múltiple debido a que el tamaño de los amplificadores es muy similar (diferencia de aproximadamente 52 pb) y no podrían ser diferenciados con facilidad en el gel. Las estrategias de utilización de los oligos y su aplicación en PCRs individuales o múltiples van a depender de la fase de estandarización de los protocolos *in vitro* para cada caso.

#### 8.4. El Modelado 3D de las proteínas Cas7 y Cas9 brinda información sobre el papel que juegan las regiones seleccionadas como blanco

El modelado 3D de la proteína Cas7 permitió visualizar los sitios en la especie *Pseudomonas aeruginosa* donde fueron diseñados los oligonucleótidos multigénero. Las regiones de la proteína codificadas por las secuencias complementarias de los oligos *forward* y *reverse* interactúan con la banda de ARN mostrada en la Figura 40b. Específicamente, el sitio donde se ubica el *reverse* es una estructura que fue reportada en la estructura cristalina de la proteína Cas7 para *Escherichia coli*, que tiene forma de pulgar y que está conformada por un bucle de 30 aminoácidos (residuos 193-223).

Las subunidades de Cas7 se unen al ARNcr en una disposición helicoidal dextrógira donde los “pulgares” de Cas7.2 a Cas7.6 se pliegan sobre la parte superior del ARNcr, enroscándose cada seis nucleótidos. La posición de cada pulgar se estabiliza mediante interacciones electrostáticas con la hélice  $\alpha 1$  de RRM en la palma de una molécula adyacente, y se ha demostrado que las mutaciones en el pulgar de las proteínas Cas7 homólogas reducen significativamente las afinidades de unión del ARN (Lintner et al., 2011). No se ha reportado la funcionabilidad del extremo N-ter (donde está ubicado el *forward*) en estudios reportados de la estructura cristalina de la proteína Cas7. Sin embargo, en el modelado de *Pseudomonas aeruginosa* obtenido se pudo observar que esa región interactúa con el ARN. Por lo tanto, los sitios donde fueron diseñados los oligonucleótidos parecen ser de gran importancia para garantizar la función de la proteína Cas7, al estar implicados en la estabilización del ARN.

Con respecto a la proteína Cas9, la visualización de la estructura de la proteína en complejo con ADN y ARNsg de *Streptococcus pyogenes* (Bravo et al., 2022) permitió comprobar los sitios donde se ubican los oligonucleótidos diseñados dentro de la proteína, confirmando que el *forward* se localiza en el lóbulo REC que media la unión del ácido nucleico. La conservación de los

aminoácidos alanina, histidina, arginina y glicina tanto en especies del género *Sphingomonas* como en la especie *Streptococcus pyogenes* podría reflejar su posible participación en garantizar la estabilización del ARNsg y la interacción REC- ARNsg. Para el caso del oligo *reverse*, se pudo evidenciar en el modelo que éste se localiza en el dominio HNH, particularmente en el lóbulo NUC. Este dominio tiene un papel fundamental ya que media la escisión de la cadena de ADN objetivo. Los aminoácidos conservados fueron treonina, leucina y asparagina, intercalados entre seis aminoácidos variables. La ubicación exacta y la disposición de estos aminoácidos podrían indicar que son parte del sitio activo de la actividad nucleasa de la proteína Cas9. En consecuencia, los oligonucleótidos diseñados se encuentran en regiones que parecen ser muy importantes a nivel funcional.

## 9. Conclusiones

- El enfoque *in silico* nos permitió identificar regiones óptimas para el diseño de oligonucleótidos para la identificación del sistema CRISPR-Cas en especies pertenecientes a géneros bacterianos reportados en sustratos del frente glacial del Pico Humboldt (*Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y *Sphingomonas*).
- Los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas* y *Rhodococcus* presentan el sistema CRISPR-Cas de clase I y subtipo I-E y su arquitectura genética muestra el orden del operón canónico reportado para este sistema, con la excepción de algunas especies pertenecientes a los géneros *Arthrobacter* y *Cryobacterium*. Para el género *Sphingomonas* la arquitectura genética obtenida corresponde al sistema CRISPR-Cas de la clase II y subtipo II-C1.
- Hemos reportado por primera vez el sistema CRISPR-Cas de clase I en el género *Arthrobacter*, demostrando que la arquitectura genética del sistema y las secuencias de nucleótidos y aminoácidos son altamente variables para las especies reportadas hasta la fecha para este género.
- La evaluación de las regiones conservadas de cada uno de los genes que conforman la matriz CRISPR-Cas demostró que no todas las proteínas Cas pueden funcionar como blanco para la identificación del sistema. Sin embargo, los genes *cas7* y *cas9* son excelentes marcadores diagnósticos para la identificación de la clase I (subtipo I-E) y la clase II (subtipo II-C1), respectivamente.

## 10. Perspectivas

- Sintetizar los oligos diseñados y confirmar su funcionamiento *in vitro* en distintas especies de los géneros estudiados y en diferentes especies provenientes de diversas regiones geográficas, en especial en aquellas aisladas de diferentes sustratos asociados al glaciar del Pico Humboldt.
- Llevar a cabo la secuenciación de genomas completos en el mayor número de especies posibles de los géneros estudiados, especialmente en representantes de los géneros *Arthrobacter*, *Cryobacterium* y *Rhodococcus* presentes en diferentes ecosistemas, pero particularmente en varios sustratos asociados al glaciar del Pico Humboldt, con el fin de analizar la presencia y configuración del sistema CRISPR-Cas, pero también la presencia de genes implicados en el metabolismo a bajas temperaturas, genes de resistencia y genes que controlan fenotipos de interés ecológico, como la promoción del crecimiento vegetal.
- Aplicar metapangenómica para poder hacer una comparación del sistema CRISPR-Cas y entender cómo varía dependiendo de la especie bacteriana, el bacteriófago y el ambiente.
- Aplicar conjuntamente virómica y metapangenómica en una cantidad masiva de muestras ambientales, especialmente en sustratos asociados a glaciares tropicales, con el fin de dilucidar las redes de interacción entre bacterias y bacteriófagos en ambientes poli-extremos, y para entender la dinámica bacteria-fago en ambientes contrastantes.
- Desarrollar *softwares* y herramientas eficientes que permitan la identificación completa de matrices CRISPR-Cas incluyendo todos los espaciadores, de manera de poder estudiar de forma más sencilla la diversidad de los sistemas hallados, y permitir el desarrollo de una clasificación más detallada y confiable que las existentes.
- Realizar pruebas *in vitro* de la funcionabilidad del sistema CRISPR-Cas en la variedad de subtipos de las clases I y II, en base a construcciones del operón en ausencia de los genes *cas1* y *cas2*, para confirmar su importancia dentro de la adquisición de espaciadores.
- Cristalizar diferentes complejos proteicos efectores (*Cascade*) de representantes de los géneros estudiados para verificar las diferencias estructurales que puedan existir entre especies filogenéticamente separadas.
- Realizar un estudio exhaustivo de todos los dominios funcionales de cada una de las proteínas Cas, generando matrices de comparación y alineamientos que permitan visualizar la diversidad de arquitecturas genéticas y aportar a la comprensión de la evolución de los tipos de sistemas CRISPR-Cas.

## 11. Bibliografía

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493-500. <https://doi.org/10.1038/s41586-024-07487-w>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Amitai, G., & Sorek, R. (2016). CRISPR–Cas adaptation: Insights into the mechanism of action. *Nature Reviews Microbiology*, 14(2), 67-76. <https://doi.org/10.1038/nrmicro.2015.14>
- Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., Chan, A. M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J. M., Mueller, J. E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C. A., & Rohwer, F. (2006). The Marine Viromes of Four Oceanic Regions. *PLoS Biology*, 4(11), e368. <https://doi.org/10.1371/journal.pbio.0040368>
- Asaf, S., Numan, M., Khan, A. L., & Al-Harrasi, A. (2020). *Sphingomonas*: From diversity and genomics to functional role in environmental remediation and plant growth. *Critical Reviews in Biotechnology*, 40(2), 138-152. <https://doi.org/10.1080/07388551.2019.1709793>
- Ball, M. M., Gómez, W., Magallanes, X., Rosales, R., Melfo, A., & Yarzabal, L. A. (2014). Bacteria recovered from a high-altitude, tropical glacier in Venezuelan Andes. *World Journal of Microbiology and Biotechnology*, 30(3), 931-941. <https://doi.org/10.1007/s11274-013-1511-1>
- Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., Qiu, D., Shan, Z., Yang, Z., Yuan, S., Zhang, C., Zhang, X., Liu, B., Kong, F., Li, X., Zhou, X., Tran, L.-S. P., & Cao, D. (2019). CRISPR/Cas9-mediated targeted mutagenesis of GmSPL9 genes alters plant architecture in soybean. *BMC Plant Biology*, 19. <https://doi.org/10.1186/s12870-019-1746-6>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242. <https://doi.org/10.1093/nar/28.1.235>
- Beylveid, G., White, K. M., Ayllon, J., & Shaw, M. L. (2013). New-generation screening assays for the detection of anti-influenza compounds targeting viral and host functions. *Antiviral Research*, 100(1), 120-132. <https://doi.org/10.1016/j.antiviral.2013.07.018>
- Bhattacharjee, R., Nandi, A., Mitra, P., Saha, K., Patel, P., Jha, E., Panda, P. K., Singh, S. K., Dutt, A., Mishra, Y. K., Verma, S. K., & Suar, M. (2022). Theragnostic application of nanoparticle and CRISPR against food-borne multi-drug resistant pathogens. *Materials Today Bio*, 15. <https://doi.org/10.1016/j.mtbio.2022.100291>
- Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research*, 14(5), 988-995. <https://doi.org/10.1101/gr.1865504>
- Bravo, J. P. K., Liu, M.-S., Hibshman, G. N., Dangerfield, T. L., Jung, K., McCool, R. S., Johnson, K. A., & Taylor, D. W. (2022). Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature*, 603(7900), 343-347. <https://doi.org/10.1038/s41586-022-04470-1>
- Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)*, 321(5891), 960-964. <https://doi.org/10.1126/science.1159689>
- Busse, H.-J. (2016). Review of the taxonomy of the genus *Arthrobacter*, emendation of the genus *Arthrobacter sensu lato*, proposal to reclassify selected species of the genus *Arthrobacter* in the novel genera *Glutamicibacter* gen. Nov., *Paeniglutamicibacter* gen. Nov., *Pseudoglutamicibacter* gen. Nov., *Paenarthrobacter* gen. Nov. And *Pseudarthrobacter* gen. Nov., and emended description of *Arthrobacter roseus*. *International Journal of Systematic and Evolutionary Microbiology*, 66(1), 9-37. <https://doi.org/10.1099/ijsem.0.000702>
- Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR–Cas systems. *Nucleic Acids Research*, 42(10), 6091-6105. <https://doi.org/10.1093/nar/gku241>
- Cousin, M. A. (1999). PSEUDOMONAS | Introduction. En R. K. Robinson (Ed.), *Encyclopedia of Food Microbiology* (pp. 1864-1867). Elsevier. <https://doi.org/10.1006/rwfm.1999.1290>
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, 46(W1), W246-W251. <https://doi.org/10.1093/nar/gky425>
- Dastager, S. G., Lee, J.-C., Ju, Y.-J., Park, D.-J., & Kim, C.-J. (2008). *Cryobacterium mesophilum* sp. Nov., a novel mesophilic bacterium. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 58(5), 1241-1244. <https://doi.org/10.1099/ijss.0.65650-0>

- De Oliveira Luz, A. C., Da Silva, J. M. A., Rezende, A. M., De Barros, M. P. S., & Leal-Balbino, T. C. (2019). Analysis of direct repeats and spacers of CRISPR/Cas systems type I-F in Brazilian clinical strains of *Pseudomonas aeruginosa*. *Molecular Genetics and Genomics*, 294(5), 1095-1105. <https://doi.org/10.1007/s00438-019-01575-7>
- De Oliveira Luz, A. C., Da Silva Junior, W. J., Do Nascimento Junior, J. B., Da Silva, J. M. A., De Queiroz Balbino, V., & Leal-Balbino, T. C. (2021). Genetic characteristics and phylogenetic analysis of Brazilian clinical strains of *Pseudomonas aeruginosa* harboring CRISPR/Cas systems. *Current Genetics*, 67(4), 663-672. <https://doi.org/10.1007/s00294-021-01173-4>
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602-607. <https://doi.org/10.1038/nature09886>
- Díaz-Muñoz, S. L., & Koskella, B. (2014). Chapter Four—Bacteria–Phage Interactions in Natural Environments. En S. Sariaslani & G. M. Gadd (Eds.), *Advances in Applied Microbiology* (Vol. 89, pp. 135-183). Academic Press. <https://doi.org/10.1016/B978-0-12-800259-9.00004-4>
- Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, 346(6213), 1258096. <https://doi.org/10.1126/science.1258096>
- Garrido-Sanz, D., Redondo-Nieto, M., Martín, M., & Rivilla, R. (2020). Comparative Genomics of the *Rhodococcus* Genus Shows Wide Distribution of Biodegradation Traits. *Microorganisms*, 8(5), 774. <https://doi.org/10.3390/microorganisms8050774>
- Garside, E. L., Schellenberg, M. J., Gesner, E. M., Bonanno, J. B., Sauder, J. M., Burley, S. K., Almo, S. C., Mehta, G., & MacMillan, A. M. (2012). Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA*, 18(11), 2020-2028. <https://doi.org/10.1261/rna.033100.112>
- Gobbetti, M., & Rizzello, C. G. (2014). *Arthrobacter*. En *Encyclopedia of Food Microbiology* (pp. 69-76). Elsevier. <https://doi.org/10.1016/B978-0-12-384730-0.00009-4>
- Gomila, M., Peña, A., Mulet, M., Lalucat, J., & García-Valdés, E. (2015). Phylogenomics and systematics in *Pseudomonas*. *Frontiers in Microbiology*, 6, 129093. <https://doi.org/10.3389/fmicb.2015.00214>
- He, L., John James, M. St., Radovic, M., Ivancic-Bace, I., & Bolt, E. L. (2020). Cas3 Protein—A Review of a Multi-Tasking Machine. *Genes*, 11(2), 208. <https://doi.org/10.3390/genes11020208>
- He, Y., Wang, M., Liu, M., Huang, L., Liu, C., Zhang, X., Yi, H., Cheng, A., Zhu, D., Yang, Q., Wu, Y., Zhao, X., Chen, S., Jia, R., Zhang, S., Liu, Y., Yu, Y., & Zhang, L. (2018). Cas1 and Cas2 From the Type II-C CRISPR-Cas System of *Riemerella anatipestifer* Are Required for Spacer Acquisition. *Frontiers in Cellular and Infection Microbiology*, 8. <https://doi.org/10.3389/fcimb.2018.00195>
- Heinrichs, M. E., Piedade, G. J., Popa, O., Sommers, P., Trubl, G., Weissenbach, J., & Rahlff, J. (2023). Breaking the Ice: A Review of Phages in Polar Ecosystems. *Methods in Molecular Biology (Clifton, N.J.)*, 2738, 31-71. [https://doi.org/10.1007/978-1-0716-3549-0\\_3](https://doi.org/10.1007/978-1-0716-3549-0_3)
- Henderson, H. (2024). *CRISPR Clinical Trials: A 2024 Update*. <https://innovativegenomics.org/news/crispr-clinical-trials-2024/>
- Heo, L., Lee, H., & Seok, C. (2016). GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. *Scientific Reports*, 6, 32153. <https://doi.org/10.1038/srep32153>
- Hille, F., & Charpentier, E. (2016). CRISPR-Cas: Biology, mechanisms and relevance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1707), 20150496. <https://doi.org/10.1098/rstb.2015.0496>
- Hochstrasser, M. L., Taylor, D. W., Bhat, P., Guegler, C. K., Sternberg, S. H., Nogales, E., & Doudna, J. A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18), 6618-6623. <https://doi.org/10.1073/pnas.1405079111>
- Hryhorowicz, M., Lipiński, D., & Zeyland, J. (2023). Evolution of CRISPR/Cas Systems for Precise Genome Editing. *International Journal of Molecular Sciences*, 24(18), 14233. <https://doi.org/10.3390/ijms241814233>
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 169(12), 5429-5433.
- Jackson, R. N., Golden, S. M., van Erp, P. B. G., Carter, J., Westra, E. R., Brouns, S. J. J., van der Oost, J., Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science (New York, N.Y.)*, 345(6203), 1473-1479. <https://doi.org/10.1126/science.1256328>
- Jauregui, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., Carbonnelle, E., Lortholary, O., Clermont, O., Denamur, E., Picard, B., Nassif, X., & Brisse, S. (2008). Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics*, 9(1), 560. <https://doi.org/10.1186/1471-2164-9-560>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), 816-821. <https://doi.org/10.1126/science.1225829>

- Ka, D., Lee, H., Jung, Y.-D., Kim, K., Seok, C., Suh, N., & Bae, E. (2016). Crystal Structure of *Streptococcus pyogenes* Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System. *Structure*, *24*(1), 70-79. <https://doi.org/10.1016/j.str.2015.10.019>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059-3066. <https://doi.org/10.1093/nar/gkf436>
- Koonin, E. V., Gootenberg, J. S., & Abudayyeh, O. O. (2023). Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. *Biochemistry*, *62*(24), 3465-3487. <https://doi.org/10.1021/acs.biochem.3c00159>
- Koskella, B., & Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*, *38*(5), 916-931. <https://doi.org/10.1111/1574-6976.12072>
- LaBauve, A. E., & Wargo, M. J. (2012). Growth and Laboratory Maintenance of *Pseudomonas aeruginosa*. *Current protocols in microbiology*, *0* 6, Unit-6E.1. <https://doi.org/10.1002/9780471729259.mc06e01s25>
- Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., & Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, *520*(7548), 505-510. <https://doi.org/10.1038/nature14302>
- Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M. J., White, M. F., & Lawrence, C. M. (2011). Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). *The Journal of Biological Chemistry*, *286*(24), 21643-21656. <https://doi.org/10.1074/jbc.M111.238485>
- Liu, Q., Yang, L.-L., & Xin, Y.-H. (2023). Diversity of the genus *Cryobacterium* and proposal of 19 novel species isolated from glaciers. *Frontiers in Microbiology*, *14*, 1115168. <https://doi.org/10.3389/fmicb.2023.1115168>
- Liu, W., Li, L., Jiang, J., Wu, M., & Lin, P. (2021). Applications and challenges of CRISPR-Cas gene-editing to disease treatment in clinics. *Precision Clinical Medicine*, *4*(3), 179-191. <https://doi.org/10.1093/pcmedi/pbab014>
- Liu, Y., Shen, L., Zeng, Y., Xing, T., Xu, B., & Wang, N. (2020). Genomic Insights of *Cryobacterium* Isolated From Ice Core Reveal Genome Dynamics for Adaptation in Glacier. *Frontiers in Microbiology*, *11*. <https://doi.org/10.3389/fmicb.2020.01530>
- Liu, Z., Jiang, W., Kim, C., Peng, X., Fan, C., Wu, Y., Xie, Z., & Peng, F. (2023). A *Pseudomonas* Lysogenic Bacteriophage Crossing the Antarctic and Arctic, Representing a New Genus of Autographiviridae. *International Journal of Molecular Sciences*, *24*(8), 7662. <https://doi.org/10.3390/ijms24087662>
- Llambí, L. D., Melfo, A., Gámez, L. E., Pelayo, R. C., Cárdenas, M., Rojas, C., Torres, J. E., Ramírez, N., Huber, B., & Hernández, J. (2021). Vegetation Assembly, Adaptive Strategies and Positive Interactions During Primary Succession in the Forefield of the Last Venezuelan Glacier. *Frontiers in Ecology and Evolution*, *9*, 657755. <https://doi.org/10.3389/fevo.2021.657755>
- Luo, Y., Zhou, M., Zhao, Q., Wang, F., Gao, J., Sheng, H., & An, L. (2020). Complete genome sequence of *Sphingomonas* sp. Cra20, a drought resistant and plant growth promoting rhizobacteria. *Genomics*, *112*(5), 3648-3657. <https://doi.org/10.1016/j.ygeno.2020.04.013>
- Madsen, E. L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Current Opinion in Biotechnology*, *22*(3), 456-464. <https://doi.org/10.1016/j.copbio.2011.01.008>
- Majidzadeh, M., & Fatahi-Bafghi, M. (2018). Current taxonomy of *Rhodococcus* species and their role in infections. *European Journal of Clinical Microbiology & Infectious Diseases*, *37*(11), 2045-2062. <https://doi.org/10.1007/s10096-018-3364-x>
- Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods in molecular biology (Clifton, N.J.)*, *1311*, 47-75. [https://doi.org/10.1007/978-1-4939-2687-9\\_4](https://doi.org/10.1007/978-1-4939-2687-9_4)
- Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2013). The basic building blocks and evolution of CRISPR-Cas systems. *Biochemical Society transactions*, *41*(6), 1392-1400. <https://doi.org/10.1042/BST20130038>
- Makarova, K. S., Wolf Yuri I., Iranzo Jaime, & Shmakov Sergey A. (2019). *Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants*. <https://doi.org/10.1038/s41579-019-0299-x>
- Margesin, R., & Collins, T. (2019). Microbial ecology of the cryosphere (glacial and permafrost habitats): Current knowledge. *Applied Microbiology and Biotechnology*, *103*(6), 2537-2549. <https://doi.org/10.1007/s00253-019-09631-3>
- Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature*, *526*(7571), 55-61. <https://doi.org/10.1038/nature15386>
- Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in *Staphylococci* by Targeting DNA. *Science (New York, N.Y.)*, *322*(5909), 1843-1845. <https://doi.org/10.1126/science.1165771>
- Mingma, T. S., Ishfaq Nabi Najar, Sayak Das, & Nagendra Thakur. (2019). The Diversity of *Pseudomonas* species from the Accumulation Zone of Kanchengayao Glacier, North Sikkim, India. *J Pure Appl Microbiol*, *5493*. <https://dx.doi.org/10.22207/JPAM.13.1.37>
- Mojica, F. J., Juez, G., & Rodríguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Molecular Microbiology*, *9*(3), 613-621. <https://doi.org/10.1111/j.1365-2958.1993.tb01721.x>
- Morita, R. Y. (1975). Psychrophilic bacteria. *Bacteriological Reviews*, *39*(2), 144-167.

- Morita, R. Y., & Moyer, C. L. (2001). Psychrophiles. En *Encyclopedia of Biodiversity* (pp. 298-303). Elsevier. <https://doi.org/10.1016/B978-0-12-384719-5.00176-3>
- Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science (New York, N.Y.)*, *345*(6203), 1479-1484. <https://doi.org/10.1126/science.1256996>
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics*, *8*(7), e1002764. <https://doi.org/10.1371/journal.pgen.1002764>
- Nam, K. H., Huang, Q., & Ke, A. (2012). Nucleic acid binding surface and dimer interface revealed by CRISPR-associated CasB protein structures. *FEBS letters*, *586*(22), 10.1016/j.febslet.2012.09.041. <https://doi.org/10.1016/j.febslet.2012.09.041>
- Navarrete, R., Aranda, S., Rodríguez, M. de L., Moya, S. L., González, M. G., Navarrete Maya, R., Aranda Ocampo, S., Rodríguez Mejía, M. de L., Moya Hernández, S. L., & González Ochoa, M. G. (2014). Bacterias Fitopatógenas en Semillas: Su Detección y Regulación. *Revista mexicana de fitopatología*, *32*(2), 75-88.
- NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8-D13. <https://doi.org/10.1093/nar/gkx1095>
- Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5269-5273.
- Nishimasu, H., Cong, L., Yan, W. X., Ran, F. A., Zetsche, B., Li, Y., Kurabayashi, A., Ishitani, R., Zhang, F., & Nureki, O. (2015). Crystal structure of Staphylococcus aureus Cas9. *Cell*, *162*(5), 1113-1126. <https://doi.org/10.1016/j.cell.2015.08.007>
- Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S., Dohmae, N., Ishitani, R., Zhang, F., & Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, *156*(5), 935-949. <https://doi.org/10.1016/j.cell.2014.02.001>
- Nishimura, K., Ikarashi, M., Yasuda, Y., Sato, M., Cano Guerrero, M., Galipon, J., & Arakawa, K. (2021). Complete Genome Sequence of Sphingomonas paucimobilis Strain Kira, Isolated from Human Neuroblastoma SH-SY5Y Cell Cultures Supplemented with Retinoic Acid. *Microbiology Resource Announcements*, *10*(6), e01156-20. <https://doi.org/10.1128/MRA.01156-20>
- Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*, *527*(7579), 535-538. <https://doi.org/10.1038/nature15760>
- Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology*, *21*(6), 528-534. <https://doi.org/10.1038/nsmb.2820>
- Ochman, H., & Wilson, A. C. (1987). Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *Journal of Molecular Evolution*, *26*(1-2), 74-86. <https://doi.org/10.1007/BF02111283>
- Palermo, G., Miao, Y., Walker, R. C., Jinek, M., & McCammon, J. A. (2016). Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Central Science*, *2*(10), 756-763. <https://doi.org/10.1021/acscentsci.6b00218>
- Papadopoulos, J. S., & Agarwala, R. (2007). COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics (Oxford, England)*, *23*(9), 1073-1079. <https://doi.org/10.1093/bioinformatics/btm076>
- Parra-Sánchez, Á., Antequera-Zambrano, L., Martínez-Navarrete, G., Zorrilla-Muñoz, V., Paz, J. L., Alvarado, Y. J., González-Paz, L., & Fernández, E. (2023). Comparative Analysis of CRISPR-Cas Systems in Pseudomonas Genomes. *Genes*, *14*(7), Article 7. <https://doi.org/10.3390/genes14071337>
- Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., & Göker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, *70*(11), 5607-5612. <https://doi.org/10.1099/ijsem.0.004332>
- Pessi, I. S., Viitamäki, S., Virkkala, A.-M., Eronen-Rasimus, E., Delmont, T. O., Marushchak, M. E., Luoto, M., & Hultman, J. (2022). In-depth characterization of denitrifier communities across different soil ecosystems in the tundra. *Environmental Microbiome*, *17*, 30. <https://doi.org/10.1186/s40793-022-00424-2>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612. <https://doi.org/10.1002/jcc.20084>
- Pittino, F., Ambrosini, R., Seeger, M., Azzoni, R. S., Diolaiuti, G., Alviz Gazitua, P., & Franzetti, A. (2023). Geographical variability of bacterial communities of cryoconite holes of Andean glaciers. *Scientific Reports*, *13*(1), Article 1. <https://doi.org/10.1038/s41598-022-24373-5>
- Reimer, L. C., Sardà Carbasse, J., Koblitz, J., Ebeling, C., Podstawka, A., & Overmann, J. (2021). BacDive in 2022: The knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Research*, *50*(D1), D741-D746. <https://doi.org/10.1093/nar/gkab961>

- Rozas, J., & Rozas, R. (1995). DnaSP, DNA sequence polymorphism: An interactive program for estimating population genetics parameters from DNA sequence data. *Computer Applications in the Biosciences: CABIOS*, 11(6), 621-625. <https://doi.org/10.1093/bioinformatics/11.6.621>
- Sävström, C., Anesio, M. A., Granéli, W., & Laybourn-Parry, J. (2007). Seasonal Viral Loop Dynamics in Two Large Ultraoligotrophic Antarctic Freshwater Lakes. *Microbial Ecology*, 53(1), 1-11. <https://doi.org/10.1007/s00248-006-9146-5>
- Semenova, E., Kuznedelov, K., Datsenko, K. A., Boudry, P. M., Savitskaya, E. E., Medvedeva, S., Beloglazova, N., Logacheva, M., Yakunin, A. F., & Severinov, K. (2015). The Cas6e ribonuclease is not required for interference and adaptation by the E. coli type I-E CRISPR-Cas system. *Nucleic Acids Research*, 43(12), 6049-6061. <https://doi.org/10.1093/nar/gkv546>
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J.-L., & Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, 31(8), 686-688. <https://doi.org/10.1038/nbt.2650>
- Soliman, M., Said, H. S., El-Mowafy, M., & Barwa, R. (2022). Novel PCR detection of CRISPR/Cas systems in *Pseudomonas aeruginosa* and its correlation with antibiotic resistance. *Applied Microbiology and Biotechnology*, 106(21), 7223-7234. <https://doi.org/10.1007/s00253-022-12144-1>
- Suzuki, K.-I., Sasaki, J., Uramoto, M., Nakase, T., & Komagata, K. (1997). *Cryobacterium psychrophilum* gen. Nov., sp. Nov., nom. Rev., comb. Nov., an Obligately Psychrophilic Actinomycete To Accommodate «*Curtobacterium psychrophilum*» Inoue and Komagata 1976. *International Journal of Systematic Bacteriology*, 47(2), 474-478. <https://doi.org/10.1099/00207713-47-2-474>
- Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, 123(3), 585-595.
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, 38(7), 3022-3027. <https://doi.org/10.1093/molbev/msab120>
- Tang, T.-H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., Elge, T., Brosius, J., & Hüttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(11), 7536-7541. <https://doi.org/10.1073/pnas.112047299>
- The UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523-D531. <https://doi.org/10.1093/nar/gkac1052>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115. <https://doi.org/10.1093/nar/gks596>
- van Erp, P. B. G., Jackson, R. N., Carter, J., Golden, S. M., Bailey, S., & Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Research*, 43(17), 8381-8391. <https://doi.org/10.1093/nar/gkv793>
- Venkatachalam, S., Kannan, V. M., Saritha, V. N., Loganathachetti, D. S., Mohan, M., & Krishnan, K. P. (2021). Bacterial diversity and community structure along the glacier foreland of Midtre Lovénbreen, Svalbard, Arctic. *Ecological Indicators*, 126, 107704. <https://doi.org/10.1016/j.ecolind.2021.107704>
- Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2023). The conserved domain database in 2023. *Nucleic Acids Research*, 51(D1), D384-D388. <https://doi.org/10.1093/nar/gkac1096>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296-W303. <https://doi.org/10.1093/nar/gky427>
- Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution*, 35(3), 773-777. <https://doi.org/10.1093/molbev/msx335>
- Weiland-Bräuer, N. (2021). Friends or Foes—Microbial Interactions in Nature. *Biology*, 10(6). <https://doi.org/10.3390/biology10060496>
- Wheatley, R. M., & MacLean, R. C. (2021). CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. *The ISME Journal*, 15(5), Article 5. <https://doi.org/10.1038/s41396-020-00860-3>
- Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., & Doudna, J. A. (2009a). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure (London, England: 1993)*, 17(6), 904-912. <https://doi.org/10.1016/j.str.2009.03.019>

- Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., & Doudna, J. A. (2009b). Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense. *Structure*, 17(6), 904-912. <https://doi.org/10.1016/j.str.2009.03.019>
- Wolf, Y. I., Makarova, K. S., Yutin, N., & Koonin, E. V. (2012). Updated clusters of orthologous genes for Archaea: A complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biology Direct*, 7, 46. <https://doi.org/10.1186/1745-6150-7-46>
- Xiao, Y., Luo, M., Dolan, A. E., Liao, M., & Ke, A. (2018). Structure Basis for RNA-guided DNA degradation by Cascade and Cas3. *Science (New York, N.Y.)*, 361(6397), eaat0839. <https://doi.org/10.1126/science.aat0839>
- Xue, C., & Sashital, D. G. (2019). Mechanisms of type I-E and I-F CRISPR-Cas systems in Enterobacteriaceae. *EcoSal Plus*, 8(2), 10.1128/ecosalplus.ESP-0008-2018. <https://doi.org/10.1128/ecosalplus.ESP-0008-2018>
- Yan, Z.-Z., Chen, Q.-L., Li, C.-Y., Thi Nguyen, B.-A., Zhu, Y.-G., He, J.-Z., & Hu, H.-W. (2021). Biotic and abiotic factors distinctly drive contrasting biogeographic patterns between phyllosphere and soil resistomes in natural ecosystems. *ISME Communications*, 1(1), 1-9. <https://doi.org/10.1038/s43705-021-00012-4>
- Yarzabal, L. A. (2020). Perspectives for using glacial and periglacial microorganisms for plant growth promotion at low temperatures. *Applied Microbiology and Biotechnology*, 104(8), 3267-3278. <https://doi.org/10.1007/s00253-020-10468-4>
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134. <https://doi.org/10.1186/1471-2105-13-134>
- Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Research*, 40(12), 5569-5576. <https://doi.org/10.1093/nar/gks216>
- Zhang, Q., & Ye, Y. (2017). Not all predicted CRISPR-Cas systems are equal: Isolated cas genes and classes of CRISPR like elements. *BMC Bioinformatics*, 18(1). <https://doi.org/10.1186/s12859-017-1512-4>
- Zhang, S., Zhang, R., Gao, J., Gu, T., Song, G., Li, W., Li, D., Li, Y., & Li, G. (2019). Highly Efficient and Heritable Targeted Mutagenesis in Wheat via the Agrobacterium tumefaciens-Mediated CRISPR/Cas9 System. *International Journal of Molecular Sciences*, 20(17), 4257. <https://doi.org/10.3390/ijms20174257>
- Zuo, Z., & Liu, J. (2017). Structure and Dynamics of Cas9 HNH Domain Catalytic State. *Scientific Reports*, 7, 17271. <https://doi.org/10.1038/s41598-017-17578-6>

www.bdigital.ula.ve

## 12. Anexos

### Especie

#### **Arthrobacter**

<i>A.sp. YJM1</i>	NZ_JAVALS010000004
<i>A.woluwensis strain NM-35</i>	PVZ001000014
<i>A.sp. MYb227</i>	NZ_PCPT01000002
<i>A.sp. HMSC06H05</i>	LWOA01000037
<i>A.sp. ES1 BJQ89_9</i>	NZ_MQTO01000009
<i>A.oryzae strain TNB02</i>	NZ_RBED01000045
<i>A.sp. isolate S7A_St_54</i>	NZ_JAUNVV010000017
<i>A.sp. Y-9</i>	CP121463
<i>A.sp. isolate ERR9968748</i>	NZ_CAUQMX010000012
<i>A.sp. efr-133-R2A-120</i>	NZ_JARVRF010000001
<i>A.sp. HMSC08H08</i>	NZ_KV789394
<i>A.sp. isolate CPoEBzCyWp</i>	CALUBV000000000
<i>A.albus</i>	JAVDXX010000001
<i>A.cumminsii</i>	QFWG01000008
<i>A.globiformis</i>	NZ_MYFE01000548
<i>A.sp. ISL-28 ISL28_10</i>	JAGGPJ010000009

#### **Cryobacterium**

<i>C.frigoriphilum</i>	SOHE01000078
<i>C.sp. TMS1-13-1</i>	SOGX01000010
<i>C.sinapicolor strain TMT1-23-1</i>	SOGQ01000054
<i>C.roopkundense strain RuG17</i>	JPXF01000100
<i>C.roopkundense strain DSM 21065</i>	JACHBQ010000002

#### **Pseudomonas**

<i>P.aeruginosa strain 11987-2-5</i>	QHLE01000005
<i>P.aeruginosa strain 62 IPC1175_3</i>	NZ_RXCB01000003
<i>P.aeruginosa strain AR442</i>	CP029090
<i>P.aeruginosa strain BT1 2.2 NODE</i>	NZ_WIXQ01000007
<i>P.aeruginosa strain DR1</i>	NZ_KN716069
<i>P.aeruginosa strain GOM1</i>	SDVY01000011
<i>P.aeruginosa strain KT962</i>	NZ_JAREJS010000002
<i>P.aeruginosa strain LMG5031 IPC878_16</i>	NZ_NSST01000016
<i>P.aeruginosa strain N15-01092</i>	CP012901

### Especie

<i>P.aeruginosa strain P37 P37_9</i>	NZ_JBALIT010000009
<i>P.aeruginosa strain RNC3</i>	NZ_JAVFAV010000639
<i>P.aeruginosa strain S1f NODE_5</i>	NZ_NRFP01000005
<i>P.aeruginosa strain SJU-S79_1</i>	NZ_NRFA01000001
<i>P.aeruginosa strain SF416</i>	CP125288
<i>P.aeruginosa strain strain KUD2</i>	NZ_CP138577
<i>P.aeruginosa strain YB01</i>	CP028132
<i>P.benzenivorans strain HNIBRBA3361</i>	CP137892
<i>P.chengduensis strain GD03775 4</i>	NZ_JAOCGN010000004
<i>P.delhiensis strain RLD-1</i>	NZ_FZPC01000008
<i>P.dryadis strain P26B</i>	NZ_QJUM01000003
<i>P.fluorescens strain NCTC10783</i>	LR134300
<i>P.fluorescens strain S613 AN403</i>	LJXB01000040
<i>P.indica isolate RTH1 c_000000000020</i>	NZ_JAHKQV010000149
<i>P.indoloxydans strain JCM 14246 75</i>	NZ_QASO01000075
<i>P.jinjuensis strain JCM 21621</i>	NZ_FNIJ01000002
<i>P.kuykendallii isolate S2_009_000_R2_77</i>	NZ_QFOH01000006
<i>P.lopnurensis strain AL-54 98</i>	NZ_JADDIX010000098
<i>P.mendocina strain EF27 EF27-38</i>	VRYE01000038
<i>P.oleovorans strain GD03705 28</i>	NZ_JAOCJD010000028
<i>P.oleovorans strain GD04132</i>	CP104579
<i>P.oligotrophica strain JM10B5a</i>	NZ_JAKJRU010000001
<i>P.otitidis strain BC12</i>	CP102328
<i>P.otitidis strain GD04077</i>	NZ_JAOEBU010000025
<i>P.otitidis strain TL17</i>	CP126342
<i>P.sp. 1-7 pse1-7-32</i>	JPRQ01000032
<i>P.sp. 3PA37B6</i>	NZ_SCNO01000052
<i>P.sp. A29</i>	NZ_JAULRW010000012
<i>P.sp. AN-1</i>	CP139195
<i>P.sp. CW003PS</i>	CP123623
<i>P.sp. ENNP23</i>	LYSP01000013
<i>P.sp. FeN3W</i>	CP136010
<i>P.sp. FRB 230</i>	NZ_QJUN01000026
<i>P.sp. GD03721</i>	CP104583
<i>P.sp. Hp2</i>	NZ_VAHJ01000004

**Especie**

*P.sp.* NBRC 100443  
*P.sp.* phDV1  
*P.sp.* s4  
*P.thermotolerans* DSM 14292 H165DRAFT  
*P.thermotolerans* isolate SRR10912534  
*P.wenzhouensis*

**Rhodococcus**

*R.sovatensis* strain 18930  
*R.yunnanensis* strain G21638-S1  
*R.sp.* ABRD24  
*R.sp.* IEGM 1366  
*R.sp.* 1168

**Sphingomonas**

*S.colombiensis* MAG\_27 (*Candidatus*)  
*S.bacterium*  
*S.changbaiensis* NBRC 104936  
*S.fennica* strain K101 C583  
*S.gilva* strain ZDH117

**Número de acceso**

NZ\_BSQE01000005  
 CP031606  
 NZ\_JAWNFO010000048  
 NZ\_KB822604  
 NZ\_CAMYXR010000024  
 NZ\_JAUDTJ010000022

CP147846  
 NZ\_JAPWGS010000015  
 CP035319  
 NZ\_JAWLRZ010000018  
 MKKY01000036

CP119315  
 NZ\_CAHJWX010000020  
 NZ\_BBWU01000015  
 PHHF01000075  
 QWLV01000002

**Especie**

*S.horti* strain MAH-20 GCM70011337\_1  
*S.koreensis* strain S12  
*S.liriopis* strain RP10  
*S.naphthae* strain KACC 18716  
*S.oligoaromativorans*  
*S.paucimobilis* strain ZJSH1  
*S.pseudosanguinis* strain DSM 19512  
*S.sanxanigenens* isolate S2\_018\_000  
*S.soli* NBRC 100801  
*S.sp.* 3P27F8  
*S.sp.* A2-49 24  
*S.sp.* CFBP8993  
*S.sp.* CGMCC 13658  
*S.sp.* FARSPH  
*S.sp.* HITSZ\_GF NODE  
*S.sp.* isolate SMAG\_U3283 k141\_449853  
*S.sp.* MMO-176  
*S.sp.* PB1R3 Y681\_8  
*S.sp.* S17  
*S.sp.* S-NIH.Pt1\_0416 NODE\_5  
*S.sp.* SRS2  
*S.sp.* UBA978  
*S.taxi* isolate S2\_005\_001\_R1\_22

**Número de acceso**

NZ\_WQMS01000001  
 RCZN01000006  
 NZ\_JAMLDY010000033  
 CP117411  
 NZ\_JAASQK010000008  
 CP070367  
 NZ\_JAFHKV010000022  
 QFNN01000139  
 NZ\_BCTZ01000003  
 NZ\_SCNL01000007  
 NZ\_JAJLPA010000024  
 NZ\_JAXCWU010000002  
 NZ\_JACEIA010000007  
 CP029985  
 NZ\_JARQXA01000003  
 NZ\_DASTMA010000014  
 NZ\_JBBCUQ010000004  
 NZ\_JAXOJH010000008  
 NZ\_AFGG01000041  
 QLJL01000005  
 NZ\_LARW01000050  
 NZ\_DBBS01000021  
 QFQI01000010

Figura SS.- Diagrama de flujo de la evaluación e identificación de sitios de interés para el diseño de oligonucleótidos para la identificación de los genes cas del sistema CRISPR-Cas.

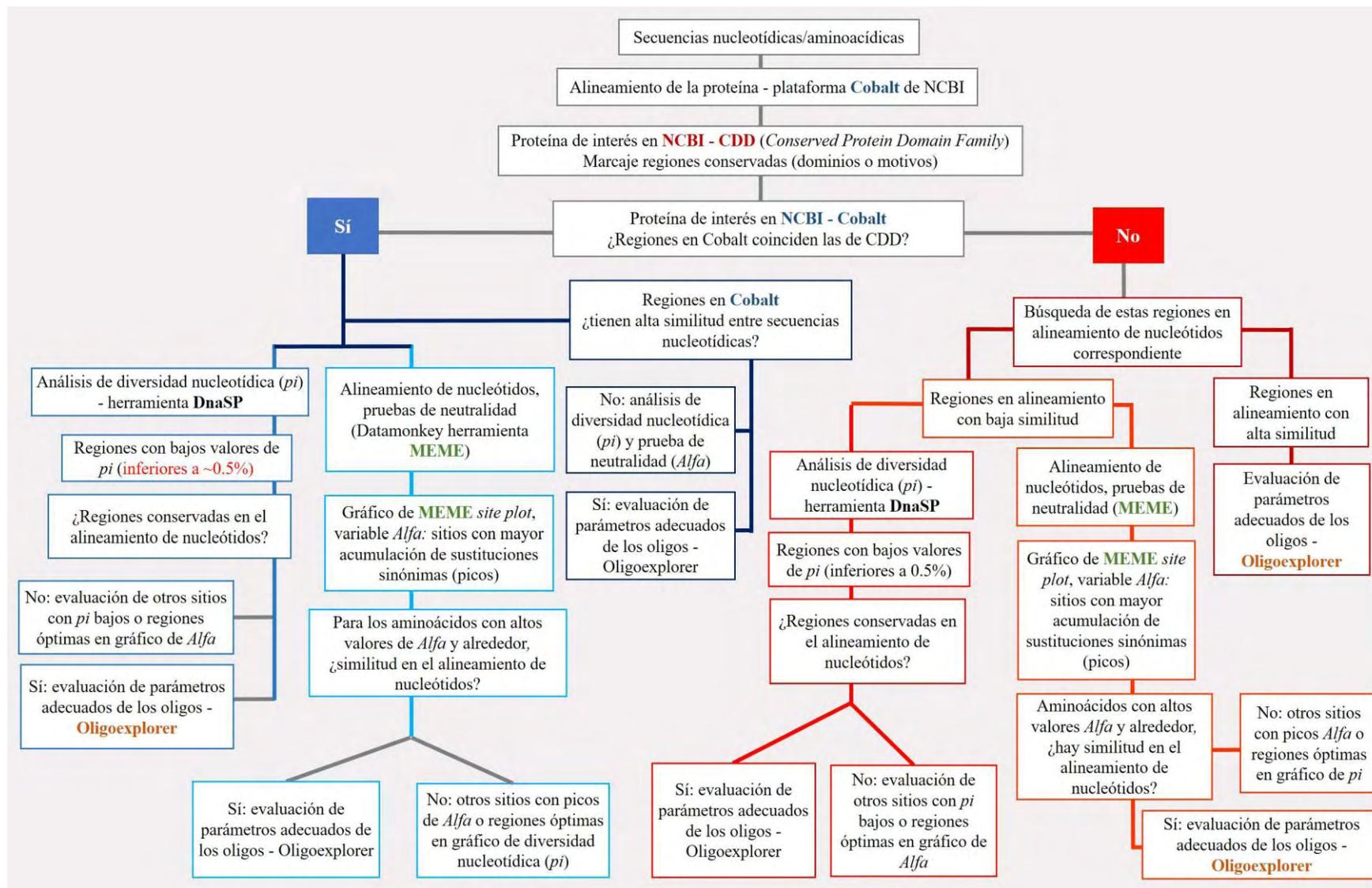


Tabla I.- Características de los oligonucleótidos diseñados para la identificación del gen *cas1* en el género *Rhodococcus*.

<b>Oligos Gen <i>cas1</i></b>										
<b>Género</b>	<b>Amplicón (pb)</b>	<b>Dímeros entre oligos</b>	<b>Forward</b>				<b>Reverse</b>			
<i>Rhodococcus</i>	307	-2,50 kcal/mol	GTGCTCTTCTCCTCGGCC				TGCTCCCGATAGCTCTTCCG			
	<b>Porcentaje de especies identificadas (%)</b>		<b>Posición (nt)</b>	<b>Temperatura de fusión (°C)</b>	<b>Dímeros de oligos</b>	<b>Porcentaje de GC (%)</b>	<b>Posición (nt)</b>	<b>Temperatura de fusión (°C)</b>	<b>Dímeros de oligos</b>	<b>Porcentaje de GC (%)</b>
	20		±167-185	63,8	-1,56 kcal/mol	68,4	±454-473	62,5	-1,56 kcal/mol	60

Tabla J.- Características de los oligonucleótidos diseñados para la identificación del gen *cas5* en los géneros *Cryobacterium*, *Pseudomonas* y *Rhodococcus*.

Oligos Gen <i>cas5</i>										
Género	Amplicón (pb)	Dímeros entre oligos	Forward				Reverse			
<i>Cryobacterium</i>	309	-2,44 kcal/mol	TCTGGGCSAGAGTCGACC				TGCGACCCARRTACGTCATG			
	Porcentaje de especies identificadas (%)		Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	75		±188-205	60,7	-3,12 kcal/mol	66,7	±477-496	58,4	-1,54 kcal/mol	50
Género	Amplicón (pb)	Dímeros entre oligos	Forward				Reverse			
<i>Pseudomonas</i>	194	-1,58 kcal/mol	CARTCCTGGGGCACCACC				GTGGCGGTCTGGTARTCGC			
	Porcentaje de especies identificadas (%)		Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	41		±37-54	60,7	-1,77 kcal/mol	66,7	±212-230	63,8	-1,77 kcal/mol	68,4
Género	Amplicón (pb)	Dímeros entre oligos	Forward				Reverse			
<i>Rhodococcus</i>	287	-1,79 kcal/mol	ACGAAGAGCGGVGTGCTCG				CCBAGGTAGAGCGGGAAKGT			
	Porcentaje de especies identificadas (%)		Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	20		±91-109	61,6	-1,55 kcal/mol	63,2	±358-378	62,5	-1,73 kcal/mol	60

Tabla K.- Características de los oligonucleótidos diseñados para la identificación de la región intermedia entre los genes *cas5* y *cas6* en el género *Pseudomonas*.

Oligos Gen <i>cas6</i>										
Género	Amplicón (pb)	Dímeros entre oligos	<i>Forward</i>				<i>Reverse</i>			
<i>Pseudomonas</i>	365	-1,53 kcal/mol	ACCTGACCAGACTGACSCCTTGAC				CTCGCCCACCAGVCCATAG			
	Porcentaje de especies identificadas (%)		Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	40		±5-27	66,4	-0,17 kcal/mol	56,5	±348-366	63,8	0	68,4

Tabla L.- Características de los oligonucleótidos diseñados para la identificación del gen *cas7* en los géneros *Arthrobacter*, *Cryobacterium*, *Pseudomonas*, *Rhodococcus* y oligonucleótidos multigénero.

Oligos Gen <i>cas7</i>										
Género	Amplicón (pb)	Dímeros entre oligos	Forward				Reverse			
<i>Arthrobacter</i>	822	-3,3 kcal/mol	CCTCAACCGKGACGACACCR				GCGGTGGMGTTCTTCTTGS			
	Porcentaje de especies identificadas (%)		Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	43,75		±45-63	63,34	-0,87 kcal/mol	65	±839-857	62,6	0	63,2
Género	Amplicón (pb)	Dímeros entre oligos	Forward				Reverse			
<i>Cryobacterium</i>	639	-1,5 kcal/mol	GCCCTGGCACAACCTSAATC				GAGCGGTAGTAGACGCCYGACG			
	Porcentaje de especies identificadas (%)		Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	60		±45-64	61,6	-1,5 kcal/mol	60	±668-689	64,6	-1,77 kcal/mol	63,6
Género	Tamaño de amplificado (nt)	Dímeros entre oligos	Forward				Reverse			
<i>Pseudomonas</i>	404	-1,13 kcal/mol	SCTCAAGCTCAAGGACGAYG				CGGARTTGAACTCCACCTGG			
	Porcentaje de especies identificadas (%)		Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	51,7		±294-313	60,5	-1,56 kcal/mol	55	±678-697	60,5	-1,3 kcal/mol	55

<b>Género</b>	<b>Amplicón (pb)</b>	<b>Dímeros entre oligos</b>	<b>Forward</b>				<b>Reverse</b>			
<i>Rhodococcus</i>	655	-1,80 kcal/mol	GAGCAATMTCAACCGTGACG				CAGGGTGGAKGAGTTGAACT			
	<b>Porcentaje de especies identificadas (%)</b>		<b>Posición (nt)</b>	<b>Temperatura de fusión (°C)</b>	<b>Dímeros de oligos</b>	<b>Porcentaje de GC (%)</b>	<b>Posición (nt)</b>	<b>Temperatura de fusión(°C)</b>	<b>Dímeros de oligos</b>	<b>Porcentaje de GC (%)</b>
	40		±42-67	58,4	-0,87 kcal/mol	50	±676-696	60,5	-0,84 kcal/mol	55
<b>Género</b>	<b>Amplicón (pb)</b>	<b>Dímeros entre oligos</b>	<b>Forward</b>				<b>Reverse</b>			
<i>Multigénero</i>	586	-2,41 kcal/mol	ACCTCAACCGNGACGACACC				TCGACGGCGGTGWAGTAGTC			
	<b>Porcentaje de especies identificadas (%)</b>		<b>Posición (nt)</b>	<b>Temperatura de fusión (°C)</b>	<b>Dímeros de oligos</b>	<b>Porcentaje de GC (%)</b>	<b>Posición (nt)</b>	<b>Temperatura de fusión (°C)</b>	<b>Dímeros de oligos</b>	<b>Porcentaje de GC (%)</b>
	67		±50-69	62,5	-0,87 kcal/mol	60	±613-632	62,5	-1,77 kcal/mol	60

Tabla M.- Características de los oligonucleótidos diseñados para la identificación del gen *cas8* en el género *Pseudomonas*.

Oligos Gen <i>cas8</i>									
Género	Amplicón (pb)	Dímeros entre oligos	Forward			Reverse			
<i>Pseudomonas</i>	774	-1,74 kcal/mol	GGSTGTTCCATCCGCAGTACC			CAGCAGCTTGGCCTGRTCRCCT			
	Porcentaje de especies identificadas (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)
	37,9	±299-319	62,5	-0,90 kcal/mol	60	±1051-1071	63,3	-3,03 kcal/mol	57,1

www.bdigital.ula.ve

Tabla N.- Características de los oligonucleótidos diseñados para la identificación del gen *cas11* en los géneros *Cryobacterium* y *Pseudomonas*.

Oligos Gen <i>cas11</i>										
Género	Amplicón (pb)	Dímeros entre oligos	<i>Forward</i>				<i>Reverse</i>			
<i>Cryobacterium</i>	400	-1,73 kcal/mol	CCGAGCTGCAAGCCTTCAC				CCGGTGTCCCAGAACGTGAG			
	Porcentaje de especies identificadas (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	
	40	±14-32	64,6	-1,92 kcal/mol	65	±406-425	64,6	=3,27 kcal/mol	65	
Género	Amplicón (pb)	Dímeros entre oligos	<i>Forward</i>				<i>Reverse</i>			
<i>Pseudomonas</i>	226	-1,73 kcal/mol	CCGGCYTATCCSTATGTGGA				ACACCGMCAGGTTTTCCG			
	Porcentaje de especies identificadas (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	
	34,48	±112-131	60,54	-3,27 kcal/mol	60	±323-340	60,59	-0,60 kcal/mol	61,1	

Tabla O.- Características de los oligonucleótidos diseñados para la identificación del gen *cas1* en el género *Sphingomonas*.

Oligos Gen <i>cas1</i>										
Género	Amplicón (pb)	Dímeros entre oligos	Forward				Reverse			
<i>Sphingomonas</i>	534	-1,77 kcal/mol	CSCACGGCRTCACCTGGTC				SGCGGAACGGYTTCGATCAVA			
	Porcentaje de especies identificadas (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	
	28,5	±134-152	66	-1,77 kcal/mol	73,7	±648-668	64,6	-1,77 kcal/mol	65	

Tabla P.-Características de los oligonucleótidos diseñados para la identificación del gen *cas1* en el género *Sphingomonas*.

Oligos Gen <i>cas9</i>										
Género	Amplicón (pb)	Dímeros entre oligos	Forward				Reverse			
<i>Sphingomonas</i>	482	-1,80 kcal/mol	TBTTCACCTCAAMCAGCGS				RAABAGCGGATGCGCCTTS			
	Porcentaje de especies identificadas (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	Posición (nt)	Temperatura de fusión (°C)	Dímeros de oligos	Porcentaje de GC (%)	
	35,7	±383-402	62,5	-0,03 kcal/mol	60	±846-864	61,6	-3,34 kcal/mol	63,2	