



UNIVERSIDAD
DE MURCIA

Práctica Análisis de Genomas Bacterianos

Analizar la presencia de sistemas CRISPR-Cas en las cepas
Vibrio metoecus YB9D03 y YB5B06

Máster en Bioinformática - Facultad de Biología



Francisco Javier López Carbonell

01/03/2024

Índice

1. Introducción locus CRISPR-Cas bacterianos	1
2. Cálculo de ANI entre las dos cepas	3
3. Tipos de sistemas CRISPR-Cas	4
3.1. Cepa V. metoecus YB9D03	5
3.2. Cepa V. metoecus YB5B06	5
4. Localización genómica sistemas CRISPR-Cas	6
4.1. Cepa V. metoecus YB9D03	6
4.2. Cepa V. metoecus YB5B06	6
5. Composición de sistemas CRISPR-Cas	7
5.1. Secuencia repetida en el locus CRISPR	7
5.2. Número de espaciadores	7
5.3. Dianas de espaciadores	7
6. Conclusiones	8
7. Referencias	9

Índice de figuras

1.	Esquema sistemas CRISPR-Cas	1
2.	Criterios de búsqueda genomas NCBI	1
3.	Selección de genomas de cepas de interés	2
4.	Genomas de cepas de interés	2
5.	<i>ANI pairwise</i>	3
6.	Página web de <i>CRISPRCasFinder</i>	4
7.	CRISPR count cepa YB9D03	4
8.	CRISPR count cepa YB5B06	4
9.	Secuencia FASTA scaffold CRISPR-Cas	5
10.	Locus YB9D03	5
11.	Detalles CRISPR-Cas	5
12.	Locus YB5B06	5
13.	Detalles CRISPR-Cas	5
14.	Localización genómica en cepa YB9D03	6
15.	Localización genómica en cepa YB5B06	6
16.	Targets de espaciadores CRISPR en cepa YB9D03	7
17.	Targets de espaciadores CRISPR en cepa YB5B06	8

Índice de tablas

1.	Cálculo del ANI de las dos cepas	3
----	--	---

1. Introducción locus CRISPR-Cas bacterianos

Los sistemas CRISPR-Cas son sistemas de defensa inmunitaria adaptativa presentes en bacterias y arqueas que les permiten defenderse contra material genético invasor, como virus y plásmidos. CRISPR es un acrónimo de *Clustered Regularly Interspaced Short Palindromic Repeats* (Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas), que son secuencias de ADN que contienen copias de ADN de secuencias virales previamente encontradas. Cas, por otro lado, se refiere a las proteínas asociadas a CRISPR, que son las responsables de degradar el ADN invasor.

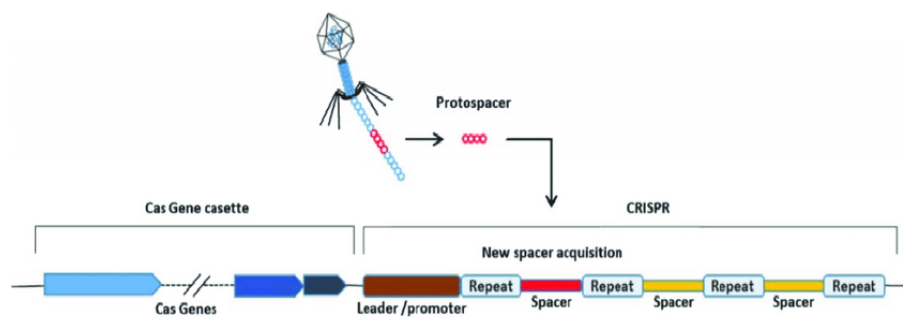


Figura 1: Esquema sistemas CRISPR-Cas

El ejercicio consistirá en conseguir la secuencia genómica de ambas cepas de la misma especie bacteriana y analizarlas a nivel de similitud. Además, analizaremos la presencia de posibles sistemas de defensa frente a fagos CRISPR-Cas contenidos en ambos genomas de interés.

Para conseguir la secuencia genómica de ambas cepas, accederemos a NCBI [1] y buscaremos por *Genome*, indicando *Vibrio metoecus* y obteniendo una lista de genomas de diferentes cepas

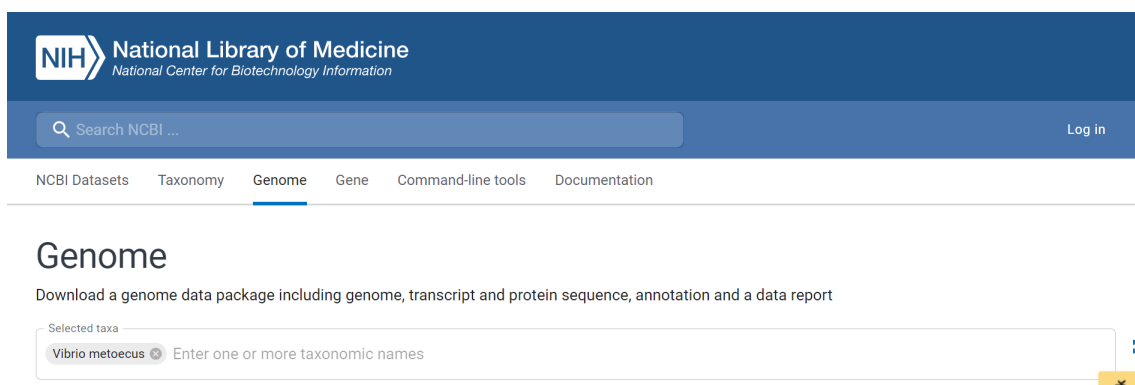


Figura 2: Criterios de búsqueda genomas NCBI

El próximo paso será descargarlas en formato FASTA, donde obtendremos dos zip con las secuencias GCF (genoma completo y finalizado) y GCA (genoma sujeto a revisiones y actualizaciones de anotación).

<input type="checkbox"/>	ASM140234v1	YB5B06 (strain)	NCBI RefSeq Submitter	3.938	Scaffold	Oct, 2015	LBGQ01
<input type="checkbox"/>	ASM140251v1	YB9D03 (strain)	NCBI RefSeq Submitter	3.963	Scaffold	Oct, 2015	LBGR01

Figura 3: Selección de genomas de cepas de interés

Una vez descargados ya tendríamos nuestros genomas de *Vibrio metoecus* YB9D03 y YB5B06 en formato FASTA.

Por otro lado, también podemos acceder de nuevo a IMG (*Integrated Microbial Genomes and Microbiomes*), donde tenemos disponibles todos los genomas secuenciados (completamente o en curso) por el DOE-JGI junto con otros genomas disponibles públicamente. El IMG proporciona un conjunto de herramientas para el análisis comparativo de genes, genomas y funciones. En esta página web podremos encontrar de nuevo el genoma de *Vibrio metoecus* YB9D03 y YB5B06, mediante la opción *Genome Search*. [2]

Select	Domain ▲	Sequencing Status	Study Name	Genome Name / Sample Name	Sequencing Center	IMG Genome ID	Genome Size * assembled	Gene Count * assembled
<input type="checkbox"/>	Bacteria	Permanent Draft	Genome sequencing of <i>Vibrio metoecus</i> and <i>Vibrio cholerae</i> strains co-occurring in the environment	Vibrio metoecus YB5B06	University of Alberta	2693429609	3938456	3626
<input type="checkbox"/>	Bacteria	Permanent Draft	Genome sequencing of <i>Vibrio metoecus</i> and <i>Vibrio cholerae</i> strains co-occurring in the environment	Vibrio metoecus YB9D03	University of Alberta	2693429636	3963180	3646

Figura 4: Genomas de cepas de interés

De igual manera, es posible obtener el genoma en formato FASTA para incorporarlo a otras páginas en próximas cuestiones planteadas.

2. Cálculo de ANI entre las dos cepas

El ANI pairwise (Average Nucleotide Identity) es una medida de similitud genómica utilizada para comparar dos genomas bacterianos. Este método se encarga de comparar las secuencias nucleotídicas de los genomas y determina el porcentaje promedio de identidad entre los nucleótidos correspondientes en ambos genomas. [2]

En este caso, ha sido realizado mediante la herramienta “ANI pairwise” de IMG, donde previamente hemos buscado los genomas de ambas cepas y los hemos añadido a lo que se denomina “*Genome CART*”.

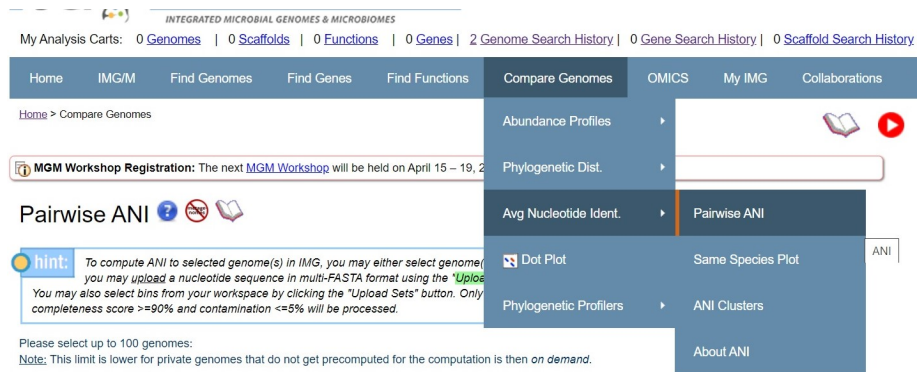


Figura 5: ANI pairwise

Los resultados son mostrados, a continuación:

ID	Genome1 ID	Genome1 Name	Genome2 ID	Genome2 Name	ANI1-2	ANI2-1	AF1-2	AF2-1	Precomputed
2,693E+09	2693429609	Vibrio metoecus YB5B06	2693429636	Vibrio metoecus YB9D03	97,7221	97,7222	92,098	92,098	No

Tabla 1: Cálculo del ANI de las dos cepas

Un valor de ANI en torno al 95-96 % o superior, se considera típicamente un umbral para definir especies bacterianas, aunque esto puede variar dependiendo del contexto y del grupo taxonómico específico.

Los resultados obtenidos de ANI revelan una alta similitud entre ambos genomas con casi un 98 % de identidad nucleotídica y un factor de cobertura de alineamiento del 92 %.

3. Tipos de sistemas CRISPR-Cas

Para encontrar los posibles sistemas CRISPR-Cas, haremos uso de la página web *CRISPRCasFinder*, una herramienta bioinformática diseñada para identificar y analizar sistemas CRISPR-Cas en secuencias de ADN. Esta herramienta, ha sido desarrollada por un equipo de científicos y bioinformáticos en el Laboratorio de Biología de Sistemas y Bioinformática (SysBio) del Instituto Pasteur de París, Francia. [3]

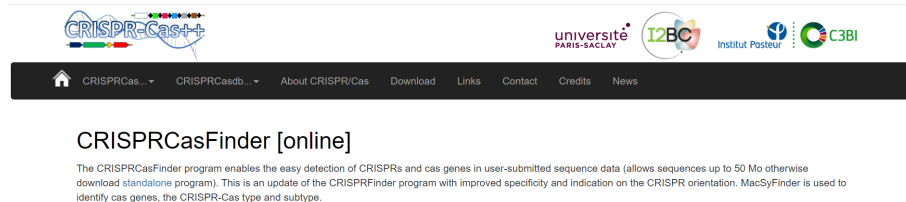


Figura 6: Página web de *CRISPRCasFinder*

La búsqueda se podría haber llevado a cabo de dos formas.

Por un lado, proporcionando el genoma descargado de ambas cepas anteriormente en formato FASTA, dejando por defecto el resto de parámetros de búsqueda y de forma automática obtendremos si contiene sistemas CRISPR-Cas y su clasificación.

En este caso, accedemos al genoma directamente en IMG y nos encontramos con varias opciones. Entramos a la opción *Statistics*, donde vamos a encontrar una sección llamada *CRISPR count*, que nos llevará al *scaffold* donde se sitúa el locus CRISPR. En ambos casos nos encontramos con un sistema CRISPR-Cas. El siguiente paso será obtener la clasificación de los sistemas a nivel del tipo de nucleasas Cas que poseen y las características de ambos locus CRISPR.

DNA scaffolds	64	100.00%
CRISPR Count	1	

Figura 7: CRISPR count cepa YB9D03

DNA scaffolds	71	100.00%
CRISPR Count	1	

Figura 8: CRISPR count cepa YB5B06

En este momento, accederemos a *Get Nucleotide Sequence For Range*, que nos facilitará la secuencia en FASTA sin cabecera (ponemos una que nos identifique correctamente cepa y scaffold). El próximo paso es llevarla a *CRISPRCasFinder*.

Scaffold Nucleotide Sequence for Range

```
>2693451267 Vibrio metoecus YB9D03 : Ga0112260_102 241170..293209
>2693451267 Vibrio metoecus YB9D03 : Ga0112260_102 241170..293209
CTGGCTTAGTGATCAGTAACCTCTTACTGCAAGTCGTGACAGTCGGTATC
GCACGCCCATGGGTGATGGTGAGAACGACGCGCTATCTGGCTGAACGCAC
AGCTGTGATTGGCGATATGGATGCACTCAAATCGACCGATCAATCTTCAG
-----
```

Figura 9: Secuencia FASTA scaffold CRISPR-Cas

En la página de CRISPRCasFinder, elegimos la opción de introducir manualmente la secuencia en formato FASTA y pegamos la secuencia obtenida en el paso anterior. Corremos CRISPRCasFinder y obtenemos los siguientes resultados.

3.1. Cepa *V. metoecus* YB9D03

Para el caso de la cepa YB9D03, encontramos un locus CRISPR, con un nivel de evidencia 4 y un cluster Cas asociado con dos genes, con una posición de comienzo y fin similares. La figura 11 muestra el tipo de proteínas Cas, en este caso según la clasificación de Makarova et al. 2020 [4], vemos proteínas Cas1 y Cas2 de clase 1 (engloba tipos I, III, IV) y clase 2 (engloba tipo II, V, VI).

Element	CRISPR Id / Cas Type	Start	End	Spacer / Gene
Cas cluster	CAS	21745	24843	2
CRISPR	2693451267_Vibrio_metoecus_YB9D03__Ga0112260_102_241170_293209_1	25001	27039	28

Figura 10: Locus YB9D03

Repeat consensus / cas genes	Direction	Evidence Level
cas1_TypeI-II-III, cas2_TypeI-II-III		
GCTTCATTAGATAAGGTCGTTGAGACCGTTGAGAC	ND	4

Figura 11: Detalles CRISPR-Cas

3.2. Cepa *V. metoecus* YB5B06

Para el caso de la cepa YB5B06, encontramos un locus CRISPR, con un nivel de evidencia 4, un cluster Cas y un cluster Cas tipo-IE, asociados con 8 y 7 genes respectivamente, con una posición de comienzo y fin cercanas. En concreto, en la figura 13 aparecen proteínas Cas 1-7, excepto 4 tanto de tipo I como de tipo IE y proteínas asociadas a CRISPR de la familia Cse.

Parámetros como la secuencia repetida y el número de espaciadores de ambos locus, serán comentados en la sección 5 de este informe.

Element	CRISPR Id / Cas Type	Start	End	Spacer / Gene
Cas cluster	CAS	16674	24957	8
Cas cluster	CAS-TypeIE	19289	24957	7
CRISPR	2693446271_Vibrio_metoecus_YB5B06__Ga0112259_114_11507_64158_1	25001	27651	43

Figura 12: Locus YB5B06

Repeat consensus / cas genes	Direction	Evidence Level
cas2_TypeIE, cas3_TypeI, cas5_TypeIE, cas6_TypeIE, cas7_TypeIE, cse1_TypeIE, cse2_TypeIE		
cas1_TypeIE, cas2_TypeIE, cas5_TypeIE, cas6_TypeIE, cas7_TypeIE, cse1_TypeIE, cse2_TypeIE		
GTCTCCCCACGACAGTGGGGGTGTTTC	ND	4

Figura 13: Detalles CRISPR-Cas

4. Localización genómica sistemas CRISPR-Cas

Para analizar la localización de los locus CRISPR-Cas en ambas cepas, debemos atender a los elementos que se sitúan aguas arriba y aguas abajo de ambos sistemas y compararlos entre si. La comprobación se ha realizado de nuevo en IMG [2], obteniendo la representación en color de dichos elementos, tras entrar a los scaffolds correspondientes, como se ha indicado en la sección anterior.

4.1. Cepa *V. metoecus* YB9D03

En este caso, encontramos genes que codifican proteínas que contribuyen en el transporte de lipopolisacaridos fuera de la célula o aminopeptidasas, aunque por lo general la anotación no es especialmente precisa en los elementos que rodean el locus CRISPR, encontrando varias *hypothetical protein*.

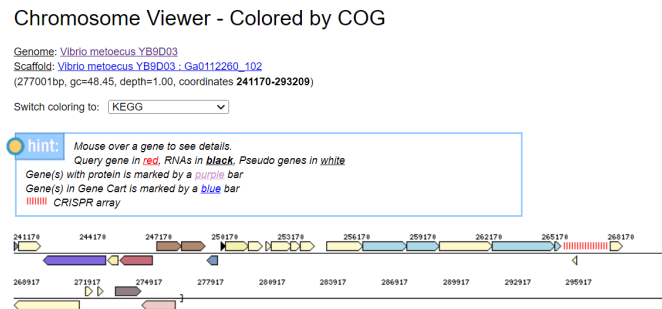


Figura 14: Localización genómica en cepa YB9D03

4.2. Cepa *V. metoecus* YB5B06

En este caso, encontramos genes que codifican proteínas de tipo integrasa, proteínas de transporte o relacionadas con el metabolismo de la lisina, entre otras.

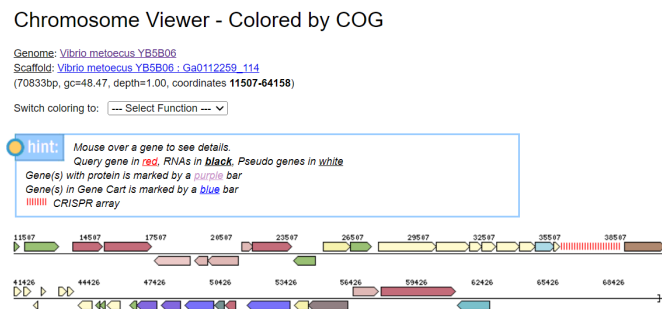


Figura 15: Localización genómica en cepa YB5B06

Por tanto, al no estar situados elementos genéticos similares alrededor de ambos locus, podemos decir que no se encuentran en la misma localización genómica.

5. Composición de sistemas CRISPR-Cas

La composición de los sistemas CRISPR-Cas viene dada por la salida de CRISPRCas-Finder como vimos anteriormente, donde nos aporta la secuencia repetida, el número de espaciadores y la secuencia en formato FASTA de esos espaciadores que será de gran utilidad para encontrar posibles dianas.

5.1. Secuencia repetida en el locus CRISPR

En la figura 11 y 13, podemos ver cuál es la secuencia repetida (DR consensus) en el locus CRISPR para ambas cepas, que corresponden a GCTTCATTA-GATAAGGTCGTTTCAGACCGTTGAGAC para YB9D03 y GTCTTCCCCACG-CAGGTGGGGGTGTTTC para YB5B06.

5.2. Número de espaciadores

Los espaciadores de los sistemas CRISPR, reconocen secuencias específicas y guían a las nucleasas Cas para cortar y degradar esos elementos génicos exógenos. En el caso de *Vibrio metoecus* YB9D03 y YB5B06, el numero total de estos espaciadores es 28 (figura 11) y 43 (figura 13) respectivamente, como se puede ver en las imágenes indicadas (Number of spacers).

5.3. Dianas de espaciadores

Para obtener las dianas, previamente nos descargaremos la secuencia de espaciadores en formato FASTA llamada *Display spacers* en la página de CRISPRCas-Finder. A continuación, accedemos CRISPR-target [5], que es una herramienta usada para la detección de dianas de los espaciadores CRISPR. En la primera sección cargamos los espaciadores descargados anteriormente y obtenemos las dianas. Nos salen un total de 21 y 83 dianas en total para las cepas YB9D03 y YB5B06 respectivamente, de las cuales veremos dos de cada cepa para simplificar el informe:

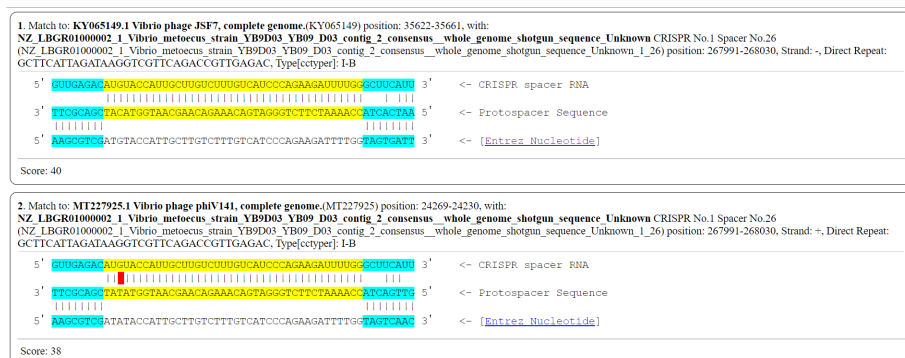


Figura 16: Targets de espaciadores CRISPR en cepa YB9D03

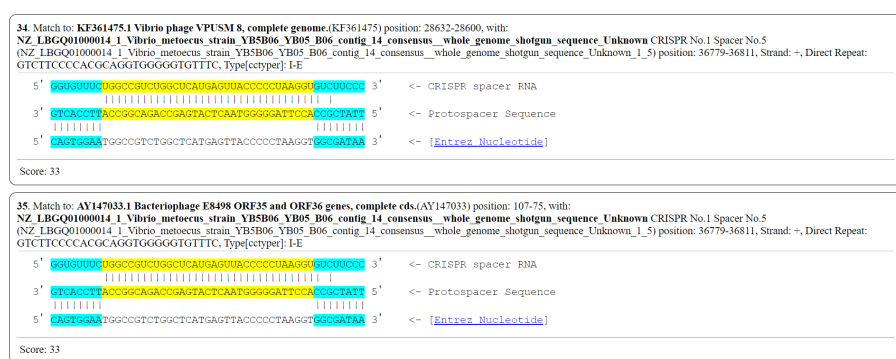


Figura 17: Targets de espaciadores CRISPR en cepa YB5B06

Como podemos observar, en ambos casos las dianas pertenecen a fagos, la mayoría a fagos no cultivados. Esto se debe a que las bacterias son atacadas por muchos virus que secuestran la célula bacteriana para propagarse, lo que puede resultar en la lisis y muerte de ellas.

Por tanto, las bacterias han desarrollado diversas estrategias para contrarrestar la depredación del virus, una de las cuales es la repetición palindrómica corta agrupada regularmente interespaciada (CRISPR) y el sistema de defensa inmunitaria de proteínas asociadas a CRISPR (Cas).

Dicho con otras palabras, los sistemas CRISPR-Cas han sido desarrollados por las bacterias para protegerse de estos ataques de bacteriófagos.

Además, esta familia de bacterias Vibrionaceae, son organismos marinos que se encuentran con una gran cantidad de fagos, por lo que necesitan diversas dianas (hasta 21 y 83 en este caso) para poder defenderse de ellos mediante este sistema. [6]

6. Conclusiones

Podemos concluir mencionando que, tras la realización de esta práctica, hemos podido familiarizarnos con páginas web muy útiles en análisis de genomas microbianos donde tenemos acceso a una gran cantidad de información y herramientas. Además, hemos explorado la capacidad de incorporar como ficheros de entrada a una herramienta, los ficheros de salida de otra, combinando ambas entre sí.

7. Referencias

- [1] F. D. Orata, N. A. S. Hussain, K. Y. H. Liang, D. Hu, and Y. F. Boucher, “Genomes of *Vibrio metoecus* co-isolated with *Vibrio cholerae* extend our understanding of differences between these closely related species,” *Gut Pathogens*, vol. 14, p. 42, Nov. 2022.
- [2] I.-M. A. Chen, K. Chu, K. Palaniappan, A. Ratner, J. Huang, M. Huntemann, P. Hajek, S. Ritter, C. Webb, D. Wu, N. Varghese, T. B. K. Reddy, S. Mukherjee, G. Ovchinnikova, M. Nolan, R. Seshadri, S. Roux, A. Visel, T. Woyke, E. Eloefadros, N. Kyrpides, and N. Ivanova, “The IMG/M data management and analysis system v.7: content updates and new features,” *Nucleic Acids Research*, vol. 51, pp. D723–D732, Jan. 2023.
- [3] D. Couvin, A. Bernheim, C. Toffano-Nioche, M. Touchon, J. Michalik, B. Néron, E. P. C. Rocha, G. Vergnaud, D. Gautheret, and C. Pourcel, “CRISPRCasFinder, an update of CRISPRfinder, includes a portable version, enhanced performance and integrates search for Cas proteins,” *Nucleic Acids Research*, vol. 46, pp. W246–W251, July 2018.
- [4] K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. J. Brouns, E. Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, R. Barrangou, and E. V. Koonin, “Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants,” *Nature Reviews. Microbiology*, vol. 18, pp. 67–83, Feb. 2020.
- [5] A. Biswas, J. N. Gagnon, S. J. Brouns, P. C. Fineran, and C. M. Brown, “CRISPRTarget,” *RNA Biology*, vol. 10, pp. 817–827, May 2013.
- [6] P. C. Kirchberger, “Population dynamics of *Vibrio cholerae* and its close relative *Vibrio metoecus* in an aquatic ecosystem,”