



CENTRO DE INVESTIGACIONES BIOLÓGICAS  
DEL NOROESTE, S.C.

---

---

Programa de Estudios de Posgrado

Genómica comparativa de bacterias patógenas  
halófilas aisladas de granjas acuícolas  
de Baja California Sur.

T E S I S

Que para obtener el grado de

Maestra en Ciencias

Uso, Manejo y Preservación de los Recursos Naturales  
(Orientación en Biotecnología)

P r e s e n t a

**Anais Lucia Lucero Olachea**

La Paz, Baja California Sur, Marzo de 2020

## ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 09 horas del día 20 del Mes de Febrero del 2020, se procedió por los abajo firmantes, miembros de la Comisión Revisora de Tesis avalada por la Dirección de Estudios de Posgrado y Formación de Recursos Humanos del Centro de Investigaciones Biológicas del Noroeste, S. C., a liberar la Tesis de Grado titulada:

**"Genómica comparativa de bacterias patógenas halófilas aisladas de granjas acuícolas de Baja California Sur"**

Presentada por el alumno:

**ANAIS LUCIA LUCERO OLACHEA**

Aspirante al Grado de MAESTRO EN CIENCIAS EN EL USO, MANEJO Y PRESERVACIÓN DE LOS RECURSOS NATURALES CON ORIENTACIÓN EN **BIOTECNOLOGIA**.

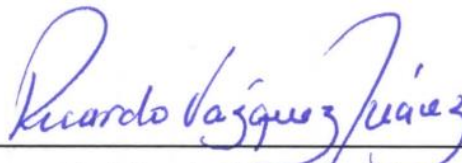
Después de intercambiar opiniones los miembros de la Comisión manifestaron su **APROBACIÓN DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA



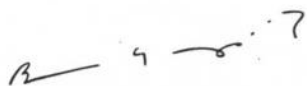
---

Dra. Norma Angélica Estrada Muñoz  
Co-Director de Tesis



---

Dr. Ricardo Vázquez Juárez  
Co-Director de Tesis



---

Dr. Bruno Gómez Gil Rodríguez Sala  
Co-Tutor



---

Dra. Gracia Alicia Gómez Anduro  
Directora de Estudios de Posgrado y Formación  
de Recursos Humanos

## **Conformación de Comités**

### **Comité Tutorial de Tesis**

Dra. Norma Angélica Estrada Muñoz

Centro de Investigaciones Biológicas del Noroeste, S.C. La Paz, B.C.S.

### **Co-Directora de Tesis**

Dr. Ricardo Vázquez Juárez

Centro de Investigaciones Biológicas del Noroeste, S.C., La Paz, B.C.S.

### **Co-Director de Tesis**

Dr. Bruno Gómez-Gil Rodríguez Sala

Centro de Investigación en Alimentación y Desarrollo, A.C., Mazatlán, Sin.

### **Co-Tutor de Tesis**

### **Comité Revisor de Tesis**

Dra. Norma Angélica Estrada Muñoz

Dr. Ricardo Vázquez Juárez

Dr. Bruno Gómez-Gil Rodríguez Sala

### **Jurado de examen**

Dra. Norma Angélica Estrada Muñoz

Dr. Ricardo Vázquez Juárez

Dr. Bruno Gómez-Gil Rodríguez Sala

### **Suplente**

M.C Julio Antonio Hernández González

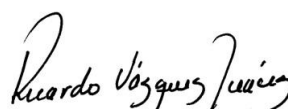
## Resumen

Algunos miembros de la familia Vibrionaceae, son bacterias patógenas que causan mortalidad en organismos marinos de cultivo. La genómica comparativa ha revolucionado el campo de estudio de estas bacterias; herramienta poderosa que permite identificar factores de virulencia, realizar análisis comparativos y filogenéticos, así como el estudio de redes metabólicas. El objetivo de este trabajo es evaluar y comparar *in silico* genomas de bacterias halófilas de la Familia Vibrionaceae aisladas en granjas acuícolas de Baja California Sur, de organismos que presentaron signos de patogenicidad en diferentes hospederos (*Crassostrea gigas*, *Litopenaeus vannamei* y *Totoaba macdonaldi*). Se eligieron 8 cepas patógenas de las especies *Vibrio campbellii*, *V. parahaemolyticus* y *Photobacterium damsela subsp. damsela*. Se extrajo el ADN genómico de cada cepa y se secuenciaron en la plataforma de Illumina Miseq (lecturas pareadas de 2 x 150 pb). Se evaluó la calidad de las lecturas con Nextera cleaner. Se ensamblaron secuencias cromosómicas (SPAdes, A5, Mix) y de profagos (Phaster). Se anotó con PROKKA y RASTtk. Se evaluó la filogenia con 16S ARNr, ANI y MLST. Con CMG-Biotools se analizaron el uso de codones y aminoácidos, y se compararon proteomas y el análisis del genoma núcleo y pangenoma. Se analizó el fenotipo con Traitat. Los factores de virulencia se analizaron usando la base de datos VFDB. Las cepas estudiadas tienen dos cromosomas circulares de diferente tamaño. El ADN cromosómico presentan diferencias en el contenido de GC agrupándose *V. parahaemolyticus* y *V. campbellii* (45%), mientras que la cepa *P.damsela subsp. damsela* tuvo un %GC menor (41%). El tamaño del genoma de las cepas de *Vibrio* tuvo un rango de 5.2-5.9 Mb mientras que el de *P.damsela subsp. damsela* contiene 4.4 Mb. Las cepas de *Vibrio* spp., muestran un genoma más grande al incrementarse el contenido de %GC, en consecuencia, tienen una tendencia a llegar a invadir varios hospederos. En las secuencias cromosómicas se identificaron diversos profagos especie-específicos del hospedero. Las cepas de *V. parahaemolyticus* 6 y M mostraron distancias filogenéticas mayores que las cepas 7L, 1A, A1 y N, resultados similares en el uso de codones y aminoácidos, así como en el análisis de proteomas. Estas últimas cepas podrían ser consideradas bacterias emergentes con poca variación evolutiva. Se confirmó con el pangenoma, que *V. parahaemolyticus* contiene un genoma plástico y diverso, ya que su genoma núcleo se conserva y el genoma variable se expande. Se identificaron diversos factores de virulencia en cada una de las cepas, sin embargo, según otros autores concluyen que la ausencia o presencia de algún factor de virulencia no determina por completo el carácter patogénico de estas bacterias patógenas. Las bacterias patógenas han desarrollado mecanismos que resultan en enfermedad en el huésped, y los genomas son cada vez más utilizados ya que dictan el comportamiento de estas bacterias, por lo que este conocimiento podría ser útil para el diagnóstico, la prevención y el control de las enfermedades en acuicultura marina.

**Palabras clave:** Genómica comparativa, bacterias patógenas, Vibrionaceae, acuicultura



Dra. Norma Estrada Muñoz  
Co-Directora de Tesis



Dr. Ricardo Vázquez Juárez  
Co-Director de Tesis

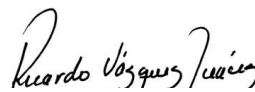
## Summary

Some members of the Vibrionaceae family are pathogenic bacteria that cause mortality in marine culture organisms. Comparative genomics has revolutionized the field of study of these bacteria that allows identifying virulence factors, perform comparative and phylogenetic analyses, as well as the study of metabolic networks. The objective of this work is to evaluate and compare *in silico* genomes of halophilic bacteria of the Vibrionaceae Family isolated from aquaculture farms in Baja California Sur, which showed signs of pathogenicity in different hosts (*Crassostrea gigas*, *Litopenaeus vannamei*, and *Totoaba macdonaldi*). We selected eight pathogenic strains of the species *Vibrio campbellii*, *V. parahaemolyticus*, and *Photobacterium damsela* subsp. *damsela*. Genomic DNA was extracted and sequenced on the Illumina Miseq platform (2 x 150 paired-end reads). The quality of the reads was evaluated with nextera cleaner. We assembled chromosomal (SPAdes, A5, Mix) and prophage (Phaster) sequences. Annotation was made with PROKKA and RASTtk. The phylogeny was evaluated with 16S RNAr, ANI, and MLST. CMG-Biotools serve for evaluating the use of codons and amino acids, comparison of proteomes, and analyzing the core genome and pangenome. The phenotype was analyzed with Traitar. Virulence factors were analyzed with VFDB. The strains studied have two circular chromosomes, from different size. Chromosomal DNA shows differences in GC content grouping *V. parahaemolyticus* and *V. campbellii* (45%), while *P.damsela* subsp. *damsela* had a lower GC (41%). The genome size of *Vibrio* strains had a range of 5.2-5.9 Mb while that of *P. damsela* subsp. *damsela* was 4.4 Mb. The *Vibrio* spp. strains show a larger genome as the% GC content increases; as a result, they tend to invade several hosts. In the chromosomal sequences, several species-specific host prophages were identified. Strains of *V. parahaemolyticus* 6 and M showed phylogenetic distances greater than strains 7L, 1A, A1, and N, similar results in the use of codons and amino acids, as well as in the analysis of proteomes. These last strains could be considered emerging bacteria with a little evolutionary variation. It was confirmed with the pangenome that *V. parahaemolyticus* contains a plastic and diverse genome since its core genome is conserved, and the variable genome expands. Virulence factors were identified in each of the strains; however, according to other authors, they conclude that the absence or presence of some virulence factor does not entirely determine the pathogenic nature of these bacteria. Pathogenic bacteria have evolved mechanisms that result in disease in the host, and genomes are increasingly used as they dictate the behavior of these bacteria so that this knowledge could be useful for the diagnosis, prevention, and control of diseases in marine aquaculture.

**Key words:** Comparative genomics, pathogenic bacteria, Vibrionaceae, aquaculture.



Dra. Norma Estrada Muñoz  
Co-Directora de Tesis



Dr. Ricardo Vázquez Juárez  
Co-Director de Tesis

## Dedicatoria

A mi familia,

“La cura para todo es siempre el agua salada: el sudor, las lagrimas o el mar”

Karen Blixie

## **Agradecimientos**

A CONACyT por la beca otorgada No. 489987.

A CONACyT por el proyecto Desarrollo Científico Para Atender Problemas Nacionales 2014 No. 247567

Al CIBNOR por las instalaciones otorgadas y el espacio en el servidor para que este proyecto pudiese realizarse.

Al CIAD unidad Mazatlan por el espacio otorgado en el Servidor, gracias al cual se pudo realizar diversos análisis del proyecto.

A mi Comité Tutorial, a la Dra. Norma Angelica Estrada Muñoz, Dr. Ricardo Vazquez y el Dr. Bruno Gómez-Gil Rodríguez Sala. A la Dra. Norma por su ayuda durante el escrito y recomendaciones con cada presentacion, al Dr. Ricardo por su orientación y al Dr. Bruno por sacarme del hoyo con cada duda.

Al M.C Julio A. Hernández, Dr. Mario Rojas Arzaluz y la Biol. Carla Muñoz, por su participación en el anteproyecto. Gracias Julio por todo su apoyo y confianza.

A mis compañeros de laboratorio Carolina Garciglia, Hector Garza, Ismael Mendoza, Natalia Palacios, Beatriz Meza y amigos de maestría, gracias por este tiempo juntos.

A mi familia, por su apoyo incondicional, por estar ahí conmigo en las buenas y en las malas, los amo y adoro.

Al Laboratorio de Biología Molecular de Plantas, Laboratorio de Genomica y Bioinformática por el apoyo y los consejos para la realización de la tesis. A la Dra. Gracia Alicia Gomez Anduro por permitirme trabajar en su laboratorio.

A Tania Veronica Nuñez Vadez y Horacio Sandoval Gomez, por su apoyo en los trámites, presentaciones y cualquier duda que surgiese durante estos dos años de maestria.

Dr. Raúl Llera Herrera por su apoyo durante la estancia en Mazatlan y en Morelia, gracias, y al Dr. Raul Martinez Rincon por su apoyo con la paquetería de R.

A todas aquellas personas que se encontraron involucradas directa e indirectamente con el proyecto, muchas gracias por participar en mi formación, ya sea de manera personal o profesional.

## Contenido

<b>Resumen</b> .....	i
<b>Summary</b> .....	ii
<b>Dedicatoria</b> .....	iii
<b>Agradecimientos</b> .....	iv
<b>Contenido</b> .....	v
<b>Lista de figuras</b> .....	viii
<b>Lista de tablas</b> .....	xi
<b>Abreviaturas</b> .....	xii
<b>1. INTRODUCCIÓN</b> .....	1
1.1. Acuicultura y enfermedades.....	1
1.2. Bacterias patógenas en acuicultura.....	2
1.3. Características generales de la familia Vibrionaceae.....	3
1.4. Genomas bacterianos.....	4
1.4.1. Cromosomas.....	5
1.4.2. Plásmidos.....	7
1.4.3. Bacteriófagos.....	7
1.4.4. Elementos genéticos móviles.....	8
1.5. Factores de virulencia.....	9
1.5.1. Adherencia y factores de colonización.....	10
1.5.2. Toxinas.....	11
1.5.3. Sistemas de secreción.....	12
1.5.4. Antifagocitosis: evasión del sistema inmune.....	13
1.5.5. Quimiotaxis.....	14
1.5.6. Obtención de nutrientes: Siderófos.....	14
1.5.7. Quorum sensing.....	15
1.5.8. Enzimas hidrolíticas.....	16
1.6. Genómica comparativa.....	16
1.6.1. Flujo del análisis informático para genomas bacterianos.....	17
<b>2. ANTECEDENTES</b> .....	20
2.1. Vibriosis y otras enfermedades de la familia Vibrionaceae en acuicultura....	20
2.2. Genomas bacterianos de la familia Vibrionaceae.....	21
2.3. Factores de virulencia en la familia Vibrionaceae.....	23
2.3.1. Adherencia.....	23
2.3.2. Toxinas.....	24
2.3.3. Sistemas de secreción.....	25
2.3.4. Anti-fagocitosis.....	26
2.3.5. Quimiotaxis.....	27
2.3.6. Sideróforos.....	28
2.3.7. Quorum sensing.....	29
2.4. Bacterias patógenas asociadas en la acuicultura en México.....	30
<b>3. JUSTIFICACIÓN</b> .....	32
<b>4. HIPÓTESIS</b> .....	33
<b>5. OBJETIVOS</b> .....	34



5.1. Objetivo general.....	34
5.2. Objetivos particulares.....	34
<b>6. MATERIAL Y MÉTODOS</b> .....	35
6.1. Extracción de ADN total y secuenciación.....	35
6.2. Calidad y limpieza de las secuencias.....	36
6.3. Ensamble de las secuencias de nucleótidos .....	36
6.3.1. Ensamble de los cromosomas.....	37
6.3.2. Validación del ensamble.....	38
6.3.3. Identificación de secuencias de bacteriófagos.....	39
6.4. Anotación de las secuencias.....	39
6.5. Filogenia.....	41
6.5.1. Identidad promedio de nucleótidos (ANI).....	41
6.5.2. Tipificación multilocus de secuencias (MLST).....	42
6.5.3. Filogenia 16S ARNr.....	43
6.6. Comparación de genomas.....	44
6.6.1. Atlas estructural de cromosomas de <i>V. parahaemolyticus</i> .....	44
6.6.2. Preferencia de uso de codones y aminoácidos.....	45
6.6.3. Comparación de proteomas.....	45
6.6.4. Análisis del genoma núcleo y pangenoma de <i>V. parahaemolyticus</i> .....	45
6.7. Comparación de fenotipos.....	46
6.8. Identificación de factores de virulencia.....	46
<b>7. RESULTADOS</b> .....	48
7.1. Calidad y limpieza de las secuencias.....	48
7.2. Validación del ensamble con QUAST.....	48
7.3. Identificación de secuencias de bacteriófagos.....	50
7.4. Anotación de los genomas.....	56
7.5. Filogenia de la familia Vibrionaceae.....	56
7.5.1. Promedio de la distancia genética (ANI).....	56
7.5.2. Tipificación multilocus de secuencias.....	62
7.5.3. Filogenia del gen 16S ARNr.....	62
7.6. Genómica comparativa de <i>V. parahaemolyticus</i> .....	63
7.6.1. Atlas estructural del ADN de los cromosomas.....	63
7.6.2. Uso de codones y aminoácidos.....	67
7.6.3. Comparación de proteomas.....	67
7.6.4. Análisis del genoma núcleo y pangenoma.....	68
7.7. Comparación de Fenotipos de la familia Vibrionaceae.....	69
7.8. Identificación de factores de virulencia.....	69
<b>8. DISCUSIÓN</b> .....	79
8.1. Secuencias cromosómicas y de bacteriófagos.....	79
8.2. Filogenia de la familia Vibrionaceae.....	85
8.3. Genómica Comparativa.....	87
8.3.1. Uso de codones y aminoácidos.....	89
8.3.2. Comparaciones de Proteomas.....	91

8.3.3. Comparación de fenotipos.....	92
8.4. Factores de Virulencia.....	93
<b>9. CONCLUSIONES.....</b>	<b>100</b>
<b>10. LITERATURA CITADA.....</b>	<b>103</b>
<b>11. ANEXOS.....</b>	<b>123</b>
Anexo A.....	123
Anexo B.....	124
Anexo C.....	125

## Lista de figuras

<b>Figura 1.</b> Estructura básica de una bacteria bacilar.....	4
<b>Figura 2.</b> Genoma bacteriano de la familia Vibrionaceae.....	6
<b>Figura 3.</b> Flujo de análisis bioinformático para genomas.....	17
<b>Figura 4.</b> Ensamble de lecturas. Composición de secuencias contiguas.....	19
<b>Figura 5.</b> Distancia genética y promedio de la identidad nucleotídica de <i>Photobacterium damsela</i> subsp. <i>damsela</i> . Se utilizaron cuatro genomas de <i>P. damsela</i> subsp. <i>damsela</i> de la base de datos NCBI y la cepa <i>P. damsela</i> subsp. <i>damsela</i> T01 aislada de B.C.S. Los dígitos dentro de la matriz indican el promedio de la identidad nucleotídica de los genomas comparados, siendo 1.00 idéntico y 0.00 diferente. Se utilizó la opción ANIm, donde utiliza MUMmer (NUCmer).....	59
<b>Figura 6.</b> Distancia genética y promedio de la identidad nucleotídica de <i>Vibrio campbellii</i> . Se utilizaron 10 genomas de la bacteria <i>Vibrio campbellii</i> de NCBI y la cepa <i>V. campbellii</i> 2 aislada de B.C.S. Los dígitos dentro de la matriz indican el promedio de la identidad nucleotídica de los genomas comparados, siendo 1.00 idéntico y 0.00 diferente. Se utilizó la opción ANIm, donde se usa MUMmer (NUCmer).....	60
<b>Figura 7.</b> Promedio de la identidad nucleotídica y distancia genética de <i>Vibrio parahaemolyticus</i> . Se utilizaron 30 genomas de <i>V. parahaemolyticus</i> y las seis cepas A1, 7L, N, 1A, 6 y M aisladas de B.C.S. Los dígitos dentro de la matriz indican el promedio de la identidad nucleotídica de los genomas comparados, siendo 1.00 idéntico y 0.00 diferente. Se utilizó la opción ANIm, donde usa MUMmer (NUCmer).....	61
<b>Figura 8.</b> Multi-locus Secuencia Tipo (MLST) de cepas de <i>Vibrio parahaemolyticus</i> . Árbol y tabla de genes constitutivos de <i>V. parahaemolyticus</i> . Se utilizaron siete genes <i>dnaE</i> , la subunidad catalítica de la ADN polimerasa III: <i>gyrB</i> , la ADN girasa subunidad B <i>recA</i> , recombinasa A; <i>dtdS</i> , deshidrogenasa D-treonina; <i>pntA</i> , deshidrogenasa NAD (P subunidad alfa; <i>pyrC</i> , Dihidrorotasa y; <i>tnaA</i> , Triptofanasa. Para genomas completos reportados en la base de datos de Genome de NCBI y los genomas de interés aislados de B.C.S.....	64
<b>Figura 9.</b> Árbol filogenético de <i>Vibrio parahaemolyticus</i> con el gen ribosomal de la subunidad pequeña 16s de genomas completos de la base de datos de NCBI. Se utilizó GBLOCKS para la eliminación de sitios poco informativos y divergentes, MAFFT para el alineamiento y Fasttree para la elaboración del árbol, utilizando el método Jukes-Cantor Joins con repeticiones de 1000. Se utilizó la secuencia de <i>Streptomyces</i> sp. como grupo externo.....	65
<b>Figura 10.</b> Atlas estructural de ADN de los cromosomas I y II de <i>Vibrio parahaemolyticus</i> . Se utilizó la cepa RIMD 2210633 como genoma de referencia. La imagen circular muestra el siguiente orden desde el círculo interno: 1) tamaño del cromosoma (cromosoma I = 3,288,558 pb y cromosoma II = 1,877,212 pb) , 2) porcentaje de AT (rojo=alto contenido AT), 3) asimetría de GC (azul=mayor cantidad de G's), 4) repeticiones invertidas y directas (color=repetición), 5) posición de preferencia, 6) energía de apilamiento, 7) curvatura intrínseca, 8-13) secuencias de las cepas de estudio (1A, 6, 7L, A1, M y N) y 14) cepa no patogénica (ATCC 17802).	

Flechas naranjas indican cambios en la asimetría de G y C, que muestran el origen y el término de la replicación. Las flechas azules indican la localización de los operones de ARNr. Las flechas de color rojo oscuro resaltan las áreas del cromosoma que muestran significativamente diferentes estructuras de ADN que el resto del genoma..... 66

**Figura 11.** Mapa de calor de la preferencia del uso de codones. Se utilizaron 41 genomas de *Vibrio parahaemolyticus*, incluyendo los genomas de B.C.S, México (M, N, 6, 7L, A1 y 1A). Se utilizó CMG-biotools para la visualización, y para la creación del mapa se utilizó el programa R. Se muestra un árbol con la agrupación del uso del triplete de codones y un histograma con la abundancia de los mismos..... 71

**Figura 12.** Mapa de calor de la preferencia de uso de aminoácidos. Se utilizaron 35 genomas completos de *Vibrio parahaemolyticus* y los genomas de *V. parahaemolyticus* de B.C.S (M, N, 6, 7L, A1 y 1A). Se utilizó CMG-biotools para la visualización, y para la creación del mapa se utilizó el programa R. Se muestra un árbol con la agrupación del uso de aminoácidos y un histograma con la abundancia de los mismos..... 72

**Figura 13.** BLAST-Matriz comparativa de proteomas. Genomas de 41 cepas de *Vibrio parahaemolyticus*. Se utilizó el algoritmo de BLAST (Basic Local Alignment Search Tool), para identificar si dos proteínas se comparten entre los genomas, utilizando el criterio 50/50 idéntico en secuencia y cubriendo el 50% de la longitud del gen. Las líneas azules indican las cepas pertenecientes a B.C.S., México, y los recuadros rojos encierran los porcentajes con mayor similitud que comparten con genomas obtenidos de la base de datos del NCBI..... 73

**Figura 14.** BLAST-Matriz comparativa de proteomas de *Vibrio parahaemolyticus* de B.C.S., México (N, A1, 7L, 1A, 6, y M) comparados con *V. campbellii* (02) y *Photobacterium damselaesupesp. Damselaes* (T01). Se utilizó el algoritmo de BLAST (Basic Local Alignment Search Tool), para identificar si dos proteínas se comparten entre los genomas, utilizando el criterio 50/50 idéntico en secuencia y cubriendo el 50% de la longitud del gen..... 74

**Figura 15.** Genoma núcleo y pangenoma de *Vibrio parahaemolyticus*. Representación de 35 cepas de *V. parahaemolyticus* obtenidas en la bases de datos y 6 genomas aislados de B.C.S., México (A1, 1A, N, M, 7L y 6). Se utilizó la herramienta de CMG-biotools, y se eligió un límite máximo del 50% de identidad, con una cobertura del 50% del gen más largo. Si dos proteínas dentro de un genoma coincidían de acuerdo con el límite del 50/50%, se agruparon en una familia de proteínas. La línea roja indica el genoma núcleo con 3,143 genes y la línea azul indica el pangenoma con 11,104 genes. Las barras indican la comparación de nuevos genes (gris intenso) y nuevas familias de genes (gris claro)..... 75

**Figura 16.** Predicción del fenotipo de cepas de la familia *Vibronaceae*. Agrupamiento de ocho cepas en base a 67 rasgos usando el programa de Traitair. Cada columna representa uno de los 67 rasgos, mientras que las filas representan las ocho cepas del presente proyecto. El esquema de color de las columnas representa 11 propiedades fenotípicas, desde la proteólisis hasta la producción de enzimas. El origen de la predicción del fenotipo (Traitair phypat y / o phypat + PGL clasificador) determina el color de las entradas del mapa de calor. Los colores del dendrograma indican

distribuciones de fenotipos similares en las muestras, según lo determinado por un agrupamiento jerárquico con SciPy.....	76
<b>Figura 17.</b> Mapa de calor que muestra la abundancia de los factores de virulencia de la familia Vibrionaceae. Comparación general de categorías relacionadas a la patogenicidad de bacterias de la anotación con RASTtk, realizada con el programa R, usando la clasificación de SEED de RASTtk.....	77
<b>Figura 18.</b> Abundancia de genes que codifican para factores de virulencia en <i>Vibrio parahaemolyticus</i> . Grafica de abundancias de genes relacionados a la virulencia de una cepa patogénica. Las secuencias fueron tomadas de la base de datos VFDB para el género <i>Vibrio</i> sp.. Se realizó un blastn con el programa BRIG para elegir aquellas secuencias con un $\geq 90\%$ de identidad, que cubrían más del 90% de la secuencia del gen	78

## Lista de tablas

<b>Tabla I.</b> Pruebas bioquímicas y grado de daño.....	31
<b>Tabla II.</b> Identificación de las bacterias y crecimiento a diferentes concentraciones de NaCl.....	<b>31</b>
<b>Tabla III.</b> Genomas de referencia para el ensamble y la validación.....	38
<b>Tabla IV.</b> Genomas de referencia utilizados para el reordenamiento de secuencias contiguas.....	40
<b>Tabla V.</b> Genes de referencia de la familia Vibrionaceae.....	43
<b>Tabla VI.</b> Genomas de referencia e interés de la especie de <i>V. parahaemolyticus</i> .....	47
<b>Tabla VII.</b> Estadísticas básicas de la limpieza de las secuencias.....	49
<b>Tabla VIII.</b> Parámetros estadísticos de la validación con QUASt de los ensambles con SPAdes, A5 y MIX.....	51
<b>Tabla IX.</b> Características de las secuencias identificadas como presuntos profagos por PHASTER.....	52
<b>Tabla X.</b> Resultados de la anotación con RASTtk y PROKKA.....	58

## Abreviaturas

TCBS	Agar tiosulfato citrato bilis sacarosa
TSB	Caldo tripticasa soya
.fastq	Formato en base texto con una secuencia nucleotidica y con un valor de score que indica la calidad de la misma
.faa	Formato fasta en aminoácidos
.fasta	Formato fasta en acidos nucleicos
.fna	Formato fasta en acidos nucleicos
.gbk	Formato GenBank
SSH	Secure Shell
N50	Medida estadística, donde la longitud de los contigs de igual o mayor tamaño a la mitad del genoma completo
L50	se toma el conteo de N50, en donde se toma el número más pequeño de contigs cuya suma de la longitud representa la mitad del tamaño del genoma
.ps	PostScript
Fw	Forward
Rv	Reverse
%GC	Porcentaje de Guanina-Citosina
tRNA	Ácido ribonucleico de transferencia
pb	Pairbase, pares de bases
Mpb	un millón de pares de bases
.xml	Archivo de lenguaje de marcado extensible
Kbp	mil pares de bases
.tab	Archivo con formato de tabla (tabulaciones)
.db	DataBase
spp	Especies
MUMmer	Sistema para alineación rápida de genomas
NUCmer	NUCleotide MUMmer, programa de alineación estándar para secuencias de ADN
ANI	Average nucleotide Identity
.list	Archivo que contiene una lista
IGs	Islands Genomic
BRIG	Blast Ring Image Generator
HGT	transferencia horizontal de genes
VFDB	Virulence Factor Database
IS	Secuencias de insercion
CRISPR-Cas	Clustered regulatory interspaced short palindromic repeats
PCR	Reaccion en cadena de la polimerasa
NGS	Next Generation Sequencing
AHPND	Enfermedad necrosis hepatopancreatica aguda

Q $\geq$ 30	indica que el factor de error seria de 1 base entre 1000 con el mismo nivel de calidad Phred, con un 99.9% de precisión
CDS	secuencias codificantes
MLST	Multilocus sequence typing
CC	Complejo clonal
16s.	Gen 16s del ARN ribosomal
BLAST	Basic Aligment Search Tool



# 1. INTRODUCCIÓN

## 1.1. Acuicultura y enfermedades

La acuicultura es un sector importante en la producción de alimentos, y uno de los campos con mayor desarrollo económico a nivel mundial (FAO, 2018). En México, la acuicultura ha experimentado un progreso significativo en las últimas décadas, particularmente en la región Noroeste, considerada como la de mayor productividad biológica en el país, incluidos peces, moluscos y crustáceos (SAGARPA, 2017). Debido al aumento en la comercialización nacional y de exportación de productos de origen acuícola, se ha favorecido la introducción, propagación y transmisión de enfermedades, entre las que destacan las producidas por hongos, protozoarios, virus y bacterias, que son las principales causas de mortalidad de las especies cultivadas (OIE, 2019).

Las enfermedades infecciosas se pueden dividir en dos categorías principales: 1) enfermedades causadas por patógenos específicos u obligados y 2) enfermedades causadas por patógenos no específicos facultativos u oportunistas. Las enfermedades infecciosas constituyen una de las principales causas de mortalidad en acuicultura debido a: 1) aparición de enfermedades emergentes, 2) reemergencia de enfermedades, y 3) persistencia de enfermedades. Diversas causas que contribuyen a que se desarrolle la enfermedad incluyen algunas relacionadas con las prácticas de cultivo, y otras que debilitan la resistencia del hospedero o aumentan de manera muy importante el número de patógenos. Estas enfermedades infecciosas existen habitualmente en forma enzootica o epizootica, tanto en animales cultivados como en poblaciones naturales, debilitando o incapacitando individuos, haciéndolos más susceptibles a depredadores o a otros tipos de agresiones ambientales, y en ciertas ocasiones provocando mortandades masivas. El estudio de la historia de las enfermedades revela las propiedades evolutivas de los patógenos y la dinámica relación entre los patógenos, huésped y medioambiente. El conocimiento de la historia y los factores asociados a nuevas infecciones es clave para la prevención y control de futuros desafíos sanitarios (Austin, 2012).

## 1.2. Bacterias patógenas en acuicultura

Entre los patógenos más recurrentes en la acuicultura están las bacterias marinas, que representan un problema significativo para el desarrollo del sector acuícola, reflejándose en pérdidas económicas severas (OIE, 2019). Diversas bacterias causan infección en organismos marinos de interés económico, como las bacterias gram positivas de los géneros *Mycobacterium*, *Streptococcus*, y *Renibacterium*, así como, enfermedades causadas por bacterias gram negativas como *Aeromonas*, *Edwardsiella*, *Pseudomonas*, *Flavobacterium* y *Vibrio* (FAO, 2017).

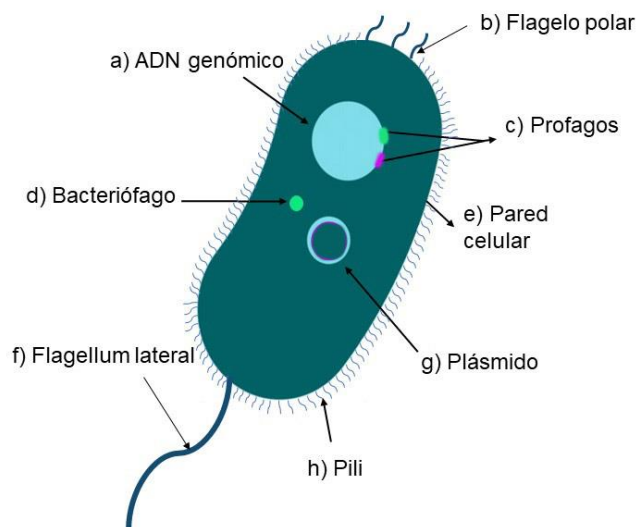
Un patógeno se define como un microorganismo, parásito o agente que puede llegar a ocasionar daño al hospedero que habita. Usualmente los patógenos para sobrevivir deben: (1) colonizar al hospedero, (2) encontrar una fuente de nutrientes y un nicho compatible en el hospedero, (3) evitar el sistema inmune del hospedero, (4) replicarse y (5) salir de su hospedero actual e invadir uno nuevo.

La diferencia entre bacterias patógenas y sus relativas no patogénicas, se debe a un grupo de genes capaces de generar una enfermedad, estos son denominados genes de virulencia, los cuales codifican para proteína que son llamadas factores de virulencia. Los genes de virulencia usualmente se pueden encontrar en el cromosoma bacteriano que contiene la información necesaria para la sobrevivencia de la bacteria, en el ADN extracromosómico (plásmido), y también se han encontrado en bacteriófagos. Entre los genes de virulencia, los más frecuentes son aquellos que codifican para proteínas que de naturaleza tóxica, que interactúan directamente con proteínas estructurales o de señalización de la célula huésped, obteniendo así una ventaja para la colonización y/o su replicación; asimismo, codifican para proteínas que forman complejos proteicos que facilitan la función de las toxinas, que se les denomina sistemas de secreción (Alberts *et al.*, 2002).

### 1.3. Características generales de la familia Vibrionaceae

Entre las enfermedades infecciosas cabe destacar las bacterianas producidas por la familia Vibrionaceae, lo que se considera un problema importante para el desarrollo del sector con graves pérdidas económicas en todo el mundo. En México las bacterias causantes de Vibriosis suelen encontrarse en la zona Noroeste del país, a este hecho se la ha atribuido las características geográficas como las temperaturas tropicales y las altas concentraciones de solutos en el mar de la costa del Pacífico (Hernández-Díaz *et al.*, 2015).

Los *Vibrios* son bacterias de forma bacilar Gram negativas (Fig.1), pertenecen a la clase *Gammaproteobacteria*, usualmente son móviles, quimiorganotrofos (oxidan compuestos orgánicos), y tienen un metabolismo fermentativo facultativo (pueden desarrollarse en ausencia o presencia de oxígeno), con flagelo polar encerrado en una vaina y puede crecer a 20°C, sin embargo, el rango de temperaturas en las que pueden crecer puede variar. Estas bacterias son acuáticas y la mayoría de las especies son oxidasa positiva, pueden reducir el nitrato a nitrito, y requieren de Na<sup>+</sup> para su crecimiento (Gómez-Gil *et al.*, 2014; Thompson *et al.*, 2004). La mayoría de las especies de *Vibrio* degradan una variedad de carbohidratos sin la producción de gas, y crecen en medio TCBS (Farmen *et al.*, 2005; Thompson *et al.*, 2009). Estos microorganismos habitan usualmente en ambientes marinos, estuarios y sistemas de agua dulce. Algunas especies de este género son patógenas para humanos y animales acuáticos, los cuales se han reportado que pueden causar gastroenteritis, infecciones en heridas, cólera y septicemia (Ceccarelli *et al.*, 2014). *Vibrio* comprende uno de los géneros bacterianos más diversos en el ambiente marino (Gomez-Gil *et al.*, 2014), la familia *Vibrionaceae* presenta 142 especies de bacterias (Ussery *et al.*, 2014).



**Figura 1.** Estructura básica de una bacteria bacilar.

#### 1.4. Genomas bacterianos

El genoma se describe como el repositorio de información genética de un organismo y/o microorganismo, compuesto por millones o miles de millones de nucleótidos de ADN (ácido desoxirribonucleico), este repositorio contiene la información necesaria para que un organismo pueda vivir en un determinado ambiente (Goldman y Landweber, 2016). Cada genoma bacteriano es una colección total de genes que posee la bacteria tanto a nivel cromosomal como en elementos genéticos extracromosomales. Los genomas específicos de ciertas especies pueden contener genes que codifican proteínas responsables de la patogenicidad o pueden ser cruciales para la supervivencia en un ambiente determinado.

Los genomas bacterianos tienen características que los distinguen entre sí, a estas características se les denomina firmas genómicas, en las que se toma en consideración el uso de codón, contenido de GC, abundancia relativa de dinucleótidos, entre otros. El uso de codón está correlacionado con el nivel de expresión de un gen, donde genes altamente expresados tienen un uso de codón que se asemeja a la abundancia relativa de ARNt, mientras que genes expresados en menor abundancia no se asemejan a este (Quax *et al.*, 2015). El contenido de

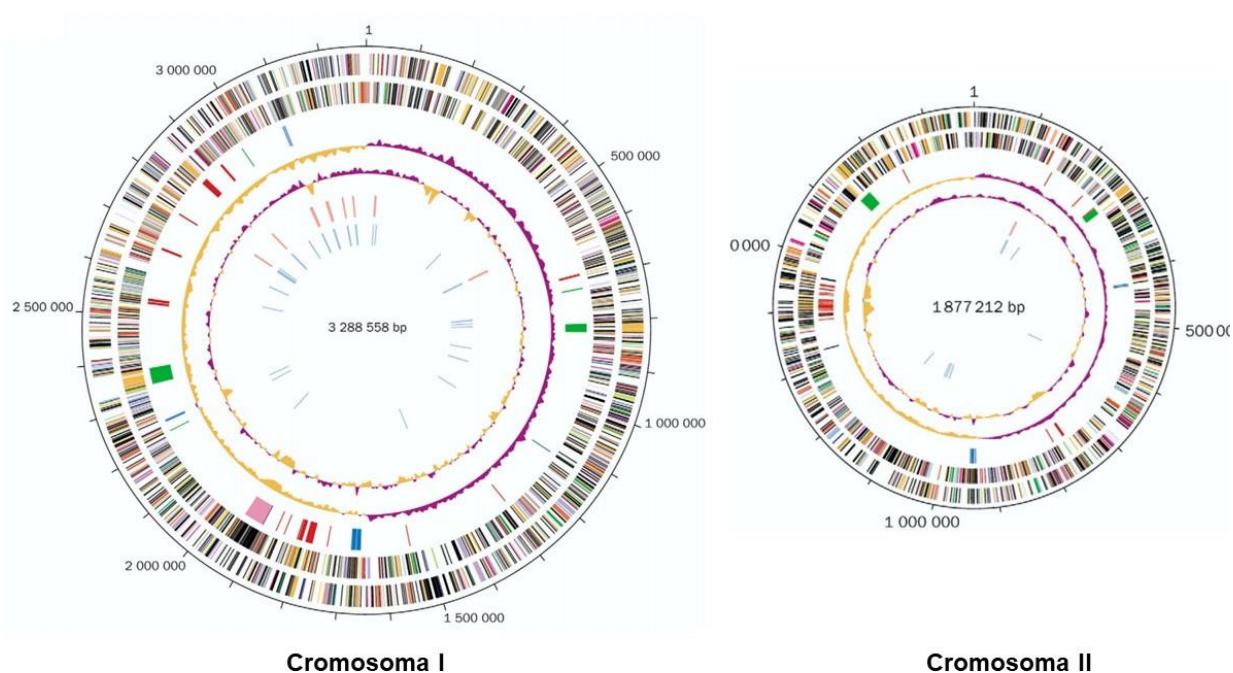
GC o porcentaje de GC, varía considerablemente entre organismos que puede ir desde un 15% a un 75% (dependiendo la especie) (Foerstner *et al.*, 2005; Mann *et al.*, 2010); varios factores pueden llegar a influir la variación de este porcentaje, como la adaptación al medio, la recombinación genómica. Este porcentaje de GC además de diferir entre especies, también puede variar dentro del mismo genoma, característica que se utiliza comúnmente para identificar elementos genéticos móviles (Mann *et al.*, 2010). Con lo que respecta a los perfiles de la abundancia de dinucleótidos son diferentes para cada genoma bacteriano (incluso dentro de la misma especie) y reflejan la filogenia de este (Kariin y Burge, 1995; van Passel *et al.*, 2006). Esta abundancia de dinucleótidos se refiere a la suma de las diferencias de cada par de dinucleótidos (GC y AT) entre los puentes de ADN (Wong y Golding, 2002).

#### 1.4.1. Cromosomas

Los genomas bacterianos se caracterizan de ser en su mayoría circulares, y tener un único origen de replicación (OriC), sin embargo se han encontrado bacterias que poseen más de un replicón, un segundo cromosoma, como lo han reportado en *Rhodobacter sphaeroides*, bacterias del género *Brucella* spp., y bacterias del género *Vibrio* spp. (Fig. 2). La localización de un gen en el cromosoma puede influir en su nivel de expresión (Bryant *et al.*, 2014), y en especies donde hay una rápida replicación el número de copias del gen aumenta (Dryselius *et al.*, 2008).

En los cromosomas bacterianos, al cromosoma I se le ha atribuido que contiene los genes necesarios para las funciones primarias, esenciales para su sobrevivencia (replicación, transcripción, traducción, biosíntesis), a lo que también se le conoce como genoma núcleo. En el primer cromosoma en promedio contienen de 3.65 a 3.46 Mpb aproximadamente, pero esto puede variar dependiendo de la especie o la cepa. El segundo cromosoma no contiene genes esenciales, por lo que es un replicón más variable que el primero y dada esta variabilidad se relaciona con su adaptación al ambiente (resistencia a antibióticos, genes relacionados a metales pesados, genes de patogenicidad). La mayoría de los genes que se encuentran en el segundo cromosoma se han adquirido a través de la transferencia horizontal de genes (HGT),

y su contenido de GC (Guanina-Citosina) y la composición de nucleótidos, es diferente del primer cromosoma (Harrison *et al.*, 2010; Makino *et al.*, 2003; Okada *et al.*, 2005; Suwanto y Kaplan, 1989).



**Figura 2.** Genoma bacteriano de la familia Vibrionaceae

Una característica remarcable de *Vibrio* es que su genoma es altamente plástico (Rowe-Magnus *et al.*, 2001). Hasta el momento se sabe que en todos los genomas de *Vibrio* hay dos cromosomas (Trucksis *et al.*, 1998). El Cromosoma I es usualmente más largo, con una constante relativa de tres millones de pares de bases, codificando alrededor de 2,700 proteínas que representan las principales funciones. En contraste el Cromosoma II es más pequeño, tiene aproximadamente un millón de pares de bases, que codifican miles de proteínas y contiene una alta variable de “super-integron” (Rowe-Magnus *et al.*, 1999). El pangenoma de *Vibrio* puede ser muy grande, con más de 17,000 de genes, y se ha encontrado más variabilidad en el cromosoma II que en el I (Lukjancenko *et al.*, 2014).

#### 1.4.2. Plásmidos

Los plásmidos son pequeñas moléculas de ADN en comparación con el ADN cromosómico (Fig.1). Son estructuras circulares de doble cadena de ADN, pero también se pueden encontrar de forma lineal; su contenido nucleótido difiere del cromosómico y puede replicarse de manera independiente al cromosoma (<https://www.nature.com/scitable/definition/plasmid-plasmids-28/>). Los plásmidos difieren en tamaño y número de copias en cada célula bacteriana (también se pueden encontrar en *Arqueas* y células eucarióticas), además que contienen genes que otorgan a la bacteria propiedades adicionales, que no son forzosamente necesarias para su vida pero aumentan su vitalidad (<https://blog.addgene.org/plasmids-101-a-brief-history-of-plasmids-and-an-improved-ebook>). Los plásmidos han sido utilizados como vectores dentro de la ingeniería genética, para la transferencia de genes, este fenómeno también se da de forma natural en el medio ambiente.

Estas moléculas extracromosómicas pueden desempeñar diferentes funciones, dentro de las más remarcables se ha reportado la sobrevivencia del organismo, ya que contiene genes que pueden llegar a matar a otros organismos o genes relacionados a mecanismos de defensa contra el huésped que habitan a través de producción de toxinas u otro tipo de moléculas (Tsang, 2017). Asimismo, participan en el proceso de conjugación que es una de las maneras más conocidas en la que las bacterias comparten material genético (Guglielmini *et al.*, 2011)

#### 1.4.3. Bacteriófagos

Los bacteriófagos también conocidos como fagos (Fig.1), son virus que pueden infectar y replicarse solamente en células bacterianas. Son ubicuas en el medioambiente y se les conoce como el agente biológico más abundante del mundo. Todos los virus, incluidos los fagos consisten de un genoma de ácidos nucleicos (ADN o ARN) encerrado en una cápside, flanqueado por proteínas, que protegen el material genético y median la transmisión del material genético a la célula hospedera (Holmes, 2000; Pham *et al.*, 2018; Simmonds y Aiewsakun, 2018).

Como todos los virus, los fagos son especie-específicos de sus hospederos, esto quiere decir que por lo general una cepa de un fago solo afecta a una sola especie bacteriana, o incluso cepas específicas dentro de una misma especie. Una vez que un virus ha ingresado a la célula bacteriana, tiene dos rutas de replicación. La primera ruta es la lítica, donde el fago inserta su genoma en el genoma del huésped, y utiliza la maquinaria de este para seguir replicándose y sintetizar sus proteínas; durante este proceso la célula hospedera va perdiendo vitalidad, consecuencia de un proceso de lisis mediado por el fago, posteriormente las partículas virales son liberadas para infectar a otra célula huésped. La segunda ruta es la lisogénica, donde el fago se integra al genoma bacteriano como en la ruta lítica, pero a diferencia de la ruta lítica en esta no lisa a la célula bacteriana, sino que se mantiene de manera episomal (ADN que puede replicarse de manera autónoma, o incorporarse al genoma y replicarse con él) y se replica junto con el genoma bacteriano por lo que puede ser heredado a las células bacterianas hijas sin matarlas, a estos fagos se les denomina lisogénicos o profagos. Estos profagos pueden pasar a ser fagos líticos (fagos como en la primera ruta) y matar a su huésped, y las causas de esta transición se atribuyen mayormente a condiciones ambientales (estrés de la célula hospedera, antimicrobianos, exposición a luz UV, etc.) (Doore y Fane, 2016).

#### 1.4.4. Elementos genéticos móviles

Se le denomina elementos genéticos móviles a segmentos de ADN móviles dentro de un genoma (movilidad intracelular) o entre células bacterianas u otra clase de organismos (movilidad intercelular). En el movimiento intercelular en bacterias hay tres mecanismos en los que la información genética se comparte, por medio de transformación, conjugación y transducción (Thomas y Nielsen, 2005) a lo que se le conoce como transferencia horizontal de genes. La transformación implica la transferencia de ADN celular entre bacterias estrechamente relacionadas. La conjugación está dada por uniones de célula-célula por medio del pili (Fig. 1), a través de este conducto el ADN pasa de célula a célula, y en los casos de conjugación se ha reportado un mayor intercambio de ADN plasmídico (Thomas y Nielsen, 2005). La transducción es la transferencia de material genético entre células bacterianas mediada por virus bacteriófagos, donde los bacteriófagos incorporan fragmentos del genoma



de la célula donadora en la célula receptora. En la transducción, el bacteriófago contiene ADN del genoma bacteriano, y reemplaza una parte o todo el complemento normal de su propio ADN.

También se han reportado segmentos de ADN que se pueden compartir a través de la transferencia horizontal de genes, a estos segmentos se le conocen como islas genómicas, que pueden variar de tamaño de menos de 4.5 Kb a 600 Kb (Vernikos, 2008). Las islas genómicas se encuentran flanqueadas por regiones repetidas (DRs, por sus siglas en inglés), además de tener sitios de inserción (genes ARNt o ARNtm). Estos segmentos juegan un papel importante en la evolución de bacterias, así como su diversificación, ya que aportan genes con diversas funciones (Bellanger *et al.*, 2014; Williams, 2002).

Asimismo, la movilidad intracelular tiene otros elementos móviles como la transposición dado por los transposones. Los transposones son secuencias cortas de ADN que se mueven dentro del ADN genómico, con secuencias de inserción (IS) catalizadas mayormente por transportasas, que reconocen e introducen específicamente nicks (paro en una molécula de ADN, en donde no hay un enlace fosfodiéster entre los nucleótidos adyacentes) en los extremos (Bourque *et al.*, 2018; Craig, 2002). Otro elemento móvil intracelular son los elementos CRISPR-Cas (Clustered regulatory interspaced short palindromic repeats), que son agrupamientos de secuencias que contienen fragmentos de ADN de fagos. Estos fragmentos son utilizados por las bacterias para detectar y destruir el ADN de los virus, también reconocen secuencias específicas y las degrada, además tienen la capacidad de mover genes dentro del genoma (Hsu *et al.*, 2014; Pickar-Oliver y Gersbach, 2019).

### 1.5. Factores de virulencia

Los factores de virulencia son elementos (por ejemplo productos génicos) con los que cuenta un microorganismo para poder entrar al huésped, invadir los tejidos y provocar enfermedad. Los patógenos bacterianos han desarrollado factores específicos de virulencia que les permiten multiplicarse en su huésped, sin ser eliminados o expulsados por las defensas del huésped.

Estos no son considerados independientes al hospedero, ya que, para la evaluación de una enfermedad, interviene la interacción de los factores de virulencia con la respuesta del huésped. Los factores de virulencia incluyen moléculas de adhesión, quimiotaxis, toxinas, sistemas de secreción, sideróforos, enzimas, entre otros. Estos factores de virulencia se encuentran codificados ya sea en ADN cromosómico, bacteriófagos, o en plásmidos. El número y tipo de factores de virulencia en una cepa bacteriana dada cambian constantemente, generalmente a través de la selección genómica de aquellos factores que favorecen la supervivencia de la bacteria en el hospedero (Brodgen *et al.*, 2000; Finlay y Falkow, 1997; Wu *et al.*, 2008; Zachary, 2017).

#### 1.5.1. Adherencia y factores de colonización

Para causar una infección, algunas bacterias pueden adherirse primero a la superficie de la mucosa del hospedero. Para establecer un sitio de infección la bacteria debe adherirse al epitelio y multiplicarse antes de que las células epiteliales sean extruidas. Para lograr esto, las bacterias han desarrollado mecanismos de adhesión como, las adhesinas. Se llama adhesina a los múltiples factores que producen las bacterias para adherirse efectivamente a sus hospederos, a través de moléculas receptoras sobre la superficie de la célula huésped, permitiendo la unión de la bacteria a la célula para colonizar. Están compuestas principalmente de polipéptidos o polisacáridos, y proteínas, que se expresan en estructuras como las fimbrias o pili (Fig. 1). La fimbria son estructuras sobresalientes alargadas en forma de vellosidades, las cuales son verdaderos orgánulos que les permiten la fijación a las superficies de las células que infectan, así como la transferencia de ADN. Las fimbrias pueden ser rígidas o flexibles, y su función principal es servir como soporte de las adhesinas, encargadas de reconocer a su receptor en la célula hospedera. Son factores de adhesión, donde la mayoría son glicoproteínas o lipoproteínas, mientras que sus receptores suelen ser azúcares que se expresan en la superficie celular (Daffe y Etienne, 1999; Donnenberg, 2000; Hahn, 1997; Merz y So, 2000).

### 1.5.2. Toxinas

En general las toxinas que producen las bacterias se clasifican en dos grupos: endotoxinas (retenidas dentro de las células bacterianas) y exotoxinas (excretadas al medio que las rodea). Las bacterias Gram negativas al estar en contacto con el hospedero causan reacciones del sistema inmune y la activación de diferentes cascadas de reacciones no celulares. Uno de los componentes de la pared celular de esta familia de bacterias, son los lipopolisacáridos, que forman complejos de alto peso molecular conocidos como endotoxinas bacterianas, por el efecto tóxico que provocan en otros organismos. Estas sustancias de diversa naturaleza se liberan durante la multiplicación bacteriana o cuando el microorganismo se destruye. Estas toxinas comúnmente consisten de residuos de oligosacáridos repetidos, un polisacárido que compone la columna vertebral de la macromolécula, y un lípido A, compuesto generalmente de glucosamina disacárido adyacente a ácidos grasos de cadena larga con grupos fosfato. Las porciones de polisacáridos son responsables de la diversidad antigénica, mientras que el lípido A confiere la toxicidad (Gill, 1982; Zivot *et al.*, 1995).

Otras toxinas producidas por bacterias son las exotoxinas, y a diferencia de las endotoxinas, estas son lipoproteínas difusibles y eliminadas por la célula productora al medio que la rodea, o al sistema circulatorio y tejidos del hospedero. Las exotoxinas incluyen varios tipos de toxinas proteicas y enzimas producidas y/o secretadas por bacterias patógenas, que la bacteria libera durante su desarrollo y que son perjudiciales para el organismo huésped. Sintetizadas por bacterias que contienen profagos, genes cromosómicos o plasmídicos que codifican la exotoxina. La mayoría de las exotoxinas al ser proteínas de alto peso molecular son lábiles a altas temperaturas, sin embargo, numerosos péptidos de bajo peso molecular son estables al calor. Las categorías principales incluyen citotoxinas, neurotoxinas y enterotoxinas (Zivot *et al.*, 1995).

### 1.5.3. Sistemas de secreción

Diferentes bacterias Gram-negativas patógenas han desarrollado maquinarias para transferir proteínas codificadas en su cromosoma al exterior o al interior de células eucariotas, y se conocen como sistema de secreción de proteínas. La secreción extracelular de proteínas es un mecanismo de virulencia determinante en la infección bacteriana. Estas bacterias emplean fimbrias para unirse a la célula y así inyectan proteínas efectoras.

En bacterias Gram negativas, donde la secreción implica la translocación a través de las membranas internas y externas, se reconocen seis clases generales de sistemas de secreción de proteínas, cada uno de los cuales muestra una diversidad considerable. Muchas proteínas secretadas por agentes patógenos son internalizadas a las células huésped para modificar la fisiología de este y así promover la colonización (principalmente toxinas y proteínas efectoras). Estos sistemas son útiles para el crecimiento de bacterias, ya que hay sistemas de secreción que son utilizados por patógenos para manipular al huésped y establecer un nicho de replicación. Existen diferentes tipos de sistemas de secreción bacteriana dependiendo la composición de la membrana bacteriana, en el caso de las bacterias Gram negativas se han encontrado el T1SS, T2SS, T3SS, T4SS y T6SS. Los sistemas de secreción tipo 1, 3, 4 y 6 utilizan un mecanismo de un solo paso, de tal manera que los sustratos se transportan directamente del citoplasma de la bacteria hacia la célula del huésped, y la composición difiere con respecto a sus sustratos proteicos (Costa *et al.*, 2015; Green *et al.*, 2016).

Los mecanismos de secreción que inyectan las proteínas en células eucariotas son III, IV y VI (T3SS, T4SS y T6SS), una vez dentro estas proteínas modulan funciones específicas de las células del huésped, promoviendo así la invasión bacteriana y la colonización (Büttner, 2012; Cornelis, 2006). El sistema de secreción tipo 6 funciona como inyector de toxinas en las células del huésped, y desempeña un papel fundamental en la patogénesis y competencia entre bacterias (Ho, 2014; Zoued, 2014).

Además del uso de complejos multiproteínicos, las bacterias Gram negativas poseen otro método para la liberación de sustancias: la formación de vesículas en la membrana externa. Partes de la membrana externa forman estructuras esféricas, compuestas de una bicapa de lípidos, que encierran sustancias periplásmicas. Las vesículas de varias especies bacterianas pueden contener sustancias patógenas, tener efectos inmunomoduladores, y también adherirse directamente e intoxicar las células huésped. Mientras que la liberación de vesículas es una respuesta general a condiciones de estrés, el proceso de cargar proteínas transportadoras parece ser selectivo (Ferrari *et al.*, 2006, Lee *et al.*, 2007).

#### 1.5.4. Antifagocitosis: evasión del sistema inmune

En una infección, el sistema inmunitario del huésped inicia respuestas inflamatorias rápidas para proteger al huésped de la propagación de microbios invasores. La inflamación se realiza cuando las células de la primera línea de defensa, tales como células epiteliales, macrófagos y neutrófilos detectan señales de alarma (Van-Avondt *et al.*, 2015), manteniendo así la homeostasis de la célula, a este sistema se le denomina fagosoma. El fagosoma tiene receptores para el reconocimiento de patógenos, los PAMPS (patrones moleculares asociados a patógenos), sin embargo, las bacterias patógenas han desarrollado mecanismos para la evasión haciéndose menos inmunogénicas (Akira *et al.*, 2016; Uribe-Querol *et al.*, 2017). Las bacterias han desarrollado varias estrategias para evitar o sobrevivir a la fagocitosis. Algunos de estos, la evasión inmune y la antifagocitosis; mientras que con la evasión inmune la bacteria se esconde del sistema inmune, con la antifagocitosis la bacteria permanece activamente extracelular, incluso después del reconocimiento por parte del sistema inmune. Algunas de las estrategias observadas ha sido la liberación de toxinas (leucocidina y  $\alpha$ -hemolisinas) que intoxican el fagosoma y tienen como objetivo membranas de células de defensa (macrófagos y leucocitos) (Alonzo y Torres, 2014). Otra manera que se ha reportado es evitando la polimerización de la actina y evitar así la fagocitosis, ya que la actina participa en la construcción de la capa del fagosoma (Flo *et al.*, 2004; Goetz *et al.*, 2002).

### 1.5.5. Quimiotaxis

Las bacterias tienen capacidad de desplazarse de un lugar a otro con la ayuda del flagelo (Fig. 1). Los flagelos son apéndices largos los cuales se encuentran fijados a la célula por uno de sus extremos y libres por el otro. La estrategia de movilidad más común entre las bacterias es el movimiento dirigido por la rotación flagelar. Existen diversos tipos de movilidad por flagelos que dependen de la localización y número de flagelos. En algunas especies las bacterias se mueven hacia adelante y hacia atrás y la reorientación parece ser debida al movimiento browniano; en otras especies las células se mueven en línea recta, y ocasionalmente se reorientan activamente.

, La quimiotaxis guía a la bacteria hacia un gradiente químico, a través del sistema de transducción de señales, lo que facilita a las bacterias el acceso a nutrientes y nichos ecológicos favorables. Las bacterias por medio de gradientes de concentraciones de solutos en su entorno avanzan hacia condiciones más favorables (Bren y Eisenbach, 2000). Este comportamiento se logra integrando las señales recibidas de los receptores que detectan el entorno y modulando la dirección de la rotación del flagelo (Fig.1) (Alon *et al.*, 1998; Eisenbach, 1996; Falke *et al.*, 1997).

### 1.5.6. Obtención de nutrientes: Siderófos

La adquisición de hierro es un proceso indispensable para todas las bacterias, ya que es un elemento fundamental en el metabolismo celular. El hierro es un elemento esencial para los microorganismos porque actúa como catalizador en procesos enzimáticos, en el metabolismo de oxígeno, transferencia de electrones, síntesis de ADN y ARN, este ion también participa en la formación de biopelículas regulando la motilidad superficial y estabilidad de la matriz de polisacáridos (Aguado-Santacruz *et al.*, 2012; Chhibber y Bansal, 2013; Weinberg, 2004). Sin embargo, a pesar de la concentración micromolar requerida para el crecimiento celular, el hierro es un compuesto biológicamente limitante debido a su baja solubilidad ( $10^{-18}$ M) en los ambientes aerobios y de pH neutro.

Por esta razón los seres vivos han desarrollado mecanismos para la asimilación de este elemento fundamental. Uno de los más importantes en las bacterias es la biosíntesis de agentes quelantes, denominados sideróforos. Estos compuestos de bajo peso molecular (300-2.000 Da) son secretados al medio circundante dónde secuestran eficientemente al  $\text{Fe}^{3+}$ . Los microorganismos adquieren luego el complejo  $\text{Fe}^{3+}$ -sideróforo mediante un transporte activo específico a través de receptores de la pared, situados en la membrana externa en las bacterias Gram-negativas. Algunas bacterias no solo producen sus propios sideróforos sino también producen receptores capaces de unir sideróforos producidos por otras bacterias (Ahmed y Holmström, 2014). Aunque la función principal de los sideróforos es la captación de  $\text{Fe}^+$  del medio, también llegan a formar complejos con otros elementos esenciales, es decir Mo, Mn, Co, y Ni del medio y así ponerlos a la disposición de las células bacterianas (Bellenger *et al.*, 2008; Braud *et al.*, 2009).

#### 1.5.7. Quorum sensing

Los conjuntos bacterianos son sistemas dinámicos complejos que interactúan entre sí, y gracias a esa interacción, las bacterias coexisten, colaboran, compiten e intercambian información de forma coordinada. Una forma de comunicación de las bacterias es el quorum sensing. Mediante este mecanismo las bacterias pueden conocer su concentración en un ambiente determinado y decidir el momento en el que se va a poner en marcha la expresión de un determinado conjunto de genes con el fin de desarrollar una respuesta concreta y de forma simultánea, como es el caso de la expresión de los factores de virulencia (March-Rosselló y Eiros-Bouza, 2013; Rutherford y Bassler, 2012). Esta comunicación implica la producción, detección y respuesta a moléculas de señalización extracelular llamadas autoinductores (IA). Los autoinductores se acumulan en el ambiente a medida que aumenta la densidad de población bacteriana, y las bacterias monitorean esta información para controlar la densidad bacteriana y la expresión de genes, además de la esporulación, bioluminiscencia, competencia, producción de antibióticos y formación de biopelículas. El quorum sensing controla señalizaciones de genes que son beneficiosas para las bacterias cuando actúan por grupos de bacterias en sincronía (Ng y Bassler, 2009; Novick y Geisinger, 2008; Williams y Cámara, 2009).

### 1.5.8. Enzimas hidrolíticas

Muchas bacterias liberan enzimas degradativas que disgregan los tejidos del huésped proporcionando así el alimento para el crecimiento de estas y facilitando la colonización. Existen diversas proteasas que son enzimas que pueden hidrolizar proteínas y péptidos, se dividen en peptidasas (exoproteasas) y proteinasas (endoproteasas) dependiendo de su actividad. Las peptidasas actúan sobre los enlaces peptídicos de los extremos de la cadena y pueden ser aminopeptidasas o carboxipeptidasas. Las proteinasas actúan en el interior de la cadena y se clasifican de acuerdo con la identidad del residuo catalítico primario. Así pueden ser: serinproteinasas, cisteinil-proteinasas, aspartilproteinasas y metalo-proteinasas (Chayakulkeeree *et al.*, 2008; Schmiel y Miller, 1999; Voigt *et al.*, 2005).

Algunas bacterias producen enzimas como la hialuronidasa, fibrinolisisina y las lipasas que modifican el medio tisular, mientras otras bacterias producen estreptocinasas y ADNasas las cuales facilitan el desarrollo de la infección y su diseminación. Un factor común utilizado por muchos patógenos, bacterianos, es la actividad de lipasa y fosfolipasa. Estas enzimas hidrolizan sustratos lipídicos, como triacilglicéridos o fosfolípidos, y cumplen diversas funciones en la patogénesis a través del daño / modulación de la célula huésped, señalización celular e inflamación (Ichinose *et al.*, 1994; Miyoshi *et al.*, 1993; Oguma *et al.*, 1995; Wretlind y Pavlovskis, 1983).

### 1.6. Genómica comparativa

La genómica comparativa es un campo de la investigación biológica en el que los investigadores usan una variedad de herramientas, entre ellas, los análisis computarizados, para comparar las secuencias del genoma completo de distintas especies. Al comparar detenidamente las características que definen a distintos organismos, los investigadores pueden localizar regiones de similitudes y diferencias (Hardison, 2003).

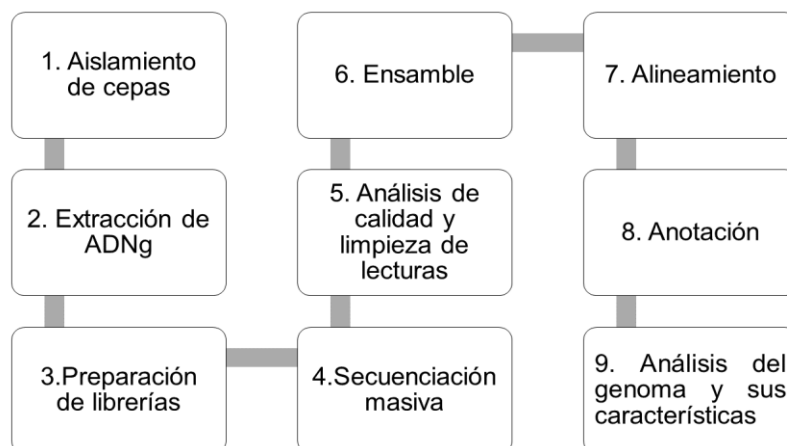


A medida que la tecnología de secuenciación del ADN se vuelve más poderosa y económica, la genómica comparativa está encontrando su lugar en aplicaciones más generalizadas en diversas áreas del conocimiento. Las tecnologías de secuenciación masiva (NGS) dan una nueva visión de lo que es la evolución microbiana, su interacción con el medio ambiente, la interacción con su hospedero, la diversidad microbiana, así como la variación genética, dando la posibilidad de generar nuevas hipótesis del comportamiento de algunos microorganismos de interés, como lo son las bacterias patógenas.

Las secuencias de los genomas de bacterias ahora pueden generarse en cuestión de horas en muchos laboratorios utilizando plataformas como Illumina MiSeq, IonTorrent, Roche, lo que permite determinar el orden de los nucleótidos (Adenina, Timina, Guanina y Citosina) de una muestra de ADN.

#### 1.6.1. Flujo del análisis informático para genomas bacterianos

Para poder determinar la secuencia de nucleótidos de una muestra de interés y las funciones que los organismos en estudio pueden desempeñar con respecto a su ambiente, se deben cumplir ciertos pasos. La Figura 3 muestra el flujo de trabajo a seguir desde la obtención de la muestra hasta el análisis genómico final.



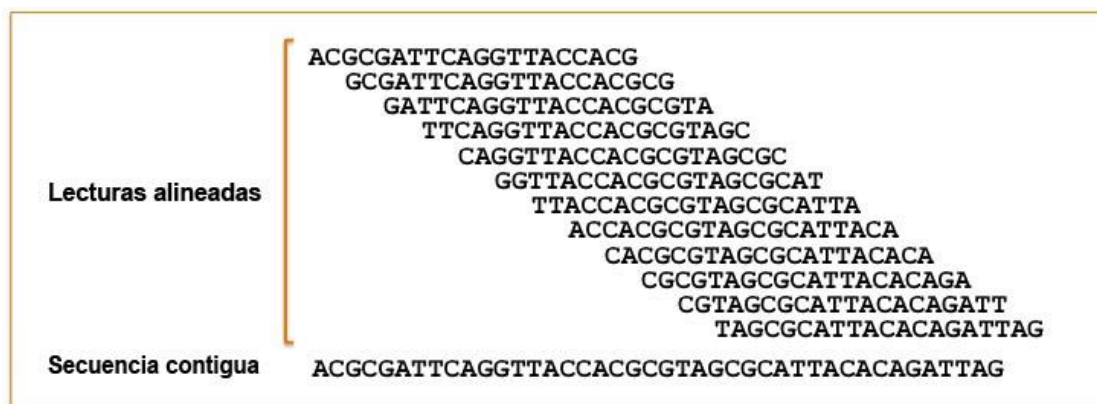
**Figura 3.** Flujo de análisis bioinformático para genomas.

Primeramente se debe obtener una cepa pura, es decir una población de células bacterianas que comparten las mismas características morfológicas y genéticas. Al conjunto de bacterias se les hace una extracción de ADN genómico para realizar la preparación de librerías, lo cual implica la generación de una colección de fragmentos de ADN para la secuenciación, que se obtienen fragmentando el ADN y ligando adaptadores específicos a los extremos de los fragmentos, los cuales se purifican, y luego se hace una amplificación por PCR. Estos fragmentos se aplican a una celda de flujo, utilizando nucleótidos marcados fluorescentemente para secuenciar millones de moléculas de ADN en la superficie de la celda en paralelo ([https://www.illumina.com/documents/products/techspotlights/techspotlight\\_sequencing.pdf](https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf)), a cada nueva hebra de ADN secuenciada se le denomina lectura.

Existen diferentes estrategias de secuenciación, siendo la más utilizada en NGS, es la de lecturas pareadas, donde se lee desde ambos extremos y en las dos direcciones. En general, en NGS se generan lecturas a partir de cada uno de los extremos de cada fragmento de ADN (el inserto), dando lugar a lecturas pareadas, y lo hacen usando dos estrategias diferentes: 1) mate pairs y 2) paired end. Las secuencias mate pairs se crean a partir de fragmentos de ADN de tamaño conocido (creando librerías con tamaños >600 pares de bases (pb), algunas 4kb), que se circularizan y se ligan usando un adaptador interno. Ya circularizados, se fragmentan al azar para luego purificar los segmentos que contienen el adaptador a partir del que se secuencia. Las secuencias paired end se generan mediante la fragmentación del ADN en pequeños segmentos (<300 pb) de los que se secuencia el final de ambos extremos. Proporcionan rangos de tamaños de inserto más estrechos, mientras que las de tipo mate pair tienen la ventaja de cubrir tamaños mayores.

Una vez obtenidas las lecturas se realiza un análisis de la calidad y la limpieza de estas, que consiste en remover los adaptadores, y lecturas de baja calidad (Phred score), así como otros parámetros como el contenido de GC y el número y longitud de secuencias. Posteriormente se realiza el ensamble de las lecturas, que consiste en el alineamiento y la fusión ordenada de los fragmentos de nucleótidos para formar una secuencia mayor mediante

superposición de lecturas, lo que produce un alineamiento de las miles de lecturas (Fig. 4), generando así una secuencia consenso de cada nucleótido alineado a lo que se le denomina secuencias contiguas (contigs) (<https://www.ncbi.nlm.nih.gov/assembly/basics/#WGA>).



**Figura 4.** Ensamble de lecturas. Composición de secuencias contiguas.

Una vez que se obtienen las secuencias consenso se realiza el alineamiento, el cual es el orden conservado de bloques genómicos alineados entre especies (un genoma de interés y un genoma de referencia de la misma especie), los alineamientos de genomas parciales o completos se pueden utilizar para determinar la conservación y/o las diferencias entre especies, para emparejar regiones que coinciden entre especies, con el objetivo de estudiar estas regiones en múltiples genomas (<https://www.ensembl.org/info/genome/compara/analyses.html>). Finalmente en la anotación de genomas se definen las propiedades funcionales de genes, promotores, pseudogenes, genes de ARN y otras regiones pequeñas de ARN, y funciones de una secuencia determinada los cuales se encargan del papel de las estructuras genéticas que codifican en la secuencia de ADN (Beckloff *et al.*, 2012).

Dadas todas las nuevas tecnologías mencionadas anteriormente se puede hacer una comparación de genomas de interés. A la comparación de varios genomas bacterianos se le ha asignado el nombre de análisis pangenómico, que representa el repertorio completo de todos

los genes presentes en todas las cepas de un clado bacteriano. Este pangenoma esta conformado por los genes del genoma nucleo, es decir aquellos genes que se encuentran en todas las cepas, más los genes accesorios (genoma dispensable) que es el repertorio de genes que se encuentran solo en algunas de las cepas del clado bajo analisis, siendo esta última característica una ventaja evolutiva para la adaptación al nicho (Tettelin *et al.*, 2005).

## 2. ANTECEDENTES

### 2.1. Vibriosis y otras enfermedades de la familia Vibrionaceae en acuicultura

Muchos miembros de la familia *Vibrionaceae* son capaces de formar asociaciones con huéspedes eucariotas (fitoplancton, algas, hongos acuáticos, invertebrados, peces y mamíferos marinos), que pueden variar desde ser dañinos, neutrales o benéficos para el hospedero (Soto *et al.*, 2010; Urbanczyk *et al.*, 2011). En el caso de las cepas patógenas provocan la enfermedad denominada Vibriosis, entre otras enfermedades causadas por dicha familia.

La Vibriosis es una enfermedad generada por organismos del género *Vibrio*, y es de las enfermedades más comunes en acuicultura. La manera en que bacterias del género *Vibrio* afectan al camarón blanco (*Litopenaeus vannamei*) es infectando y destruyendo su hepatopáncreas, que es el órgano principal digestivo en el camarón y otros crustáceos (Jory, 1995). Las cepas patógenas de *Vibrio* pueden ser o no luminiscentes (como es el caso de *Vibrio campbellii*), y afectan al camarón en cualquier etapa de crecimiento y desarrollo, llegando a causar signos de septicemia, infecciones localizadas o lesiones en varios órganos (corazón, branquias, hepatopáncreas, epidermis u órgano linfoide) (Gil, 2000). Otro de los organismos de importancia para la acuicultura es el molusco bivalvo *Crassostrea gigas*, el cual es afectado por bacterias del género *Vibrio*, las cuales se unen al epitelio del manto resultando en necrosis (Elston *et al.*, 1999), llegando a atravesar la cavidad celómica (cavidad donde se encuentran los tejidos blandos), durante este proceso el epitelio es también infectado, esto se debe a la producción de exotoxinas que son producidas por estas bacterias (Nottage *et*

*al.*, 1989). De las especies más importantes en cultivo de peces es el de *Totoaba macdonaldi*, y una de las enfermedades que causa mayor pérdida es la enfermedad fotobacteriosis causada por una bacteria Gram negativa *Photobacterium damsela*, causando una fascitis necrotizante, siendo esta una infección inflamatoria rápida y progresiva de la fascia, provocando una necrosis secundaria de los tejidos subcutáneos (Barber *et al.*, 2000).

## 2.2. Genomas bacterianos de la familia Vibrionaceae

Debido a que las bacterias patógenas ocasionan pérdidas en cultivos de organismos de interés, se ha intensificado la investigación sobre el comportamiento de estos patógenos. Las técnicas de secuenciación masiva para la obtención de los genomas y el análisis de estos, ha permitido un avance significativo en el estudio de bacterias patógenas. Dentro de los genomas secuenciados de *Vibrio* spp., se encuentran *Vibrio cholerae*, *V. parahaemolyticus*, *V. campbelli* entre otras especies de *Vibrio*, y se ha encontrado que el genoma de la familia Vibrionaceae se caracteriza por tener dos cromosomas.

Los dos cromosomas de *Vibrio* se encuentran conformados por características como la transferencia horizontal de genes, lo que puede explicar su capacidad de resistencia a antibióticos, virulencia y adaptación al nicho (Kirkup *et al.*, 2010; Rowe-Magnus *et al.*, 2001). Un ejemplo claro es *V. cholera* epidémica en Angola, donde al realizar un análisis se encontró variabilidad genética entre las variantes, destacando las islas genómicas, fagos y elementos conjugativos integrativos en la diversidad genética en un solo evento de epidemia. Otra especie es *V. campbellii*, una bacteria marina bioluminiscente a la cual se le ha analizado el fenotipo pigmentado de piomelanina, que tiene resistencia al estrés UV, oxidativo y una mejor adaptación en la colonización intestinal (Wang *et al.*, 2013). Hay indicios de que los *Vibrio* tiene un papel importante en el ciclo de nutrientes en el ambiente marino por medio de la ingesta de materia orgánica disuelta (Sherr *et al.*, 2000; 2002).

Los *Vibrios* pueden otorgar ácidos grasos poliinsaturados a organismos acuáticos que no pueden producirlos *de novo* (Cottrell *et al.*, 2003; Nichols *et al.*, 2003). Estos microorganismos pueden degradar quitina (un homopolímero de N-acetil-D-glucosamida), siendo uno de los polímeros más abundantes en los océanos (Cottrell *et al.*, 2000 a, b, c; Riemann *et al.*, 2002); *Vibrio harveyi* excreta por lo menos diez enzimas diferentes que degradan quitina (Suginta *et al.*, 2004; Svitil *et al.*, 1997). Debido a esta característica, se podría explicar porque la ocurrencia de *Vibrio* es ubicua en ambientes acuáticos (Riemann *et al.*, 2002).

En el caso de *V. cholerae* se realizó una investigación en la que se compararon 247 cepas de esta especie (Dutilh *et al.*, 2014), la cual logró asociar grupos de genomas de cada cepa a nichos específicos (tomando en cuenta tiempo, espacio y hábitat) recolectando metadatos del nicho, incluyendo 39 cepas de *V. cholera* que no se habían secuenciado. Los análisis revelaron variaciones en los genomas, principalmente en las funciones móviles, en particular fagos, profagos, elementos como transposones y plásmidos. Los resultados indican que los elementos móviles propician más rápidamente la evolución en los genomas bacterianos, creando así endemidad local (espacio), divergencia temporal (tiempo) y la capacidad de invadir nuevos hábitats (hábitat).

En otro trabajo realizado con *Vibrio*, se utilizaron dos genomas de la especie *V. parahaemolyticus*. Los resultados de la comparación mostraron que los *Vibrios* abarcan un genoma núcleo constante, mientras que el pangenoma es más amplio con ganancia de genes debido a eventos de transferencia horizontal. El análisis evolutivo indicó que el genoma núcleo surgió de *V. fischeri* aproximadamente 536 millones de años, junto con la ocurrencia de cefalópodos y el florecimiento de peces. En lo que respecta al tamaño del genoma es relativamente grande, esto puede deberse al alto número de copias de genes de 16s ARNr, la presencia de sistemas R-M (Restricción-Modificación) y a los sistemas de CRISPR, lo que indica la capacidad de *V. parahemolyticus* (y otros *Vibrios*) a vivir en diversos ambientes marinos. Se ha encontrado que los genomas de este género presentan genes relacionados a la degradación de quitina, y se ha observado que el número de genes que degradan quitinasa se

ha ampliado en comparación al genoma del ancestro más común; se estima que estos genes han ido evolucionando junto con el género, y han sufrido una fuerza selectiva de purificación para conservar el estado ancestral (Lin *et al.*, 2018).

### 2.3. Factores de virulencia en la familia Vibrionaceae

En la identificación de factores de virulencia de 28 cepas de *V. anguillarum* se caracterizaron e identificaron genes que confieren resistencia a antibióticos, y codifican para toxinas que mejoran el fitness del microorganismo. Lo anterior puede indicar que estos genes se obtuvieron por medio de la transferencia lateral, además la diversidad del genoma núcleo indicó que las cepas más virulentas se agrupan en conglomerados genéticos diferentes a aquellas cepas menos virulentas. Así, la virulencia de *V. anguillarum* es multifactorial, y tanto el genoma núcleo como el genoma flexible son fuentes significativas de genes de virulencia asociados a *Klebsiella pneumoniae*, *E. coli*, *S. aureus* y *P. aeruginosa*. Los autores sugieren que este patógeno al ser parte de la comunidad microbiana de un nicho específico como lo es el pez, existe una presión selectiva que permitió la adquisición de nuevos rasgos genéticos para aumentar su adecuación biológica y así potenciar su virulencia. Los autores respaldan dicha hipótesis con los datos sobre el hallazgo de islas genómicas, las cuales poseen un conjunto de genes de virulencia y otros elementos, impulsando así, la evolución de dicho patógeno (Castillo *et al.*, 2015).

#### 2.3.1. Adherencia

La unión de patógenos a las células del huésped es un requisito para la translocación eficiente de los efectores que suprimen el sistema inmunitario del huésped y/o modulan las vías de señalización celular para ayudar a su infección. La adherencia también garantiza la persistencia de los patógenos a formar parte de un nicho en el huésped. En *V. parahaemolyticus* se conocen diversos mecanismos de unión como la molécula de adhesión multivalente 7 (MAM7), la hemaglutina sensible a manosa (MSHA), el pilus, la enolasa, polisacárido capsular, y dos sistemas de secreción tipo VI (Hsieh *et al.*, 2003; Jiang *et al.*,

2014; Krachler *et al.*, 2011; O'Boyle *et al.*, 2013; Yu *et al.*, 2012). Sin embargo, de todas las estructuras mencionadas solo MAM7 se encontró patogénica (para un modelo en gusanos). MAM7 es una proteína transmembrana, y junto con el dominio N-terminal y otros siete dominios permite la unión con la fibronectina (Fn), y al ácido fosfatídico en la etapa inicial de infección (Bardiau *et al.*, 2010); Con la ayuda de la MSHA se da una cascada de señalización para los sistemas de secreción tipo 3, además se ha demostrado que el pili tipo IV (TFP) juega un papel crucial en la unión a células epiteliales (Giltner *et al.*, 2006). Liu y Chen (2015), identificaron un nuevo gen de adhesina vp1767, que se denominó VpadF (factor adhesivo de *V. parahaemolyticus*), contribuyendo a la unión y citotoxicidad en las células epiteliales. Este gen es esencial para el efecto letal de *V. parahaemolyticus* en ratones, y VpadF se une a los receptores de la superficie celular, fibronectina (Fn) y fibrinógeno (Fg), que pueden ayudar a contribuir a su colonización y patogénesis del huésped.

### 2.3.2. Toxinas

Kim *et al.*, (1999), identificaron el gen que codifica para una toxina denominada toxR, el cual parece estar bien conservado entre las especies de *Vibrio*. El grado de homología del gen toxR entre *V. parahaemolyticus* y *V. cholerae* (52% de identidad) es mucho más bajo que el del gen rRNA (91 a 92% de identidad). *V. parahaemolyticus* también tiene en su genoma las toxinas hemolisina directa termoestable (TDH) y la hemolisina relativa (TRH). La toxina TDH se une a las membranas de los glóbulos rojos de las células del huésped, y forma un poro en la superficie de la membrana causando un desbalance osmótico en las células. La toxina TRH causa niveles similares de hemólisis *in vitro* (Ceccarelli *et al.*, 2013; Honda *et al.*, 1988; Matsuda *et al.*, 2010). Otras de las toxinas identificadas en *V. parahaemolyticus* es la toxina binaria pirAB que se ha relacionado con la enfermedad de la necrosis hepatopancreática aguda (AHPND) en camarón. PirAB es sintetizada en forma de dos subunidades, denominadas PirA (14.0 kDa) y PirB (50.0 kDa). PirAB es una toxina binaria y es codificada por el plásmido PVA1. Los genes de PirA y PirB son homólogos de las toxinas producidas por la bacteria *Photobacterium luminescens*, las cuales se producen en simbiosis con Nematodos entomopatógenos (relacionadas a insectos, PIR, por sus siglas en inglés) y tienen propiedades



insecticidas (Lee *et al.*, 2015). Asimismo, Dong *et al.*, (2017), argumentan que *V. campbellii* puede albergar las toxinas PirA y PirB, y causar AHPND en camarones.

### 2.3.3. Sistemas de secreción

*V. parahaemolyticus* cuenta con dos sistemas de secreción de diferente tipo, dos T3SS (sistema de secreción tipo 3) y dos T6SS (sistema de secreción tipo 6). El sistema T3SS tiene como función la entrega directamente a citosol de los efectores (en su mayoría toxinas) (Tandhavanant *et al.*, 2018). Estos efectores una vez dentro del citosol de la célula huésped modifican las funciones de esta. El T3SS1 se ha encontrado mayormente en el cromosoma uno, y se encuentra codificado por la primera isla de patogenicidad en la mayoría de las cepas clínicas y ambientales (Paranjpye *et al.*, 2012). Varios estudios han demostrado que T3SS1 es citotóxico, causa autofagia, lo que provoca la muerte celular. Se han identificado cuatro efectores VopQ, VopS, VPA0450 y VopR. VopQ forma poros y se une a la V-ATPasa causando autofagia, lisis celular, activación de MAPK y secreción de IL-8 (interleucina quimiocina 8) (Shimohata y Takahashi, 2010). El efector VopS provoca la muerte de macrófagos inhibiendo la actividad de NF- $\kappa$ B, y este cambio permite la inhibición de la fagocitosis por los macrófagos (Higa *et al.*, 2013). El VPA0450 es una fosfatidilinositida fosfatasa que hidroliza fosfatidilinositida (4,5)- bídfofato, uno de los componentes principales de membranas celulares, donde el producto de esta unión interrumpe la homeostasis de la actina del citoesqueleto (Broberg *et al.*, 2010). El efector VopR se une a la membrana del hospedero por su dominio N-terminal, desempeñando una función importante en el plegado de T3SS, después de que entrega sus efectores en el citosol (Hicks y Galán, 2013).

Asimismo, la función del T6SS1 es competir contra otras bacterias, mientras la función del T6SS2 se relaciona con la adherencia a las células del huésped (Salomon *et al.*, 2013a). El sistema T6SS es funcionalmente análogo, y puede tener una función crítica en el proceso de infección. Se ha demostrado que el T6SS es necesario para la adhesión de *V. parahaemolyticus* a las células del huésped, así como, se involucra en el transporte vesicular (Salomon *et al.*, 2013b). Li *et al.*, (2017) estudiaron diversas cepas de *V. parahaemolyticus*

donde se comparó los genomas de bacterias que provocaban la enfermedad en camarón AHPND con otras que no la producían. Encontraron que las cepas que producían la enfermedad tienen el sistema T6SS1, y las que no producen la enfermedad no lo presentan, lo que confiere una ventaja para competir con otras bacterias y para llevar a cabo la infección en el camarón.

#### 2.3.4. Anti-fagocitosis

Algunas bacterias pueden evitar el sistema inmune del huésped y, por lo tanto, la fagocitosis, lo que se conoce como anti-fagocitosis. A pesar del impacto de la enfermedad bacteriana en los sistemas de acuicultura, solo unos pocos estudios directos han caracterizado la evasión del sistema inmune del huésped de bacterias del género *Vibrio*. Los *Vibrios* han desarrollado tropismo en los tejidos epiteliales, llegan a colonizar el epitelio, así como el tracto gastrointestinal. Las bacterias evaden el sistema inmune mediante la eliminación, inhibición o degradación de células del huésped, y por otro lado, algunas bacterias se rodean de una gruesa cápsula de polisacárido, que es un escudo muy eficiente contra el sistema inmunitario del huésped.

Lindell *et al.*, (2012) describieron en *V. anguillarum* un antígeno O-lipopolisacárido que cuando está presente le permite a la bacteria evadir la fagocitosis y el transporte del antígeno O requiere el operón *wzm-wzt-wbhA*, que codifica dos proteínas transportadoras de polisacárido ABC y una metiltransferasa. El pretratamiento de las células epiteliales con manosa evitó la fagocitosis de *V. anguillarum*, lo que sugiere que un receptor de manosa está involucrado en el proceso de absorción. Del mismo modo, la membrana externa de bacterias gramnegativas sirve como barrera de permeabilidad que no se atraviesa fácilmente por compuestos hidrofóbicos dañinos para las bacterias, por ejemplo, sales biliares de tracto gastrointestinal. Los lipo-polisacáridos de los patógenos gramnegativos pueden proteger a las células de lisis mediada por moléculas como la lisozima. La membrana externa y los componentes capsulares de bacterias gramnegativas pueden proteger la capa de

peptidoglucano de la actividad lítica de la lisozima (Gao *et al.*, 2003; Hughey y Johnson, 1987; Martínez de Tejada *et al.*, 1995).

### 2.3.5. Quimiotaxis

Muchas bacterias son móviles como *V. cholera* y *V. parahaemolyticus*, capaces de explorar su entorno. Uno de los principales movimientos usado por bacterias es la motilidad por flagelos. Los flagella son filamentos de proteínas helicoidales que sobresalen del cuerpo celular que giran con un motor flagelante que impulsa a la bacteria a través de un entorno circundante. Responden activamente a una variedad de estímulos modulando la dirección de rotación de su flagelo y ajustando así su comportamiento de natación (Berg y Purcell, 1977; Colin y Sourjik, 2017; Minamino e Imada, 2015). Debido a su pequeño tamaño, se sugiere que las bacterias no son capaces de medir las diferencias de concentración a lo largo de espacio, y por lo tanto han desarrollado una manera de notar los estímulos de manera temporal. Uno de los principales medios como detectan las diferencias de concentración es por el comportamiento quimiotáctico, en particular la quimiotaxis (Colin y Sourjik, 2017; Sourjik y Wingreen, 2012).

Muchas bacterias, como *Vibrio*, son monoflageladas y tienen este único flagelo en un polo de la célula. El flagelo bacteriano es un organelo de locomoción ampliamente utilizado entre los organismos procariontes, constituye una ventaja adaptativa y es un factor muy importante tanto para la colonización de nichos como para la patogénesis. En bacterias del género *Vibrio*, utilizan un flagelo polar cuya rotación está acoplada a un flujo de iones sodio ( $\text{Na}^+$ ) y algunas especies del género como *V. alginolyticus*, *V. parahaemolyticus* o *V. shilonii*, son capaces de expresar, adicionalmente al desplazarse en medio sólido, un sistema de flagelos laterales cuya rotación está acoplada a un flujo de protones ( $\text{H}^+$ ). La expresión del flagelo polar es constitutiva y está presente en la célula todo el tiempo, por otro lado, la expresión de los flagelos laterales es inducible y se encuentra asociada con la alta densidad del medio (Atsumi *et al.*, 1992). Bacterias de *V. parahaemolyticus*, son sociales y se agrupan en grupos en las superficies. Esta bacteria segrega el autoinductor de señal S, que indica la degradación intracelular de c-di-GMP. La señal S también dirige la motilidad a nivel de comportamiento al

servir como quimioatrayente. Los datos demuestran que *V. parahaemolyticus* detecta la señal S usando SscL y SscS, proteínas de quimiotaxis homólogas que aceptan metilo. La adaptación de estos quimiorreceptores depende de la metilación y desmetilación del quimiorreceptor, y los receptores son conocidos como proteínas de quimiotaxis que aceptan metilo (MCP) (Lamb *et al.*, 2019) *V. parahaemolyticus* expresa alrededor de 30 MCPs, y cabe mencionar que el gen que codifica para SscS está co-expresado como parte del flagelo lateral, junto con el operon scrABC responsable para la producción de la señal S (Gode-Potratz *et al.*, 2011).

### 2.3.6. Sideróforos

La colonización de un huésped presenta desafíos para la adquisición de hierro en los microorganismos. En animales, el nivel de hierro libre disponible es inferior al requerido para la multiplicación bacteriana. Para contrarrestar esta estrategia, las bacterias se han adaptado a un entorno de este tipo mediante el desarrollo de mecanismos para obtener hierro de los hospederos, dentro de estos mecanismos se encuentran lo que son sideróforos y xenosideróforos, proteasas y receptores de proteína de hierro (Payne *et al.*, 2016).

Los *Vibrios* al igual que otras bacterias han desarrollado una variedad de mecanismos para adquirir hierro de cada ambiente que habitan, algunos de estos sistemas de adquisición están estrechamente relacionados entre todos los *Vibrios*, lo que refleja su ascendencia en común. Los sideróforos son secretados al ambiente unidos a la superficie de *Vibrio spp.*, y sirven como captadores del hierro, una vez que el complejo ferrisideróforo es reconocido por un receptor específico en la superficie de la membrana externa (Neilands, 1982; Wyckoff *et al.*, 2007). El transporte a través de la membrana externa por medio del receptor requiere energía, que es suministrada por la proteína interna de membrana TonB y proteínas asociadas ExbB y ExbD (Postle *et al.*, 2007). Los vibrios típicamente tienen más de un conjunto de proteínas TonB/Exb, y estas proteínas tienen funciones específicas y redundantes (Occhino *et al.*, 1998). Una vez en el periplasma el complejo se asocia con una proteína de unión periplasmática (PBP) y se entrega a una membrana citoplasmática permeasa para su transporte al citoplasma (Braun y Hantke, 2011; Chu *et al.*, 2011). Las proteínas de unión periférica y las permeasas

citoplasmáticas son menos específicas que los receptores de membrana externa, y reconocen una clase más amplia de complejos, generalmente adicionales de la misma clase de sideróforos. Una vez dentro del citoplasma bacteriano, el hierro se libera del sideróforo mediante la reducción del hierro y/o la escisión enzimática del sideróforo, por lo que se encuentra disponible para las necesidades celulares (Butterton *et al.*, 1994).

Sandy *et al.*, (2010), reportan varios sideróforos tipo amida en *Vibrio* sp. como la trivancrobactina, divancrobactina, vancrobactina y anguibactona. La cantidad de hierro requerida por los *Vibrios* varía dependiendo de la fisiología y metabolismo de la célula, pero las concentraciones en el medio van de un rango de 0.1  $\mu\text{M}$  - 5.0  $\mu\text{M}$  proporcionando el suficiente hierro para el crecimiento óptimo de las bacterias (bajo condiciones de laboratorio).

### 2.3.7. Quorum sensing

En *V. parahaemolyticus* la comunicación del QS (Quorum sensing) se da a través de la secreción del autoinductor-2 (AI-2), una vez que el AI-2 ha alcanzado una concentración crítica, el sistema QS comienza a inducir la expresión de varios factores de virulencia, incluyendo la formación de biopelículas, motilidad, adhesión, T3SS y producción de toxinas (Mizan *et al.*, 2016; Zhu *et al.*, 2002; Zhu *et al.*, 2013).

A baja densidad celular, las concentraciones de autoinductores también son bajas, las quinasas fosforilan el regulador LuxO dependiente de LuxU, y Luxo-P induce la transcripción de pequeños ARN reguladores del QS (Qrrs), estos a su vez desestabilizan el ARNm para el regulador de salida. Aunque el sistema central del QS parece funcionar de manera similar en todos los *Vibrios*, los organismos difieren con respecto al número, tipos de autoinductores, quinasas de receptores y sensores afines, el número de genes Qrr y los tipos de genes que regulan la salida (Gode-Potratz y McCarter, 2011; Milton, 2006).

En un estudio comparando genomas de cepas clínicas y cepas del ambiente, se llegó a las conclusiones de que filogenéticamente las cepas clínicas son distintas a las cepas ambientales,

los análisis mostraron un nuevo agrupamiento de proteínas asociadas a la formación de biofilms de *V. parahaemolyticus*, sugiriendo que esta nueva agrupación es una característica de las cepas ambientales, lo que las distingue de las cepas clínicas. Se encontró además una nueva isla genómica (VPaI-8), que frecuentemente se distribuye en todas las cepas clínicas, se observó además de las características de factores de virulencia y profagos que una de las cepas clínicas contiene un nuevo tipo de *toxRS* y una variedad de elementos de profagos nueva en las cepas ambientales (Li *et al.*, 2014).

#### 2.4. Bacterias patógenas asociadas en la acuicultura en México

El presente trabajo parte de un proyecto en donde se aislaron ocho cepas bacterianas de diferentes organismos marinos (*L. vannamei*, *T. macdonaldii* y *C. gigas*) que se encontraban en granjas acuícolas de Baja California Sur, los cuales presentaban signos de enfermedad (nado errático, necrosis, etc.) (Rojas *et al.* 2017). A estas bacterias se le realizaron diferentes pruebas bioquímicas e identificación molecular con el gen 16s del ARNr (Tabla I y II). Todas las cepas aisladas dieron verdes en el medio selectivo TCBS, lo que indicaría que dichos aislados son *V. parahaemolyticus*, todas las bacterias dieron positivas a movilidad en el medio de MIO, a excepción de las cepas 1A y N, lo cual puede indicar la presencia de un flagelo que permita a estas bacterias desplazarse en el medio. Todas las bacterias dieron negativas a Indol, prueba que se utiliza para identificar patógenos entéricos. Para la prueba de Citrato de Simmons dieron positivas para A1 y M, las demás cepas dieron negativas, esta prueba es positiva para *Enterobacterias*. La máxima temperatura a la que mostraron crecimiento fue a 42°C. Estas bacterias se utilizaron para un reto de infección con post-larvas de camarón, dichos camarones mostraron los signos de infección, siendo las cepas A1 y M las cepas que causaron un mayor grado de daño, seguido por las cepas 6, 1A y T01, mientras que las cepas 7L y N no se pudo determinar su grado de daño (Tabla I).

**Tabla I.** Pruebas bioquímicas y grado de daño.

Cepa	H	TCBS	Movilidad	Indol	Citrato de Simmons	Crecimiento a 42°C	Patogenicidad
T01	F	G	+	-	-	+	+++
A1	O	G	+	-	+	+	++++
02	S	G	+	-	-	+	++
7L	O	G	+	-	-	+	ND
1A	S	G	-	-	-	+	+++
6	S	G	+	-	-	+	+++
M	S	G	+	-	+	+	++++
N	S	G	-	-	-	+	ND

H= Huésped, F= Pez *T. macdonaldii*, O= Ostión *C. gigas*, S= Camarón *L. vannamei*, G= Verde, + =positivo, -= negativo, ND= No determinado, += grado de daño.

**Tabla II.** Identificación de las bacterias y crecimiento a diferentes concentraciones de NaCl.

Cepa	H	16s ARNr	Crecimiento en NaCl					
			0%	1%	3%	6%	8%	10%
T01	F	<i>Photobacterium damsela</i>	+	+	+	+	-	-
A1	O	<i>Vibrio parahaemolyticus</i>		+	+	+	-	-
02	S	<i>Vibrio campbellii</i>	+	+	+	+	+	+
7L	O	<i>Vibrio parahaemolyticus</i>	+	+	+	+	-	-
1A	S	<i>Vibrio parahaemolyticus</i>	-	+	+	+	+	+
6	S	<i>Vibrio parahaemolyticus</i>	-	+	+	+	+	+
M	S	<i>Vibrio parahaemolyticus</i>	-	+	+	+	+	+
N	S	<i>Vibrio parahaemolyticus</i>	-	+	+	+	+	+

H= Huésped, F= Pez *T. macdonaldii*, O= Ostión *C. gigas*, S= Camarón *L. vannamei*

### 3. JUSTIFICACIÓN

Diversas enfermedades provocan daños graves y pérdidas económicas en la industria acuícola a nivel mundial, y las granjas acuícolas de Baja California Sur forman parte de este problema. Para una mejor comprensión y prevención de las enfermedades en las granjas acuícolas, es esencial caracterizar los agentes causales. Las cuales pueden ser de carácter microbiano, las bacterias patógenas han desarrollado una serie de mecanismos diferentes, que resultan en enfermedad en el huésped. Una manera efectiva de evaluar a las bacterias que provocan enfermedad, es conocer su contenido de ADN genómico, ya que la estructura del genoma de un organismo dicta parte del comportamiento de este.

El estado de Baja California Sur es considerado una zona seca-desértica por su clima, lo que conlleva a que tenga características áridas y semiáridas, por lo que los microorganismos que se encuentran viviendo bajo esta presión selectiva ambiental particular se diferencian de otros organismos en su estructura del genoma, aunado a lo anterior los cultivos hipertensivos en acuicultura pueden tener variables de estrés para organismos de interés económico (*L.vannamei*, *T.macdonaldii* y *C.gigas*), derivando así en la invasión de bacterias patógenas.

Por lo que la ampliación del conocimiento en el comportamiento de estos patógenos da pie al desarrollo de nuevas tecnologías para su temprana detección y/o prevención. Las tecnologías de secuenciación de nueva generación (NGS) han revolucionado el campo de la genómica, permitiendo la secuenciación de un gran número de genomas en muy poco tiempo, en el presente proyecto se hace uso de dichas herramientas.



#### **4. HIPÓTESIS**

Las bacterias patógenas en la Familia Vibrionaceae presentan diferencias en la información genética dependiendo la procedencia del organismo, la función que cumple en su hospedero, así como, variaciones inter e intraespecíficamente. Identificar y caracterizar los genomas de bacterias patógenas aisladas de camarones, peces y ostiones, nos permitirá definir que las características genómicas se comparten entre cepas de la misma especie, pero no de la misma Familia.

## **5. OBJETIVOS**

### **5.1. Objetivo general**

Evaluar y comparar los genomas de bacterias halófilas de la Familia Vibrionaceae aisladas de granjas acuícolas de Baja California Sur que presentan signos de patogenicidad en diferentes hospederos (*Crassostrea gigas*, *Litopenaeus vannamei* y *Totoaba macdonaldi*).

### **5.2. Objetivos particulares**

- 1) Obtener y evaluar las secuencias nucleotídicas de los cromosomas y bacteriófagos de las bacterias aisladas.
- 2) Comparar e identificar variaciones particulares y regiones conservadas entre genomas.
- 3) Identificar *in silico* los mecanismos de virulencia de las cepas aisladas, y compararlos en los diferentes hospederos

## 6. MATERIAL Y MÉTODOS

### 6.1. Extracción de ADN total y secuenciación

Se obtuvieron organismos (*Totoaba macdonaldi*, *Litopenaeus vannamei* y *Crassostrea gigas*) con signos de infección por *Vibriosis*, de los cuales se realizaron aislamientos bacterianos haciendo raspados en el tracto digestivo para *L. vannamei* y *C. gigas*, y de estómago e intestino para *T. macdonaldi*. Los aislamientos se realizaron en medios selectivos como CHROMagar y TCBS. Los aislamientos y selección de bacterias se realizaron de acuerdo a los resultados obtenidos por Rojas *et al.*, (2017), eligiendo las cepas que crecieron a una concentración  $\geq 6\%$  de NaCl, un crecimiento a 42°C, y que mostraron características patogénicas en un reto de infección con post-larvas de *L. vannamei* (Tablas I y II)

Una vez obtenidas las bacterias se llevó a cabo la extracción de ADN total mediante el método de extracción orgánico tradicional de Sambrook *et al.* (1989), y el ADN se resuspendió en buffer TE 10 mM pH 8. La pureza y concentración del ADN se estimaron utilizando el espectrofotómetro Nanodrop 1000 (Thermo Fisher Scientific) y el bioanalizador Agilent 2100 (Agilent EUA). La integridad del ADN también se evaluó por electroforesis en un gel de agarosa al 1.5% y se visualizó en el sistema de imágenes UVP Biodoc-It 2 (Analytik Jena), teñido con el colorante fluorescente GelRed (Biotium).

La secuenciación se realizó por parte de la empresa GeneWiz (South Plainfield NJ, EUA), con la plataforma de Illumina MI Seq System v2, con lecturas pareadas (Fw y Rv), dichas lecturas con tamaños de 150 pb por las dos vías. Como resultado de la secuenciación masiva se obtuvieron archivos FASTQ por medio de la empresa GeneWiz (South Plainfield NJ, EUA), los cuales se depositaron en el servidor bioinformático Biobacter del CIAD unidad Mazatlán (System76 Starling Pro ARM con 96 núcleos, 128 MB RAM, 16 TB HDD. Con sistema operativo Ubuntu server 16.02 LTS 64-bit.), para analizarlos vía remota a través de SSH (Secure Shell).

## 6.2. Calidad y limpieza de las secuencias

Los archivos .fastq depositados, fueron analizados con el programa `nextera_cleaner.sh` ([https://github.com/GenomicaMicrob/nextera\\_cleaner](https://github.com/GenomicaMicrob/nextera_cleaner)), el cual se encuentra compuesto de varios programas: 1) Cutadapt (Martin, 2011), que tiene como propósito remover las secuencias de adaptadores que fueron añadidas durante la secuenciación; 2) Flash (Magoč y Salzberg, 2011), es una herramienta de ajuste rápido de longitud de lecturas cortas que está diseñado para combinar pares de lecturas cuando los fragmentos de ADN son más cortos que el doble de la longitud de las lecturas; 3) programa FastQC (Trivedi *et al.*, 2014) el cual se encarga de analizar la calidad de las lecturas en base al nivel de la calidad de Phred (Q), valor que indica la probabilidad de que una base determinada sea nombrada incorrectamente por el secuenciador, que generalmente se toma como parámetro  $Q \geq 30$ , donde Q30 indica que el factor de error sería de 1 base entre 1000 con el mismo nivel de calidad Phred, con un 99.9% de precisión; 4) FASTX-Toolkit (`fastx_trimmer`) ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)), el cual permite mejorar la calidad de las secuencias mediante la eliminación de los nucleótidos con mala calidad o mediante la limitación de la longitud de las secuencias, en el caso de nuestras lecturas se removieron los primeros 15 nucleótidos, y los últimos 15 nucleótidos. Posteriormente se utilizó el programa PEAR (Zhang *et al.*, 2013), para unir secuencias de cada par de archivos (sentido y antisentido), esto con la finalidad de obtener lecturas de un tamaño más grande y conseguir así un mejor ensamble. Este programa evalúa todas las posibles superposiciones de los extremos de las lecturas pareadas. Este programa genera tres archivos con dos extensiones diferentes `notmerged` y `merged`; las lecturas que lograron unirse se encuentran en el archivo `merged` y aquellos fragmentos cortos que no lograron unirse se encuentran en los archivos `notmerged`.

## 6.3. Ensamble de las secuencias de nucleótidos

Para poder obtener un genoma completo es necesario unir estas secuencias que previamente fueron fragmentadas, para lo que existen diversos programas que realizan este trabajo, lo más recomendable es utilizar varios programas y comparar entre estos para poder obtener así el

mejor ensamble, por lo que para esto se utilizaron diversos programas (A5-miseq, SPAdes, MIX).

### 6.3.1. Ensamble de los cromosomas

Para realizar los diversos ensambles se utilizó el programa `genome_assembler` que se encuentra en el servidor Biobacter (CIAD, Mazatlán), el cual despliega un menú con varios programas para ensamblar genomas, una vez seleccionado el programa a utilizar se introdujeron las secuencias en formato fasta que el ensamblador utiliza. Los dos ensambladores usados fueron A5-Miseq (Coil, *et.al.*, 2014) y SPAdes (Nurk *et al.*, 2013). A5-Miseq es un ensamblador que trabaja con secuencias generadas por la plataforma de Illumina y con secuencias con longitudes mayores a 80 nucleótidos. Este ensamblador tiene un conjunto de programas organizados para analizar la calidad de las lecturas, cortar adaptadores, correcciones de errores, generación de secuencias contiguas y detección de malos ensambles. Por otro lado, SPAdes es un programa que funciona con lecturas de Illumina y IonTorrent, y es capaz de realizar ensambles híbridos, es decir lecturas con diferentes plataformas como PacBio, Oxford Nanopore y Sanger. Puede utilizar lecturas pareadas, no pareadas y lecturas mate-paired, además de tomarlas como entrada simultáneamente. Este programa fue diseñado para genomas pequeños (como los de bacterias).

Se utilizó además un ensamble de ensambles con el programa MIX (Soueidan *et.al.*, 2013). Este programa combina dos ensambles previamente realizados, para lo cual se utilizaron los archivos fastos generados del resultado del ensamble con A5-Miseq y SPAdes. Este programa no necesita un genoma de referencia, tiene como objetivo en reducir la fragmentación de secuencias contiguas y así obtener un genoma más completo. El algoritmo del programa crea un gráfico de extensión donde los vértices representan las extremidades de las secuencias contiguas y los bordes representan alineamientos existentes entre estas extremidades; estas aristas de alineación se utilizan para la extensión de las secuencias contiguas.

### 6.3.2. Validación del ensamble

La validación de los ensambles realizados se llevó a cabo con el programa QUAST v4.6.0, 22f3f69 (Gurevich *et al.*, 2013). QUAST evalúa los ensambles de los genomas, utilizando varias características métricas que necesitan diferentes usuarios, sin embargo, el número de características métricas no es tan grande ya que eso dificultaría su interpretación. QUAST puede evaluar la calidad del ensamble incluso sin un genoma de referencia. A pesar de que no es necesario un genoma de referencia, el utilizar alguno da una mejor visión de que tan exitoso fue el ensamble, así que se utilizaron genomas de referencia completamente cerrados y anotados (Tabla III).

Los genomas fueron descargados en formato gbk (GenBank), el cual contiene las características y funciones de los genes del genoma, así como en formato fasta. Se utilizó el programa `assembly_download.sh` ver.0.0.4 el cual se encuentra disponible en el servidor de Biobacter, y este script baja un archivo txt de genomas de procariotas, y al ingresar una palabra clave (el nombre de la cepa) extrae el número de genomas que coincidan con la palabra ingresada. Una vez obtenidos los genomas se procedió a validar y comparar los ensambles.

**Tabla III.** Genomas de referencia para el ensamble y la validación.

<b>Especie</b>	<b>Cepa</b>	<b>GenBank</b>
<i>Vibrio parahaemolyticus</i>	RIMD 2210633 O3:K3	NC_004603.1
<i>Photobacterium damsela subsp. damsela</i>	KC.Na-1	NZ_CP021151.1
<i>Vibrio campbellii</i>	LMB29	NZ_CP019293.1

Quast generó una carpeta en una ruta indicada previamente, `/quast/results`, con los resultados de la comparación de los ensambles, dando un reporte final en formato html. Para elegir el mejor ensamble se siguieron ciertos criterios como el número de lecturas (N50), longitud de

las lecturas (L50), número de secuencias contiguas, secuencia contigua más larga, ensamble con un comportamiento parecido al genoma de referencia, etc.

Adicionalmente, se utilizó el programa Mauve para ordenar las secuencias contiguas en la interface grafica basada en Java (Darling *et al.*, 2004), Los errores de secuenciación se pueden corregir mediante la elaboración de un consenso de un alineamiento de múltiples secuencias solapadas. El algoritmo de ordenamiento de Mauve, utiliza una aproximación de un mapeo interactivo para encontrar el mejor ajuste de las secuencias contiguas contra un genoma de referencia. Mauve toma como entrada el genoma de referencia en formato fasta junto con el ensamblaje en formato multifasta, y emite otro archivo multifasta que contiene las secuencias contiguas ordenadas (5'→3'). La Tabla IV muestra los genomas de referencia utilizados para llevar a cabo el ordenamiento de las secuencias obtenidas.

### 6.3.3. Identificación de secuencias de bacteriófagos

Para la identificación de posibles secuencias provenientes de fagos, se utilizó la página web PHASTER (Arndt *et al.*, 2016) (<http://phaster.ca/>). Este programa identifica y anota secuencias de fagos, y acepta datos de secuenciación masiva en formato GenBank y fasta. Realiza búsquedas utilizando BLAST en una base de datos de fagos/profagos de NCBI; los genes identificados como-fagos es por medio de DBSCAN, y los genes identificados como no-fagos se anotan mediante una segunda búsqueda de BLAST contra una base de proteínas no redundantes, y a las regiones de profagos se les asigna una puntuación de integridad basada en la proporción de genes de fagos en la región identificada.

### 6.4. Anotación de las secuencias

Una vez obtenido el ordenamiento de los contigs (secuencias contiguas), se prosiguió a asignarle un nombre y función a las secuencias, para lo cual se utilizaron dos programas diferentes uno por línea de comando llamado PROKKA ver 1.14-dev (Seemann, 2014), y una página web RASTtk (<http://rast.nmpdr.org/rast.cgi>). El programa de PROKKA (Prokaryotic

Genomes Automatic Annotation Pipeline), Prokka (Seemann, 2014) utiliza secuencias de ADN genómicas ensambladas en contigs o en scaffolds (contigs unidos). Este programa se basa en herramientas externas de predicción de características para identificar los segmentos de estas características genómicas dentro del ensamble. Prokka se basa en varias herramientas para la predicción de características genómicas dentro del ensamble, todas excepto Prodigal otorgan coordenadas y etiquetas adecuadas para describir dicha entidad. Prokka utiliza Prodigal (Hyatt *et.al.*, 2010) para identificar secuencias codificantes (CDS), RNAmmer (Lagesen *et al.*, 2007) para genes ARN ribosomales, Aragorn (Laslett *et al.*, 2004) para genes de ARN de transferencia, SignalP (Petersen *et al.*, 2011) para péptidos líderes de señal, e Infernal (Kolbe y Eddy, 2011) para ARN no codificantes.

**Tabla IV.** Genomas de referencia utilizados para el reordenamiento de secuencias contiguas.

Cepas de interés	Cepas de referencia	GenBank numero de acceso
<i>Photobacterium damsela</i> <i>subsp. damsela</i> T01	<i>Photobacterium damsela</i> <i>subsp. damsela</i> Phdp Wu-1	NZ_CP010883.1 <sup>a</sup> NZ_CP010884.1 <sup>b</sup>
<i>V. campbellii</i> 2	<i>V. campbellii</i> 20130629003S01	NZ_CP020076.1 <sup>a</sup> NZ_CP020077.1 <sup>b</sup>
<i>V. parahaemolyticus</i> A1	<i>V. parahaemolyticus</i> FORC_023	NZ_CP012950.1 <sup>a</sup> NZ_CP012951.1 <sup>b</sup>
<i>V. parahaemolyticus</i> N	<i>V. parahaemolyticus</i> FORC_023	NZ_CP012950.1 <sup>a</sup> NZ_CP012951.1 <sup>b</sup>
<i>V. parahaemolyticus</i> M	<i>V. parahaemolyticus</i> CHN25	NZ_CP010883.1 <sup>a</sup> NZ_CP010884.1 <sup>b</sup>
<i>V. parahaemolyticus</i> 1A	<i>V. parahaemolyticus</i> 20130629002S01	NZ_CP020034.1 <sup>a</sup> NZ_CP020035.1 <sup>b</sup>
<i>V. parahaemolyticus</i> 7L	<i>V. parahaemolyticus</i> 20130629002S01	NZ_CP020034.1 <sup>a</sup> NZ_CP020035.1 <sup>b</sup>
<i>V. parahaemolyticus</i> 6	<i>V. parahaemolyticus</i> 10329	NZ_AFBW00000000.1

a = No. de acceso correspondiente a las secuencias del cromosoma I

b= No. de acceso correspondiente a las secuencias del cromosoma II



El programa RASTtk (Aziz *et al.*, 2008; Brettin *et al.*, 2015; Overbeek *et al.*, 2013), funciona proyectando anotaciones genéticas relacionadas manualmente de la base de datos de SEED que se organiza en subsistemas estructurados de roles funcionales lógicamente. RASTtk cuenta con Prodigal, que además de asignar los CDS da posiciones de inicio y termino, y diferencias en el porcentaje de GC, contiene programas para genes de ARNr y ARN de transferencia, regiones repetidas, CRISPRs, secuencias de inserción.

En el caso de RASTtk se ingresaron los genomas en formato .fasta ya ordenados previamente. Para los requerimientos se utilizó para las cepas de *V.parahaemolyticus* el ID taxonómico (txid) 670, obtenido de la base de datos TaxonomyBrowser (<https://www.ncbi.nlm.nih.gov/taxonomy/>). Para la cepa *V. campbellii* (2) su txid 680 y para *P. damsela subsp. damsela* (T01) se utilizó el txid85581. En el caso de *V.parahaemolyticus* y *V.campbellii* se utilizó el linaje *Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio; grupo Vibrio harveyi* debido a que estas dos bacterias pertenecen al mismo clado; y para la cepa de *Photobacterium* T01 se utilizó *Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Photobacterium; Photobacterium damsela*. Se utilizó el código genético 11, ya que este corresponde a bacterias. Las opciones para la anotación que se seleccionaron fueron: el esquema de anotación RASTtk, siendo este el último esquema actualizado, se seleccionaron todos los parámetros para la customización del pipeline (tubería de programas), arreglo de cambios de lectura, y construcción del modelo metabólico.

## 6.5. Filogenia

### 6.5.1. Identidad promedio de nucleótidos (ANI)

La identidad promedio de nucleótidos (ANI en sus siglas en ingles), es una medida de la similitud genómica a nivel nucleotídico entre las regiones codificantes de dos o más genomas. Esta herramienta realiza una alineación de las secuencias identificando regiones coincidentes,

para obtener un promedio de todas las regiones coincidentes. En este caso se utilizó el método ANIm, donde utiliza MUMmer (NUCmer) para alinear las secuencias (genomas) ingresados. ANI se ha propuesto para la sustitución de hibridación DNA-DNA, en donde se marca como umbral la relación a nivel especie con un 95% (Goris *et al.*, 2007; Richter y Rosselló-Móra, 2009).

En este método se descargaron todos los genomas completos de nuestras especies de interés *V. parahaemolyticus*, *V. campbellii* y *P. damsela subsp. damsela* (Anexo A) por medio de línea de comando. Se generó una carpeta para cada especie, en la cual se depositaron los genomas descargados y los genomas de interés, y un archivo .txt con las etiquetas de los genomas a analizar. ANI crea al final una carpeta de salida en donde se almacenan los resultados, en la cual se encuentra un archivo .png donde muestra una matriz tipo mapa de calor y un archivo .xls donde se encuentra dicha matriz con los valores en porcentaje de los alineamientos.

#### 6.5.2. Tipificación multilocus de secuencias (MLST)

Se llevó cabo el análisis de tipificación multilocus de secuencias (MLST), el cual proporciona una mayor resolución de las relaciones filogenéticas dentro de una familia. Para generar árboles filogenéticos se utilizan secuencias parciales de genes que codifican para proteínas con funciones conservadas (a los que se les denomina comúnmente genes conservados o housekeeping genes en inglés).

Se analizaron nueve genes de referencia utilizados comúnmente en la familia *Vibrionaceae* para hacer el análisis (Tabla V). Se obtuvieron las secuencias de dichos genes de genomas descargados de la base de datos Genome del NCBI (<https://www.ncbi.nlm.nih.gov/genome/genomes/>) de *V. parahemolyticus*, donde se ingresaron los genomas en formato fasta, y dichas secuencias se compararon con la base de datos PubMLST (<https://github.com/tseemann/mlst>).

**Tabla V.** Genes de referencia de la familia Vibrionaceae.

<b>Gen</b>	<b>Nombre</b>	<b>Función</b>
<i>gyrB</i>	ADN girasa subunidad B	Se encarga de la unión del ADN bicatenario y disminuye la asociación de la DnaA con el OriC inhibiendo así la replicación.
<i>pyrC</i>	Dihidrotasa	Cataliza la ciclación reversible del carbomil del aspartato a dihidrotato.
<i>recA</i>	Proteína recA, recombinasa A.	Participa en la recombinación homologa y reparación de ADN por mutágenos.
<i>dnaE</i>	ADN polimerasa subunidad $\alpha$	III Enzima responsable de la mayor parte de la síntesis durante la replicación de las bacterias.
<i>recA</i>	Proteína RecA	Participa en la recombinación homologa de ADN
<i>pntA</i>	NAD transdeshidrogenasa subunidad $\alpha$	(P) Participa en la transhidrogenación entre NADH y NADP, acoplándose en la respiración e hidrolisis de ATP como una bomba de protones a través de la membrana.
<i>tnaA</i>	Triptofanasa	Proteína involucrada en la síntesis de indol y piruvato a partir de L-triptófano.

### 6.5.3. Filogenia 16S ARNr

La comparación de secuencias del gen 16 ARNr permite establecer relaciones filogenéticas existentes entre bacterias. Este gen es de aproximadamente 1500 nucleótidos y se encuentra altamente conservado en todas las bacterias. Para dicho gen se obtuvo su secuencia en los genomas a analizar (Anexo B) con RNAmmer (Lagesen *et al.*, 2007) el cual predice genes de ARNr ribosomal y se prosiguió a su alineamiento con MAFT (Kato y Standley, 2013). MAFT utiliza un MSA (Multiple sequence alignment) que tiene varias estrategias de alineamiento, métodos progresivos, método de refinamiento iterativo y alineamiento para ARNs estructurales. La limpieza se realizó con GBlocks (Castresana, 2000) el cual elimina las posiciones mal alineadas y las regiones divergentes del ADN o proteínas para que sea más adecuada para el análisis filogenético. Posteriormente se realizó la construcción del árbol con Fasttree (Price *et al.*, 2009), el cual infiere arboles filogenéticos basados en el método de Maxima-Verosimilitud.

## 6.6. Comparación de genomas

### 6.6.1. Atlas estructural de cromosomas de *V. parahaemolyticus*

Las lecturas secuenciadas fueron mapeadas para identificar los cromosomas en las cepas estudiadas de *Vibrio parahaemolyticus* con la base de datos CBS Genome Atlas (<https://www.cbs.dtu.dk/services/GenomeAtlas/>) (Vesth *et al.*, 2013), con el fin de llevar a cabo un análisis estructural para hacer una descripción básica de los elementos que componen el genoma y de cómo están organizados. Se hizo una comparación de los genomas de *V. parahaemolyticus* (cepas A1, 1A, M, 7L, N y 6) del proyecto, con el genoma de *V. parahaemolyticus* ATCC 17802 reportada cepa no patogénica como referencia (Klein *et al.*, 2014), y como genoma de referencia se utilizó la cepa clínica RIMD 2210633, para encontrar elementos que se encuentren ausentes en la cepas A1, 1A, M, 7L, N y 6. Por lo que se creó un mapa de cada cepa patógena y se obtuvieron las secuencias de aminoácidos en formato fsa de todas las cepas, por medio de Prodigal (Hyatt *et al.*, 2010). Se generó un archivo con la configuración del genoma de referencia (cepa patogénica del proyecto), para posteriormente agregar los genomas de referencia a dicha configuración. Una vez realizado el análisis el mapa se almacenó en un archivo ps (postscript). A partir de la información de los ficheros de secuencias del GenBank, el sistema representa, en forma de un mapa circular, una serie de parámetros: 1) Tamaño del genoma; 2) Identidad entre secuencias seleccionadas; 3) Curvatura, proporcional a la propension del DNA a curvarse; 4) Energía de apilamiento, cuanto más negativa, menor es la propension de las hebras del DNA a separarse; 5) Posición preferencial, proporcional a la rigidez del DNA; 6) Secuencias codificantes, en las cadenas (+) y (-); repeticiones directas; 7) repeticiones inversas; 8) Porcentaje de AT y asimetría de GC, entre otros.

### 6.6.2. Preferencia de uso de codones y aminoácidos

Para la creación de mapas de calor, para el cálculo de la preferencia del uso de codones y aminoácidos, se utilizaron las herramientas del CMG-biotools (<http://www.cbs.dtu.dk/staff/dave/CMGtools/>). Se utilizaron los archivos faa (archivo en fasta con secuencias de aminoácidos), posteriormente de los archivos de salida se utilizaron en el programa de R, para la visualización de los mapas con el programa de gplots (<https://cran.r-project.org/web/packages/gplots/index.html>), los genomas utilizados fueron descartados de la base de datos NCBI (Anexo B).

### 6.6.3. Comparación de proteomas

Este análisis tiene el propósito de comparar el porcentaje del genoma que se comparte entre cepas, además de identificar eventos de duplicación por medio de sus secuencias de aminoácidos. Todas las secuencias se comparan entre sí, y si comparten un 50% de identidad y en la longitud de la secuencia en la alineación con BLAST estas dos secuencias son similares. Lo anterior se realizó con los programas de CMG-Biotools (<http://www.cbs.dtu.dk/staff/dave/CMGtools/>) (Vesth *et al.*, 2013). Para realizar el análisis se creó una carpeta y se almacenaron los genomas a analizar (Anexo B) en formato fsa obtenido previamente en la anotación. Posteriormente se generó una matriz xml, y se realizó un blastmatrix.

### 6.6.4. Análisis del genoma núcleo y pangenoma de *V. parahaemolyticus*

El genoma núcleo consiste en familias de proteínas con representantes encontradas en todos los genomas investigados. El pangenoma es el conjunto de todas las familias de proteínas de todos los genomas en la comparación. El primer genoma en el análisis tiene un genoma núcleo igual al pangenoma. La adición de un segundo genoma reduce el genoma núcleo de los dos genomas y aumenta el pangenoma. Cada secuencia de un nuevo genoma se compara con un representante de cada una de las familias de genes existentes. Si la nueva secuencia coincide,

la familia es una familia núcleo, si la secuencia no coincide con una familia se convierte en una nueva familia de proteínas. Cuando las nuevas secuencias son comparadas con las familias de genes existentes, las familias núcleo que no tienen representación en el genoma, son removidas del genoma núcleo. El cambio en el pangenoma y el genoma núcleo se representan como dos líneas (azul y rojo, respectivamente). El número de nuevas proteínas, junto con cuantas nuevas proteínas corresponden a este, se indica también en el análisis. Una vez analizado esto se genera una gráfica y una tabla con el programa pancoreplot. Los genomas utilizados fueron descartados de la base de datos NCBI (Anexo B) y genomas de interés de *V. parahaemolyticus*.

### 6.7. Comparación de fenotipos

Con el programa Traitair versión 1.1.2 (Weimann *et al.*, 2016) se caracterizó el fenotipo *in silico*, utilizando secuencias nucleotídicas. Traitair analiza rasgos microbianos de manera automatizada para la predicción de 67 fenotipos que provienen de secuencias de genomas. Se utilizan datos de la sección de microbiología de la Red de Enfermedades, Infecciones y Epidemiología en línea (GIDEON). Traitair utiliza dos métodos diferentes para clasificar: 1) la clasificación phyplat, basado en la presencia y ausencia de etiquetas de proteínas y fenotipo de 234 bacterias; 2) la clasificación phyplat +PGL en donde se utilizan los mismos datos sobre las proteínas y además información sobre la evolución de las mismas, las pérdidas y ganancias del fenotipo. Para esto se utilizaron los archivos fasta generados de los ensamblajes, a estos se les extrajo los genes utilizando prodigal con extensión faa, con los que se creó un archivo de etiquetas, el cual fue utilizado posteriormente.

### 6.8. Identificación de factores de virulencia

Para la identificación de los factores de virulencia se utilizó la página web VFDB, por sus siglas en inglés Virulence Factor DataBase (<http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi>) (Liu *et al.*, 2019), en esta base se escogieron únicamente aquellas pertenecientes al género *Vibrio* (Anexo C), posteriormente se creó un archivo multifasta (un archivo txt con varias

secuencias nucleotídicas), para ser utilizado en el programa BRIG (Alikhan *et al.*, 2011) utilizando la opción blastn, el cual da como resultado archivos de salida en formato .tab con los resultados del alineamiento (porcentaje de identidad, cobertura, e-value, bitscore, etc.), con los parámetros del alineamiento se eligieron aquellas secuencias que tuviesen 90% de identidad y un alineamiento (cobertura) del  $\geq 90\%$ , con dichas secuencias se realizó un histograma de frecuencias. Para complementar la identificación de factores de virulencia, se utilizó los resultados de las abundancias de los subsistemas de la anotación con RASTtk, la cual se graficó en R como un mapa de calor utilizando ggplot.

**Tabla VI.** Genomas de referencia e interés de la especie de *V. parahaemolyticus*.

Cepa	Hospedero	No. de GenBank
<i>V. parahaemolyticus</i> 10329	<i>Homo sapiens</i>	NZ_AFBW01000024.1
<i>V. parahaemolyticus</i> 20130629002S01	<i>L. vannamei</i>	NZ_CP020034.1 <sup>a</sup> NZ_CP020035.1 <sup>b</sup>
<i>V. parahaemolyticus</i> FORC_023	<i>Homo sapiens</i>	NZ_CP012950.1 <sup>a</sup> NZ_CP012951.1 <sup>b</sup>
<i>V. parahaemolyticus</i> CHN25	<i>Homo sapiens</i>	NZ_CP010883.1 <sup>a</sup> NZ_CP010884.1 <sup>b</sup>
<i>V. parahaemolyticus</i> RIMD 2210633 O3:K6	<i>Homo sapiens</i>	NC_004603.1 <sup>a</sup> NC_004605.1 <sup>b</sup>
<i>V. parahaemolyticus</i> ATCC 17802	<i>Shirasu (Young sardine)</i>	NZ_LATW01000001.1
<i>V. parahaemolyticus</i> A1	<i>C. gigas</i>	-
<i>V. parahaemolyticus</i> N	<i>L. vannamei</i>	-
<i>V. parahaemolyticus</i> M	<i>L. vannamei</i>	-
<i>V. parahaemolyticus</i> 1A	<i>L. vannamei</i>	-
<i>V. parahaemolyticus</i> 7L	<i>C. gigas</i>	-
<i>V. parahaemolyticus</i> 6	<i>L. vannamei</i>	-

a = No. de acceso correspondiente a las secuencias del cromosoma I  
b = No. de acceso correspondiente a las secuencias del cromosoma II

## 7. RESULTADOS

### 7.1. Calidad y limpieza de las secuencias

En la Tabla VII se muestran los resultados que se obtuvieron de la limpieza de las lecturas con el programa *nextera\_cleaner*. El total de lecturas que se obtuvieron después de la secuenciación van de un rango de 1,460,000 a 1,720,000 pares de bases. Después de la limpieza, las secuencias disminuyeron en cantidad, siguiendo los parámetros de calidad de Phred (*P30*), que es equivalente a la probabilidad de que una base sea llamada incorrectamente 1 en 1000 veces.

En cuanto a la longitud de las lecturas de las secuencias, constan de un rango de 35 pb a 151 pb. Después de eliminar pares de bases mal colocados (15 pb al principio y 20 pb al final), las secuencias quedaron en un rango de 20 a 131 pb. Todos los genomas de *Vibrio* presentan un porcentaje de Guanina-Citosina (%GC) de 45%, a diferencia del genoma de *Photobacterium*, que presentó un 41% de contenido de GC.

### 7.2. Validación del ensamble con QUAST

Para el ensamble de las lecturas se utilizaron diversos programas (*A5*, *SPAdes* y *MIX*), (Tabla VIII). Para elegir el mejor ensamble a diversos parámetros estadísticos se utilizó QUAST. Para el valor estadístico de número de lecturas (N50) se utilizó un umbral de 100,000 y se tomó en consideración el N50 más alto, es la longitud de contigs tal, que usando contigs de igual o mayor tamaño recuperamos la mitad de las bases de ese ensamblaje. Asimismo, la longitud de las lecturas (L50) se eligió la más pequeña; este parámetro se considera como el número mínimo de secuencias contiguas que cubren la mitad del ensamblaje. Además se eligió el ensamblaje que tuviese el menor número de secuencias contiguas formadas, ya que lo ideal es que coincidan el número de replicones con el número de secuencias, y en el caso de la familia *Vibrionaceae* son dos, por el número de cromosomas, sin contar elementos extracromosómicos (plásmidos).



**Tabla VII.** Estadísticas básicas de la limpieza de las secuencias.

Características	<i>Pho</i> T01	<i>Vc</i> 02	<i>Vp</i> A1	<i>Vp</i> N	<i>Vp</i> 1A	<i>Vp</i> 6	<i>Vp</i> M	<i>Vp</i> 7L
Total de lecturas antes de la limpieza	1,503,564	1,715,023	1,544,923	1,693,539	1,905,854	1,462,008	1,562,115	1,719,966
Secuencias después de la limpieza	1,472,599	1,672,517	1,495,475	1,656,039	1,863,231	1,426,791	1,524,032	1,677,857
Rango de longitud de las secuencias antes de la limpieza	35-151	35-151	35-151	35-151	35-151	35-151	35-151	35-151
Lecturas de baja calidad	0	0	0	0	0	0	0	0
Rango de longitud de las secuencias después de la limpieza	20-131	20-131	20-131	20-131	20-131	20-131	20-131	20-131
%GC	41	45	45	45	45	45	45	45
Lecturas eliminadas	30,965	42,506	49,448	37,500	42,623	35,217	42,109	38 083

*Pho*= *Photobacterium damsela* subsp. *Damsela*, *Vc*= *Vibrio campbellii*, *Vp*= *Vibrio parahaemolyticus*

Para el caso de la longitud se decidió por el ensamble que se asemeja más a la longitud con el genoma de referencia, y esta característica también se tomó en consideración con el porcentaje de GC (Tabla VII). Se ha reportado que una buena cobertura es de 50X, ya que a partir de este no se muestran diferencias significativas en el número de secuencias contiguas formadas, el cual fue calculado con el programa `coverage_calculator.v0.0.1.sh`.

### 7.3. Identificación de secuencias de bacteriófagos

Para la identificación de secuencias de profagos se utilizó la página PHASTER (<https://phaster.ca/>). La Tabla IX muestra los posibles fagos por replicón (molécula cromosómica), indicando la posición del profago en la secuencia contigua, y la longitud de este en pares de bases. El criterio de que un bacteriófago esté completo, es la predicción de si la región contiene un profago intacto (>90), cuestionable (70-90) o incompleto (>70) basado en tres diferentes criterios: 1) si cierto número de bacteriófagos, comparado con la base de bacteriófagos de PHASTER, es mayor o igual al 100% del número total de CDS (coding sequence, por sus siglas en inglés) de la región, entonces la región se encuentra marcada con una puntuación de 150, y si es inferior a 100 se utilizan los otros dos métodos; 2). si el número de ciertos bacteriófagos en esta base de datos es más del 50% del número total de CDS de la región, este bacteriófago se considera como el principal bacteriófago potencial para esta región; 3) si alguna de las palabras claves relacionadas a bacteriófagos (cápside, cabeza, integrasa, cola, lisina, proteasa, transportasa, etc.) están presentes, la puntuación se incrementará en 10 para cada palabra clave encontrada, así, si el tamaño de la región es mayor que 30 Kbp, la puntuación incrementará un 10, y si hay por lo menos 40 proteínas identificadas en la región, la puntuación incrementará un 10. La cepa de *V. campbelli* 2 en su cromosoma II, no presenta evidencia de profagos. Asimismo, las cepas de *V. parahaemolyticus* A1, 7L y M no presentan profagos en el cromosoma I. El bacteriófago (s) con el mayor número de proteínas similares a los de la región, se le considera como el bacteriófago más común., y el %GC indica la cantidad de Guanina-Citosina en porcentaje de la región indicada del profago.

**Tabla VIII.** Parámetros estadísticos de la validación con QUASt de los ensamblajes con SPAdes, A5 y MIX.

Características	Pho T01	Vc 02	Vp A1	Vp N	Vp 1A	Vp 6	Vp 7L	Vp M
Ensamblador	SPAdes	Mix	SPAdes	A5	A5	SPAdes	A5	A5
N50	148,671	771,291	313,600	437,488	546,021	289,639	409,731	240,742
L50	11	3	6	5	3	7	4	6
Contig más largo	306,721	1,117,723	645,114	869,578	1,128,563	601,867	645,114	869,910
# de contig	75	30	62	38	30	77	39	69
Longitud total / genoma de referencia	4,451,386/ 4,485,055	5,891,719/ 6,195,978	5,406,785/ 5,165,770	5,418,937/ 5,165,770	5,086,640/ 5,165,770	5,345,442/ 5,165,770	5,425,990/ 5,165,770	5,283,624/ 5,165,770
% GC / % GC del genoma de referencia	40.50/ 40.59	45.45/ 45.06	45.15/ 45.37	45.15/ 45.37	45.16/ 45.37	45.34/ 45.37	45.20/ 45.37	45.35/ 45.37
N's/100 kbp	22.46	20.34	20.34	13.12	5	9.35	11.70	8.16
Cobertura	66.3 X	67.2X	58.7X	72.2X	55.2X	58.5X	64.04X	63.2X

Pho= *Photobacterium damsela* subsp. *damsela*, Vc= *Vibrio campbellii*, Vp= *Vibrio parahaemolyticus*

**Tabla IX.** Características de secuencias identificadas como presuntos profagos por PHASTER.

Cepas	Posición de la región	Criterio de completo	Puntuación	# Total de proteínas	Fago más común	% GC	Tamaño de la región (Kbp)
<i>PhoT01_chI</i>	73-6404	Incompleto	10	7	Enterobacter phi92 (NC_023693.1)	34.8	6.3
	27681-36473	Incompleto	20	9	Enterobacteria P4 (NC_001609.1)	38.2	8.7
<i>PhoT01_chII</i>	14102-50071	Cuestionable	70	35	Enterobacteria 933W (NC_000924.1)	38.8	35.9
<i>Vc 02_chI</i>				NRA			
<i>Vc 02_chII</i>	253514-278279	Incompleto	40	14	Erwinia ENT90 (NC_019932.1)	44.3	24.7
	180556-190619	Incompleto	20	12	Salmonella RE-2010 (NC_019488.1)	40.7	10
<i>V.p A1_chI</i>				NRA			
<i>Vp A1_chII</i>	487-11999	Intacto	107	17	Vibrio VGHphi (NC_004736.1)	42.9	11.5
	121578-141063	Incompleto	50	22	Gordonia GRU3 (NC_030935.1)	39.7	19.4
	209-41019	Intacto	126	56	Vibrio vB_VpaM_MAR (NC_019722.1)	50.9	40.8
	78-6820	Intacto	93	11	Vibrio VfO3K6 (NC_002362.1)	42.6	6.7

**Continuación Tabla IX**

<i>Vp</i> N_chI	72096-96056	Incompleto	40	6	Salisaeta icosahedral 1 (NC_017983.1)	43.1	23.9
	185857-199461	Intacto	128	22	Vibrio VGHphi (NC_004736.1)	43.1	13.6
<i>Vp</i> N_chII	95744-115254	Incompleto	50	22	Gordonia GRU3 (NC_030935.1)	40.3	19.5
	601-41411	Intacto	126	55	Vibrio vB_VpaM_MAR (NC_019722.1)	51.3	40.8
<i>Vp</i> 1A_chI	73656-83398	Intacto	105	20	Vibrio phage VCY-phi (NC_016162.1)	43.8	9.7
	188001-193056	Intacto	107	7	Vibrio phage VGJphi (NC_004736.1)	43.7	5
<i>Vp</i> 1A_chII	24916-44401	Incompleto	50	22	Gordonia phage GRU3 (NC_030935.1)	40.2	19.4
	334-36951	Intacto	150	48	Vibrio phage VP882 (NC_009016.1)	49.9	36.6
<i>Vp</i> 6_chI	18264-45961	Incompleto	30	7	Fago phi OH2 DNA (NC_021784.1)	45.1	27.6

Continuación **Tabla IX**

	569797-597112	Cuestionable	90	29	Mesorhizobium vB_MIoP_Lo5R7ANS (NC_025431.1)	49.3	27.3
<i>Vp</i> 6_chII	599751-717436	Intacto	100	112	Vibrio VFJ (NC_021562.1)	45.3	117.6
	26015-46090	Incompleto	50	32	Escherichia D108 (NC_013594.1)	47.1	20
	13055-21920	Incompleto	40	7	Enterobacteria Sf6 (NC_005344.1)	48.5	8.8
	16360-37788	Intacto	150	29	Vibrio vB_VpaM_MAR (NC_019722.1)	54.5	21.4
<i>Vp</i> 7L_chI				NRA			
	93147-112744	Incompleto	50	22	Rhizobium RR1-A (NC_021560.1)	40.2	19.5
	185945-199549	Intacto	128	22	Vibrio VGJphi (NC_004736.1)	43.1	13.6
<i>Vp</i> 7L_chII	210-36827	Intacto	150	48	Vibrio VP882 (NC_009016.1)	49.8	36.6
	516-41326	Intacto	126	55	Vibrio vB_VpaM_MAR (NC_019722.1)	51.3	40.8

Continuación **Tabla IX**

<i>Vp M_chI</i>				NRA			
	48029-68859	Incompleto	40	32	Bacillus BalMu-1 copia 1 (NC_030945.1)	47.2	20.8
<i>Vp M_chII</i>							
	59-14463	Incompleto	40	23	Rhizobium RR1-B (NC_021557.1)	46.8	14.4

NRA = No Resultado Aparente, chI= cromosoma I, chII= cromosoma II, % GC = porcentaje de Guaninca-Citosina, Kbp= 1,000 pares de base

#### 7.4. Anotación de los genomas

En la Tabla X se muestran los resultados de la anotación con RASTtk y PROKKA. Los genomas de *Vibrio* poseen un tamaño de alrededor de 5.2-5.9 Mb, mientras que la cepa de *Photobacterium* presenta un menor tamaño con 4.5 Mb. Se buscaron elementos genéticos móviles como transposones en la anotación con PROKKA, y elementos como CRISPR, siendo la cepa 6 la única en la que se encontraron elementos de CRISPR en su anotación. Los transposones se identificaron en todas las cepas excepto en la T01 y la 7L.

#### 7.5. Filogenia de la familia Vibrionaceae

##### 7.5.1 Promedio de la distancia genética (ANI)

La distancia genética en genomas anotados como completos de la bacteria de *Photobacterium damsela* subsp. *damsela* se muestra en la Figura 5. De los genomas utilizados, el genoma más cercano a nuestra cepa T01, es la cepa PhdpWu-1 con un 98.98%, aislada de una muestra de sangre de corazón del pez de agua salada *Scophthalmus maximu*, mientras que el genoma *Photobacterium damsela* subsp. *piscicida* 91-197 es la cepa más diferente genéticamente con un 97.09%.

La distancia genética de *V. campbellii*, de genomas anotados como completos se indica en la Figura 6. La bacteria con mayor similitud a la cepa *V. campbellii* 2 es la *V. campbellii* 20130629003S01 con un 97.78% de distancia genética, esta última cepa fue aislada del hospedero *Penaeus vannamei*. El genoma más distanciado genéticamente a la cepa 2 es la DS40M4 con un 96.53%, esta última cepa aislada de mar abierto.

Para las cepas de *V. parahaemolyticus* (Fig. 7) también se realizó un mapa de calor y un dendograma indicando las cepas seleccionadas como las más cercanas a nuestras cepas de interés (A1, N, 1A, 7L, 6 y M). Para nuestras cepas A1 y N su mayor similitud es con la cepa FORC\_023 aisladas del hospedero *Homo sapiens* con porcentajes de 98.54% y 98.53%



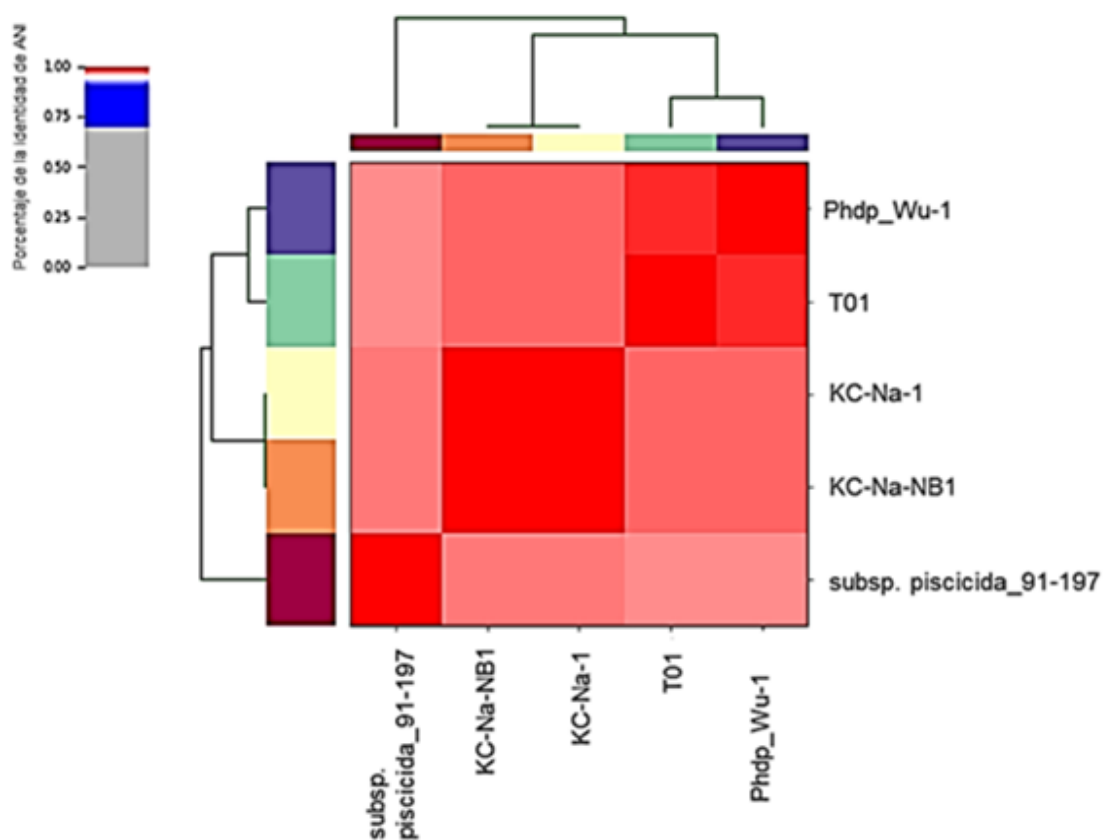
respectivamente; el genoma más distante de estas cepas, es el genoma O1:Kuk.str.FDA\_R31 (esta cepa es tdh/ tlh positiva). Para las cepas 1A y 7L la cepa con mayor similitud fue la 20130629002S01 aislada de *Penaeus vannamei* con porcentajes de similitud de 98.53% y 98.52% respectivamente, y los genomas más distanciados genéticamente de las cepas A1 y la 7L es la cepa O1: Kuk.str.FDA\_R31 (aislada de *C. gigas*) con 98.31% para ambas cepas. Para la cepa M la cepa CHN25 fue su correspondiente con un 98.51% aislada de *Penaeus vannamei*, de igual manera la cepa O1:Kuk.str.FDA\_R31 fue la cepa con menor similitud con el genoma de *V. parahaemolyticus* M. referente a la cepa 6 el porcentaje más alto corresponde con la cepa S107-1 con un 98.51% se desconoce la fuente de origen de esta última cepa, y la cepa más distanciada genéticamente de la cepa 6 fue la cepa O1:Kuk.str.FDA\_R31 con un 98.31%.

También en la Figura 7 se puede observar que las cepas A1, 7L, N y 1A, se agrupan y se encuentran distanciadas de las cepas M y 6 donde estas dos últimas fueron aisladas de *Penaeus vannamei*, al igual que las cepas N y 1A, mientras que la cepa A1 y 7L fueron aisladas de *C. gigas*.

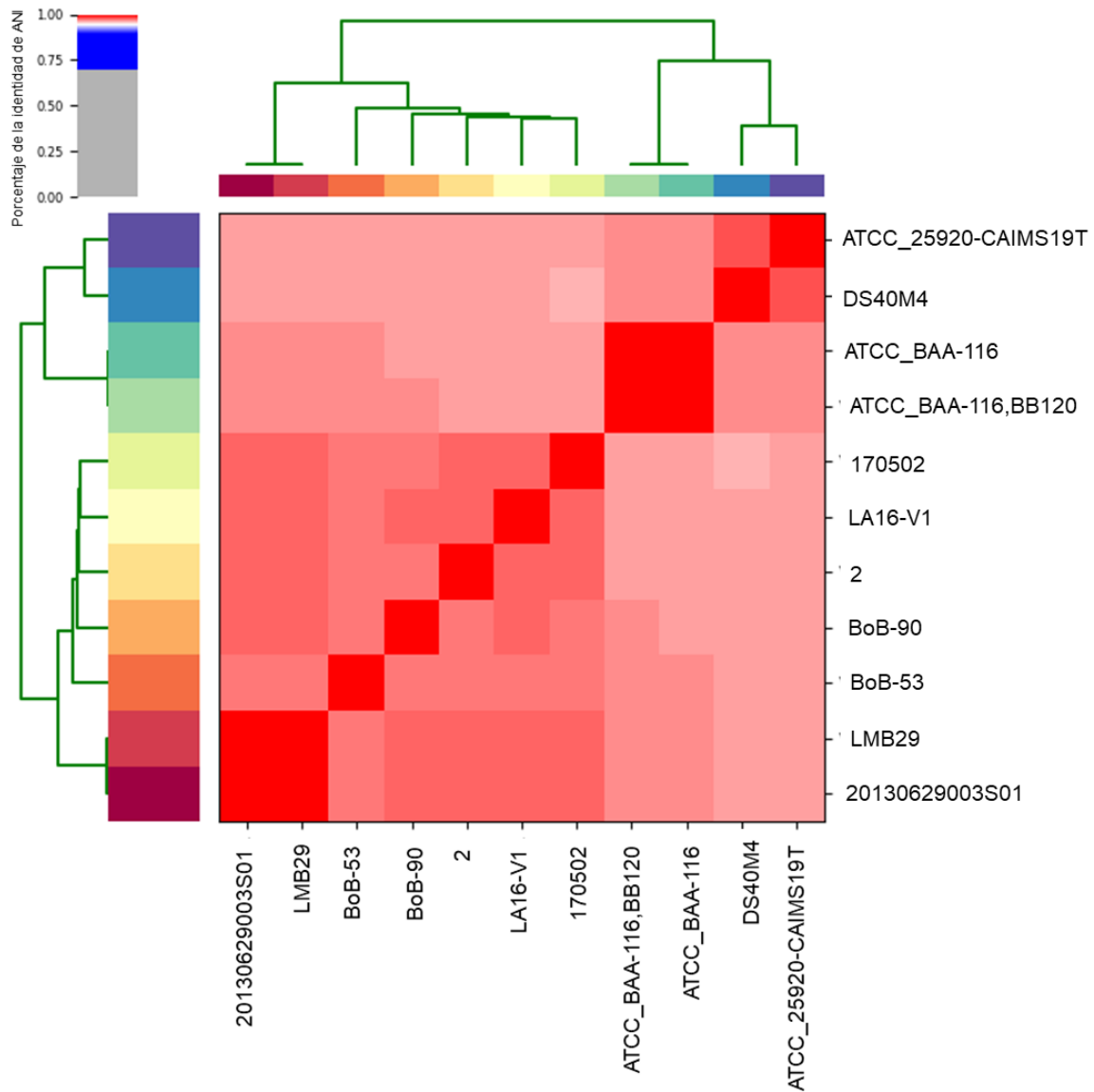
**Tabla X.** Resultados de la anotación con RASTtk y PROKKA.

Características	Pho T01	Vc 02	Vp A1	Vp N	Vp 1A	Vp 6	Vp 7L	Vp M
# de genes	4,011	5,318	5,044	5,019	5,004	4,969	5,079	5,054
# de proteínas	3,882	5,199	4,922	4,900	4,888	4,851	4,591	4,933
Tamaño (Mb)	4.5	5.89	5.43	5.39	5.36	5.28	5.43	5.42
%GC	40.57	45.4	45.15	45.16	45.34	45.35	45.2	45.15
# tARN	123	105	116	110	107	110	117	112
# CRISPR	0	0	0	0	0	6	0	0
# transposones	0	5	7	7	7	4	0	11

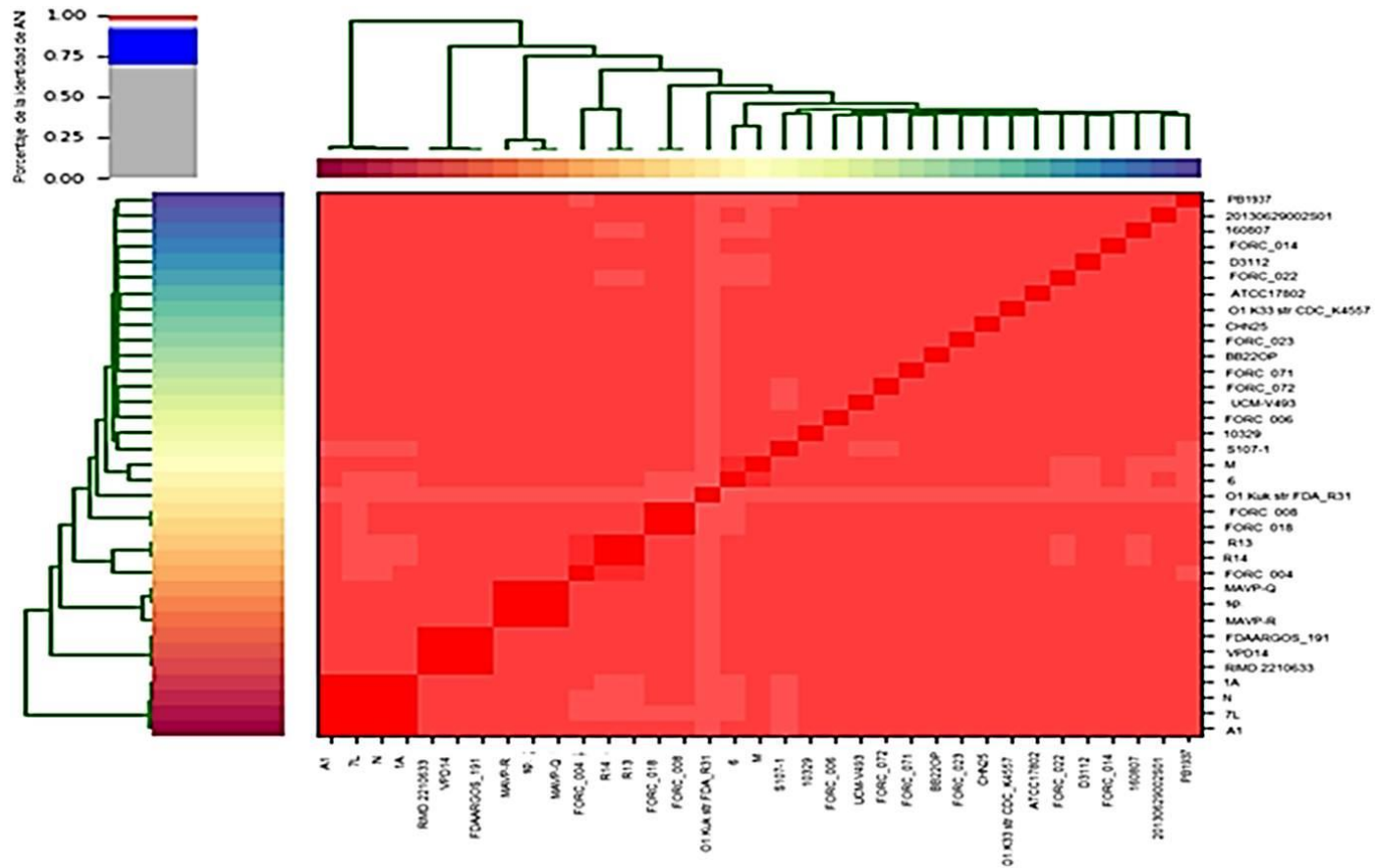
Pho= *Photobacterium damsela* subsp. *Damsela*, Vc= *Vibrio cabpbellii*, Vp= *Vibrio parahaemolyticus*, tARN= ARN de transferencia, CRISPR= Clustered Regularly Interspaced Short Palindromic Repeats



**Figura 5.** Distancia genética y promedio de la identidad nucleotídica de *Photobacterium damsela subsp. damsela*. Se utilizaron cuatro genomas de *P. damsela subsp. damsela* de la base de datos NCBI y la cepa *P. damsela subsp. damsela* T01 aislada de B.C.S. Los dígitos dentro de la matriz indican el promedio de la identidad nucleotídica de los genomas comparados, siendo 1.00 idéntico y 0.00 diferente. Se utilizó la opción ANIm, donde utiliza MUMmer (NUCmer).



**Figura 6.** Distancia genética y promedio de la identidad nucleotídica de *Vibrio campbellii*. Se utilizaron 10 genomas de la bacteria *Vibrio campbellii* de NCBI y la cepa *V. campbellii* 2 aislada de B.C.S. Los dígitos dentro de la matriz indican el promedio de la identidad nucleotídica de los genomas comparados, siendo 1.00 idéntico y 00 diferente. Se utilizó la opción ANIm, donde se usa MUMmer (NUCmer).



**Figura 7.** Promedio de la identidad nucleotídica y distancia genética de *Vibrio parahaemolyticus*. Se utilizaron 30 genomas de *V. parahaemolyticus* y las seis cepas A1, 7L, N, 1A, 6 y M aisladas de B.C.S. Los dígitos dentro de la matriz indican el promedio de la identidad nucleotídica de los genomas comparados, siendo 1.00 idéntico y 00 diferente. Se utilizó la opción ANIm, donde usa MUMmer (NUCmer).

### 7.5.2 Tipificación multilocus de secuencias

La Figura 8 muestra una tabla y un árbol filogenético realizado con siete genes constitutivos (*dnaE*, *gyrB*, *recA*, *dtdS*, *pntA*, *pyrC*, *tnaA*) que se encuentran en todas las bacterias de *Vibrio sp.* Se evaluaron las variaciones en nucleótidos de estas siete secuencias constitutivas las cuales se representan por una numeración de alelos en orden de descubrimiento, y cada uno de los alelos se combinan para formar perfiles alélicos únicos asignados a los tipos de secuencias denominadas ST (sequence type). Cuando se encuentran aislados estrechamente relacionados tienen una secuencia tipo (ST) idéntica o con locis de diferencia muy cercanos, se toma de cuatro locis idénticos (o más) para considerar las cepas dentro de un mismo complejo clonal.

Como se puede observar en la Figura 8, las cepas de *V. parahaemolyticus* N, 7L, 1A y A1 se agrupan en la misma secuencia tipo 1257, por lo que se puede considerar estas cepas dentro del mismo complejo clonal (CC), sin embargo para el ST1257 no hay reportes de este complejo, solo se encuentra un reporte con el mismo ST (1257) de una bacteria aislada en Venecia, Italia (IT\_CR\_VP34), de una muestra ambiental, la cual es negativa a los genes *trh/tdh*. Asimismo, la cepa M obtuvo una secuencia tipo 1167, perteneciente a CC114 el cual da resultados con otros dos reportes dentro de la base PubMLST, de los aislados FIM-S1708+ y FIM-S1708 de origen ambiental, aisladas en Hermosillo, Sonora en México. La cepa 6 no se identificó una secuencia tipo, sin embargo, por medio de su perfil alélico se pudo obtener su secuencia tipo en la base de datos PubMLST para *V. parahaemolyticus*, dando como resultado una ST de 351, que indica que no pertenece a ningún complejo clonal, pero se encuentra una cepa aislada con el mismo ST, la cual fue obtenida en China de una fuente ambiental.

### 7.5.3. Filogenia del gen 16S ARNr

En la Figura 9 se muestra un árbol filogenético con el gen de la subunidad pequeña ribosomal (16S ARNr) de diversas bacterias de la especie *Vibrio parahaemolyticus* incluyendo las de este trabajo (6, 7L, A1, N, M, 1A). Las cepas 6, 7L, A1 y N se agrupan en un mismo clado, teniendo un ancestro en común con las cepas Isolate-R13 obtenida del huésped *Penaeus*

vannamei y PB1937 también de camarón. La cepa M se agrupa con las cepas BB22OP de fuente desconocida y la bacteria FORC\_018 aislada de un pez de agua salada *Dicentrarchus labrax*. La cepa 1A se encuentra agrupada sola en un solo clado.

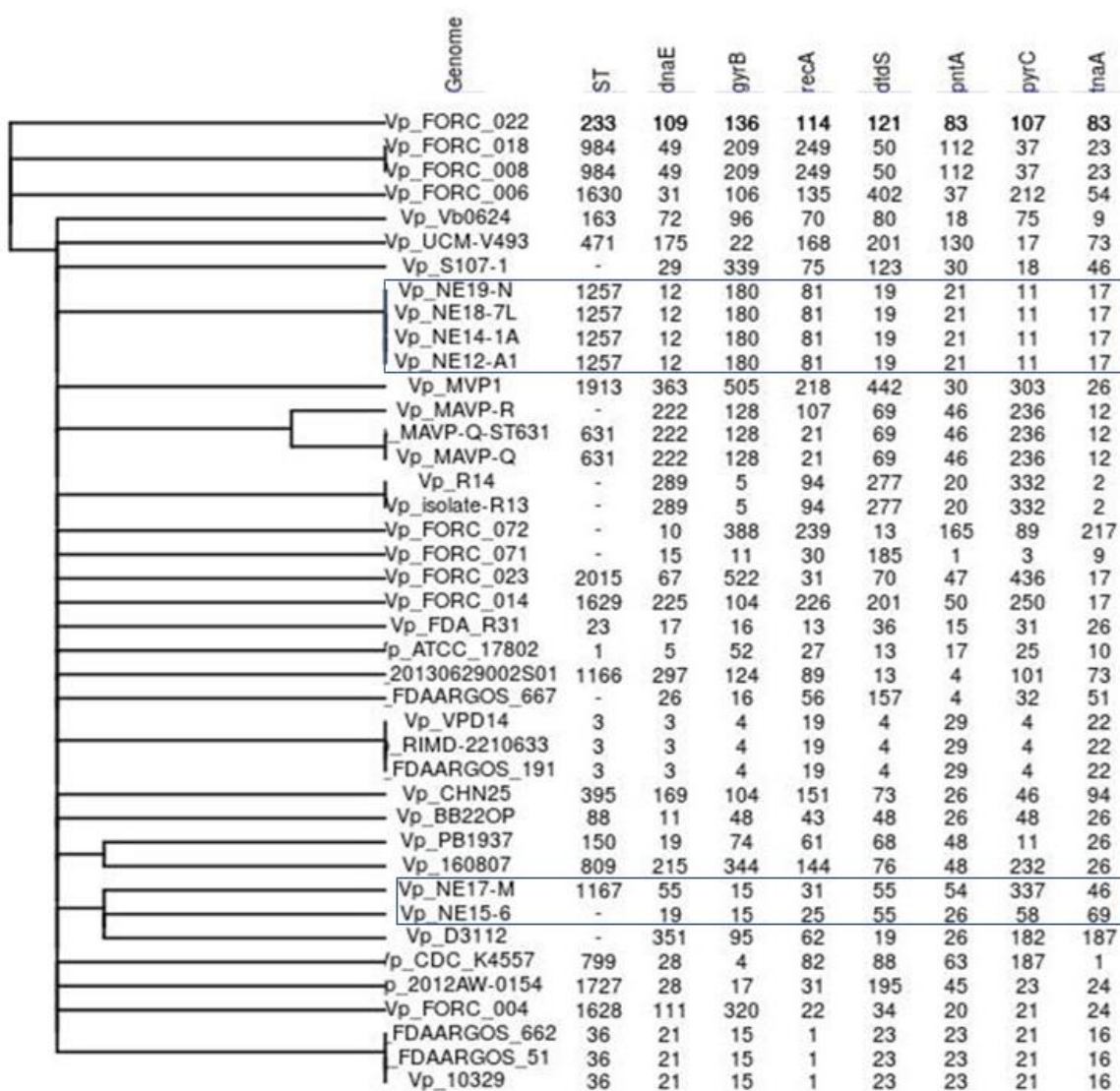
Como grupo externo se utilizó el genoma de la cepa de *Streptomyces* sp. S1A1\_8, el cual se agrupa junto con la cepa de *V. parahaemolyticus* FORC\_008, estas dos cepas se ven distanciadas de las demás bacterias por la longitud de las líneas de esta agrupación.

## 7.6. Genómica comparativa de *V. parahaemolyticus*

### 7.6.1. Atlas estructural del ADN de los cromosomas

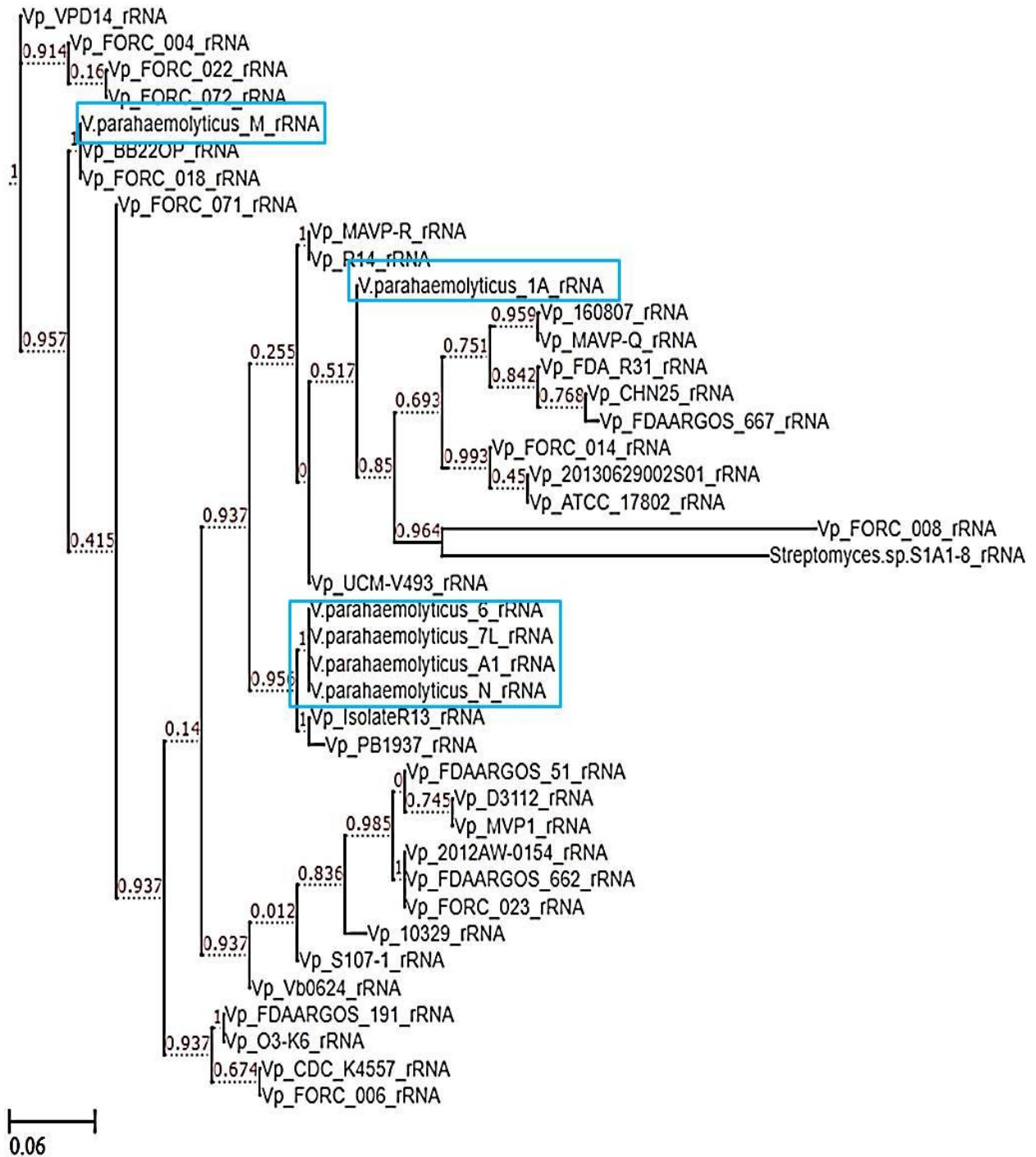
A partir de un sistema automático para generar atlas estructurales de genomas, con la información de los ficheros de secuencias del GenBank, la Figura 10 representa en forma de un mapa circular, de la representación gráfica de la estructura del ADN en los cromosomas I y II de *V. parahaemolyticus*, en el cual se utilizó la cepa RIMD 2210633 como genoma de referencia. El atlas muestra tres tipos de información: composición base (contenido AT, asimetría GC), repeticiones globales dentro del replicon (directo e invertido) y las propiedades estructurales del ADN (preferencia de posición, energía de apilamiento de ADN, y curvatura). El cromosoma I presenta un tamaño de 3, 288,558 pb, mientras que el cromosoma II es más pequeño con un tamaño de 1, 877,212 pb.

En el gen de referencia se observan áreas con elevada curvatura, energía de apilamiento de ADN y preferencia de posición que no están presentes en las cepas de este estudio, como lo indican las flechas rojas oscuro en el atlas. Las flechas naranjas indican los sitios de origen y termino de la replicación, las cuales están colocadas de acuerdo a los cambios en mayor cantidad de G's (azul) hacia más cantidad de C's (rosa). En ambos cromosomas también se observan diversas regiones asociadas con secuencias de ARNr y ARNt. Asimismo las flechas azules indican la localización de los operones del ARNr, donde se observa que el cromosoma I tiene varios operones mientras que en el cromosoma II solo se encuentra un operón. Cabe resaltar que en el cromosoma I la cepa A1

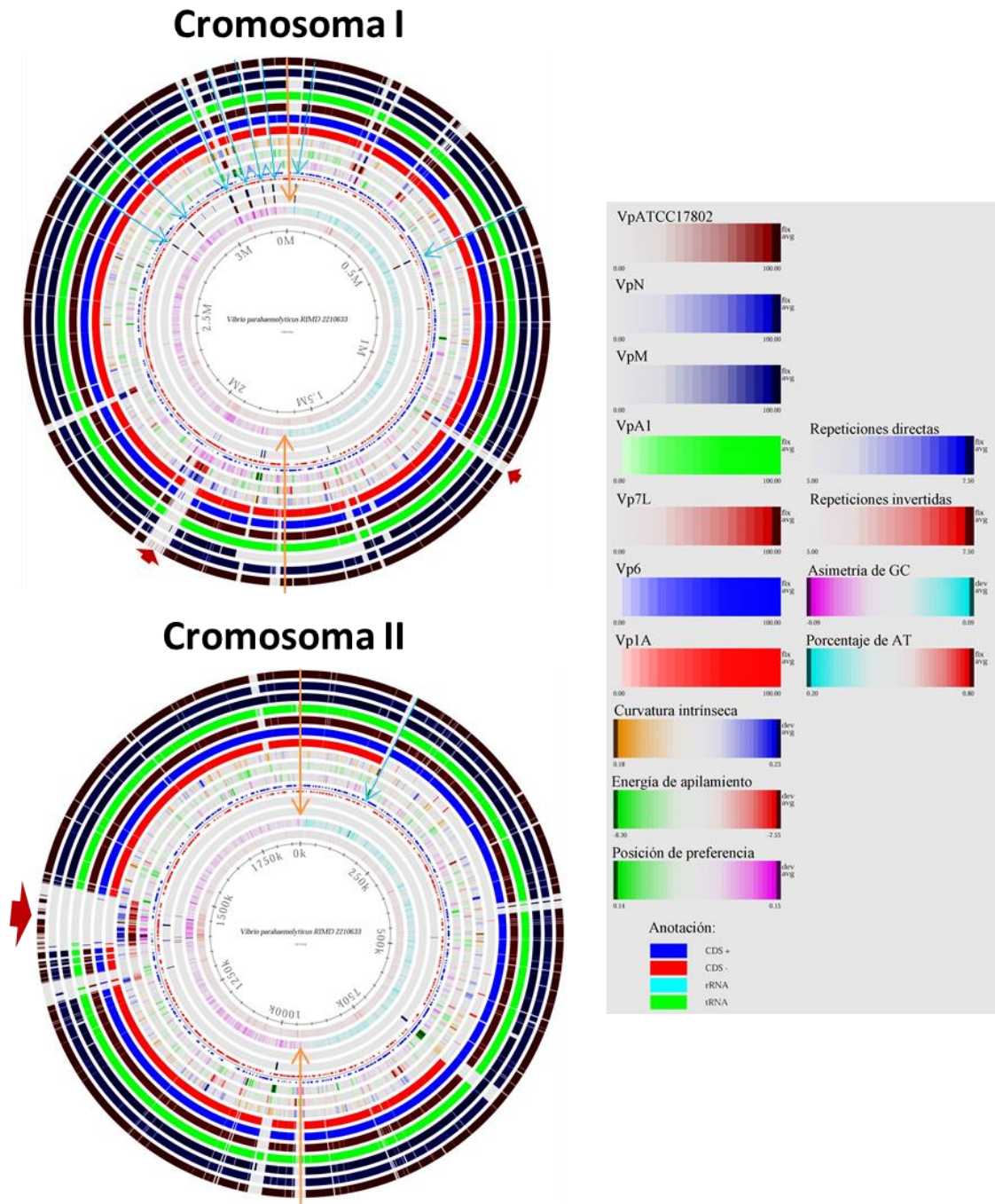


**Figura 8.** Tipificación multilocus de secuencias (MLST) de cepas de *Vibrio parahaemolyticus*. Árbol y tabla de genes constitutivos de *V. parahaemolyticus*. Se utilizaron siete genes: dnaE, la subunidad catalítica de la ADN polimerasa III; gyrB, la ADN girasa subunidad B; recA, recombinasa A; dtdS, deshidrogenasa D-treonina; pntA, deshidrogenasa NAD (P) subunidad alfa; pyrC, Dihidrorotasa y; tnaA, Triptofanasa. Para genomas completos reportados en la base de datos de Genome de NCBI y los genomas de interés aislados de B.C.S.





**Figura 9.** Árbol filogenético de *Vibrio parahaemolyticus* con el gen ribosomal de la subunidad pequeña 16s de genomas completos de la base de datos de NCBI. Se utilizó GBlocks para la eliminación de sitios poco informativos y divergentes, MAFFT para el alineamiento y Fasttree para la elaboración del árbol, utilizando el método Jukes-Cantor Joins con repeticiones de 1000. Se utilizó la secuencia de *Streptomyces* sp. como grupo externo.



**Figura 10.** Atlas estructural de ADN de los cromosomas I y II de *Vibrio parahaemolyticus*. Se utilizó la cepa RIMD 2210633 como genoma de referencia. La imagen circular muestra el siguiente orden desde el círculo interno: 1) tamaño del cromosoma (cromosoma I = 3,288,558 pb y cromosoma II = 1,877,212 pb) , 2) porcentaje de AT (rojo=alto contenido AT), 3) asimetría de GC (azul=mayor cantidad de G's), 4) repeticiones invertidas y directas (color=repetición), 5) posición de preferencia, 6) energía de apilamiento, 7) curvatura intrínseca, 8-13) secuencias de las cepas de estudio (1A, 6, 7L, A1, M y N) y 14) cepa no patogénica (ATCC 17802). Flechas naranjas indican cambios en la asimetría de G y C, que muestran el origen y el término de la replicación. Las flechas azules indican la localización de los operones de ARNr. Las flechas de color rojo oscuro resaltan las áreas del cromosoma que muestran significativamente diferentes estructuras de ADN que el resto del genoma.

### 7.6.2. Uso de codones y aminoácidos

Se analizó el uso de codones y la composición de aminoácidos de las 6 cepas de *V. parahaemolyticus* de este estudio provenientes de B.C.S., y de otras 35 cepas de referencia de la misma especie, con los genomas disponibles en el GenBank. Las Figuras 11 y 12 muestran un mapa de calor del uso de codones y aminoácidos, respectivamente. Todas las 41 cepas presentan un uso de codones y composición de aminoácidos similar. En la Figura 11 se muestra que las cepas N, 1A, A1, 7L y 6 se agrupan en el mismo grupo (I), mientras la cepa M se agrupa en otro grupo (II). Sin embargo en la composición de aminoácidos la cepa 6 se encuentra en el grupo I, la cepa M en el grupo II, y las cepas N, 1A, A1 y 7L en el grupo III (Fig. 12). En la Figura 12 las cepas N, 1A, A1 y 7L se agrupan con la cepa CHN25 (aislada de camarón). La cepa M se agrupa con los genomas FORC\_071 (fuente desconocida), FDA\_R31 (obtenida de ostiones), R14 (aislada de *Penaeus vannamei*) y BB22OP (fuente desconocida). La cepa 6 de este trabajo se agrupa con las cepas PB1937 (camarón) y FORC\_004 (obtenida de agua de un acuario). Las agrupaciones de uso de codones y aminoácidos coinciden con la agrupación con el MLST, donde las cepas N, 7L, 1A y A1 se agrupan por similitud, mientras las cepas 6 y M se muestran más distanciadas; estos análisis no coinciden con el análisis del 16s.

### 7.6.3. Comparación de proteomas

Las Figuras 13 y 14 muestran una matriz comparativa de proteomas, donde la parte principal de la matriz consiste en comparaciones de genoma por pares; con fracciones de proteínas compartidas sombreado en verde (más verde, más familias de proteínas compartidas). El sombreado en rojo indica comparaciones dentro del mismo proteoma que indica homologías internas (paralogos internos), es decir, proteínas con un alto grado de identidad pero que ambas se encuentran dentro del proteoma de una especie. La Figura 13 muestra una matriz comparativa de proteomas de 35 genomas de *V. parahaemolyticus* de la base de datos del GenBank y los 6 genomas aislados de B.C.S., México. Se observa que las 41 cepas analizadas tienen un alto porcentaje de genomas conservados (67.7-99.6%), donde las intensidades de

color se basan en el porcentaje relativo, cuanto más intenso sea el color verde, mayor es el número de familias de genes compartidos. Para la cepa 6 el mayor porcentaje fue de 78.7% con la bacteria FORC\_004; la cepa A1 con un 77.9% con la CDC\_K4557; la cepa M con la cepa FORC\_004 con 79.9%; el genoma 7L obtuvo dos porcentajes con FORC\_008 y FORC\_018, 86.7 y 86.8%, respectivamente; la cepa CDC\_K4557 obtuvo el mayor porcentaje con las cepas 1A y N con 78.3% y 77.9% respectivamente. La Figura 14 muestra una matriz comparativa donde se muestran los genomas evaluados en este estudio con 6 representantes de *V. parahaemolyticus* (M, N, 6, 7L, A1 y 1A), y las cepas de *V. campellii* (02) y *Photobacterium damsela* subsp. *damsela* (T01). Entre las cepas de *V. parahaemolyticus* se observa entre el 73.7-99.3% de homología, y la semejanza de estas cepas con la otra cepa de *V. campbellii* es de 58.3-59%. Cuando se comparan las cepas de *Vibrio* spp. con *P. damsela* subsp. *damsela* estas solo tienen un 29.8-30.6% de semejanza. Asimismo, Las Figuras 13 y 14 representan la homología interna (cuadros rojos) con un valor de 2.4-4.7% en las 41 cepas de *V. parahaemolyticus* analizadas (Fig. 13), y entre 2.5-5.6% dentro de los genomas evaluados en este estudio (Fig. 14).

#### 7.6.4. Análisis del genoma núcleo y pangenoma

El análisis del pangenoma y genoma núcleo fue otra forma de comparar proteomas. La Figura 15 muestra el conjunto de familias de proteínas de todos los genomas de las 41 cepas de *V. parahaemolyticus* (pangenoma) con 11,104 familias. También se muestra el genoma núcleo con 3,143 familias de proteínas que se comparten en todos los genomas.

Para obtener esta información las secuencias se comparan utilizando BLAST y el límite es del 50%/50% (50% de identidad entre las secuencias, y que se presente un alineamiento de al menos el 50% de la secuencia del gen). Para el primer genoma, el conjunto de genes tanto en el pangenoma como en el núcleo son idénticos, el núcleo se hace más pequeño con la adición de un segundo genoma, ya que los genomas en este conjunto ahora necesitan ser encontrados en ambos genomas, en cambio el pangenoma con cada genoma agregado va en aumento, por lo que se considera que es un pangenoma “abierto”.

### 7.7. Comparación de Fenotipos de la familia Vibrionaceae

La Figura 16 representa las predicciones basadas en el ensamblaje conjunto de los genomas de este estudio, para analizar el fenotipo de cada cepa. Se utilizaron dos métodos para analizar los rasgos microbianos: 1) la clasificación phypat, basada en la presencia y ausencia de etiquetas de proteínas y fenotipos de bacterias; y 2) la clasificación phypat +PGL en donde se utilizan los mismos datos sobre las proteínas, así como, información sobre la evolución de las mismas, las pérdidas y ganancias del fenotipo.

Se identifican dos grandes grupos de fenotipos similares como lo indican los colores del dendograma (Fig. 16). Se observó que las 8 cepas estudiadas comparten diversos procesos metabólicos (azul intenso), indicando que los dos métodos de predicción dieron positivos para diversas características. Sin embargo se presentan diferencias fenotípicas en las cepas de *P. damsela* subsp. *Damsela* y *V. campbellii* comparadas con las de *V. parahaemolyticus*. Características como celobiosa, hidrólisis de gelatina, crecimiento en 6.5% de NaCl y L-Arabinosa dieron negativas para *P. damsela* subsp. *damsela* T01, mientras que para *V. campbellii* 02 la L-Arabinosa y el crecimiento en cianuro de potasio también dieron negativo. Asimismo, *V. campbellii* dio positivo para susceptibilidad a polimixina-colistina, hidrólisis de urea, utilización de acetato y myo-inositol, mientras que *P. damsela* subsp. *damsela* dio positivo para susceptibilidad a polimixina-colistina, hidrólisis de urea, utilización de acetato, acetoina (Voges proskauer), gas de glucosa y salicina.

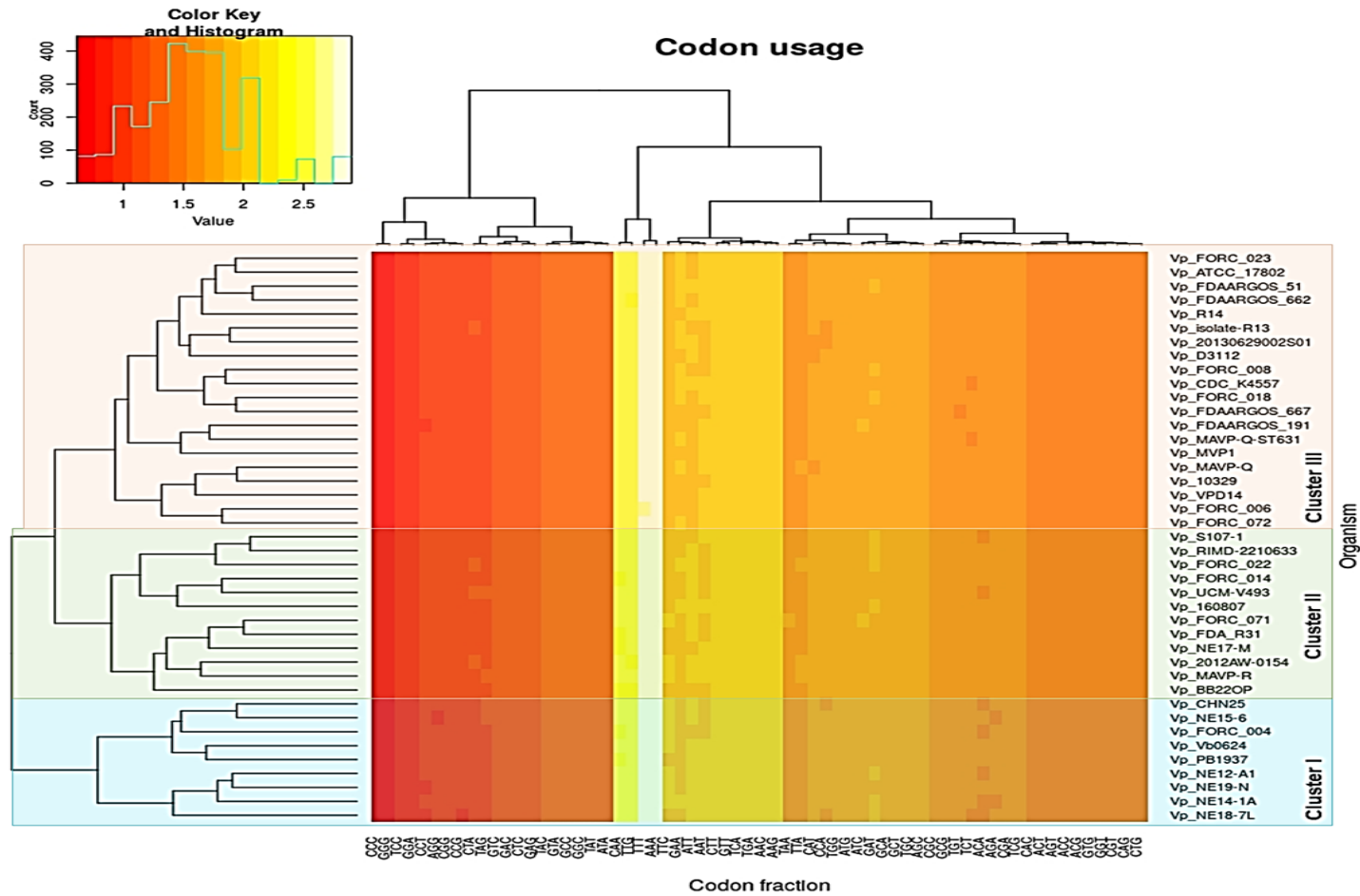
También se observa que las cepas 6 y M de *V. parahaemolyticus*, fueron positivas a la susceptibilidad a polimixina-colistina, a diferencia de las demás cepas de la misma especie. Cabe recalcar que todas las cepas presentaron la característica de D-manisitol.

### 7.8. Identificación de factores de virulencia

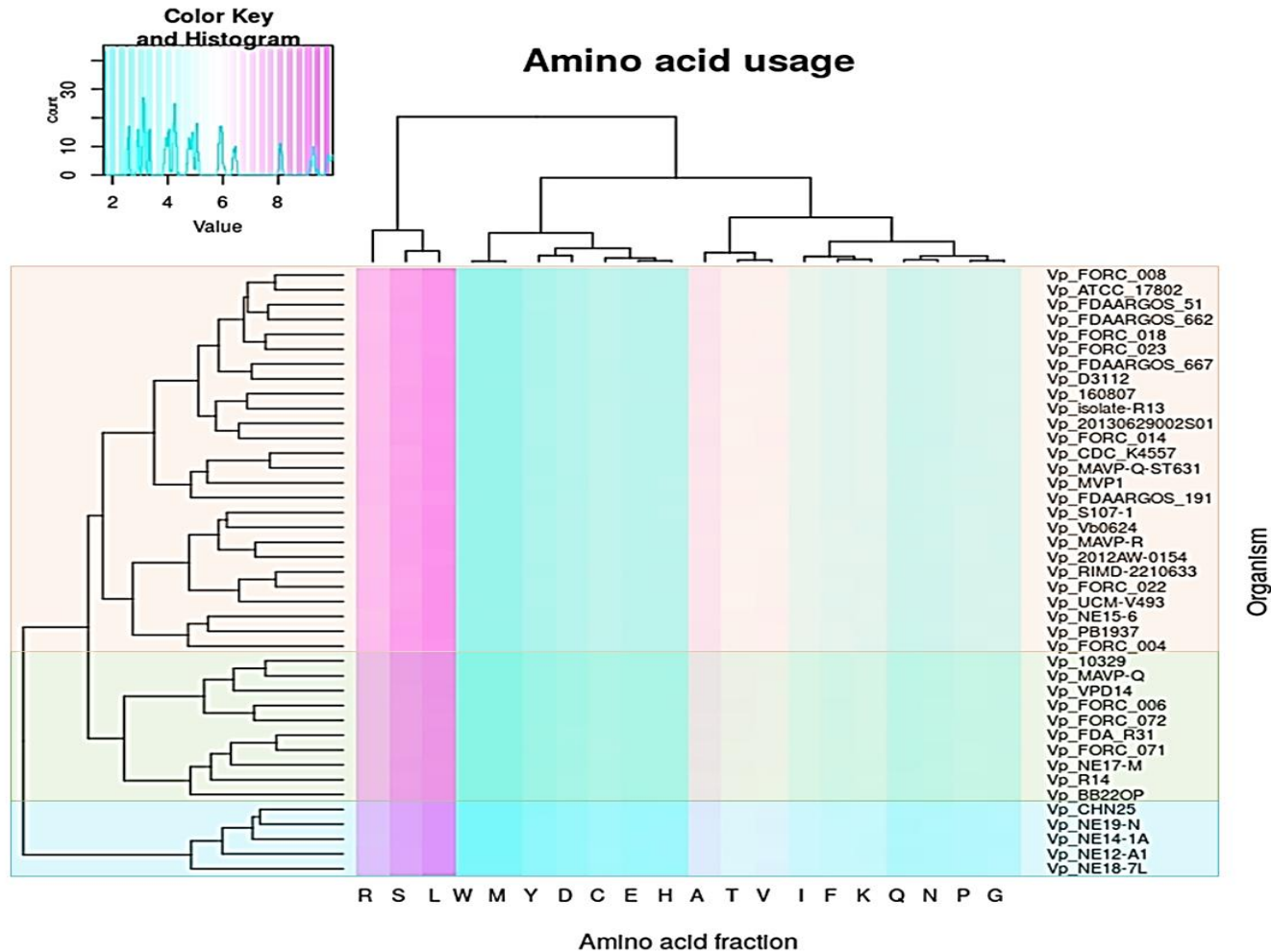
En la Figura 17 se muestra un mapa de calor general de categorías relacionadas con la patogenicidad en bacterias de la familia Vibrionaceae de este estudio con la anotación de

RASTtk. La categoría con mayor abundancia es la de pared celular y capsula, siendo las cepas con mayor cantidad *V. campbellii* 2, seguidas de las cepas de *V. parahaemolyticus* 6, 7L, N, 1A, M, 6 y *P. damselae subsp. damselae* T01. La segunda categoría con mayor abundancia es la de virulencia, enfermedad y defensa donde la cepa 2 de *V. campbellii* es la más abundante, seguida por las cepas de *V. parahaemolyticus* 6, M, 1A, N y 7L. La tercera categoría más abundante es la de respuesta a estrés, y la cuarta es la de motilidad y quimiotaxis en orden de abundancia se encuentran las cepas de *V. campbellii* 2, *V. parahaemolyticus* 7L, N y A1, M y 6, *P. damselae subsp. damselae* T01 y *V. parahaemolyticus* A1. La categoría menos abundante es la de elementos genéticos móviles (plásmidos, profagos y transposones). Se puede observar también que la cepa de *V. parahaemolyticus* A1 se agrupa con menor abundancia de todas las categorías en comparación de las otras cepas de *V. parahaemolyticus*.

La Figura 18 muestra un histograma de las abundancias en secuencias de genes de las cepas aisladas de B.C.S., que dieron 90% de identidad con secuencias para factores de virulencia que se han reportado en la base de datos VFDB (<http://www.mgc.ac.cn/cgi-bin/VFs/compvfs.cgi?Genus=Vibrio>) para el género *Vibrio sp.*. Las secuencias utilizadas se obtuvieron de realizar un alineamiento de nucleótidos (blastn) con el programa de BRIG, de los diversos factores de virulencia reportados en esta base de datos (adherencia, antifagocitosis, quimiotaxis y motilidad, enzimas, sideróforos, quorum sensing, sistemas de secreción, toxinas). Las secuencias génicas referentes a adherencia las más representativas fueron las de pilus tipo IV MSHA (Hemaglutina sensible a manosa) y pilus tipo IV. Para quimiotaxis los genes (che, fil, fla, flg, flh, fli, flr, mot) encontrados fueron los que codifican para el sistema flagelar. En quorum sensing todas las cepas contenían los genes LuxS y CqsA, a excepción de la cepa A1 en donde solo se encontró el gen LuxS. Para los sistemas de secreción, se encontraron genes que codifican para el sistema de secreción eps tipo II y T3SS1 (sistema de secreción 1 tipo 3), y en toxinas todas las cepas presentaron la presencia de la toxina tlh a excepción de la cepa 1A, la cual presentó la presencia de la toxina vvha, una hemolisina/ citolisina.

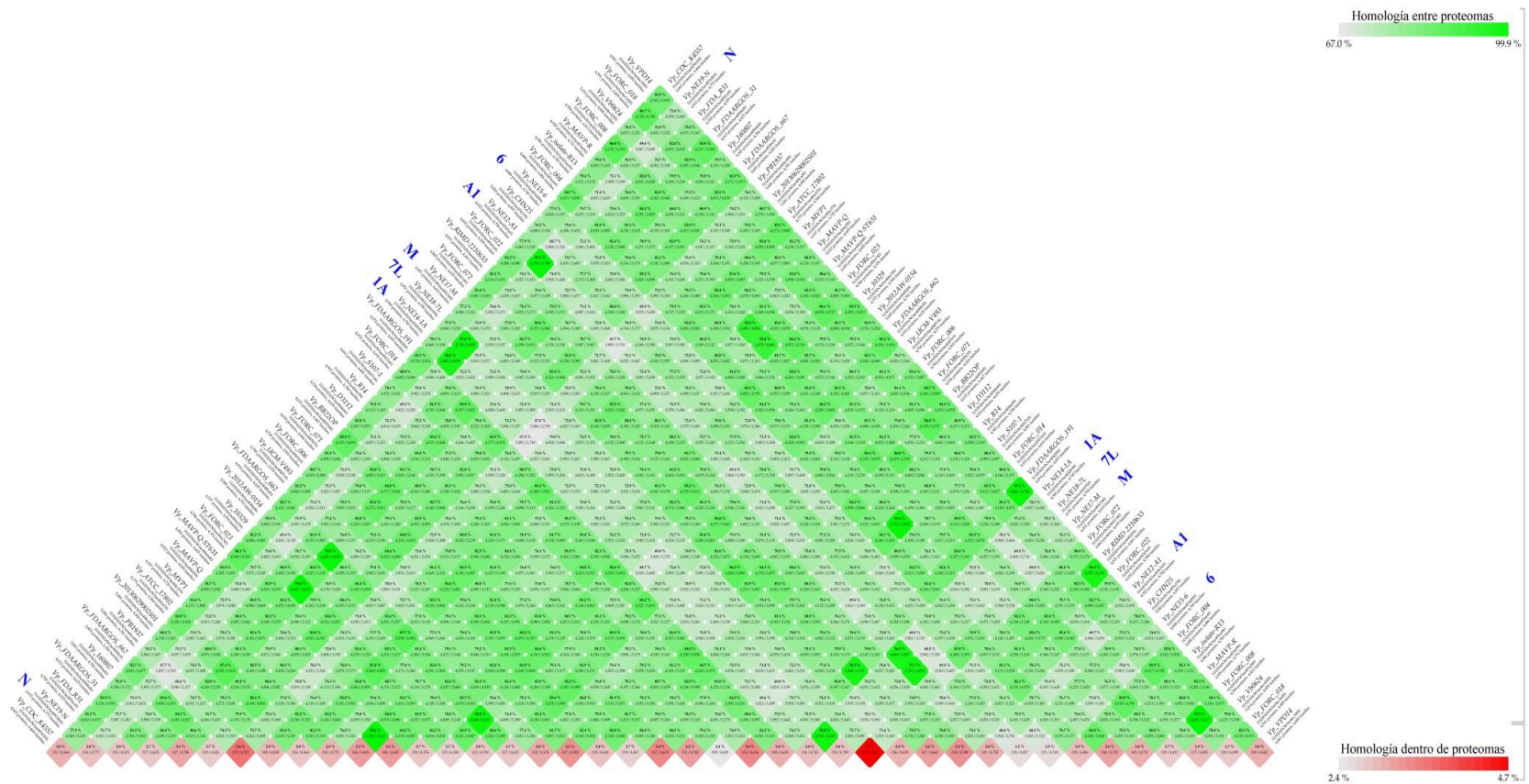


**Figura 11.** Mapa de calor de la preferencia del uso de codones. Se utilizaron 41 genomas de *Vibrio parahaemolyticus*, incluyendo los genomas de B.C.S, México (M, N, 6, 7L, A1 y 1A). Se utilizó CMG-biotools para la visualización, y para la creación del mapa se utilizó el programa R. Se muestra un árbol con la agrupación del uso del triplete de codones y un histograma con la abundancia de los mismos.

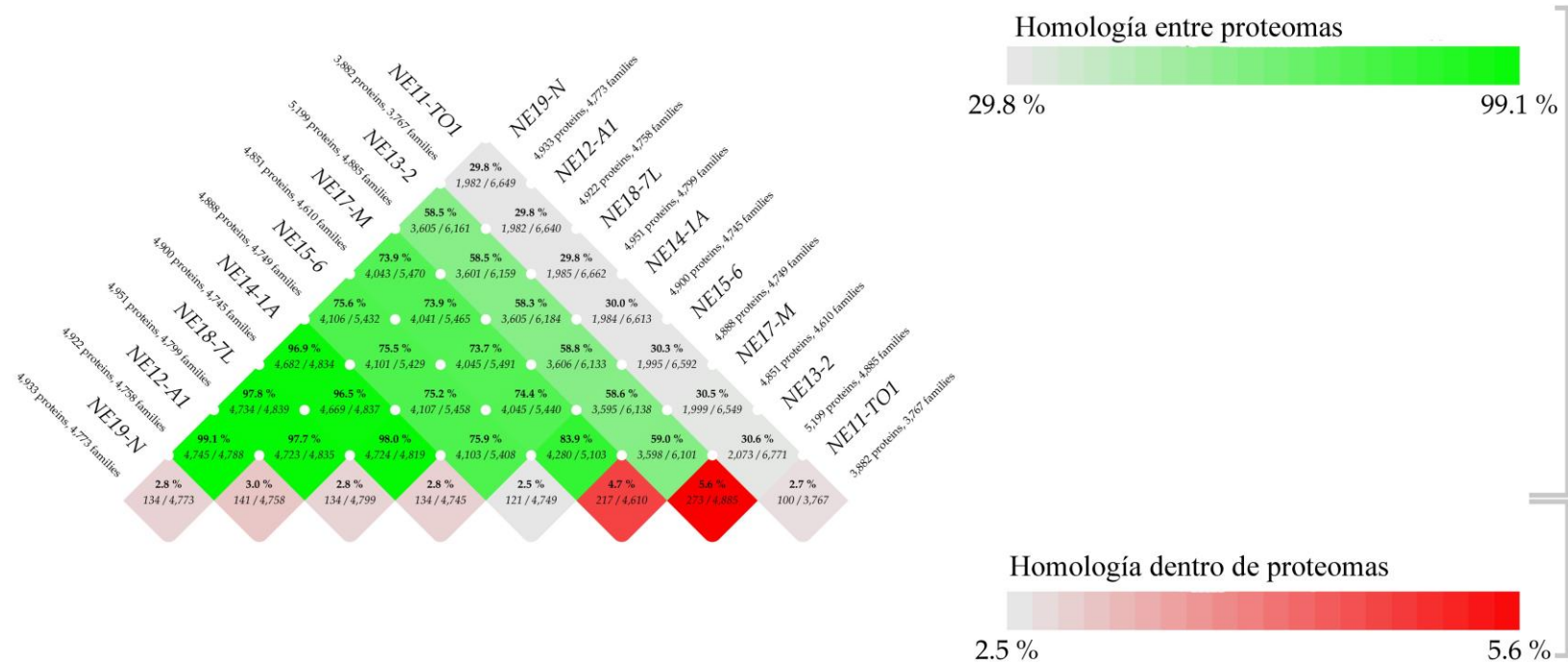


**Figura 12.** Mapa de calor de la preferencia de uso de aminoácidos. Se utilizaron 35 genomas completos de *Vibrio parahaemolyticus* y los genomas de *V. parahaemolyticus* de B.C.S (M, N, 6, 7L, A1 y 1A). Se utilizó CMG-biotools para la visualización, y para la creación del mapa se utilizó el programa R. Se muestra un árbol con la agrupación del uso de aminoácidos y un histograma con la abundancia de los mismos.

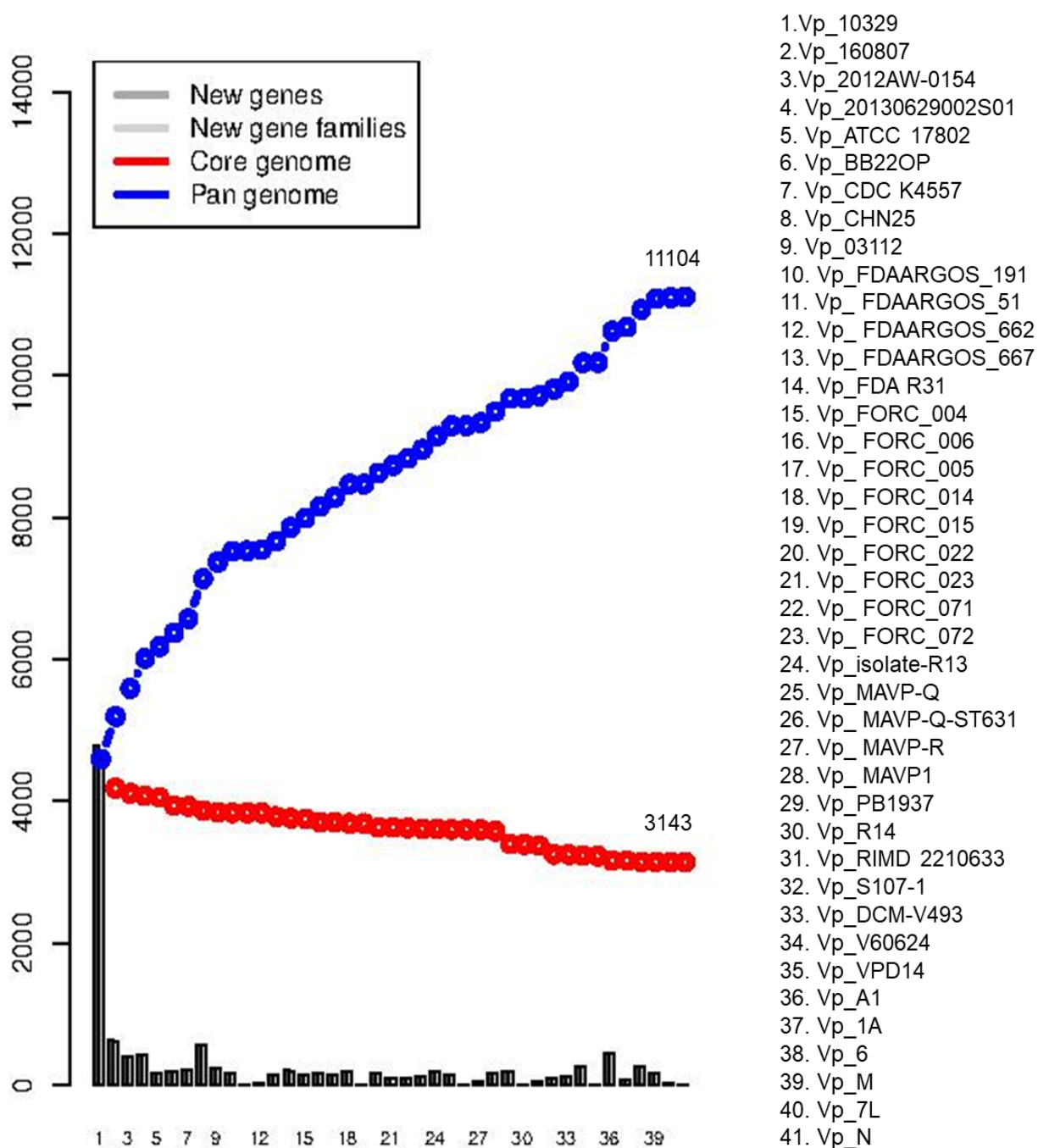




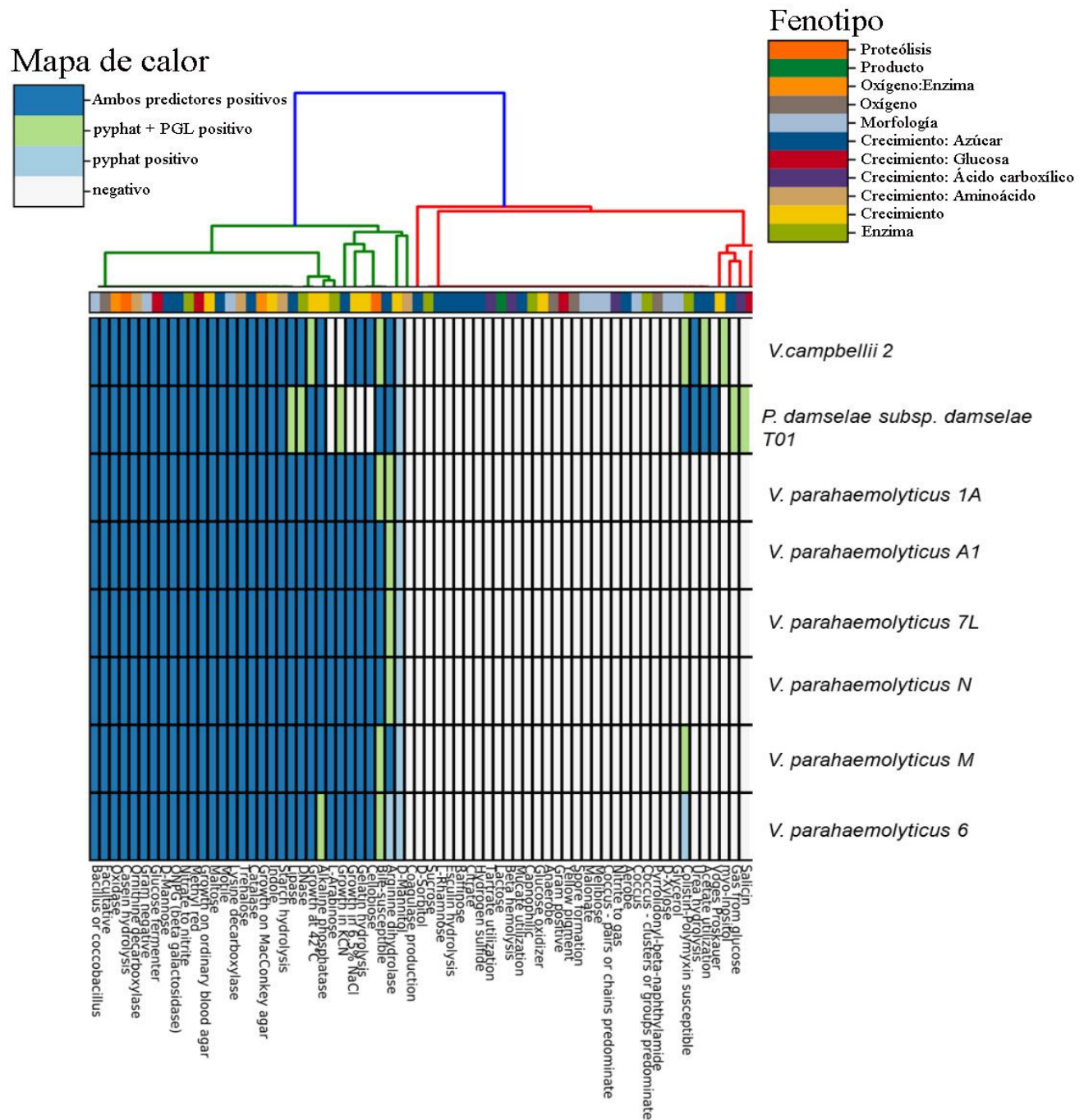
**Figura 13.** BLAST-Matriz comparativa de proteomas. Genomas de 41 cepas de *Vibrio parahaemolyticus*. Se utilizó el algoritmo de BLAST (Basic Local Alignment Search Tool), para identificar si dos proteínas se comparten entre los genomas, utilizando el criterio 50/50 idéntico en secuencia y cubriendo el 50% de la longitud del gen. Las etiquetas azules indican las cepas pertenecientes a B.C.S., México.



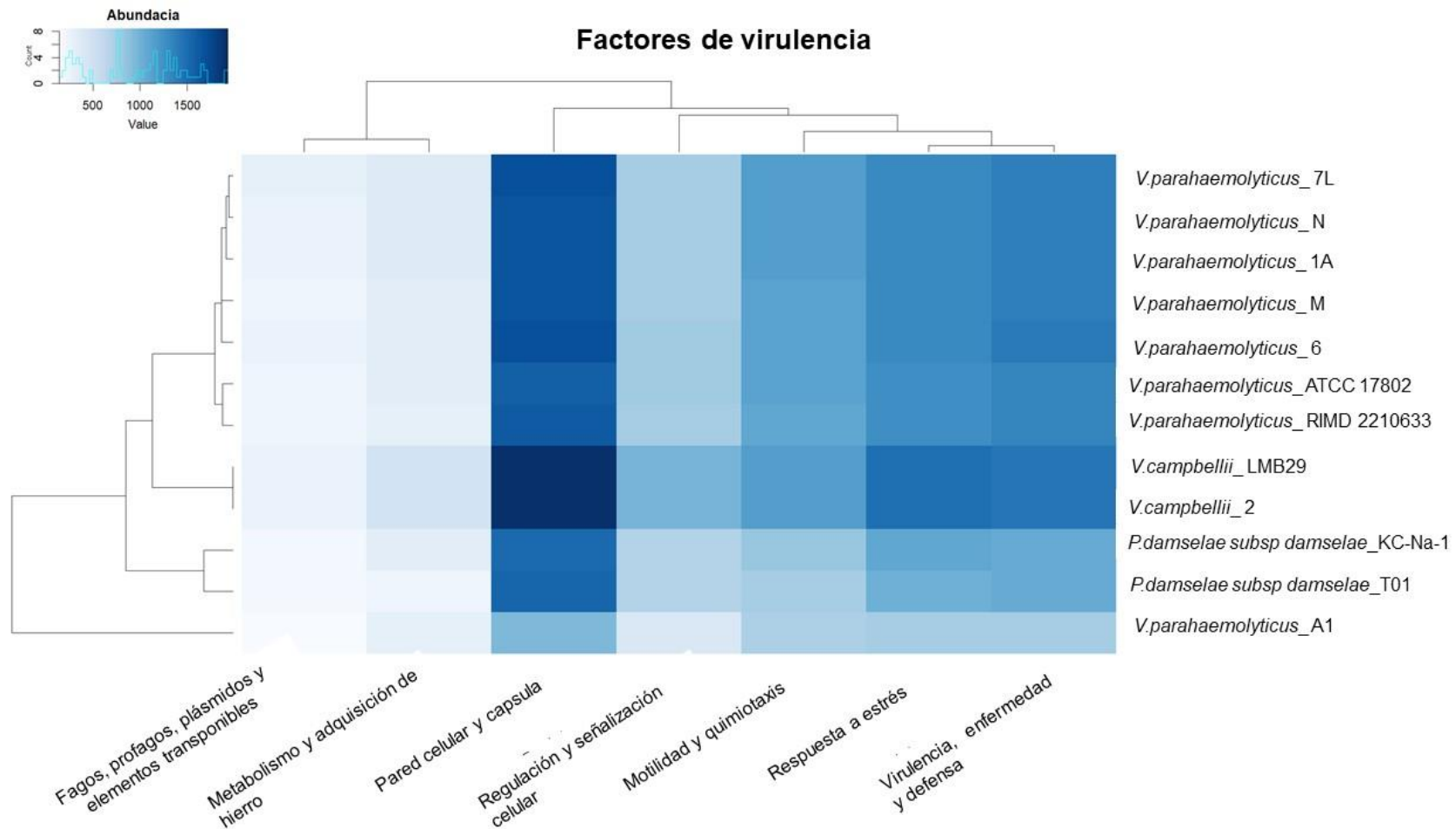
**Figura 14.** BLAST-Matriz comparativa de proteomas de *Vibrio parahaemolyticus* de B.C.S., México (N, A1, 7L, 1A, 6, y M) comparados con *V. campbellii* (02) y *Photobacterium damselaе* supesp. *Damselaе* (T01). Se utilizó el algoritmo de BLAST (Basic Local Alignment Search Tool), para identificar si dos proteínas se comparten entre los genomas, utilizando el criterio 50/50 idéntico en secuencia y cubriendo el 50% de la longitud del gen.



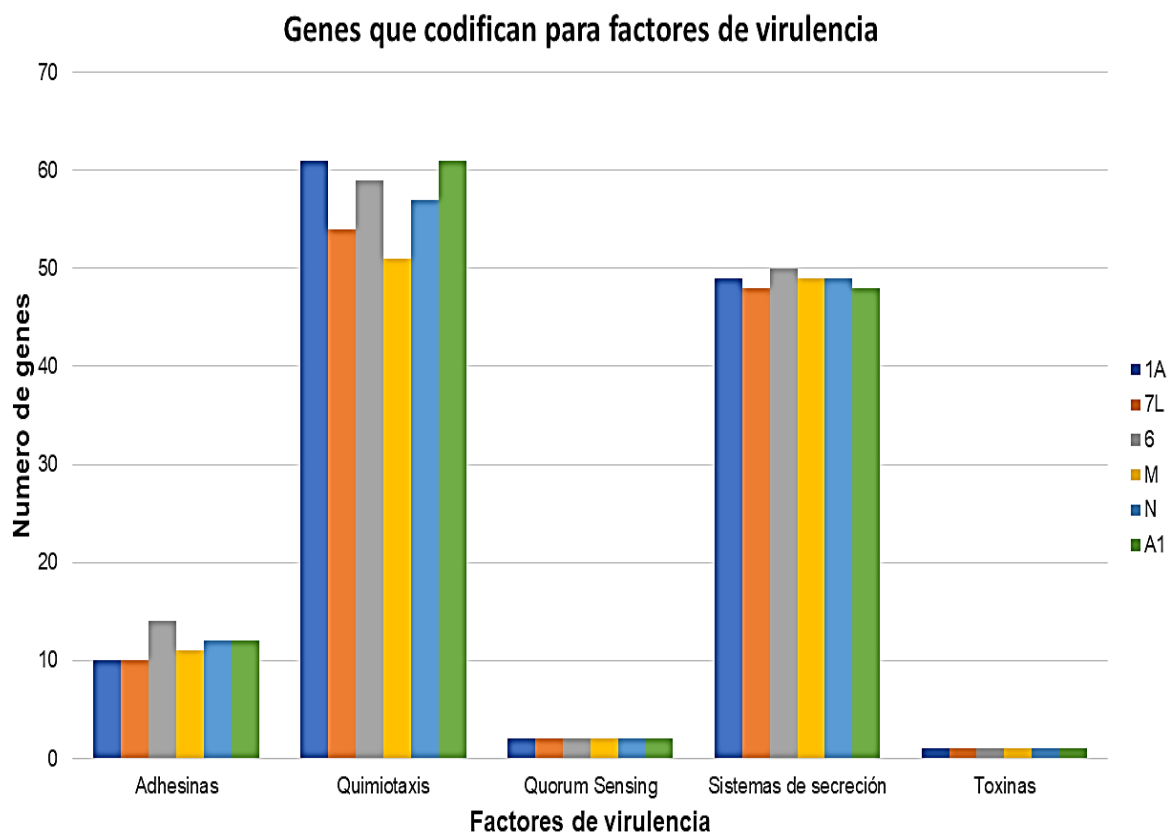
**Figura 15.** Genoma núcleo y pangenoma de *Vibrio parahaemolyticus*. Representación de 35 cepas de *V. parahaemolyticus* obtenidas en la bases de datos y 6 genomas aislados de B.C.S., México (A1, 1A, N, M, 7L y 6). Se utilizó la herramienta de CMG-biotools, y se eligió un límite máximo del 50% de identidad, con una cobertura del 50% del gen más largo. Si dos proteínas dentro de un genoma coincidían de acuerdo con el límite del 50/50%, se agruparon en una familia de proteínas. La línea roja indica el genoma núcleo con 3,143 genes y la línea azul indica el pangenoma con 11,104 genes. Las barras indican la comparación de nuevos genes (gris intenso) y nuevas familias de genes (gris claro).



**Figura 16.** Predicción del fenotipo de cepas de la familia *Vibronaceae*. Agrupamiento de ocho cepas en base a 67 rasgos usando el programa de Traitair. Cada columna representa uno de los 67 rasgos, mientras que las filas representan las ocho cepas del presente proyecto. El esquema de color de las columnas representa 11 propiedades fenotípicas, desde la proteólisis hasta la producción de enzimas. El origen de la predicción del fenotipo (Traitair phypat y / o phypat + PGL clasificador) determina el color de las entradas del mapa de calor. Los colores del dendrograma indican distribuciones de fenotipos similares en las muestras, según lo determinado por un agrupamiento jerárquico con SciPy.



**Figura 17.** Mapa de calor que muestra la abundancia de los factores de virulencia de la familia Vibrionaceae. Comparación general de categorías relacionadas a la patogenicidad de bacterias de la anotación con RASTtk, realizada con el programa R, usando la clasificación de SEED de RASTtk.



**Figura 18.** Abundancia de genes que codifican para factores de virulencia en *Vibrio parahaemolyticus*. Grafica de abundancias de genes relacionados a la virulencia de una cepa patogénica. Las secuencias fueron tomadas de la base de datos VFDB para el género *Vibrio* sp.. Se realizó un blastn con el programa BRIG para elegir aquellas secuencias con un  $\geq 90\%$  de identidad, que cubrían más del 90% de la secuencia del gen.

## 8. DISCUSIÓN

### 8.1. Secuencias cromosómicas y de bacteriófagos

La aparición de enfermedades en animales acuáticos de cultivo es una amenaza económica permanente y un desafío para la industria acuícola. El impacto de enfermedades puede afectar la sanidad y bienestar de los animales acuáticos, el comercio y la salud humana. En la península de Baja California Sur al igual que otros estados de la zona Noroeste de México, los cultivos de peces, camarones y ostiones, continuamente están siendo devastados por enfermedades provocadas principalmente por protozoarios, bacterias y virus. En este trabajo nos dimos a la tarea de utilizar técnicas moleculares para evaluar bacterias patógenas de la familia *Vibrionaceae* con herramientas como son la secuenciación y comparación de genomas. Estas herramientas han revolucionado diversas áreas como la patología obteniendo resultados muy relevantes, que permiten extraer conclusiones no solo sobre la biología del organismo, sino también sobre aspectos comparativos de gran importancia para comprender las relaciones que tienen, por ejemplo, las bacterias patógenas con sus hospederos.

Se secuenciaron ocho genomas de bacterias patógenas aisladas de granjas de camarón, ostión y peces de la península de Baja California Sur (*Vibrio parahaemolyticus*, camarón; *V. campbellii*, ostión; y *Photobacterium damselaesupesp. damselaes*, pez). Estas bacterias además de que fueron patógenas en un reto con juveniles de camarón, crecieron en un medio hipersalino (Rojas *et al.*, 2017), lo que generó el interés de trabajar con ellas. Se identificó que los genomas secuenciados en este trabajo poseen dos cromosomas circulares, uno más pequeño que el otro, con una pronunciada asimetría en la distribución de los genes entre los dos cromosomas, como se ha identificado en otros organismos de la familia *Vibrionaceae* (Heidelberg *et al.*, 2000; Rowe-Magnus *et al.*, 2001; Trucksis *et al.*, 1998). Además estos genomas también poseen otras secuencias de ADN pertenecientes a plásmidos, bacteriófagos, transposones, factores de virulencia, entre otros.

Para poder utilizar la información dentro de la secuencia de un genoma es fundamental anotarla con datos biológicos relevantes. Las anotaciones de un genoma ayudan a los científicos a entender qué significa su secuencia, cómo está estructurada y cómo funciona. Los genomas del género *Vibrio* están compuestos de alrededor de 5,000 genes, a diferencia de *P. damsela* *supesp. damsela* que presentó un genoma más pequeño con alrededor de 4,000 genes. También el porcentaje de contenido de Guanina-Citosina (GC) fue mayor en las especies estudiadas del género *Vibrio* con un 45% mientras que *P. damsela* *subsp. damsela* contiene un 40.57% de GC.

Diversos estudios han sugerido una correlación entre el contenido de GC y el tamaño del genoma, propuesto primeramente por Sueoka (1962). Se ha documentado que las bacterias de vida libre tienen un contenido de GC mayor que los organismos patógenos obligados y facultativos (Rocha y Danchin, 2002). En el caso de *Vibrio* y *Photobacterium* ambos son considerados patógenos oportunistas que producen enfermedad sólo cuando el sistema inmune del hospedero se deprime por alguna causa, sin embargo, un genoma y contenido de GC más pequeño como el de *Photobacterium* podría indicar que tiende a comportarse más como un patógeno obligado que un patógeno oportunista, siendo más selectivos y manteniendo una relación más estrecha con sus hospederos (Moran, 2002; Ochman y Davalos, 2006), además esta especie posee diversos factores de virulencia que suponen una ventaja para invadir nuevos hospederos (Osorio *et al.*, 2018; Rivas *et al.*, 2013).

Por otro lado las especies de *Vibrio* spp., comúnmente transitan entre patógenos oportunistas y un ambiente de vida libre (mar abierto, estanques, diversos animales incluido el humano, sedimentos), lo que indica que su genoma puede tener más apertura para el intercambio de elementos genéticos móviles, y además de permitirle la invasión de diversos hábitats (Lin *et al.*, 2018). También se ha documentado lo contrario con un estudio realizado recientemente con diversos genomas bacterianos por Almpanis *et al.* (2018), quienes sugieren que no se puede otorgar una correlación lineal al comportamiento del genoma y el contenido de GC debido a la heterocedasticidad de los datos que presentan los genomas, por lo que es necesario



realizar un meta-análisis para encontrar el factor o los factores ambientales claves que influyen en estos genomas.

Otro elemento que nos permite caracterizar el genoma de un organismo es el ARN de transferencia (ARNt). El ARN de transferencia (ARNt) es una pequeña molécula de ARN que participa en la síntesis de proteínas, y es una molécula clave en la traducción del ARN. Los genomas de *Vibrio* presentaron entre 105 y 117 ARNt, mientras que *P. damsela* subsp. *damsela* cuenta con 123 ARNt. Varias bacterias clasificadas como patógenas tienen un grado de inestabilidad genómica superior a la media con mutaciones frecuentes en puntos críticos (condiciones de crecimiento deficiente o ambientes hostiles), como se ha observado en las secuencias repetitivas de microsatélites, y genes que codifican para ARNt. En el ARNt existen diversas modificaciones post-transcripcionales en los nucleosidos, y estas modificaciones en el ARNt son muy dinámicas y su prevalencia varía según el estado metabólico de la célula (Björk y Hagervall, 2014). Se ha identificado que estas modificaciones del ARNt pueden presentarse debido al estrés fisiológico como oxidación, temperatura, pH, salinidad, inanición de nutrientes, etc. Asimismo, se ha destacado que las modificaciones de ARNt son reguladores importantes de la expresión de factores de virulencia, siendo que estas modificaciones, constituyen una respuesta rápida a estímulos externos repentinos, como cambios en el ambiente o una infección patogénica. También los ARNt son sitio de unión para los genomas virales y sirven como elementos que flanquean las islas genómicas (Hou, 1999; Koh y Sarin, 2018).

Las llamadas secuencias CRISPR, también presentes en el material genético de las bacterias, son secuencias palindrómicas cortas repetidas, y se ha identificado que entre estas a menudo procede la secuencia de un virus, lo que indica, que estas secuencias repetidas son utilizadas como un sistema inmune adaptativo que protege al huésped contra virus invasores. Los sistemas CRISPR están asociados a unos genes denominados Cas, los cuales constituyen el elemento esencial de la defensa antivírica. Los elementos CRISPR en nuestro trabajo, solo se encontraron en la anotación del genoma de la cepa *V. parahaemolyticus* 6, y se conoce que este sistema pueden influir en la virulencia de las bacterias (Louwen *et al.*, 2014). Los datos

de un análisis comparativo de genomas de la familia *Vibronaceae*, sugieren que existen muchas variaciones del contenido de proteínas Cas dentro de los diferentes tipos, y la adquisición de los sistemas CRISP-Cas en elementos genéticos móviles, es una característica común en este grupo (McDonald *et al.*, 2019).

En la familia *Vibronaceae* no hay trabajos que indiquen una correlación directa entre la presencia de los sistemas CRISPR-Cas y los factores de virulencia, pero se ha demostrado que la presencia de estos permite la inclusión de factores de virulencia como sistemas de secreción tipo 3 y la participación de Transposones, como los Tn7, Tn3 y Tn10 (McDonald *et al.*, 2019). Estos transposones se encontraron en todas las cepas estudiadas a excepción de *V. parahaemolyticus* 7L y *P. damsela* subsp. *damsela* T01. Los transposones Tn7 se han visto involucrados con el sistema subtipo I-F CRISPR-Cas, el cual promueve la unión de elementos genéticos móviles (Peters *et al.*, 2017). La cepa *V. parahaemolyticus* A1 contiene el transposon Tn7, además de los genes tnsABCDE, que codifican para una transportasa TnsAB, un regulador de TnsC, TnsD y TnsE necesarios para la inserción de un sitio de unión, el cual es el responsable directo de dirigir elementos a otros elementos genéticos móviles (McDonald *et al.*, 2019).

En las secuencias analizadas también se evaluó la presencia de profagos en ambos cromosomas. Se ha demostrado que los bacteriófagos pueden otorgar características a sus hospederos para adaptarse a nuevos ambientes y también pueden potenciar algunas características como la patogénesis de una bacteria, la cual utilizan para invadir a sus huéspedes. En la identificación de bacteriófagos se utilizó la base de datos PHASTER, la cual muestra los resultados de la posible presencia de profagos incluidos en cada cromosoma de nuestras cepas de interés (*Photobacterium damsela* subsp. *damsela* T01, *Vibrio campbellii* 02 y *Vibrio parahaemolyticus* A1, N, 6 1A, M y 7L). Como se observa en la Tabla IX, los profagos que presentan la característica de intactos son aquellos para *Vibrio* spp., que su puntuación es  $\geq 90$  siendo los bacteriófagos, *Vibrio* VGHphi, *Vibrio* vB\_VpaM\_MAR, *Vibrio* VfO3K6, *Vibrio* phage VCY-phi, *Vibrio* phage VGJphi, *Vibrio* phage VP882, *Vibrio*

VFJ que mostraron estar intactos, esto quiere decir que presuntamente el profago se encuentra completo en el cromosoma (esto calculado de manera *in silico*).

El vibriofago Vibrio phage VP882 que se ha reportado para *V. cholera*, se encuentra en el cromosoma II de las cepas de *V. parahaemolyticus* 1A y 7L. Este bacteriófago también se ha reportado que infecta a *V. parahaemolyticus*, y tiene la capacidad de responder a un AI (Autoinductor) del QS (Quorum Sensing) producida por el hospedero (*Vibrio*). Este vibriofago codifica una proteína Gp56, la cual se denomina VqmA<sub>Phage</sub> (Silpe y Bassler, 2019), lo que indicaría que las cepas 1A y 7L, pueden tener la característica de estimular la formación del QS debido a la presencia de dicho profago.

En cuanto al vibriofago VGJ-phi se encontró en el cromosoma II en los genomas de *V. parahaemolyticus* 7L, N, A1 y 1A. Este profago se ha reportado para *V. cholerae* 0139 cepa SG25-1, el cual está estrechamente relacionado evolutivamente con el bacteriófago VF33 de *V. parahaemolyticus* (Campos *et al.*, 2003). Este último es el posible responsable de la interacción con ADN cromosómico y plasmídico de las células huésped (*Vibrio*), para así desempeñar un papel en la movilización y dinámica de genes virulentos de *V. parahaemolyticus* (Chang *et al.*, 1998). Se ha encontrado también que VGJ-phi utiliza como receptor el MSHA (hemaglutinina sensible a manosa), el cual está implicado con la formación de biofilms y adherencia a superficies plantónicas (Campos *et al.*, 2003).

Para el profago Vibrio vB\_VpaM\_MAR se encontró en el cromosoma II de las cepas *V. parahaemolyticus* A1, N, 6 y 7L. Este vibriofago tiene la capacidad de lisar un 76% de las cepas de *V. parahaemolyticus* probadas, pertenece a la familia Myoviridae y se encuentra relacionado con fagos de Vibrio como VP882, VHML y VP58.5 (Villa *et al.*, 2012). El bacteriófago vB\_VpaM\_MAR fue obtenido de San Felipe, Baja California, México, se comprobó que tiene un alta especificidad hacia la invasión de su hospedero, y su contenido de GC es de 51.3%, lo que corresponde a los resultados encontrados en las cepas N y 7L en el cromosoma II (Tabla IX). Asimismo, los vibriofagos de las cepas 6 y A1 contienen un contenido de GC de 54.5 y 50.9% respectivamente, lo que podría indicar que los profagos de

estas cepas podrían estar incompletos y/o contaminados, o podría presentar modificaciones en su genoma ya que el %GC no corresponde con lo reportado (Villa *et al.*, 2012)

En cuanto al bacetriófago VCY-phi es un fago filamentoso que se encontró en el cromosoma I de la cepa de *V. parahaemolyticus* 1A. Se ha reportado en *V. cholera* en estanques de ostión, y a finales del verano tienen una transición de su forma replicativa (como los plásmidos) e integrativa en el genoma del hospedero, la cual se debe a factores ambientales (Xue *et al.*, 2012). El VFJ-phi se localizó en el cromosoma II de la cepa de *V. parahaemolyticus* 6. Es un bacteriófago filamentoso que también se ha encontrado en *V. cholerae* ICDC-4470, que es homólogo al vibriofago VCY-phi, por una región de 700 pb. Se ha documentado que el bacteriófago VFJ-phi existe de forma replicativa (como un plásmido) y que no se integra al cromosoma de su hospedero. Después de una infección con este bacteriófago a *V. cholerae*, se mostró una inhibición en el crecimiento, y en la formación del flagelo, así como, un alto nivel de resistencia al antibiótico Kanamicina. Se ha comprobado que los bacetriófagos filamentosos proporcionan diferentes fenotipos clave en *V. cholerae*, como atributos de virulencia (Wang *et al.*, 2013).

Con lo que respecta al vibriofago VfO3K6 encontrado en el cromosoma II de la cepa *V. parahaemolyticus* A1 se ha reportado en otras cepas como la *V. parahaemolyticus* serotipo O3:K6, causante de gastroenteritis y diarrea, asociada al consumo de alimentos marinos, sin embargo, no se encontraron más reportes sobre la influencia de este fago sobre su hospedero.

En cuanto a los profagos cuestionables y/o incompletos que se encontraron como resultados en los demás cromosomas de las cepas, debido a que las regiones de estos muestran similitud con las secuencias de dichos profagos de la base de datos de PHASTER, está la posibilidad de que el profago no tenga su genoma completo y las secuencias con el mayor grado de identidad son las que se muestran en la Tabla IX.

## 8.2. Filogenia de la familia Vibrionaceae

Las relaciones filogenéticas entre organismos, nos permiten obtener datos importantes sobre la organización, función y evolución de estos y su ADN a lo largo de toda la escala filogenética. Estas relaciones se pueden describir en mayor detalle en forma de un árbol filogenético que captura la estructura jerárquica de las distintas relaciones evolutivas. Para realizar arboles filogenéticos uno de los genes que comúnmente se utilizan es el 16S ADN<sub>r</sub>, que codifica para la subunidad pequeña del ARN ribosomal (16S ARN<sub>r</sub>), el cual es un gen muy conservado en la escala evolutiva y parece que reproduce bien la relación entre organismos (Olsen y Woese 1993).

Sin embargo las filogenias de especies derivadas de las comparaciones de genes individuales rara vez son consistentes entre sí, debido a la transferencia horizontal de genes, paralogías desconocidas y las tasas de evolución altamente variables (Snel *et al.*, 1999). Además debido a que las observaciones sobre diversidad intra-específica han generado necesidad de herramientas más sensibles y precisas para evaluar distancias, la herramienta de 16S ARN<sub>r</sub> se vuelve irrelevante para cepas estrechamente relacionadas (Rosselló-Mora y Amann, 2001). Así, en este trabajo además de utilizar como herramienta de análisis el gen del 16S ARN<sub>r</sub> también utilizamos otros análisis como la identidad promedio de nucleótidos (ANI) y la tipificación multilocus de secuencias (MLST). ANI nos permite comparar la similitud de nucleótidos entre dos genomas bacterianos para determinar si pertenecen o no a la misma especie, donde cepas de la misma especie deben tener una similitud superior al 95-96%. Por su parte MLST implica el concatenar las secuencias de una misma cepa y luego alinearlos con los de otras cepas o especies para obtener una mejor representación de sus relaciones filogenéticas.

Los análisis con ANI se llevaron a cabo para cada una de las especies estudiadas por separado, mientras que los análisis de MLST y 16S ARN<sub>r</sub> se llevaron a cabo solo con las cepas de *V. parahaemolyticus*. Los análisis de ANI muestran que cuando se comparan los genomas de *Photobacterium* y *Vibrio* con otros organismos de la misma especie el porcentaje

de identidad es mayor al 95% como lo han demostrado diversos autores (Han *et al.*, 2016; Konstantinidis y Tiedje, 2005). La cepa de *P. damsela* *subsp. damsela* T01 en nuestro estudio (aislada del pez *Totoaba macdonaldi*) se agrupa con la cepa PhdpWu-1 aislada también de un pez de agua salada, así como, también el análisis muestra que están estrechamente relacionadas con la especie *Photobacterium damsela* *subsp. piscicida*. Estas dos especies han sido relacionadas por la similitud del gen 16S, y en nuestro estudio presentan un elevado porcentaje de similitud genotípica, pero estas dos subespecies se distinguen claramente por diversas características fenotípicas (Botella *et al.*, 2002).

Por otra parte *V. campbellii* 2 de nuestro estudio aislada de camarón, presentó la mayor similitud con la cepa *V. campbellii* 20130629003S01, aislada de mar abierto, representadas en un clado, y los genomas de otras bacterias de la misma especie se agrupan en otro clado, como las bacterias de referencia de la ATCC 25920 y BAA-116. Esta especie ha sido clasificada continuamente como *V. harveyi* por su similitud genética y fenotípica, así como también ha sido clasificada como no-patógena y no-bioluminiscente (de la Pena *et al.*, 2001; Lin *et al.*, 2010). Sin embargo, diversos estudios indican que esta bacteria es un patógeno importante que contribuye a diversas enfermedades en organismos marinos y a partir de estudios genómicos se ha observado que presenta diferencias con *V. harveyi* (Gómez-Gil *et al.*, 2004; Thompson *et al.*, 2004; Wang *et al.*, 2015).

En lo que respecta a las cepas de *V. parahaemolyticus* (A1, N, 1A, 7L, 6 y M), en el árbol realizado con el 16S ARNr se muestra que la cepa M diverge antes que las demás cepas de interés del proyecto, y la cepa A1 también está agrupada en otro clado diferente a las demás, que probablemente se deba a que esta cepa fue aislada de un ostión, mientras que las demás fueron aisladas de camarón. Sin embargo, con el análisis de ANI se muestra una similitud de nucleótidos entre los genomas de *V. parahaemolyticus* estudiados del 98-99%, pero el dendograma muestra que las cepas M y 6, tienen una distancia genética mayor que el resto de las cepas estudiadas, lo que concuerda con el análisis de MLST. Con el análisis de MLST las cuatro cepas pertenecen a la misma secuencia tipo 1257, y usando diversos marcadores moleculares (*gyrB*, *recA*, *dtdS*, *pntA*, *pyrC*, *tnaA* y *dnaE*) contienen el mismo perfil alélico.

Cabe recalcar que todas las cepas provienen de un ancestro común y que evolutivamente todas las cepas, excepto la M y la 6, son de reciente aparición. Esto indica que las cepas M y 6 provenientes de camarón, han estado sometidas a presiones selectivas similares y han estado presentes en el ambiente marino por un largo periodo de tiempo, a diferencia de las otras cepas (N, A1, 1A, y 7L) que de acuerdo a los dendogramas, de los tres análisis que se realizaron, evolutivamente son más recientes. Esto nos podría dar la idea de considerarlos patógenos emergentes ya sea porque han aparecido recientemente en la península de Baja California Sur, además es claro que estas 4 cepas se agrupan en un clado por separado debido a que fueron aisladas del mismo hábitat y de la misma región geográfica, por lo que tienen una relación evolutiva cercana.

### 8.3. Genómica Comparativa

El creciente número de genomas secuenciados de bacterias patógenas pertenecientes a la familia *Vibrionaceae*, hace posible la comparación y el análisis genómico, que puede revelar importantes mecanismos de evolución y variación para entender la patogénesis de estas bacterias. La reconstrucción y la comparación de genomas de especies bacterianas han revelado que la organización cromosómica no se conserva a lo largo de la escala evolutiva (Fremez *et al.*, 2007). Un análisis genómico puede ayudar a inferir relaciones entre las características de organización de los cromosomas y la fisiología celular.

En el presente estudio unas de las herramientas de comparación de genomas que utilizamos fue la que nos permitió crear un atlas estructural de los cromosomas I y II de las cepas de *Vibrio parahaemolyticus*. Además del tamaño del cromosoma, el atlas nos muestra tres tipos de información: composición base (contenido AT, asimetría GC), repeticiones globales dentro del replicón (directo e invertido) y las propiedades estructurales del ADN (preferencia de posición, energía de apilamiento de ADN, y curvatura) (Vesth *et al.*, 2013).

El atlas estructural mostró que el cromosoma I presenta un tamaño de 3, 288,558 pb, mientras que el cromosoma II es más pequeño con un tamaño de 1, 877,212 pb. Se ha argumentado que

el cromosoma I es más grande, en promedio contienen de 3.46 a 3.65 Mpb aproximadamente, pero esto puede variar dependiendo la especie o la cepa. Al cromosoma I se le ha atribuido que contiene los genes necesarios para las funciones primarias (replicación, transcripción, traducción, biosíntesis), a lo que también se le conoce como genoma núcleo, aquellos genes esenciales para su sobrevivencia. El segundo cromosoma no contiene genes esenciales, por lo que es un replicón más variable que el primero y dada esta variabilidad se relaciona con su adaptación al ambiente (Harrison *et al.*, 2010). Varias evidencias sugieren que este cromosoma II, fue originalmente un megaplásmido, capturado por una especie ancestral de *Vibrio*. La estructura de los dos cromosomas se encuentra en otras especies de *Vibrio*, sugiriendo que los genes contenidos en el megaplásmido proporcionaron al *Vibrio* una ventaja evolutiva en el ecosistema acuático (Heidelberg *et al.*, 2000).

El atlas también muestra una región que representa la asimetría de GC, que indica la ubicación del origen y término de la replicación, y cambios de la mayoría de los Gs (azul) a más Cs (rosa) (Fig. 10). También se identificaron los operones de ARNr dentro de los cromosomas, y como era de esperarse el cromosoma I contiene mayor número de operones que el cromosoma II (8 y 1, respectivamente), por el tipo de función que desempeña. El número de copias del operón ribosómico por genoma bacteriano varía considerablemente, de 1 a 15, siendo relativamente constante a nivel de especie, género e incluso familia (Klappenbach *et al.* 2001). También se ha considerado que el número de operones de ARNr puede dictar la tasa de síntesis de ribosomas y la respuesta a cambios favorables en un entorno de crecimiento (Klappenbach *et al.*, 2000).

Para el atlas genómico se utilizó una cepa de referencia de *V. parahaemolyticus* clínica que provoca gastroenteritis en el humano (RIMD 2210633), que mostró en ambos cromosomas elementos genéticos móviles que tienen una composición base diferente, y en el atlas estructural se pueden inferir por áreas de diferente curvatura, energía de apilamiento y preferencia de posición, en comparación con el contenido cromosómico promedio (gris) (flechas rojo oscuro, Fig. 10). Esta cepa de referencia en el cromosoma I presenta dos



pequeñas regiones con alta curvatura y elevada energía de apilamiento, mientras que en el cromosoma II se muestra una región muy grande con estas características.

Algo relevante que observamos es que estas regiones no están presentes en los genomas estudiados en este trabajo. Los elementos genéticos móviles no tienen capacidad de replicación y dependen de la integración en la célula del hospedero, lo que indica que nuestras cepas al ser aisladas de organismos marinos en granjas acuícolas, probablemente no presenten los mismos elementos móviles que la cepa clínica que afecta a otro tipo de hospedero, y que podría tener alguna relación con la especialización del ambiente marino. Otra explicación podría ser que estas regiones vacías son regiones de divergencia, que también se explica por el tipo de hospedero. Asimismo, por otra parte se puede argumentar que los genomas estudiados en este trabajo están incompletos y trabajar con genomas cerrados representa una ventaja, puesto que en nuestro análisis también se integró otra cepa que se analizó junto con las nuestras (ATCC 17802), la cual presenta un genoma completo, y en el cromosoma II de esta cepa se presentan indicios de estas regiones de alta curvatura y elevada energía de apilamiento. Lukjancenko y Ussery (2014), se observaron que el número de genes del genoma núcleo de los genomas incompletos disminuyeron bruscamente a solo la mitad en comparación con los genomas completos, lo que se debe principalmente a la falta de información de los genomas incompletos. Asimismo, diferencias entre la energía de apilamiento y la curvatura intrínseca son estructuras que pueden potencialmente intervenir en la expresión de los genes, su reorganización e incluso en puntos críticos de la evolución (Vesth *et al.*, 2013).

### 8.3.1. Uso de codones y aminoácidos

Otra característica importante del genoma es el uso de codones. En el análisis realizado con nuestras cepas se muestra una preferencia por los codones TTT (Serina) y AAA (Fenilalanina), estos se asocian al contenido de AT, y también a la evolución de las bacterias para el uso de estos codones. Los experimentos de laboratorio, los métodos estadísticos y bioinformáticos sugieren que la mutación hacia la riqueza de AT en los procariotas puede

deberse a la pérdida de ciertos genes de reparación (García-González et al., 2012) y/o a la falta de presiones selectivas (Hershberg y Petrov, 2010).

Entre los aminoácidos más representativos en las cepas estudiadas encontramos a la Arginina (R), Serina (S), y Leucina (L). Una de las principales estrategias de bacterias patógenas es la producción de efectores extracelulares tóxicos para persistir en un nicho, interviniendo con el crecimiento/viabilidad de otros microorganismos del nicho. Uno de los efectores que cumple esta función son los D- aminoácidos (Alvarez *et al.*, 2017). Las especies de *Vibrio* usualmente coexisten en diversos ecosistemas marinos, de agua dulce y terrestre, por lo que hay una alta probabilidad que hayan establecido estrategias cooperativas secretando aminoácidos. Aliashkevich *et al.*(2018), demostraron que la producción de D- Arginina podría ser una estrategia cooperativa de las comunidades de *Vibrio*, utilizando dicho aminoácido para controlar poblaciones de bacterias, y que este aminoácido muestra una gran potencia y amplia toxicidad en términos del número de especies de bacterias que puede afectar. Referente al aminoácido S, este se ha relacionado en bacterias del género *Vibrio* y otras bacterias marinas con la participación de proteasas, como las proteasas-serina (metaloproteasas) como un factor de virulencia (Salamone *et al.*, 2015).

Otra de las características que se remarca con el uso de aminoácidos es en relación al %GC y %AT. En un análisis realizado por Bohlin et al. (2013), utilizaron 2000 genomas de bacterias y encontraron que el aminoácido S se encuentra entre los aminoácidos más utilizados por los genomas con un alto contenido de AT y el aminoácido R en los genomas con un contenido alto de GC, mientras que Leucina (L) se distribuye en genomas con altos contenido de GC y AT. Por lo anterior las especies de *Vibrio* presentan cierta plasticidad en la composición de estos aminoácidos lo que hace que estas bacterias puedan adoptar tanto el comportamiento de patógenos oportunistas como bacterias de vida libre.

### 8.3.2. Comparaciones de Proteomas

En este trabajo se llevaron a cabo dos análisis para comparar los proteomas de los genomas estudiados mediante algoritmos de BLAST: matriz comparativa del proteoma y análisis del genoma núcleo y pangenoma. En el análisis del proteoma mediante matriz comparativa se llevaron a cabo dos matrices, una con 41 cepas de *V. parahaemolyticus* incluyendo nuestras cepas, y otra matriz donde además de las cepas de *V. parahaemolyticus* utilizadas en este estudio se incluyeron las cepas de *V. campbellii* 02 y *P. damsela subsp. damsela* T01. Cuando se analizaron los proteomas de todas las cepas de *V. parahaemolyticus* se presentó un alto porcentaje de familias de genes compartidos entre estas (67.7-99.6%), Un estudio realizado en genomas de la familia *Vibrionaceae* mostró que diferentes cepas de *V. cholera* comparten entre el 70 y 80% de proteínas, mientras la similitud con organismos de otra especie pero de la misma familia, oscilo entre 30 y 45% (Vesth *et al.*, 2010). Estas observaciones son similares a las que reportamos en este trabajo puesto que la homología de proteínas de *V. parahaemolyticus* comparada con *V. campbellii* es de 58.3-59%, y cuando se comparan las cepas de *Vibrio* spp. con *P. damsela subsp. damsela* estas solo tienen un 29.8-30.6% de semejanza.

Asimismo, las matrices muestran la homología interna entre cepas con un valor de 2.4-4.7% en las 41 cepas de *V. parahaemolyticus* analizadas y entre 2.5-5.6% dentro de los genomas evaluados en este estudio. En el estudio realizado por Vesth *et al.* (2010) encontró que la homología interna en *V. cholera* varía de 1.3 a 5.3% y entre organismos de la familia *Vibrionaceae* varía entre el 1.8 a 5%.

Por otra parte el genoma núcleo y pangenoma se obtuvieron analizando el conjunto acumulativo de todos los genes compartidos entre genomas (pangenoma) y el conjunto conservado de familias de genes en todos los genomas (genoma núcleo), de las cepas de *V. parahaemolyticus*. El pangenoma analizado con 41 cepas de *V. parahaemolyticus* en este trabajo contiene 11,104 familias y el genoma núcleo contiene 3,143 familias, datos similares a los reportados por Li *et al.*, (2014). Estos autores analizaron nueve genomas de *V.*

*parahaemolyticus* (4 cepas clínicas y 5 cepas ambientales), y reportan un pangenomas de alrededor de 10,000 familias y el genoma núcleo de alrededor de 3,000 familias. Asimismo, Vesth *et al.*, (2010) analizaron varios genomas de la familia *Vibrionaceae* y reportaron que el tamaño de su genoma núcleo es de 1000 y su pangenoma de 20,200. La diferencia en tamaño del pangenoma se debe al genoma variable que presenta *V. parahaemolyticus* en comparación con las otras especies de la familia *Vibrionaceae*. También nuestro análisis mostró que el número de nuevas familias aumenta con el número de genomas analizados, lo que indica que *V. parahaemolyticus* tiene un pangenoma abierto, lo cual coincide con otros reportes realizados (Li *et al.*, 2014). Además de que la transferencia horizontal de genes junto con la plasticidad genómica han sido considerados como un método importante para adquirir nuevos genes y acelerar la evolución (Boucher *et al.*, 2011).

### 8.3.3. Comparación de fenotipos

La identificación fenotípica de los *Vibrios* sigue siendo una tarea difícil particularmente para algunas especies hermanas. Las especies hermanas de *Vibrio* tienen genomas muy similares (alrededor del 70%) y fenotipos casi indistinguibles. A pesar de tener genomas similares, estas especies pueden ser reconocidas como diferentes unidades evolutivas en la naturaleza, con lo que para comprender la ecología de estos microorganismos, es importante obtener información fenotípica basada en todas las secuencias del genoma. Referente a la predicción *in silico* del fenotipo de las ocho bacterias del presente proyecto, se observa que se conservan en común varias características metabólicas, las cuales van desde morfología, oxígeno, crecimiento (azúcar glucosa, ácido carboxílico y aminoácidos). La particularidad más marcada de estas comparaciones es que las cepas de *V. parahaemolyticus* 1A, A1, 7L y N dieron negativo a la susceptibilidad a colistina-polimixina. La colistina y la polimixina B forman parte de la última línea de antibióticos para bacterias Gram negativas multirresistentes (Poirel *et al.*, 2017), lo que podría indicar que estas cuatro cepas bacterianas presentan resistencia a dicho antibiótico, dándole así una característica más con respecto a su patogenicidad, mientras que

las cepas 6 y M de la misma especie, junto con *P. damsela* *supesp. damsela* y *V. campbellii*, si mostraron susceptibilidad a este antibiótico.

También se mostraron diferencias de estas dos últimas especies con las cepas de *V. parahaemolyticus* como la ausencia de L-arabinosa, y presentan hidrólisis de urea y utilización de acetato. La L-Arabinosa es un monosacárido que se usa en medios de cultivo celular, que sirve como fuente de carbono bacteriano, y se puede usar para distinguir entre bacterias en función de sus capacidades de fermentación. Por su parte la hidrólisis de urea sirve para identificar la presencia de la enzima ureasa, así como la utilización de acetato sirve para identificar bacterias que presentan un metabolismo aerobio capaces de realizar fermentación homoacética, lo cual se basa en la capacidad para convertir lactato a acetato (Giono *et al.*, 1998).

El mioinositol es una característica fenotípica que puede ayudar a distinguir entre *V. parahaemolyticus* y *V. campbellii* por la capacidad de esta última para fermentar mioinositol. Asimismo los fenotipos de salicina, gas de glucosa y Voges Proskauer fueron características fenotípicas únicas de *P. damsela* *subesp. damsela*, donde la salicina se utiliza para diferenciar a las bacterias en base a su habilidad para fermentar carbohidratos, si la bacteria usa la glucosa producirá ácido y/o gas, y el test de Voges Proskauer nos indica la capacidad de la bacteria para fermentar la glucosa (Cercenado y Cantón, 2010).

#### 8.4. Factores de Virulencia

Las diferencias entre bacterias patógenas y sus relativas no patogénicas, es debido a un grupo de genes capaces de generar una enfermedad, estos son denominados genes de virulencia, los cuales codifican para proteína que son llamadas factores de virulencia. Los genes de virulencia usualmente se pueden encontrar en el cromosoma bacteriano que contiene la información necesaria para la sobrevivencia de la bacteria, en el ADN extracromosómico (plásmido), y también se han encontrado en bacteriófagos (Alberts *et al.*, 2002).

Una manera de analizar los factores de virulencia es por medio de la presencia/ausencia y la abundancia de estos, y los podemos agrupar de acuerdo las características. En las cepas estudiadas de la familia *Vibrionaceae* encontramos que la característica más abundante es la de pared celular y capsula. La cepa que presentó mayor abundancia en esta categoría fue *V. campbellii* 02, seguida por las cepas *Photobacterium damsela* subsp. *damselae* T01, *V. parahaemolyticus* 6 y 7L, M, y luego las cepas de *V. parahaemolyticus* 1A y 6. La capsula bacteriana confiere efectos protectores contra una amplia gama de presiones ambientales esto debido a los polisacáridos que sirven como barrera física, en especial contra el sistema inmunitario durante la infección de los patógenos a sus hospederos, ya que contiene antígenos para el reconocimiento de su hospedero y estos son específicos (Cress *et al.*, 2014).

En la característica de respuesta a estrés se encuentran de mayor a menor la cepa *V. campbellii* 2, *V. parahaemolyticus* 6, M, 1A, N y 7L y *Photobacterium damsela* subsp. *damselae* T01. Se ha reportado para *Vibrio* que una pre-adaptación a una concentración de NaCl al 6% aumenta su supervivencia bajo condiciones de ácidos letales (Kalburge *et al.*, 2014), lo que coincide para las cepas de *V. parahaemolyticus* 6, 1A y M donde hay un crecimiento del 1% al 10% de NaCl (Tabla II) y provocan un daño severo en juveniles de camarón (Rojas *et al.*, 2017), lo que indicaría que pueden llevar su proceso de infección evadiendo diversos agentes estresantes del ambiente.

Otra característica analizada es la motilidad y quimiotaxis, donde la expresión génica de este grupo fue muy abundante en las especies estudiadas. Esta característica ayuda a la adaptación al medio ambiente, ya que las bacterias son capaces de movilizarse a ambientes más favorables o les permite evadir sustancias desfavorables. La cepa con mayores características de motilidad y quimiotaxis fue *V. campbellii* 2 lo cual coincide con la prueba de movilidad (Tabla I) indicando que esta cepa tiende a desplazarse a ambientes más favorables. Cabe recalcar que en las pruebas bioquímicas realizadas por Rojas *et al.* (2017) las cepas de *V. parahaemolyticus* N y 1A no presentaron motilidad, pero se encontraron diversos genes responsables para quimiotaxis (Anexo C). En los resultados que obtuvimos con las cepas de *V. parahaemolyticus*, se identificó que la cepa 1A presenta la mayor abundancia de genes

relacionados a esta característica, seguida por las cepas A1, 6, N, 7L y M. La estrategia de movilidad más común entre las bacterias es el movimiento dirigido por la rotación flagelar. Se ha documentado que todos los *Vibrios* presentan motilidad, y encontramos en este trabajo varios genes que codifican para los flagelos, sin embargo, también se ha documentado que hay bacterias que sobreviven sin un flagelo.

En bacterias del género *Vibrio*, utilizan un flagelo polar cuya rotación está acoplada a un flujo de iones sodio ( $\text{Na}^+$ ) y algunas especies del género como *V. alginolyticus*, *V. parahaemolyticus* o *V. shilonii*, son capaces de expresar, adicionalmente al desplazarse en medio sólido, un sistema de flagelos laterales cuya rotación está acoplada a un flujo de protones ( $\text{H}^+$ ). La expresión del flagelo polar es constitutiva y está presente en la célula todo el tiempo, por otro lado, la expresión de los flagelos laterales es inducible y se encuentra asociada con la alta densidad del medio (Atsumi *et al.*, 1992; Ringgaard *et al.*, 2018). Aunque el papel de la motilidad durante la infección se ha examinado en varios patógenos bacterianos, la importancia de la quimiotaxis en este proceso no se ha estudiado tan extensamente. A priori, se predeciría que la quimiotaxis funciona de la mano con la motilidad para permitir que las bacterias naden hacia los sitios de colonización preferidos. Sin embargo, en el caso de los patógenos entéricos, el requisito de motilidad y quimiotaxis varía de ser crucial a ser prescindible para la infección. Por ejemplo, las especies de *Shigella* no son móviles, pero son altamente infecciosas, y claramente, la ausencia de motilidad en este organismo no es un impedimento para la infección (Cossart y Sansonetti, 2004).

Esta motilidad es dictada por un sistema de quimiotaxis que emplea diversos receptores que detectan señales extracelulares llamadas proteínas de quimiotaxis transmembrana que aceptan metilo (MCP) (Lamb *et al.*, 2019). Al recibir la señal, estos se agrupan para formar matrices de detección quimiosensoriales. Los quimioefectores se unen a los MCP en su dominio periplásmico y transmiten la señal de unión a sus puntas citoplasmáticas, donde interactúan con diversas proteínas reguladoras y adaptadoras como las proteínas Che A, CheB, CheY y otras más como las que se muestran en el Anexo C.

Algunas bacterias pueden evitar el sistema inmune del huésped y, por lo tanto, la fagocitosis, lo que se conoce como anti-fagocitosis. A pesar del impacto de la enfermedad bacteriana en los sistemas de acuicultura, solo unos pocos estudios directos han caracterizado la evasión del sistema inmune del huésped de bacterias del género *Vibrio*. En una infección el sistema inmunitario del huésped inicia respuestas inflamatorias rápidas para proteger al huésped de la propagación de microbios invasores. La inflamación se realiza cuando las células de la primera línea de defensa, tales como células epiteliales, macrófagos y neutrófilos detectan señales de alarma (Van-Avondt *et al.*, 2015). Las cepas estudiadas mostraron una gran variedad de genes de antifagocitosis descritos en otras especies de *Vibrio*. Los más abundantes fueron los que codifican para capsulas de polisacáridos (CPS) (Anexo C). La membrana externa de bacterias gram negativas sirve como barrera de permeabilidad que no se atraviesa fácilmente por compuestos hidrofóbicos dañinos para las bacterias, por ejemplo, sales biliares de tracto gastrointestinal. Los lipo-polisacáridos de los patógenos gramnegativos pueden proteger a las células de lisis mediada por moléculas como la lisozima. La membrana externa y los componentes capsulares de bacterias gram negativas pueden proteger la capa de peptidoglucano de la actividad lítica de la lisozima (Gao *et al.*, 2003; Hughey y Johnson, 1987; Martínez de Tejada *et al.*, 1995).

Además de los factores de virulencia mencionados también está la adhesión que es un factor clave de virulencia en los patógenos que puede ser afectado por el ambiente. Se identificaron diversos genes de adherencia, siendo la cepa 6 donde se encuentran en mayor abundancia. Los genes de la familia MSHA (hemaglutinina sensible a manosa) y pilus tipo IV (Anexo C), están presentes en estas cepas y en las demás cepas de *V. parahaemolyticus* estudiadas. Estas familias de genes sirven para la colonización del hospedero (O'Boyle y Boyd, 2014). Los genes de MSHA junto con los genes ChiRP (genes del pilus para la regulación-quitina, estos no fueron analizados en el presente proyecto) son usados para la regulación del pilus en presencia de quitina, uno de los principales biopolímeros presentes en el ambiente marino, además de que en camarón (*Penaeus monodon*) contiene del 14-27% de peso seco de quitina (Islam *et al.*, 2016), y se ha reportado que para desechos de ostión contienen un 69.65% (Alabaraoye *et al.*, 2018). Se ha demostrado que tanto el pilus y el MSHA participan en la



colonización intestinal en roedores y humanos (O'Boyle y Boyd, 2014), y se han realizado pruebas con cepas mutantes de MSHA de *V. parahaemolyticus*, en donde la adherencia de esta bacteria a células Caco-2, disminuye los niveles de acumulación celular que indica que la adherencia a las células huésped es un requisito previo a la colonización (O'Boyle *et al.*, 2013).

Los seres vivos han desarrollado complejos mecanismos para la asimilación de hierro. La adquisición de hierro es un proceso indispensable para todas las bacterias, ya que es un elemento fundamental en el metabolismo celular (Aguado-Santacruz *et al.*, 2012; Chhibber y Bansal, 2013; Weinberg, 2004). Uno de los mecanismos más importantes en las bacterias es la biosíntesis de agentes quelantes, denominados sideróforos. Las cepas de *V. parahaemolyticus* estudiadas presentan una gran variedad de sideróforos (Anexo C). Los *Vibrios* al igual que otras bacterias han desarrollado una variedad de mecanismos para adquirir hierro de cada ambiente que habitan, algunos de estos sistemas de adquisición están estrechamente relacionados entre todos los *Vibrios*, lo que refleja su ascendencia en común. Los sideróforos son secretados al ambiente unidos a la superficie de *Vibrio spp.*, y sirven como captadores del hierro, una vez que el complejo ferrisideróforo es reconocido por un receptor específico en la superficie de la membrana externa (Neilands, 1982; Wyckoff *et al.*, 2007). Sandy *et al.*, (2010), reportan varios sideróforos tipo amida en *Vibrio sp.* como la trivancrobactina, divancrobactina, vancrobactina y anguibactona. La cantidad de hierro requerida por los *Vibrios* varía dependiendo de la fisiología y metabolismo de la célula, pero las concentraciones en el medio van de un rango de 0.1  $\mu\text{M}$  - 5.0  $\mu\text{M}$  proporcionando el suficiente hierro para el crecimiento óptimo de las bacterias (bajo condiciones de laboratorio).

Otros de los factores de virulencia ampliamente considerados dañinos para diversos organismos son las toxinas que poseen ciertas bacterias. En nuestro trabajo, las cepas de *V. parahaemolyticus* N, 7L, 6 y A1 contienen la toxina TLH, mientras que las cepas M y 1A la toxina RTX (repeats in toxin). La toxina TLH es una hemolisina termolabil, la cual causa hemólisis en ambientes clínicos y en ambientes marinos, y el gen se sobre-regula durante condiciones de infección intestinal, causando una severa toxicidad en células del huésped, y su

efecto depende de la dosis-tiempo (Klein *et al.*, 2014), sin embargo se ha reportado que las hemolisinas (TLH/TDH) de *V. parahaemolyticus* con deleciones no afectan en el efecto citotóxico en células cultivadas (Zhang y Orth, 2013). En cuanto a la toxina RTX, se ha reportado mayormente para bacterias *V. cholerae*, sin embargo, esta toxina se encuentra presente en una amplia gama de bacterias Gram negativas (Linhartová *et al.*, 2010), capaz de romper los filamentos de actina, aunque se desconoce su papel en la patogénesis de la enfermedad. Asimismo, la toxina RTX se ha relacionado con la evasión de la fagocitosis por el hospedero, interviniendo en la señalización de moléculas involucradas (receptores como Toll, quinasas como las Src y FAK, Pyk2, y quinasas MAP) en la inducción de la fagocitosis (Chen *et al.*, 2017). Lo anterior indica que la cepa 1A y M podrían utilizar este mecanismo para causar citotoxicidad en su huésped.

Recientemente se ha reportado la enfermedad de necrosis hepatopancreática aguda (AHPND) que es una enfermedad grave del camarón que ha surgido y ha causado grandes pérdidas a la industria mundial del cultivo del camarón desde 2010 (Zhang *et al.*, 2012; Tran *et al.*, 2013; Lee *et al.*, 2015). Se encontró que el causante de la enfermedad era la bacteria *V. parahaemolyticus* (Zhang *et al.*, 2012; Tran *et al.*, 2013). Se ha demostrado que VP<sup>AHPND</sup> alberga un plásmido que expresa una toxina mortal Pirvp (constituida por PirAvp y PirBvp), que es homóloga a la toxina binaria (Pir) relacionada con el insecto *Photorehabdus* (Lee *et al.*, 2015). Las cepas de este trabajo obtenidas de camarón presentaban signos de esta enfermedad (Rojas, comunicación personal), sin embargo esta toxina no se encontró en nuestros genomas ni la toxina TDH. Esto puede deberse a que en el presente trabajo no obtuvimos las secuencias de los plásmidos solo las secuencias cromosómicas.

En cuanto al QS (Quorum Sensing) que se encuentra asociado a la formación de biofilm, se ha visto que dependiendo de la especie la formación de biofilm regula la transcripción de manera positiva o negativa de algunas toxinas, como TDH y TLH (Hammer y Bassler, 2003; Makino *et al.*, 2003). Las especies de *Vibrio* producen y secretan moléculas de señal llamados Autoinductores (AI), estos activan la transcripción de TDH/TLH y la formación de biofilm por

la unión de receptores (Lux-S, CqsA). Con relación a lo anterior se ha demostrado que LuxS regula positivamente la regulación de TDH (Guo *et al.*, 2018; Zhu *et al.*, 2002).

Asimismo, *V. parahaemolyticus* cuenta con dos sistemas de secreción de diferente tipo, dos T3SS (sistema de secreción tipo 3) y dos T6SS (sistema de secreción tipo 6). El sistema T3SS tiene como función la entrega directamente a citosol de los efectores (en su mayoría toxinas) (Tandhavanant *et al.*, 2018). Estos efectores una vez dentro del citosol de la célula huésped modifican las funciones de esta. El T3SS1 se ha encontrado mayormente en el cromosoma uno, y se encuentra codificado por la primera isla de patogenicidad en la mayoría de las cepas clínicas y ambientales (Paranjpye *et al.*, 2012). Por otro lado, el sistema T6SS es funcionalmente análogo, y puede tener una función crítica en el proceso de infección. Se ha demostrado que el T6SS es necesario para la adhesión de *V. parahaemolyticus* a las células del huésped, así como, se involucra en el transporte vesicular (Salomon *et al.*, 2013b). En las cepas del presente proyecto se encontraron los genes de sistemas de secreción tipo II EPS, efectores secretados T3SS1 y T3SS2, T3SS1 y T3SS2. En un experimento de infección realizado con la cepa de *V. parahaemolyticus* FORC\_022 en cangrejos en salsa marinada, se analizó el transcriptoma de esta cepa y dentro de las categorías sobre representadas están los sistemas de secreción tipo II EPS y T3SS1, entre otras categorías como MSHA, polisacáridos de la capsula, receptores Heme y TLH (Chung *et al.*, 2018). El sistema de secreción tipo II EPS utiliza un componente EpsE citoplasmático del sistema de secreción tipo 2 en *V. cholerae* y *V. parahaemolyticus*. Mediante la hidrólisis de ATP y una interacción con la proteína de membrana citoplasmática EpsL, el componente EpsE secreta una toxina a través de la membrana plasmática (Zhang y Orth, 2013), mientras que las proteínas efectoras T3SS (*vopQ*, *vopR*, *vopS*) transitan hacia la célula del huésped. Esto se lleva a cabo mediante un translocon especializado, el cual una vez ensamblado forma un poro y un canal que atraviesa la membrana citoplasmática del hospedero permitiendo así el paso de varios efectores (Ham y Orth, 2012).

## 9. CONCLUSIONES

En el presente trabajo se identificó que el genoma de *Vibrio campbellii* (aislada de camarón) fue el más grande con 5.9 Mb, para las cepas de *V. parahaemolyticus* (aisladas de camarón y ostión) tuvieron genomas muy similares con alrededor de 5.2 Mb, mientras que *Photobacterium damsela* subsp. *damsela* (aislada de pez) presentó un genoma menor de 4.5 Mb. De igual forma el contenido de GC para los *Vibrio* fue de 45%, mientras que para *Photobacterium damsela* subsp. *damsela* se identificó un contenido de GC del 40%. Los miembros de esta familia poseen dos cromosomas, donde el cromosoma I es más grande que el cromosoma II, esto debido a que en el cromosoma I se encuentra la información para las necesidades básicas de la bacteria. Asimismo, se identificaron diversos profagos los cuales son especie-específicos de sus hospederos.

Para tener una idea de la diversidad evolutiva y genética de las cepas de *V. parahaemolyticus*, se utilizaron diversos análisis que mostraron que la presión selectiva del nicho influye en la agrupación de estas cepas. Las cepas 6 y M son muy parecidas presentando un largo periodo evolutivo, mientras que las cepas N, 1A, A1 y 7L muestran una distancia evolutiva menor pero comparten diversas características. Estas últimas podrían considerarse que aparecieron recientemente en la región de Baja California Sur.

Debido a los análisis de genómica comparativa se ha corroborado que *V. parahaemolyticus* contiene un genoma diverso y con mucha plasticidad, ya que aunque conservan muchas características en común (genoma núcleo) se corroboró que contienen un pangenoma abierto, lo que les da la facilidad de aceptar nuevos elementos genéticos. El genoma núcleo constante y los pangenomas abiertos hacen que los *Vibrios* posean los materiales genéticos necesarios y los genes accesorios flexibles para ser cosmopolitas.

La composición de la base genómica tiene un efecto sustancial en las frecuencias de aminoácidos y codones en los genomas bacterianos. Asimismo, los organismos de la familia Vibrionaceae que son patógenos marinos parecen representar un amplio segmento de la

diversidad de este grupo y se caracterizan, en general, por tener características fenotípicas y genotípicas bien definidas que dependen del ambiente de donde fueron aislados. Un mayor conocimiento acerca de las regiones y proteínas exclusivas podría conducir a un mejor entendimiento de la patogénesis de estos microorganismos y a la posible identificación de nuevos blancos para tratamientos.

Los genes relacionados a virulencia otorgan la patogenicidad a *Vibrio*, y pueden otorgar ventajas para invadir e infectar a sus hospederos, sin embargo, la ausencia o presencia de algún factor de virulencia no muestra aparentemente cambios significativos en la disminución de la citotoxicidad en sus huéspedes. Por lo que, aunado a la presencia de estos factores de virulencia, la adaptación al medio (hospedero), puede verse relacionado a la adquisición de nuevos elementos, como lo son los elementos genéticos móviles o re-arreglos de los cromosomas. Las bacterias pueden generar información genética nueva mediante duplicación de genes, inserción de elementos genéticos transponibles y transferencia horizontal de genes, proceso gracias al cual dos especies bacterianas pueden intercambiar información genética a lo largo de la evolución.

#### Recomendaciones:

- Realizar un reto de infección con las ocho cepas del presente proyecto con los diferentes hospederos, y hacer extracción de ARN para analizar el transcriptoma de cada cepa
- Repetir el análisis de blastMatrix agregando genomas de *Vibrio campbellii* y *Photobacterium damsela subsp. damsela*, con genomas completos.
- Realizar el análisis para ver agrupamientos de ortólogos
- Realizar un cgMLST (core genoma MLST) con los genomas de interés en comparación con otros genomas de *Vibrio parahaemolyticus* ya reportados, para una mayor resolución.
- Hacer filogenia con el coregenoma de los 41 genomas utilizados, los genomas de NCBI y los aislados de B.C.S.

- De buscar elementos móviles completar los genomas de las bacterias aisladas de B.C.S., ya que algunos elementos genéticos móviles (islas genómicas, transposones, CRISPR-cas) se encuentran flanqueados por secuencias repetidas cortas, lo cual es un problema en la secuenciación-ensamble
- Determinar los ORF (Open Reading Frame)
- Realizar una extracción de ADNp (ADN plasmídico) y realizar una secuenciación masiva.
- Realizar una caracterización de los presuntos vibriofagos encontrados como intactos en los genomas de *Vibrio parahaemolyticus*.
- Identificar las secuencias CRISPR y CRISPR-Cas
- Incluir más factores de virulencia al archivo multifasta, como las secuencias de MAM7 (Molécula de adhesión multivalente 7), ChiRP (genes del pilus para la regulación-quitina).
- Bajar los criterios de selección para las secuencias de factores de virulencia a 50% de similitud y 50% de cobertura.
- Localizar los operones donde se encuentran los factores de virulencia

## 10. LITERATURA CITADA

- Aguado-Santacruz, G.A., B. Moreno-Gómez, B. Jiménez-Francisco, E. Garcia-Moya, R.E. Preciado-Ortiz. 2012. Impact of the microbial siderophores and phytosiderophores on the iron assimilation by plants: a synthesis. *Fitotecnia Mexicana*, 35:9-21.
- Ahmed, E., Holmström, S. J. 2014. Siderophores in environmental research: roles and applications. *Microbial biotechnology*, 7: 196-208.
- Alberts B, Johnson A, Lewis J., *Molecular Biology of the Cell*. 4th edition. New York: Garland Science. 2002. Introduction to Pathogens. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK26917/>
- Aliashkevich, A., Alvarez, L., Cava, F. 2018. New insights into the mechanisms and biological roles of D-amino acids in complex eco-systems. *Frontiers in microbiology*, 9:683-693.
- Alikhan, N. F., Petty, N. K., Zakour, N. L. B., Beatson, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics*, 12: 402.
- Alon, U., Camarena, L., Surette, M. G., y Arcas, B. A., Liu, Y., Leibler, S., Stock, J. B. 1998. Response regulator output in bacterial chemotaxis. *The EMBO journal*, 17: 4238-4248.
- Alonzo, F., Torres, V. J. 2014. The biocomponent pore-forming leucocidins of *Staphylococcus aureus*. *Microbiol. Mol. Biol. Rev.*, 78:199-230.
- Alvarez, L., Aliashkevich, A., de Pedro, M. A., Cava, F. 2018. Bacterial secretion of D-arginine controls environmental microbial biodiversity. *The ISME journal*, 12: 438-450.
- Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., Pevzner, P. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics*, 32:3380-3387.
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D. S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, 44: W16-W21.
- Arthur, J. R., Subasinghe, R. P. 2002. Potential adverse socio-economic and biological impacts of aquatic animal pathogens due to hatchery-based enhancement of inland open-

water systems, and possibilities for their minimisation. FAO FISHERIES TECHNICAL PAPER, 406:113-126.

Assembly Information, NCBI (<https://www.ncbi.nlm.nih.gov/assembly/basics/#WGA>) Fecha de consulta: 23/02/2020

Atsumi T, McCarter L, Imae Y.1992. Polar and lateral flagellar motors of marine *Vibrio* are driven by different ion-motive forces. *Nature* 355:182–4.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Svetlana G. F., Folker M., Gary J. Olsen, Robert Olson, Ross A Overbeek, Leslie K McNeil, Daniel Paarmann, Tobias Paczian, Bruce Parrello, Gordon D P., Claudia Reich, Rick Stevens, Olga Vassieva, Veronika Vonstein, Andreas Wilke Olga Zagnitko. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9: (75) 1-15.

Barber, G. R., Swygert, J. S. 2000. Necrotizing fasciitis due to *Photobacterium damsela* in a man lashed by a stingray. *New England Journal of Medicine*, 342: 824-824.

Bardiau, M., Szalo, M., Mainil, J. G. 2010. Initial adherence of EPEC, EHEC and VTEC to host cells. *Veterinary Research*, 41: 41-57.

Beckloff, N., Starkenburg, S., Freitas, T., Chain, P. 2012. Bacterial genome annotation. In *Microbial Systems Biology*, pp. 471-503.

Bellanger, X., Payot, S., Leblond-Bourget, N., Guédon, G. 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiology Reviews*, 38: 720-760.

Bellenger, J. P., Wichard, T., Kustka, A. B., Kraepiel, A. M. L. 2008. Uptake of molybdenum and vanadium by a nitrogen-fixing soil bacterium using siderophores. *Nature Geoscience*, 1: 243-246.

Berg, H. C., Purcell, E. M. 1977. Physics of chemoreception. *Biophysical journal*, 20: 193-219.

Bohlin, J., Brynildsrud, O., Vesth, T., Skjerve, E., Ussery, D. W. 2013. Amino acid usage is asymmetrically biased in AT- and GC-rich microbial genomes. *PLoS One*, 8: e69878.



- Bohlin, J., Eldholm, V., Pettersson, J. H., Brynildsrud, O., Snipen, L. 2017. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC genomics*, 18: 151-165.
- Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault M., Izsvák, Z., Levin, L. H., Macfarlan, T.S., Mager, D. L. Feschotte, C. 2018. Ten things you should know about transposable elements. *Genome Biology*, 19: (199) 1-15.
- Braud, A., Hoegy, F., Jezequel, K., Lebeau, T., Schalk, I. J. 2009. New insights into the metal specificity of the *Pseudomonas aeruginosa* pyoverdine–iron uptake pathway. *Environmental microbiology*, 11:1079-1091
- Bren, A., Eisenbach, M. 1998. The N terminus of the flagellar switch protein, FliM, is the binding domain for the chemotactic response regulator, CheY. *Journal of Molecular Biology*, 278: 507-514.
- Bren, A., Eisenbach, M. 2000. How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. *Journal of bacteriology*, 182: 6865-6873.
- Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, D. G., Shukla M., Thomason III A. J., Stevens, R., Vonstein, V., Wattam, R. A. Xia, F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific reports*, 5: 8365-8379.
- Broberg, C. A., Zhang, L., Gonzalez, H., Laskowski-Arce, M. A., Orth, K. 2010. A *Vibrio* effector protein is an inositol phosphatase and disrupts host cell membrane integrity. *Science*, 329:1660-1662.
- Brogden, K.A., J.A. Roth, T.B. Stanton, C.A. Bolin, F.C., Minion, M.J. Wannemuehler. 2000. *Virulence Mechanisms of Bacterial Pathogens*, 3rd edn. ASM Press, Washington DC, chapter 2 pp 19-40.
- Bryant, J. A., Sellars, L. E., Busby, S. J., Lee, D. J. 2014. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic acids research*, 42: 11383-11392

- Butterton, J. R., Calderwood, S. B. 1994. Identification, cloning, and sequencing of a gene required for ferric vibriobactin utilization by *Vibrio cholerae*. *Journal of Bacteriology*, 176: 5631-5638.
- Büttner, D. 2012. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant-and animal-pathogenic bacteria. *Microbiol. Mol. Biol. Rev.*, 76: 262-310.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, 17: 540-552.
- Ceccarelli, D., Colwell, R. R. 2014. *Vibrio* ecology, pathogenesis, and evolution. *Frontiers in microbiology*, 5: 1.
- Ceccarelli, D., Hasan, N. A., Huq, A., Colwell, R. R. 2013. Distribution and dynamics of epidemic and pandemic *Vibrio parahaemolyticus* virulence factors. *Frontiers in cellular and infection microbiology*, 3: 97-105.
- Chayakulkeeree, M., Sorrell TC, Siafakas AR, Wilson CF, Pantarat N, Gerik KJ, Boadle R, Djordjevic JT .2008. Role and mechanism of Phosphatidylinositol-specific Phospholipase C in survival and virulence of *Cryptococcus neoformans*. *Mol Microbiol*. 69: 809- 826.
- Chhibber, S., Nag, D., Bansal, S. 2013. Inhibiting biofilm formation by *Klebsiella pneumoniae* B5055 using an iron antagonizing molecule and a bacteriophage. *BMC microbiology*, 13: 174.
- Chu, B. C., Vogel, H. J. 2011. A structural and functional analysis of type III periplasmic and substrate binding proteins: their role in bacterial siderophore and heme transport. *Biological chemistry*, 392: 39-52.
- Coil, D., Jospin, G., Darling, A. E. 2014. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*, 31: 587-589.
- Colin, R., Sourjik, V. 2017. Emergent properties of bacterial chemotaxis pathway. *Current opinion in microbiology*, 39:24-33.
- CONAPESCA, 2017, México evoluciona como una potencia acuícola <https://www.gob.mx/conapesca/articulos/mexico-evolucion-a-como-una-potencia-en-produccion-acuicola?idiom=es>

- Cornelis, G. R. 2010. The type III secretion injectisome, a complex nanomachine for intracellular 'toxin' delivery. *Biological chemistry*, 391: 745-751.
- Costa, T. R., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., Waksman, G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology*, 13: 343-359.
- Cottrell, M. T., David, K. L. 2003. Contribution of major bacterial groups to bacterial biomass production (thymidine and leucine incorporation) in the Delaware estuary. *Limnology and Oceanography*, 48: 168-178.
- Cottrell, M. T., Kirchman, D. L. 2000. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Applied and Environmental Microbiology*, 66: 5116-5122A.
- Cottrell, M. T., Kirchman, D. L. 2000. Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low-and high-molecular-weight dissolved organic matter. *Applied and Environmental Microbiology*, 66: 1692-1697B.
- Cottrell, M. T., Wood, D. N., Yu, L., Kirchman, D. L. 2000. Selected chitinase genes in cultured and uncultured marine bacteria in the  $\alpha$ - and  $\gamma$ -subclasses of the proteobacteria. *Applied and environmental microbiology*, 66: 1195-1201C.
- Craig, N. L. 2002. Mobile DNA: an introduction. In *Mobile DNA II*, American Society of Microbiology, pp. 3-11.
- Daffe, M., Etienne, G. 1999. The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. *Tubercle and lung disease*, 79: 153-169.
- Darling, A. C., Mau, B., Blattner, F. R., Perna, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, 14:1394-1403.
- de Jesús Hernández-Díaz, L., Leon-Sicairos, N., Velazquez-Roman, J., Flores-Villaseñor, H., Guadron-Llanos, A. M., Martinez-Garcia, J. J., Vidal, J. E., Canizalez-Roman, A. 2015. A pandemic *Vibrio parahaemolyticus* O3: K6 clone causing most associated diarrhea cases in the Pacific Northwest coast of Mexico. *Frontiers in microbiology*, 6: 221
- Delmont, T. O., Eren, A. M. 2018. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*, 6: e4320.

- Dong, X., Wang, H., Xie, G., Zou, P., Guo, C., Liang, Y., Huang, J. 2017. An isolate of *Vibrio campbellii* carrying the pir VP gene causes acute hepatopancreatic necrosis disease. *Emerging Microbes Infections*, 6: 1-3.
- Donnenberg, M. S. 2000. Pathogenic strategies of enteric bacteria. *Nature*, 406: 768.
- Doore, S. M., Fane, B. A. 2016. The microviridae: diversity, assembly, and experimental evolution. *Virology*, 491: 45-55.
- Dryselius, R., Izutsu, K., Honda, T., Iida, T. 2008. Differential replication dynamics for large and small *Vibrio* chromosomes affect gene dosage, expression and location. *BMC genomics*, 9:559.
- Dunlap, P. V. 2011. Phylogeny, genomics, and symbiosis of *Photobacterium*. *FEMS microbiology reviews*, 35: 324-342.
- Eisenbach, M. 1996. Control of bacterial chemotaxis. *Molecular microbiology*, 20: 903-910.
- Elston, R. A., Frelier, P., Cheney, D. 1999. Extrapallial abscesses associated with chronic bacterial infections in the intensively cultured juvenile Pacific oyster *Crassostrea gigas*. *Diseases of Aquatic Organisms*, 37: 115-120.
- Erken, M., Lutz, C., McDougald, D. 2013. The rise of pathogens: predation as a factor driving the evolution of human pathogens in the environment. *Microbial ecology*, 65:860-868.
- Falke, J. J., Bass, R. B., Butler, S. L., Chervitz, S. A., Danielson, M. A. 1997. The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annual review of cell and developmental biology*, 13:457-512.
- FAO, 2017, Major bacterial diseases affecting aquaculture [http://www.fao.org/fi/static-media/MeetingDocuments/WorkshopAMR/presentations/07\\_Haenen.pdf](http://www.fao.org/fi/static-media/MeetingDocuments/WorkshopAMR/presentations/07_Haenen.pdf) Fecha de consulta: 23/02/202
- Farmer, III, J. J., Janda, J. M., Brenner, F. W., Cameron, D. N., Birkhead, K. M. 2005. Genus I. *Vibrio* Pacini 1854, 411AL, in *Bergey's Manual of Systematic Bacteriology*, 2nd Edn., Vol. 2, eds D. J. Brenner, N. R. Krieg, and J. T. Staley (New York, NY: Springer): 494–546.
- Ferrari, G., I. Garaguso, J. Adu-Bobie, F. Doro, A.R. Taddei, A. Biolchi, B. Brunelli, M.M. Giuliani, M. Pizza, N. Norais. 2006 Outer membrane vesicles from group B *Neisseria*

- meningitidis* delta *gna33* mutant: proteomic and immunological comparison with detergent-derived outer membrane vesicles. *Proteomics*, 6:1856-1866.
- Finlay, B.B., S. Falkow. 1997. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* 61:136-169.
- Flo, T. H., Smith, K. D., Sato, S., Rodriguez, D. J., Holmes, M. A., Strong, R. K., Akira S Aderem, A. 2004. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestering iron. *Nature*, 432: 917-921.
- Foerster, K. U., Von Mering, C., Hooper, S. D., Bork, P. 2005. Environments shape the nucleotide composition of genomes. *EMBO reports*, 6: 1208-1213.
- Genomic alignment, Emsembl, (<https://www.ensembl.org/info/genome/compara/analyses.html>)<https://www.ensembl.org/info/genome/compara/analyses.html>) Fecha de consulta: 23/02/2020
- Genomic Sequencing, Illumina, 2010, DataSheet: Sequencing ([https://www.illumina.com/documents/products/datasheets/datasheet\\_genomic\\_sequencing.pdf](https://www.illumina.com/documents/products/datasheets/datasheet_genomic_sequencing.pdf)) Fecha de consulta: 23/02/2020
- Gill, D. M. 1982. Bacterial toxins: a table of lethal amounts. *Microbiological reviews*, 46: 86-94.
- Gill, T. A. 2000. Waste from processing aquatic animals and animal products: implications on aquatic animal pathogen transfer No-956: FAO Fisheries Circular (FAO).
- Giltner, C. L., Van Schaik, E. J., Audette, G. F., Kao, D., Hodges, R. S., Hassett, D. J., Irvin, R. T. 2006. The *Pseudomonas aeruginosa* type IV pilin receptor binding domain functions as an adhesin for both biotic and abiotic surfaces. *Molecular microbiology*, 59: 1083-1096.
- Gode-Potratz, C.J. and McCarter, L.L. 2011. Quorum sensing and silencing in *Vibrio parahaemolyticus*. *Journal of Bacteriology*, 193: 4224–4237.
- Gode-Potratz, C.J., Kustus, R.J., Breheny, P.J., Weiss, D.S. and McCarter, L.L. 2011. Surface sensing in *Vibrio parahaemolyticus* triggers a programme of gene expression that promotes colonization and virulence. *Molecular Microbiology*, 79:240–263

- Goetz, D. H., Holmes, M. A., Borregaard, N., Bluhm, M. E., Raymond, K. N., Strong, R. K. 2002. The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Molecular cell*, 10: 1033-1043.
- Goldman, A. D., Landweber, L. F. 2016. What is a genome? *PLoS genetics*, 12: e1006181
- Gomez-Gil, B., Thompson, C. C., Matsumura, Y., Sawabe, T., Iida, T., Christen, R., et al. 2014. "Family Vibrionaceae (Chapter 225)," in *The Prokaryotes*, 4th Edn. eds E. Rosenberg, E. DeLong, F. L. Thompson's. Lory, and E. Stackebrandt New York, NY.
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., Tiedje, J. M. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International journal of systematic and evolutionary microbiology*, 57: 81-91.
- Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de La Cruz, F., Rocha, E. P. 2011. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS genetics*, 7: e1002222.
- Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. 2013. QCAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29: 1072-1075.
- Hahn, H.P. 1997. The type-4 pilus is the major virulence-associated adhesin of *Pseudomonas aeruginosa*—a review. *Gene*. 192:99-108
- Hardison, R.C. 2003. Comparative Genomics. *PLoS Biology*, 1:e58.
- Harrison, P. W., Lower, R. P., Kim, N. K., Young, J. P. W. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends in microbiology*, 18:141-148.
- Hazelbauer, G. L., Falke, J. J., Parkinson, J. S. 2008. Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends in biochemical sciences*, 33: 9-19.
- Hicks, S. W., Galán, J. E. 2013. Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nature Reviews Microbiology*, 11: 316-326.
- Higa, N., Toma, C., Koizumi, Y., Nakasone, N., Nohara, T., Masumoto, J. Suzuki, T. 2013. *Vibrio parahaemolyticus* effector proteins suppress inflammasome activation by interfering with host autophagy signaling. *PLoS pathogens*, 9: e1003142.
- Ho, B. T., Dong, T. G., Mekalanos, J. J. 2014. A view to a kill: the bacterial type VI secretion system. *Cell host and microbe*, 15: 9-21.

- Holmes, R. K. 2000. Biology and molecular epidemiology of diphtheria toxin and the tox gene. *Journal of Infectious Diseases*, 181: S156-S167.
- Honda, T. A. K. E. S. H. I., Ni, Y. X., Miwatani, T. O. S. H. I. O. 1988. Purification and characterization of a hemolysin produced by a clinical isolate of Kanagawa phenomenon-negative *Vibrio parahaemolyticus* and related to the thermostable direct hemolysin. *Infection and Immunity*, 56: 961-965.
- Hsieh, Y. C., Liang, S. M., Tsai, W. L., Chen, Y. H., Liu, T. Y., Liang, C. M. 2003. Study of capsular polysaccharide from *Vibrio parahaemolyticus*. *Infection and immunity*, 71: 3329-3336.
- Hsu, P. D., Lander, E. S., Zhang, F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157: 1262-1278.
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., Hauser, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11: 119-129.
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11: 119-129.
- Ichinose, Y., Ehara, M., Honda, T., Miwatani, T. 1994. The effect on enterotoxicity of protease purified from *Vibrio cholerae* O1. *FEMS microbiology letters*, 115:265-271.
- Illumina Sequencing Technology, 2010  
([https://www.illumina.com/documents/products/techspotlights/techspotlight\\_sequencing.pdf](https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf)) Fecha de consulta: 23/02/2020
- Jiang, W., Han, X., Wang, Q., Li, X., Yi, L., Liu, Y., Ding, C. 2014. *Vibrio parahaemolyticus* enolase is an adhesion-related factor that binds plasminogen and functions as a protective antigen. *Applied microbiology and biotechnology*, 98: 4937-4948.
- Jory, D. E. 1997. Necrotizing hepatopancreatitis and its management in shrimp ponds. *Aquaculture Magazine*, 23: 98-101.
- Kariin, S., Burge, C. 1995. Dinucleotide relative abundance extremes: a genomic signature. *Trends in genetics*, 11: 283-290.

- Katoh, K., Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30:772-780.
- Kinchen, J. M., Ravichandran, K. S. 2008. Phagosome maturation: going through the acid test. *Nature reviews Molecular cell biology*, 9: 781-795.
- Klein, S. L., West, C. K. G., Mejia, D. M., Lovell, C. R. 2014. Genes similar to the *Vibrio parahaemolyticus* virulence-related genes *tdh*, *tlh*, and *vscC2* occur in other *Vibrionaceae* species isolated from a pristine estuary. *Appl. Environ. Microbiol.*, 80: 595-602.
- Kolbe, D. L., Eddy, S. R. 2011. Fast filtering for RNA homology search. *Bioinformatics*, 27: 3102-3109.
- Krachler, A. M., Ham, H., Orth, K. 2011. Outer membrane adhesion factor multivalent adhesion molecule 7 initiates host cell binding during infection by gram-negative pathogens. *Proceedings of the National Academy of Sciences*, 108: 11614-11619.
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., Ussery, D. W. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*, 35: 3100-3108.
- Lamb, E., Trimble, M. J., McCarter, L. L. 2019. Cell-cell communication, chemotaxis, and recruitment in *Vibrio parahaemolyticus*. *Molecular Microbiology*. 112:99-113.
- Laslett, D., Canback, B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic acids research*, 32: 11-16.
- Lee, C. T., Chen, I. T., Yang, Y. T., Ko, T. P., Huang, Y. T., Huang, J. Y. Lightner, D. V. 2015. The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proceedings of the National Academy of Sciences*, 112: 10798-10803.
- Lee, E.Y., J.Y. Bang, G.W. Park, D.S.Choi, J.S. Kang, H.J. Kim, K.S. Park, J.O. Lee, Y.K. Kim, K.H. Kwon. 2007. Global proteomic profiling of native outer membrane vesicles derived from *Escherichia coli*. *Proteomics*. 7:3143-3153



- Lee, S. Y., Cho, H. S., Pelton, J. G., Yan, D., Henderson, R. K., King, D. S. Wemmer, D. E. 2001. Crystal structure of an activated response regulator bound to its target. *Nature Structural y Molecular Biology*, 8: 52-56.
- Li, L., Wong, H. C., Nong, W., Cheung, M. K., Law, P. T. W., Kam, K. M., Kwan, H. S. 2014. Comparative genomic analysis of clinical and environmental strains provides insight into the pathogenicity and evolution of *Vibrio parahaemolyticus*. *BMC genomics*, 15: 1135.
- Lin, H., Yu, M., Wang, X., Zhang, X. H. 2018. Comparative genomic analysis reveals the evolution and environmental adaptation strategies of vibrios. *BMC genomics*, 19:135.
- Liu, B., Zheng, D., Jin, Q., Chen, L., Yang, J. 2018. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic acids research*, 47: D687-D692.
- Liu, M., S. Chen. 2015. A novel adhesive factor contributing to the virulence of *Vibrio parahaemolyticus*. *Scientific Reports*. 5: 1-10.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., Darnell, J. 2000. *Molecular cell biology* 4th edition. National Center for Biotechnology Information, Bookshelf, pp 9.3.
- Lukjancenko, O., Ussery, D. W. 2014. *Vibrio* chromosome-specific families. *Frontiers in microbiology*, 5:73
- Magoč, T., Salzberg, S. L. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27: 2957-2963.
- Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Najima, M., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, Y., Honda, T., Shinagawa, H., Hattori, M., Lida, T. Kubota, Y. 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *The Lancet*, 361: 743-749.
- Mann, S., Chen, Y. P. P. 2010. Bacterial genomic G+ C composition-eliciting environmental adaptation. *Genomics*, 95: 7-15
- March-Rosselló, G.A., J.M. Eiros-Bouza, 2013. Quorum sensing en bacterias y levaduras. *Medicina Clínica*, 141:353–357.

- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17: 10-12.
- Matsuda, S., Kodama, T., Okada, N., Okayama, K., Honda, T., Iida, T. 2010. Association of *Vibrio parahaemolyticus* thermostable direct hemolysin with lipid rafts is essential for cytotoxicity but not hemolytic activity. *Infection and immunity*, 78: 603-610.
- Merz, A. J., So, M. 2000. Interactions of pathogenic neisseriae with epithelial cell membranes. *Annual review of cell and developmental biology*, 16: 423-457.
- Milton, D. L. 2006. Quorum sensing in vibrios: complexity for diversification. *International Journal of Medical Microbiology*, 296: 61-71.
- Minamino, T., Imada, K. 2015. The bacterial flagellar motor and its structural diversity. *Trends in microbiology*, 23: 267-274.
- Miyoshi, S. I., Oh, E. G., Hirata, K., Shinoda, S. 1993. Exocellular toxic factors produced by *Vibrio vulnificus*. *Journal of Toxicology: Toxin Reviews*, 12: 253-288.
- Mizan, M. F. R., Jahid, I. K., Kim, M., Lee, K. H., Kim, T. J., Ha, S. D. 2016. Variability in biofilm formation correlates with hydrophobicity and quorum sensing among *Vibrio parahaemolyticus* isolates from food contact surfaces and the distribution of the genes involved in biofilm formation. *Biofouling*, 32: 497-509.
- Neilands, J. B. 1982. Microbial envelope proteins related to iron. *Annual Reviews in Microbiology*, 36: 285-309.
- Ng, W. L., Bassler, B. L. 2009. Bacterial quorum-sensing network architectures. *Annual review of genetics*, 43: 197-222.
- Nichols, D. S. 2003. Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. *FEMS Microbiology Letters*, 219: 1-7.
- Nottage, A. S., Sinclair, P. D., Birkbeck, T. H. 1989. Role of low-molecular-weight ciliostatic toxins in vibriosis of bivalve mollusks. *Journal of Aquatic Animal Health*, 1: 180-186.
- Novick, R. P., Geisinger, E. 2008. Quorum sensing in staphylococci. *Annual review of genetics*, 42: 541-564.
- Novriadi, R. 2016. Vibriosis in aquaculture. *Omni-Akuatika*, 12: 1-12.
- Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., McLean, J.,

- Lasken, R., Clingenpeel, R. S. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads. In Annual International Conference on Research in Computational Molecular Biology (pp. 158-170). Springer, Berlin, Heidelberg.
- O'Boyle, N., Houeix, B., Kilcoyne, M., Joshi, L., Boyd, A. 201. The MSHA pilus of *Vibrio parahaemolyticus* has lectin functionality and enables TTSS-mediated pathogenicity. International Journal of Medical Microbiology, 303: 563-573.
- Occhino, D. A., Wyckoff, E. E., Henderson, D. P., Wrona, T. J., Payne, S. M. 1998. *Vibrio cholerae* iron transport: haem transport genes are linked to one of two sets of tonB, exbB, exbD genes. Molecular microbiology, 29:1493-1507.
- Oguma, K., Fujinaga, Y., Inoue, K. 1995. Structure and function of *Clostridium botulinum* toxins. Microbiology and immunology, 39: 161-168.
- OIE. 2019. World Organisation for Animal Health; Manual de las Pruebas de Diagnóstico para los Animales Acuáticos. <https://www.oie.int/es/normas/manual-acuatico/acceso-en-linea/>
- Okada, K., Iida, T., Kita-Tsukamoto, K., Honda, T. 2005. *Vibrios* commonly possess two chromosomes. Journal of bacteriology, 187: 752-757.
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A. Vonstein, V. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic acids research, 42: D206-D214.
- Panday, A., Sahoo, M. K., Osorio, D., Batra, S. 2015. NADPH oxidases: an overview from structure to innate immunity-associated pathologies. Cellular molecular immunology, 12: 5-23.
- Paranjpye, R., Hamel, O. S., Stojanovski, A., Liermann, M. 2012. Genetic diversity of clinical and environmental *Vibrio parahaemolyticus* strains from the Pacific Northwest. Appl. Environ. Microbiol. 78: 8631-8638.
- Patrick, M. 2015. Plasmid 101: Brief history of plasmids and an improved eBook (<https://blog.addgene.org/plasmids-101-a-brief-history-of-plasmids-and-an-improved-ebook>) Fecha de consulta: 23/02/2020

- Paul, K., Brunstetter, D., Titen, S., Blair, D. F. 2011. A molecular mechanism of direction switching in the flagellar motor of *Escherichia coli*. Proceedings of the National Academy of Sciences, 108:17171-17176.
- Payne, S. M., Mey, A. R., Wyckoff, E. E. 2016. Vibrio iron transport: evolutionary adaptation to life in multiple environments. Microbiol. Mol. Biol. Rev., 80: 69-90.
- Pham, T. D., Nguyen, T. H., Iwashita, H., Takemura, T., Morita, K., Yamashiro, T. 2018. Comparative analyses of CTX prophage region of *Vibrio cholerae* seventh pandemic wave 1 strains isolated in Asia. Microbiology and immunology, 62: 635-650.
- Pickar-Oliver, A., Gersbach, C. A. 2019. The next generation of CRISPR–Cas technologies and applications. Nature Reviews Molecular Cell Biology, 20: 490-507.
- Plasmids, 2014, Nature Educacion (<https://www.nature.com/scitable/definition/plasmid-plasmids-28/>) Fecha de consulta: 23/02/2020
- Poirel, L., Jayol, A., Nordmann, P. 2017. Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. Clinical microbiology reviews, 30: 557-596.
- Postle, K., Larsen, R. A. 2007. TonB-dependent energy transduction between outer and cytoplasmic membranes. Biometals, 20: 453.
- Price, M. N., Dehal, P. S., Arkin, A. P. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Molecular biology and evolution, 26:1641-1650.
- Quax, T. E., Claassens, N. J., Söll, D., van der Oost, J. 2015. Codon bias as a means to fine-tune gene expression. Molecular cell, 59: 149-161.
- Richter, M., Rosselló-Móra, R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences, 106: 19126-19131.
- Riemann, L., Azam, F. 2002. Widespread N-acetyl-D-glucosamine uptake among pelagic marine bacteria and its ecological implications. Applied and environmental microbiology, 68: 5554-5562.

- Riemann, L., Azam, F. 2002. Widespread N-acetyl-D-glucosamine uptake among pelagic marine bacteria and its ecological implications. *Applied and environmental microbiology*, 68: 5554-5562
- Rocha, E. P. 2008. The organization of the bacterial genome. *Annual review of genetics*, 42: 211-233.
- Rowe-Magnus, D. A., Guérout, A. M., Mazel, D. 1999. Super-integrans. *Research in microbiology*, 150: 641-651.
- Rowe-Magnus, D. A., Guerout, A. M., Ploncard, P., Dychinco, B., Davies, J., Mazel, D. 2001. The evolutionary history of chromosomal super-integrans provides an ancestry for multiresistant integrans. *Proceedings of the National Academy of Sciences*, 98: 652-657.
- Rutherford, S. T., Bassler, B. L. 2012. Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harbor perspectives in medicine*, 2: a012427.
- SAGARPA. 2017. Secretaría de Agricultura, Ganadería, Desarrollo rural, Pesca y Alimentación. Anuario Estadístico de Acuicultura y Pesca 2017. CONAPESCA, México.  
[https://www.conapesca.gob.mx/work/sites/cona/dgppe/2017/ANUARIO\\_ESTADISTICO\\_341\\_2017.pdf](https://www.conapesca.gob.mx/work/sites/cona/dgppe/2017/ANUARIO_ESTADISTICO_341_2017.pdf)
- Salamone, M., Nicosia, A., Bennici, C., Quatrini, P., Catania, V., Mazzola, S. Cuttitta, A. 2015. Comprehensive analysis of a *Vibrio parahaemolyticus* strain extracellular serine protease VpSP37. *PLoS One*, 10: e0126349.
- Salomon, D., Gonzalez, H., Updegraff, B. L., Orth, K. 2013b. *Vibrio parahaemolyticus* type VI secretion system 1 is activated in marine conditions to target bacteria, and is differentially regulated from system 2. *PloS one*, 8: e61086.
- Salomon, D., Guo, Y., Kinch, L. N., Grishin, N. V., Gardner, K. H., Orth, K. 2013a. Effectors of animal and plant pathogens use a common domain to bind host phosphoinositides. *Nature communications*, 4:2973-2982.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd Edition. Vol.1,2 and 3. Cold Spring Harbor Laboratory Press. Pp. 1669.

- Sarkar, M. K., Paul, K., Blair, D. 2010. Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*. Proceedings of the National Academy of Sciences, 107: 9370-9375.
- Schmiel, D.H., V.L. Miller. 1999. Bacterial phospholipases and pathogenesis. *Microbes Infect.* 1:1103–1112.
- Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30:2068-2069.
- Segarra, A., Pépin, J. F., Arzul, I., Morga, B., Faury, N., Renault, T. 2010. Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. *Virus research*, 153: 92-99.
- Sherr, E. B., Sherr, B. F. 2002. Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek*, 81: 293-308.
- Sherr, E., Sherr, B. 2000. Marine microbes: an overview. *Microbial ecology of the oceans*, 6: 13-46.
- Shimohata, T., Takahashi, A. 2010. Diarrhea induced by infection of *Vibrio parahaemolyticus*. *The Journal of Medical Investigation*, 57: 179-182.
- Simmonds, P., Aiewsakun, P. 2018. Virus classification—where do you draw the line?. *Archives of virology*, 163: 2037-2046.
- Snyder, E. E., Kampanya, N., Lu, J., Nordberg, E. K., Karur, H. R., Shukla, M. Zhang, F. 2006. PATRIC: the VBI pathosystems resource integration center. *Nucleic acids research*, 35: D401-D406
- Soto, W., Lostroh, C. P., Nishiguchi, M. K. 2010. Physiological responses to stress in the Vibrionaceae. Springer, Dordrecht. In *Symbioses and Stress*, pp. 407-426
- Soueidan, H., Maurier, F., Groppi, A., Sirand-Pugnet, P., Tardy, F., Citti, C. Nikolski, M. 2013. Finishing bacterial genome assemblies with Mix. In *BMC bioinformatics*, 4: S16-S26.
- Sourjik, V., Wingreen, N. S. 2012. Responding to chemical gradients: bacterial chemotaxis. *Current opinion in cell biology*, 24: 262-268.

- Suginta, W., Vongsuwan, A., Songsiriritthigul, C., Prinz, H., Estibeiro, P., Duncan, R. R., Fothergill-Gilmore, L. A. 2004. An endochitinase A from *Vibrio carchariae*: cloning, expression, mass and sequence analyses, and chitin hydrolysis. *Archives of Biochemistry and Biophysics*, 424: 171-180.
- Suwanto, A. N. T. O. N. I. U. S., Kaplan, S. A. M. U. E. L. 1989. Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4. 1 genome: presence of two unique circular chromosomes. *Journal of Bacteriology*, 171: 5850-5859.
- Svitil, A. L., Chadhain, S. M. N. R., Moore, J. A., Kirchman, D. L. 1997. Chitin Degradation Proteins Produced by the Marine Bacterium *Vibrio harveyi* growing on Different Forms of Chitin. *Applied and Environmental Microbiology*, 63: 408-413.
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L. DeBoy, R. T. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences*, 102: 13950-13955.
- Thomas, C. M., Nielsen, K. M. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews microbiology*, 3: 711-721.
- Thompson C. C., Vicente A. C., Souza R. C., Vasconcelos A. T., Vesth T., Alves N., Jr., Ussery D. W., et al. 2009. Genomic taxonomy of *Vibrios*. *BMC Evol. Biol.* 9:258- 273.
- Thompson, F. L., Iida, T., Swings, J. 2004. Biodiversity of vibrios. *Microbiology and molecular biology reviews*, 68: 403-431.
- Toker, A. S., Macnab, R. M. 1997. Distinct regions of bacterial flagellar switch protein FliM interact with FliG, FliN and CheY. *Journal of molecular biology*, 273: 623-634.
- Trivedi, U. H., Cézard, T., Bridgett, S., Montazam, A., Nichols, J., Blaxter, M., Gharbi, K. 2014. Quality control of next-generation sequencing data without a reference. *Frontiers in genetics*, 5: 111-123.
- Trucksis, M., Michalski, J., Deng, Y. K., and Kaper, J. B. 1998. The *Vibrio cholerae* genome contains two unique circular chromosomes. *Proc. Natl. Acad. Sci. U.S.A.*, 95: 14464–14469.

- Tsang, J. 2017. Bacterial plasmid addiction systems and their implications for antibiotic drug development. *Postdoc journal: a journal of postdoctoral research and postdoctoral affairs*, 5: 3-9.
- Urbanczyk, H., Ast, J. C.,
- Uribe-Querol, E., Rosales, C. 2017. Control of phagocytosis by microbial pathogens. *Frontiers in immunology*, 8: 1368-1390.
- Ussery, D. W. 2014. Chromosome-specific families in *Vibrio* genomes. Name: *Frontiers in Microbiology*, 5: 73.
- van Passel, M. W., Bart, A., Luyf, A. C., van Kampen, A. H., van der Ende, A. 2006. Compositional discordance between prokaryotic plasmids and host chromosomes. *BMC genomics*, 7: 26-33.
- Vernikos, G. 2008. In silico prediction of genomic islands in microbial genomes (Doctoral dissertation, University of Cambridge).
- Vesth, T., Lagesen, K., Acar, Ö., Ussery, D. 2013. CMG-biotools, a free workbench for basic comparative microbial genomics. *PloS one*, 8: e60120.
- Vesth, T., Wassenaar, T. M., Hallin, P. F., Snipen, L., Lagesen, K., Ussery, D. W. 2010. On the origins of a *Vibrio* species. *Microbial ecology*, 59: 1-13.
- Voigt, C.A., W. Schafer, S. Salomon. 2005. A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. *Plant J.* 42:364–375.
- Walker, P. J., Winton, J. R. 2010. Emerging viral diseases of fish and shrimp. *Veterinary research*, 41: 41-51.
- Ward, P. P., Uribe-Luna, S., Conneely, O. M. 2002. Lactoferrin and host defense. *Biochemistry and Cell Biology*, 80: 95-102.
- Washintong State Department,  
<https://www.doh.wa.gov/CommunityandEnvironment/Shellfish/RecreationalShellfish/Illnesses/Vibriosis> Fecha de consulta: 17 de Octubre 2019
- Weimann, A., Mooren, K., Frank, J., Pope, P. B., Bremges, A., McHardy, A. C. 2016. From genomes to phenotypes: TraitAr, the microbial trait analyzer. *MSystems*, 1: 101-16.
- Weinberg, E. D. 2004. Suppression of bacterial biofilm formation by iron limitation. *Medical hypotheses*, 63: 863-865.



- Welch, M., Oosawa, K., Aizawa, S. L., Eisenbach, M. 1993. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proceedings of the National Academy of Sciences*, 90: 8787-8791.
- What is shotgun sequencing? (<https://www.yourgenome.org/facts/what-is-shotgun-sequencing>) Fecha de consulta: 23/02/2020
- Williams, K. P. 2002. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic acids research*, 30: 866-875.
- Williams, P., Cámara, M. 2009. Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. *Current opinion in microbiology*, 12: 182-191.
- Wong, K., Finan, T. M., Golding, B. G. 2002. Dinucleotide compositional analysis of *Sinorhizobium meliloti* using the genome signature: distinguishing chromosomes and plasmids. *Functional integrative genomics*, 2: 274-281.
- Wretling, B., Pavlovskis, O. R. 1983. *Pseudomonas aeruginosa* elastase and its role in pseudomonas infections. *Reviews of infectious diseases*, 5: S998-S1004.
- Wu, H.J., A.H. Wang, M.P. Jennings. 2008. Discovery of virulence factors of pathogenic bacteria. *Curr. Opin. Chem. Biol.* 12:93–101.
- Yu, Y., Yang, H., Li, J., Zhang, P., Wu, B., Zhu, B. Fang, W. (2012). Putative type VI secretion systems of *Vibrio parahaemolyticus* contribute to adhesion to cultured cell monolayers. *Archives of microbiology*, 194: 827-835.
- Zachary, J.F. 2017. Mechanisms of Microbial Infections. En: Zachary, J.F., M.D. McGavin (eds.) *Pathologic Basis of Veterinary Disease*, Ed. Mosby. pp 132-241.
- Zhang, J., Kobert, K., Flouri, T., Stamatakis, A. 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30: 614-620.
- Zhu, J., Miller, M. B., Vance, R. E., Dziejman, M., Bassler, B. L., Mekalanos, J. J. 2002. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proceedings of the National Academy of Sciences*, 99: 3129-3134.
- Zhu, S., Kojima, S., Homma, M. 2013. Structure, gene regulation and environmental response of flagella in *Vibrio*. *Frontiers in microbiology*, 4: 410-418.

Zoued, A., Brunet, Y. R., Durand, E., Aschtgen, M. S., Logger, L., Douzi, B., Cascales, E. 2014. Architecture and assembly of the Type VI secretion system. *Biochimica Biophysica Acta (BBA)-Molecular Cell Research*, 1843: 1664-1673.

## 11. ANEXOS

**Anexo A.** Genomas utilizados en ANI. Genomas de referencia de *Photobacterium damsela* subsp. *damselae*, *Vibrio campbellii*, *V. parahaemolyticus*.

<i>Photobacterium damsela</i> subsp. <i>damselae</i>	
Cepa	ID
P. damsela KC-Na-NB1	GCF_004135105.1
P. damsela KC-Na-1	GCF_002142615.1
P. damsela Phdp Wu-1	GCF_003130755.1
P. damsela subsp. piscicida	GCF_002356235.1
<i>Vibrio campbellii</i>	
Cepa	ID
V. campbellii 170502	GCF_003691485.1
V. campbellii DS40M4	GCF_003312585.1
V. campbellii BoB-53	GCF_002906475.1
V. campbellii BoB-90	GCF_002906455.1
V. campbellii ATCC 25920, CAIM 519T	GCF_002163755.1
V. campbellii LA16-V1	GCF_002142655.1
V. campbellii 20130629003S01	GCF_002140055.1
V. campbellii LMB29	GCF_001969325.1
V. campbellii ATCC BAA-1116	GCF_000464435.1
V. campbellii ATCC BAA-1116; BB120	GCF_000017705.1
<i>Vibrio parahaemolyticus</i>	
Cepa	ID
V.parahaemolyticus.RIMD.2210633	GCF_000196095.1
V.parahaemolyticus.BB22OP	GCF_000328405.1
V.parahaemolyticus.O1:Kuk.str.FDA_R31	GCF_000430405.1
V.parahaemolyticus.O1:K33 str.CDC_K4557	GCF_000430425.1
V.parahaemolyticus.UCM-V493	GCF_000568495.1
V.parahaemolyticus.10329	GCF_001188185.2
V.parahaemolyticus.FORC_008	GCF_001244315.1

**Anexo B.** Uso de codones y aminoácidos, Pangenoma y comparación de proteomas de *Vibrio parahaemolyticus*

Cepa	ID
V.parahaemolyticus.2012AW-0154	GCA_009665495.1
V.parahaemolyticus.10329	GCA_009649015.1
V.parahaemolyticus.160807	GCA_003691525.1
V.parahaemolyticus.20130629002S01	GCA_003085735.1
V.parahaemolyticus. ATCC_17802	GCA_001558495.2
V.parahaemolyticus.BB22OP	GCA_000328405.1
V.parahaemolyticus.CDC_K4557	GCA_000430425.1
V.parahaemolyticus. CHN25	GCA_001700835.1
V.parahaemolyticus. D3112	GCA_004194515.1
V.parahaemolyticus.FDA_R31	GCA_000430405.1
V.parahaemolyticus.FDAARGOS_51	GCA_001188185.2
V.parahaemolyticus.FDAARGOS_191	GCA_002073775.2
V.parahaemolyticus.FDAARGOS_662	GCA_008693745.1
V.parahaemolyticus.FDAARGOS_667	GCA_008693625.1
V.parahaemolyticus.FORC_004	GCA_001433415.1
V.parahaemolyticus.FORC_006	GCA_001304775.1
V.parahaemolyticus.FORC_008	GCA_001244315.1
V.parahaemolyticus.FORC_014	GCA_001636035.1
V.parahaemolyticus.FORC_018	GCA_001636035.1
V.parahaemolyticus.FORC_022	GCA_001879585.1
V.parahaemolyticus.FORC_023	GCA_001758605.1
V.parahaemolyticus.FORC_071	GCA_003612715.1
V.parahaemolyticus.FORC_072	GCA_003612695.1
V.parahaemolyticus.IsolateR13	GCA_003119375.1
V.parahaemolyticus.MAVP-Q	GCA_002209725.2
V.parahaemolyticus.MAVP-R	GCA_002220985.3
V.parahaemolyticus.MVP1	GCA_001996365.2
V.parahaemolyticus.O3:K6.RIMD.2210633	GCA_000196095.1
V.parahaemolyticus.PB1937	GCA_003351885.1
V.parahaemolyticus. R14	GCA_003076895.1
V.parahaemolyticus. S107-1	GCA_003047085.1
V.parahaemolyticus. UCM-V493	GCA_000568495.1
V.parahaemolyticus. Vb0624	GCA_006517795.1
V.parahaemolyticus. VPD14	GCA_004006515.1

**Anexo C.** Factores de virulencia. Genes utilizados como consulta de factores de virulencia reportados para el género *Vibrio* en la base de datos VFDB.

Factor de Virulencia	Gen	Cepa	ID
Adherencia	<i>acfA, acfB, acfC, acfD, tcpA, tcpB, tcpC, tcpD, tcpE, tcpF, tcpH, tcpI, tcpJ, tcpN/toxT, tcpP, tcpQ, tcpR, tcpS, tcpT</i>	<i>V. cholerae</i> O1 biovar El Tor str. N16961 (O1 biovar eltor) chromosome I	NC_002505
	<i>mshA, mshB, mshC, mshD, mshE, mshF, mshG, mshH, mshI, mshJ, mshK, mshL, mshM, mshN, pilA, pilB, pilC, pilD</i>	<i>V. parahaemolyticus</i> RIMD 2210633 chromosome 1	NC_004603
Antifagocitosis	<i>cpsA, cpsB, cpsC, cpsD, cpsE, cpsF, cpsG, cpsH, cpsI, cpsJ</i>	<i>V. parahaemolyticus</i> RIMD 2210633 chromosome 2	NC_004605
	<i>hpl, wbfV/wcvB, wbfY, wbjD/wecB, wbuB, rmlD, wcaJ</i>	<i>V. vulnificus</i> CMCP6 chromosome I	NC_004459
	<i>rmlA, rmlB, rmlC, wbfB, wbfC, wbfT, wbfU, wecA, wza, wzb, wzc</i>	<i>V. fischeri</i> ES114 chromosome I	NC_006840
	<i>wecC</i>	<i>V. cholerae</i> O1 biovar El Tor str. N16961 chromosome I	NC_002505
Quimiotaxis y movilidad	<i>cheA, cheB, cheR, cheV, cheW, cheY, cheZ, filM, flaA, flaB, flaC, flaD, flaE, flaG, flaI, flgA, flgB, flgC, flgD, flgE, flgF, flgG, flgH, flgI, flgJ, flgK, flgL, flgM, flgN, flhA, flhB, flhF,</i>	<i>V. parahaemolyticus</i> RIMD 2210633 chromosome 1	NC_004603

	<i>flhG, fliA, fliD, fliE, fliF, fliG, fliH, fliI, fliJ, fliK, fliL, fliN, fliO, fliP, fliQ, fliR, fliS, flrA, flrB, flrC, motA, motB, motX, motY</i>		
Enzimas	<i>hap/vvp</i>	Vibrio cholerae O1 biovar El Tor str. N16961 (O1 biovar eltor)	NC_002506
	<i>nanH</i>		
Sideróforos	<i>hasR, hutA, hutR, vctC, vctD, vctG, vctP</i>	Vibrio parahaemolyticus RIMD 2210633 chromosome 2	NC_004605
	<i>hasR</i>	Vibrio cholerae O1 biovar El Tor str. N16961 (O1 biovar eltor) chromosome 2	NC_002506
	<i>viuC, viuD, viuG, viuP, vibA, vibB, vibC, vibD, vibE, vibF, vibH, viuA, viuB</i>	Vibrio cholerae O1 biovar El Tor str. N16961 (O1 biovar eltor) chromosome 1	NC_002506
Quorum Sensing	<i>luxS</i>	Vibrio parahaemolyticus RIMD 2210633 chromosome 1	NC_004603
	<i>cqsA</i>	Vibrio parahaemolyticus RIMD 2210633 chromosome 2	NC_004603
	<i>epsN, epsM, epsL, epsK, epsJ, epsI, epsH, epsG, epsF, epsE, gspD, epsC, VopD, VopB, VcrH, VcrV, VcrG, VcrR, VcrD, VscY, VscX, sycN, tyeA, VopN, VscN, YscO, VscP,</i>	Vibrio parahaemolyticus RIMD 2210633 chromosome 1	NC_004603

Sistemas de secreción	<i>VscQ, VscR, VscS, VscT, VscU, VscL, VscK, VscJ, VscI, VscH, VscG, VscF, VscD, VscC, vscB, ExsD, ExsA, VscW, ExsC, VopS, VopR, VopQ</i>		
	<i>VscS2, VscN2, VscC2, VscT2, VscR2, VscQ2, VopB2, VscJ2, OspB, VopA, VopT, VopC, VopL,</i>	Vibrio parahaemolyticus RIMD 2210633 chromosome 2	NC_004603
	<i>hcp-1, vgrG-1,</i>	Vibrio cholerae O1 biovar El Tor str. N16961 (O1 biovar eltor) chromosome 1	NC_002506
	<i>Hcp-2, VgrG-2, VgrG-3, VasA, VasB, VasC, VasD, VasE, VasF, VasH, VasI, VasJ, VasK</i>	Vibrio cholerae O1 biovar El Tor str. N16961 (O1 biovar eltor) chromosome 2	NC_002506
Toxinas	<i>ace</i>	Vibrio fischeri ES114 chromosome II	NC_006841
	<i>ctxA, ctxB</i>	Vibrio cholerae O1 biovar El Tor str. N16961 (O1 biovar eltor) chromosome I	NC_002505
	<i>vvhA</i>	Vibrio vulnificus CMCP6 chromosome 2	NC_004460
	<i>rtxA, rtxC, rtxB, rtxD</i>	Vibrio vulnificus CMCP6 chromosome 2	NC_004460
	<i>TLH</i>	Vibrio parahaemolyticus RIMD 2210633 chromosome 2	NC_004605
	<i>tdh</i>	Vibrio parahaemolyticus RIMD 2210633 chromosome 2	NC_004605
	<i>VCC</i>	Vibrio cholerae O1	NC_002506

		biovar El Tor str. N16961 (O1 biovar eltor) chromosome II	
	<i>zot</i>	Vibrio cholerae O1 biovar El Tor str. N16961 (O1 biovar eltor) chromosome II	NC_002506