



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**

SEÑALES DE ADAPTACIÓN  
MOLECULAR EN LOS FACTORES DE  
VIRULENCIA DE *Bartonella bacilliformis*

TESIS PARA OPTAR EL GRADO DE  
MAESTRA EN BIOQUÍMICA Y BIOLOGÍA  
MOLECULAR

CAMILA MARLEN CASTILLO VILCAHUAMAN

LIMA - PERÚ

2025



**ASESOR:** PABLO TSUKAYAMA CISNEROS, PHD.

**CO-ASESOR:** PEDRO EDUARDO ROMERO CONDORI, DR. RER. NAT.

**JURADO DE TESIS**

DR. JORGE LUIS AREVALO ZELADA

PRESIDENTE

DRA. ISKRA TUERO OCHOA

VOCAL

MG. IRENE ROSA MARIA DELGADO DE LA FLOR MONTAUBAN

SECRETARIO (A)

## **DEDICATORIA.**

A mi familia, quienes apoyaron mi desarrollo académico desde un inicio.  
Y a cualquier estudiante que en algún momento se encontrará perdido en el desarrollo de su proyecto de investigación. Yo también estuve ahí.

## **AGRADECIMIENTOS.**

Al Laboratorio de Genómica Microbiana en UPCH, cuyo maravilloso equipo fue crucial en el desarrollo de este trabajo. Con el juego en la mente.  
En especial a mi co-asesor, Pedro Romero, quien estuvo presente en discusiones nocturnas de este trabajo y quien alentó sutilmente el desarrollo de este proyecto.  
Y por último, al Laboratorio de Ecofisiología Microbiana en Santiago, Chile, en especial a la Dra. Raquel Quatrini, cuyo trabajo inspiró especialmente una sección importante de este manuscrito y cuyos comentarios fueron de gran ayuda.

## **FUENTES DE FINANCIAMIENTO.**

Tesis Autofinanciada

<b>DECLARACIÓN DE AUTOR</b>			
<b>FECHA</b>	<b>21</b>	<b>ENERO</b>	<b>2025</b>
<b>APELLIDOS Y NOMBRES DEL EGRESADO</b>	<b>CASTILLO VILCAHUAMAN CAMILA MARLEN</b>		
<b>PROGRAMA DE POSGRADO</b>	<b>MAESTRÍA EN BIOQUÍMICA Y BIOLOGÍA MOLECULAR</b>		
<b>AÑO DE INICIO DE LOS ESTUDIOS</b>	<b>2020</b>		
<b>TÍTULO DEL TRABAJO DE INVESTIGACIÓN DE GRADO</b>	<b>“Señales de adaptación molecular en los factores de virulencia de <i>Bartonella bacilliformis</i>”</b>		
<b>MODALIDAD DE TRABAJO DE GRADO</b>	<b>Tesis</b>		
<b>Declaración del Autor</b>			
<p>El presente Trabajo de Grado es original y no es el resultado de un trabajo en colaboración con otros, excepto cuando así está citado explícitamente en el texto. No ha sido ni enviado ni sometido a evaluación para la obtención de otro grado o diploma que no sea el presente.</p>			
<b>Teléfono de contacto (fijo / móvil)</b>	<b>986606938</b>		
<b>E-mail</b>	<a href="mailto:castillo.vilcahuaman@gmail.com"><b>castillo.vilcahuaman@gmail.com</b></a>		

Firma del Egresado

DNI 76187685



UNIVERSIDAD PERUANA  
CAYETANO HEREDIA

SEÑALES DE ADAPTACIÓN  
MOLECULAR EN LOS FACTORES DE  
VIRULENCIA DE *Bartonella bacilliformis*

TESIS PARA OPTAR EL GRADO DE  
MAESTRA EN BIOQUÍMICA Y BIOLOGÍA  
MOLECULAR

CAMILA MARLEN CASTILLO VILCAHUAMAN



Informe estándar

Informe en inglés no disponible Más información

11% Similitud estándar

Fuentes

Mostrar las fuentes solapadas

1 Internet

hdl.handle.net

12 bloques de texto 231 palabra que coinciden

2 Internet

journals.plos.org

8 bloques de texto 201 palabra que coinciden

3 Internet

d-nb.info

8 bloques de texto 182 Palabra que coinciden

4 Internet

www.biorxiv.org

6 bloques de texto 144 palabra que coinciden

## ÍNDICE

RESUMEN

ABSTRACT

I. INTRODUCCIÓN .....	13
II. OBJETIVOS.....	16
A. Objetivo general .....	16
B. Objetivos específicos.....	16
III. MARCO TEÓRICO .....	17
A. Historia y actualidad de la enfermedad de Carrion en el Perú.....	17
B. Epidemiología de la enfermedad de Carrion en el Perú .....	19
C. Características de la enfermedad de Carrion .....	20
D. Diagnóstico de la enfermedad de Carrion.....	22
E. Vector de la enfermedad de Carrion .....	23
F. Características de <i>Bartonella bacilliformis</i> .....	24
1) Proceso de infección de <i>Bartonella bacilliformis</i> .....	27
G. Evolución del género <i>Bartonella</i> .....	29
1) Genómica en el estudio de agentes patógenos .....	30
2) Pangenoma en estudios microbiológicos.....	31
3) Genómica de <i>Bartonella bacilliformis</i> .....	33
4) Genes de virulencia asociados a <i>Bartonella bacilliformis</i> .....	35
H. Determinación de la selección positiva en estudios evolutivos .....	39
IV. PLANTEAMIENTO DEL PROBLEMA.....	42
i) Pregunta de investigación.....	43
ii) Justificación.....	43
V. METODOLOGÍA .....	45

A.	Cultivo, extracción de ADN y generación de librerías genómicas ....	45
B.	Ensamblaje e identificación de nuevos genomas de <i>Bartonella</i> .....	45
C.	Genomas de repositorios públicos.....	46
D.	Control de calidad .....	46
E.	Anotación de genomas y construcción del pangenoma .....	47
F.	Filogenia de género y especie.....	49
G.	Genómica comparativa .....	49
H.	Extracción de genes de virulencia .....	50
I.	Alineamiento de genes de virulencia .....	50
J.	Filogenia de genes de virulencia .....	51
K.	Prueba de selección positiva para genes de virulencia.....	52
VI.	RESULTADOS.....	54
A.	Filogenia, pangenómica y virulencia de <i>B. bacilliformis</i> .....	54
B.	Filogenia y pangenómica en el género <i>Bartonella</i> .....	78
C.	Relaciones filogenéticas entre genes de virulencia en el Linaje 1 de <i>Eubartonella</i> .....	86
1.	Hbp (Heme binding protein).....	89
2.	Adherencia.....	91
3.	Motilidad.....	92
D.	Selección positiva en genes de virulencia de <i>B. bacilliformis</i> .....	95
VII.	DISCUSIÓN .....	112
A.	Divergencia de Ver097 del resto de <i>B. bacilliformis</i> .....	112
B.	Los genes de virulencia de la especie <i>B. bacilliformis</i> son parte del genoma <i>core</i> de la especie más no del género <i>Bartonella</i> .....	114
C.	<i>B. ancashensis</i> es una especie muy diferenciada de <i>B. bacilliformis</i> a partir de distintos índices .....	117

<b>D.</b>	<b>Varios genes exclusivos a <i>B. bacilliformis</i> y <i>B. bacilliformis</i> Ver097 están relacionados a funciones de membrana y adherencia.....</b>	<b>118</b>
<b>E.</b>	<b>Existe una presión selectiva sobre genes relacionados a la adherencia y la motilidad .....</b>	<b>120</b>
<b>VIII.</b>	<b>CONCLUSIÓN .....</b>	<b>124</b>
<b>IX.</b>	<b>LIMITACIONES .....</b>	<b>126</b>
<b>X.</b>	<b>RECOMENDACIONES.....</b>	<b>128</b>
<b>XI.</b>	<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>129</b>
<b>XII.</b>	<b>ANEXOS .....</b>	<b></b>

## LISTA DE TABLAS

<b>Tabla 1.</b>	Factores de virulencia de <i>B. bacilliformis</i> según la literatura.	<b>18</b>
<b>Tabla 2.</b>	Características de los ensamblajes de <i>Bartonella</i> obtenidos en el laboratorio.	<b>29</b>
<b>Tabla 3.</b>	Características de los ensamblajes de <i>B. bacilliformis</i> disponibles en GenBank.	<b>24</b>
<b>Tabla 4.</b>	Genes únicamente presentes en las cepas Ver097.	<b>33</b>
<b>Tabla 5.</b>	Genes únicamente presentes en el grupo conformado por UPCH_1469_B06, Cond044, UPCH_1466_B03, UPCH_1474_B14 y UPCH_1472_B10.	<b>34</b>
<b>Tabla 6.</b>	Genes únicamente presentes en el genoma 1416.	<b>38</b>
<b>Tabla 7.</b>	Estadísticas del pangenoma de <i>B. bacilliformis</i> en donde Ver097 se incluye.	<b>41</b>
<b>Tabla 8.</b>	Regiones profágicas encontradas en 4 cepas de <i>B. bacilliformis</i> .	<b>41</b>
<b>Tabla 9.</b>	Características de los ensamblajes de <i>Bartonella</i> disponibles en GenBank.	<b>45</b>
<b>Tabla 10.</b>	Estadísticas del pangenoma del género <i>Bartonella</i> en donde Ver097 se incluye.	<b>50</b>
<b>Tabla 11.</b>	Genes de virulencia hallados en los genomas de las cepas KC584 y Ver097.	<b>50</b>

**Tabla 12.** Valores de prueba de selección positiva. **57**

**Tabla 13.** Resumen de posiciones bajo selección positiva. **63**

## LISTA DE FIGURAS

<b>Figura 1.</b>	Mapa de incidencias acumuladas de casos de bartonelosis desde el 2015 hasta el 2020.	<b>5</b>
<b>Figura 2.</b>	Foto de microscopía electrónica de <i>Bartonella bacilliformis</i> .	<b>9</b>
<b>Figura 3.</b>	Filogenia del género <i>Bartonella</i> según Wagner et. al.	<b>10</b>
<b>Figura 4.</b>	Comparación de genomas de especies del género <i>Bartonella</i> por Segers et. al.	<b>12</b>
<b>Figura 5.</b>	Dominios presentes en los genes Brp identificados en distintas especies de <i>Bartonella</i> spp según O'Rourke et. al.	<b>18</b>
<b>Figura 6.</b>	Esquema del alineamiento realizado por BLASTP y TBLASTX.	<b>29</b>
<b>Figura 7.</b>	Matriz ANI entre las cepas de <i>B. bacilliformis</i> y <i>B. ancashensis</i>	<b>32</b>
<b>Figura 8.</b>	Gráfica tipo UpSet de la presencia y ausencia de genes en el pangenoma de <i>B. bacilliformis</i> .	<b>40</b>
<b>Figura 9.</b>	Categorías COG asignadas en el servidor web eggNOG Mapper para <i>B. bacilliformis</i> KC584 y Ver097.	<b>43</b>
<b>Figura 10.</b>	Árbol filogenético del Linaje 1 del género <i>Bartonella</i>	<b>44</b>
<b>Figura 11.</b>	Árbol del clado Eubartonellae basado en el alineamientos de genes <i>core</i> del género <i>Bartonella</i>	<b>46</b>
<b>Figura 12.</b>	Alineación genómica entre especies del género <i>Bartonella</i> .	<b>47</b>

<b>Figura 13.</b>	Filogenia del género <i>Bartonella</i> junto con Ver097.	<b>48</b>
<b>Figura 14.</b>	Pangenoma del género <i>Bartonella</i> en dónde Ver097 se incluye.	<b>49</b>
<b>Figura 15.</b>	Árbol de genes concatenados pertenecientes a la familia Hbp.	<b>53</b>
<b>Figura 16.</b>	Árbol de genes concatenados pertenecientes a los genes de adherencia.	<b>54</b>
<b>Figura 17.</b>	Árbol de genes concatenados pertenecientes a los genes de motilidad.	<b>55</b>
<b>Figura 18.</b>	Alineamiento de las secuencias aminoacídicas de las proteínas que figuran como bajo selección positiva.	<b>61</b>
<b>Figura 19.</b>	Modelamiento y alineamiento estructural de las 6 proteínas bajo selección positiva	<b>67</b>

## RESUMEN

La enfermedad de Carrión es una enfermedad poco conocida que afecta principalmente a comunidades en los Andes. Esta patología se caracteriza por presentar una fase aguda, conocida como fiebre de La Oroya, y una fase crónica, conocida como verruga peruana. Aunque se transmite principalmente a través de la picadura de artrópodos hembras del género *Lutzomyia*, se sospecha que existen otros vectores. Esta enfermedad es causada por la bacteria gram negativa *Bartonella bacilliformis*. Este patógeno pertenece al género *Bartonella*, que presenta tres grupos diferenciados, siendo *B. bacilliformis* parte del grupo de las Eubartonella, en el linaje L1, que comparte con *B. ancashensis*. Este estudio tiene como objetivo identificar señales de adaptación en la virulencia de *B. bacilliformis*. Se utilizó un conjunto de 39 genomas de *Bartonella* disponibles en el NCBI, así como 13 genomas adicionales de *Bartonella* del Linaje 1 secuenciados en el Laboratorio de Genómica Microbiana UPCH. A partir de estos genomas, se realizaron análisis de genómica comparativa y se extrajeron genes de virulencia de interés que procedieron a ser evaluados mediante análisis filogenéticos y de evolución molecular. Los resultados de genómica comparativa apuntan a variantes dentro de la especie *B. bacilliformis* previamente inexploradas, e indican que los factores de virulencia identificados, que pertenecen al genoma *core* de la especie, podrían presentar adaptaciones moleculares de interés. El género *Bartonella* ha llegado a ocupar diferentes nichos, indicando que las adaptaciones de *B. bacilliformis* serían únicas para sus requerimientos, ya que es la única especie dentro del género que tiene al ser humano como hospedero final. Esto tendría que ser explorado a nivel molecular. Con estos resultados, se espera poder conocer mejor el proceso de

patogenicidad de *B. bacilliformis*, así como elaborar estrategias que tomen en cuenta sus adaptaciones evolutivas.

### **PALABRAS CLAVE**

*Bartonella bacilliformis*, Virulencia, Genómica comparativa, Adaptación molecular

## ABSTRACT

Carrión's disease is a little-known illness that primarily affects communities in the Andes. This disease is characterized by an acute phase, known as La Oroya fever, and a chronic phase, known as *Verruga Peruana*. Although it is primarily transmitted through the bite of female arthropods of the genus *Lutzomyia*, other vectors are suspected. This disease is caused by the gram-negative bacterium *Bartonella bacilliformis*. This pathogen belongs to the genus *Bartonella*, which has three distinct groups. *B. bacilliformis* is part of the Eubartonella group, in the L1 lineage, which it shares with *B. ancashensis*. This study aims to identify adaptive signatures in the virulence of *B. bacilliformis*. A set of 39 *Bartonella* genomes available at NCBI was used, as well as 13 additional *Bartonella* genomes from Lineage 1 sequenced at the Microbial Genomics Laboratory at UPCH. From these genomes, comparative genomics analyses were performed, and virulence genes of interest were extracted and evaluated through phylogenetic and molecular evolutionary analyses. The comparative genomics results point to previously unexplored variants within the *B. bacilliformis* species and indicate that the identified virulence factors, which belong to the core genome of the species, could present molecular adaptations of interest. The genus *Bartonella* has occupied different niches, indicating that the adaptations of *B. bacilliformis* would be unique to its requirements, as it is the only species within the genus that has humans as its final host. This would need to be explored at the molecular level. With these results, it is hoped to better understand the pathogenicity process of *B. bacilliformis*, as well as develop strategies that take its evolutionary adaptations into account.

**KEYWORDS**

*Bartonella bacilliformis*, Virulence, Comparative genomics, Molecular adaptation

## I. INTRODUCCIÓN

La enfermedad de Carrión es una enfermedad infecciosa y olvidada que afecta principalmente a poblaciones en los Andes, siendo endémica en Perú, Ecuador y Colombia. Esta patología se caracteriza por una presentación bifásica, que incluye una fase aguda, conocida como fiebre de La Oroya, y una fase crónica, denominada verruga peruana. La etiología de esta enfermedad está asociada a la bacteria gram negativa *Bartonella bacilliformis*, un patógeno intracelular facultativo perteneciente al género *Bartonella*, el cual se transmite principalmente mediante la picadura de artrópodos hembras del género *Lutzomyia* (1). Aunque la bartonelosis es una enfermedad históricamente registrada en la región desde la época prehispánica, la comprensión de los mecanismos moleculares y evolutivos que subyacen a la virulencia de *B. bacilliformis* aún es escasa.

La evolución del género *Bartonella* se ha caracterizado por múltiples eventos de adquisición de virulencia. La especie más cercana al ancestro común del género es *Bartonella apis*, un simbiote de la abeja melífera, que convive en su microbioma intestinal. El resto de especies son todas patógenas, exhibiéndose dos tipos de eventos de introducción de virulencia en el género: El primero, con la adquisición de virulencia por parte de *B. tamiae*, y el segundo, a partir de la radiación adaptativa que dio origen al grupo *Eubartonellae* y a la aparición de múltiples linajes, dentro de los que está *B. bacilliformis* (2).

*Bartonella bacilliformis*, perteneciente al linaje L1 dentro del grupo *Eubartonellae*, se caracteriza por su capacidad de infectar y sobrevivir dentro de los eritrocitos, lo que le confiere un carácter intracelular facultativo. Esta adaptación a un estilo de vida intracelular es fundamental para su patogenicidad y le permite evadir la respuesta inmunológica del hospedero. Una vez que las bacterias son inoculadas en el hospedero a través de la picadura de los artrópodos vectores, logran penetrar en las células endoteliales de los vasos sanguíneos (3). Dentro de estas células, *B. bacilliformis* se replica, utilizando productos metabólicos conferidos por el eritrocito infectado (característica común en las bacterias intracelulares).

*B. bacilliformis* cuenta con varios tipos de factores de virulencia relacionados a su adherencia, motilidad y unión con el grupo heme. Estos factores son de importancia durante el proceso de infección. Estas categorías contienen factores de virulencia tales como porinas, del aparato flagelar, proteínas de invasión y proteínas de unión al grupo Heme. Estos grupos suelen estar relacionados con la adherencia del patógeno al eritrocito. Otro factor, como GroEL, favorece la proliferación de células vasculares. Sin embargo, son pocos los factores de virulencia que estén propiamente caracterizados (1).

En este estudio, se busca identificar los factores de virulencia de *B. bacilliformis*, comparándolos con los homólogos presentes en otras especies de *Bartonella* mediante métodos de genómica comparativa. Esto permitirá describir mejor aquellos factores de virulencia que son esenciales para su adaptación a su huésped humano. Se espera poder establecer una base sólida para futuros estudios que

podrían explorar en mayor alcance los mecanismos de interacción entre la bacteria y el hospedero. Asimismo, estos hallazgos podrían tener implicancias en el desarrollo de estrategias terapéuticas y de control de brotes de este patógeno.

## **II. OBJETIVOS**

### **A. Objetivo general**

Analizar las señales de selección natural en los factores de virulencia de *B. bacilliformis*, con énfasis en los determinantes moleculares que facilitan la invasión del eritrocito humano, para comprender la evolución de su arsenal patogénico en comparación con otras especies del mismo género.

### **B. Objetivos específicos**

- a) Obtener secuencias de alta calidad de *Bartonella bacilliformis* a partir de nuevos aislados de este patógeno para que puedan estar disponibles en repositorios públicos.
- b) Identificar genes de virulencia en *Bartonella bacilliformis* y contrastar dichos genes identificados con los reportados por la literatura.
- c) Evaluar la presión selectiva en los genes de virulencia relacionados a la interacción con el hospedero humano.

### **III. MARCO TEÓRICO**

#### **A. Historia y actualidad de la enfermedad de Carrión en el Perú**

La enfermedad de Carrión es una enfermedad infecciosa principalmente restringida a los Andes. Esta patogenia ha sido registrada desde los inicios de las civilizaciones asentadas en territorio peruano. Existen registros cerámicos de 2000 años de antigüedad que representan la fase verrucosa de la enfermedad de Carrión y se sabe que poblaciones preincaicas entre los siglos IV y VIII d.C. la padecieron. Se ha descrito incluso una momia perteneciente a un hombre de la cultura Huari que presentó la fase verrucosa de la enfermedad y que aparentemente fue sacrificado con fines rituales (4). Durante el inicio del colonialismo español, se reportaron casos también en 1531 y 1547 que describen verrugas en la piel que duraban de 4 a 5 meses.

En 1868 se firma el contrato para el inicio de la construcción de seis ferrocarriles en el Perú, siendo uno de ellos el ferrocarril que cubriría la ruta desde el Callao hasta La Oroya. La construcción se inició en el mismo año, necesitando alrededor de 10 000 obreros. En 1870, una misteriosa enfermedad atacaría a los obreros de esta construcción, matando a cerca de 7000 de ellos. La enfermedad fue bautizada como fiebre de La Oroya ya que, a pesar de que la ciudad de La Oroya se encontraba a varios kilómetros de distancia, en un inicio se supuso que la tierra y el gravel traídos desde dicha ciudad producían la fiebre (5).

Durante estos tiempos, la aparición de verrugas y la fiebre mostrada en los hechos anteriormente registrados parecían no guardar relación alguna. Se asumió que

ambas patologías eran distintas. La medicina tradicional de la región ya tenía sospechas de que la fase febril y la fase verrucosa podrían ser la misma enfermedad, como fuera registrado en 1764. Sería el 27 de Agosto de 1885 que Daniel Alcides Carrión, un estudiante de medicina peruano, se inoculó secreciones inmediatamente extraídas por rasgadura de una verruga peruana de un paciente. Su intención en un inicio fue reproducir los síntomas de esta verruga. El principal aporte del experimento de Carrión fue la evidencia científica de la teoría unicista, es decir, que ambas enfermedades tenían un mismo origen. Por otra parte, también se demostró que la verruga peruana era una enfermedad inoculable (6). Esto en una época en pleno auge de la idea de microbios como agentes causales de enfermedades (7).

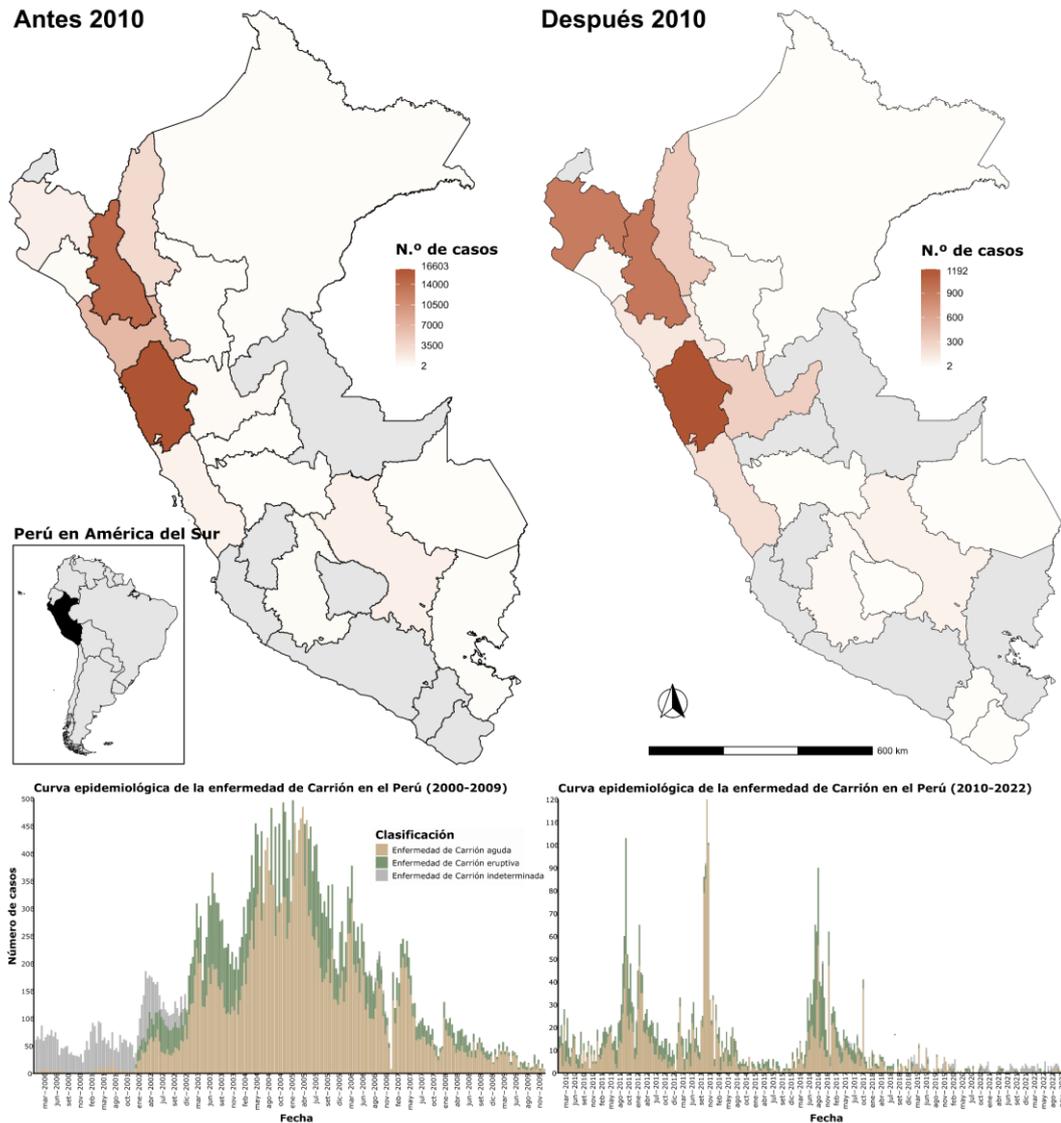
La identificación del agente patógeno de esta enfermedad fue realizada por el microbiólogo peruano Alberto Barton, quien, en 1905 y luego de observar a distintos pacientes que llegaban de La Oroya afectados por la enfermedad de Carrión, describiría unos bacilos diminutos en los glóbulos rojos de dichos pacientes. Al observarse la mejoría de estos pacientes, estos bacilos diminutos se transformaban en cuerpos redondos. Y en la fase verrucosa, se vería que estos corpúsculos globulares desaparecían de la sangre. Alberto Barton describiría estos como cuerpos endoglobulares y sería en el Instituto de Medicina Tropical de Londres (del cual había sido alumno) donde bautizaron a estos como *Bartonella bacilliformis* (8).

## **B. Epidemiología de la enfermedad de Carrión en el Perú**

La enfermedad de Carrión está limitada a áreas específicas dentro del Perú, Ecuador y Colombia. Si consideramos la altura de distintas regiones andinas, podremos observar que esta enfermedad está presente en i) la vertiente occidental de los Andes, entre los 800 y los 3200 m.s.n.m., ii) en los valles interandinos, entre los 800 y los 3200 m.s.n.m. y iii) en la selva alta, entre los 800 y 2400 m.s.n.m. Se considera como zonas endémicas a localizaciones en los departamentos de Piura, La Libertad, Ancash, Lima, Cajamarca, Amazonas, Junín y Huancavelica, siendo la mayor zona endémica el departamento de Ancash, específicamente el callejón de Huaylas y el de Conchucos. Según la Oficina General de Epidemiología del Ministerio de Salud del Perú, este departamento presentó el 71% de los casos entre 1943 y el 2000, porcentaje muy superior al de otros departamentos como Cajamarca (10,4%), Amazonas (6,5%), Lima (3,3%) entre otros.

La última incidencia de una epidemia de gran envergadura sería en el siglo XIX, durante la construcción del ferrocarril hacia La Oroya. Después de este evento, la tasa de incidencia descendió entre 1945 y 1969, de 9,60 a 0,25 por 100 000 habitantes. Entre 1969 y 1974, la tasa se estabilizó, sin mayor crecimiento. Es desde 1974 en donde hay un incremento progresivo. Al año hay alrededor de 500 y 700 casos, y en épocas epidémicas, 1500. El año 2010, se reportaron tan sólo 225 casos y, para el año 2012, se informaron de 445 casos (9). En el año 2017, se reportaron 142 casos de enfermedad de Carrión, teniendo así una incidencia de 0.45 por 100 mil habitantes. Esta cifra fue muy superior a la de años anteriores (23 y 54 casos en los años 2015 y 2016 respectivamente) y a la de años posteriores (41 y 43 casos en

los años 2018 y 2019 respectivamente) (Figura 1). Esto ha representado un pico en el seguimiento epidemiológico de los últimos años (10).



**Figura 1.** Casos acumulados de enfermedad de Carrion por distritos en el Perú desde el 2015 hasta el 2020 (10).

### C. Características de la enfermedad de Carrion

Las infecciones producidas por el género *Bartonella* se conocen como bartonelosis. Varias enfermedades se agrupan bajo este término, entre las que se tienen a la fiebre de trincheras, siendo su agente etiológico *B. quintana* (común en la guerra de

trincheras durante la Primera Guerra Mundial), la fiebre por arañazo de gato y la angiomatosis bacilar, ambas producidas por *B. henselae*, y la enfermedad de Carrión, producida por *B. bacilliformis*. La infección se da mediante un vector, *Lutzomyia verrucarum*, artrópodo que se alimenta de sangre. Solo la hembra es capaz de transmitir la enfermedad. Una vez infectado, el hospedero puede permanecer bacterémico por varios meses (11).

La enfermedad de Carrión tiene carácter bifásico. Ambas fases pueden ocurrir independiente o secuencialmente. Una de las fases, la fiebre de la Oroya, también conocida como la fase aguda, se da luego de la picadura del vector. La patología se caracteriza por una bacteremia aguda de entre 10 a 210 días. Se manifiestan síntomas como dolor de cabeza, fiebre, somnolencia, etc. Sin embargo, el síntoma más característico es una anemia hemolítica severa. En la actualidad, la tasa de mortalidad es del 0.7%, observado en un brote contemporáneo en una población no-endémica. Pero se ha observado también que la fase aguda suele incluir infecciones secundarias y puede ser fatal para el 88% de pacientes si es que no se inicia una terapia antibiótica (11).

La fase eruptiva, más conocida como la verruga peruana o la fase crónica, se caracteriza por la aparición de lesiones nodulares pleomórficas, principalmente en brazos y piernas. Esta aparición está acompañada de síntomas tales como fiebre o dolor osteoarticular. Las erupciones se caracterizan por ser propensas al sangrado. Suelen curarse de forma espontánea y las recurrencias son raras (12).

#### **D. Diagnóstico de la enfermedad de Carrión**

Debido a que es una enfermedad restringida a un área específica en Sudamérica, el desarrollo de técnicas de detección de la enfermedad de Carrión ha sido lento. Dependiendo de la fase en la que se encuentre la enfermedad, se escoge el método adecuado de diagnóstico. Uno común, de bajo costo pero poco sensible es el frotis de sangre periférica, que a través de la tinción consigue la visibilización de las formas bacilares, que es la forma tóxica de la bacteria. Este método es utilizado en la fase aguda. Sin embargo, si bien es el más simple, no es necesariamente el más confiable. Otro método implica el aislamiento de la bacteria a partir de muestras de sangre venosa obtenida mediante una asepsia rigurosa de la zona de punción. No se requiere de medios ni equipamiento sofisticado, pero se debe de tomar en cuenta que el crecimiento de los aislados requiere de medio de cultivo bifásico con sangre y su crecimiento toma entre 2 a 8 semanas. La identificación se realiza mediante PCR, utilizando un fragmento marcador específico o el fragmento 16S rRNA, muy usado en identificación de procariontes (13). La prueba de ELISA también ha sido utilizada para el diagnóstico de *B. bacilliformis*, especialmente en estudios epidemiológicos. Sin embargo, en un trabajo de estandarización, su sensibilidad fue de 68,6%, existiendo también una reacción cruzada con otras bacterias. Por ello, no se recomienda su uso (14).

El primer reporte de un diagnóstico por ensayo de inmunoblot fue reportado en 1988 por Knobloch. Sin embargo, debido al alto costo y a las limitaciones propias de las zonas endémicas de esta enfermedad, se desarrolló un inmunoblot basado en

sonicación, ideal para uso en las regiones endémicas. Este ensayo reportó una sensibilidad del 70% para la fase aguda y 94% para la fase crónica (15).

### **E. Vector de la enfermedad de Carrión**

La transmisión de *B. bacilliformis* se da de tres formas: i) vía la picadura de insectos hembras del género *Lutzomyia*, ii) mediante la transfusión de sangre y iii) de madre a hijo (3). *Lutzomyia verrucarum* es el principal vector de *B. bacilliformis* y *Leishmania peruviana*. Se la conoce coloquialmente como mosca de la arena. Se encuentra ampliamente distribuida en el Perú, específicamente en los Andes peruanos. Se sabe que únicamente la hembra requiere de sangre para la maduración de sus huevos (16). Se han podido identificar diferencias genéticas entre poblaciones de Ancash, Amazonas, Cajamarca, Piura y Lima, con poco movimiento migratorio entre departamentos. Es la propia cordillera la que restringe el movimiento de estas poblaciones, haciendo posible la identificación de estos grupos hasta el nivel de subespecies (17).

*Lutzomyia peruensis* también ha sido descrita como un vector de la enfermedad de Carrión, ya que tiene presencia en las áreas endémicas. Esta especie ha sido reportada en las localidades de Piura, La Libertad, Ancash, Lima, entre otros departamentos. Normalmente se encuentra a alturas entre los 1600 a 3250 m.s.n.m. (18).

Debido a que se han ubicado casos de la enfermedad de Carrión en regiones donde no se encuentran ninguna de estas dos especies, se ha considerado también otras

especies como posibles vectores. Se evaluó, por ejemplo, la posibilidad de que *Lutzomyia maranonensis* fuera también un vector, ya que circula en zonas que *L. verrucarum* y *L. peruensis* no ocupan y, sin embargo, tienen presencia de la enfermedad. Se pudo comprobar la presencia de ADN de *B. bacilliformis* en *L. maranonensis* mediante pruebas PCR, utilizando el marcador 16S rRNA (19).

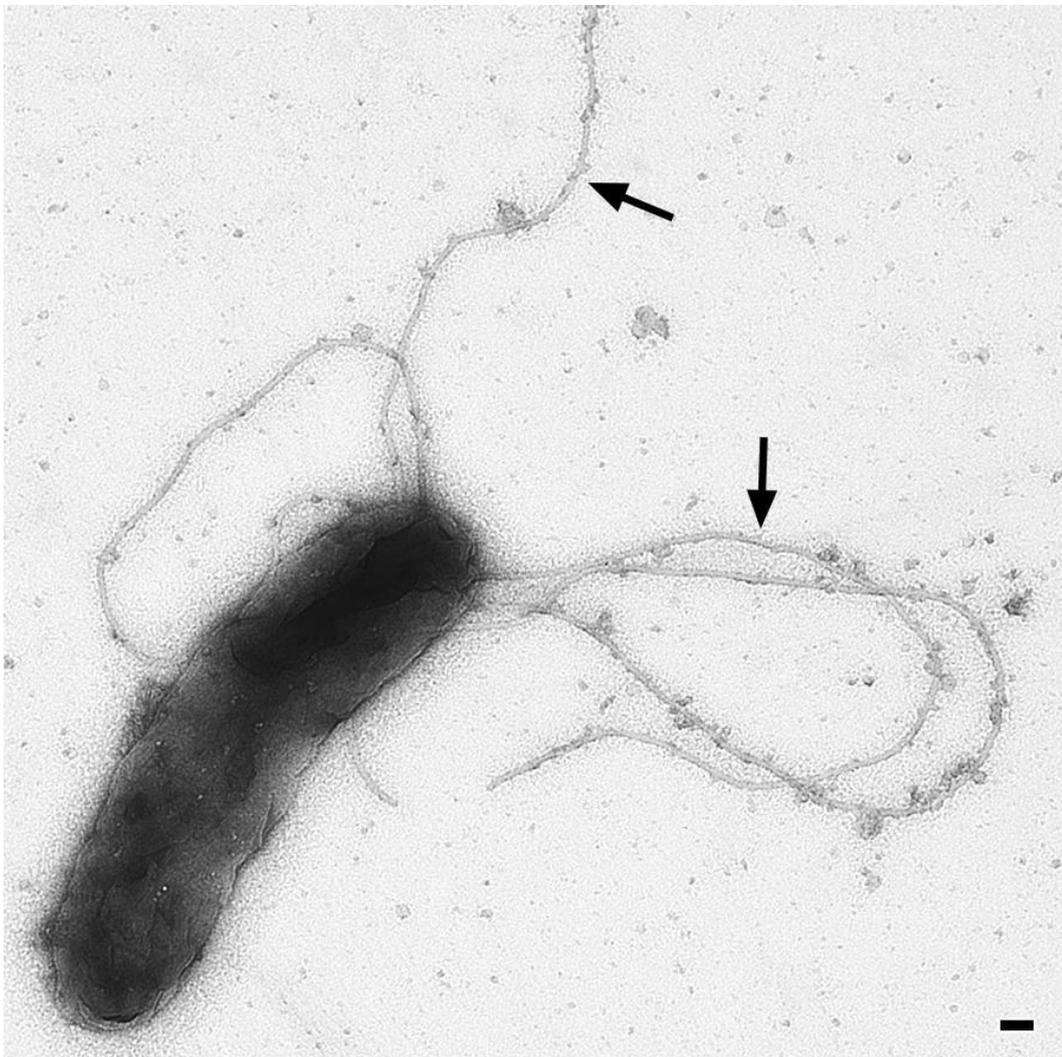
Los mecanismos de interacción entre *B. bacilliformis* y sus vectores aún no se comprenden completamente. Se han realizado experimentos utilizando *B. bacilliformis* que expresan proteína verde-fluorescente (GFP) en un vector competente (*L. verrucarum*) y no competente (*L. longipalpis*), y se ha visto que, en *L. verrucarum*, luego de 72 horas, *B. bacilliformis* es capaz de colonizar la membrana peritrófica en el lumen del intestino medio (20).

Hasta la fecha, no se han identificado reservorios animales confirmados para *B. bacilliformis*. A diferencia de otras especies del género *Bartonella*, que sí cuentan con hospedadores naturales conocidos, se han planteado hipótesis sobre posibles reservorios animales (como *Tapirus terrestris* y *Pecari tajacu*) y vegetales (como especies de euforbias). Sin embargo, la distribución de estos organismos no coincide con las zonas endémicas de la enfermedad y los intentos de detección del patógeno en estas especies no han sido exitosos (21).

## **F. Características de *Bartonella bacilliformis***

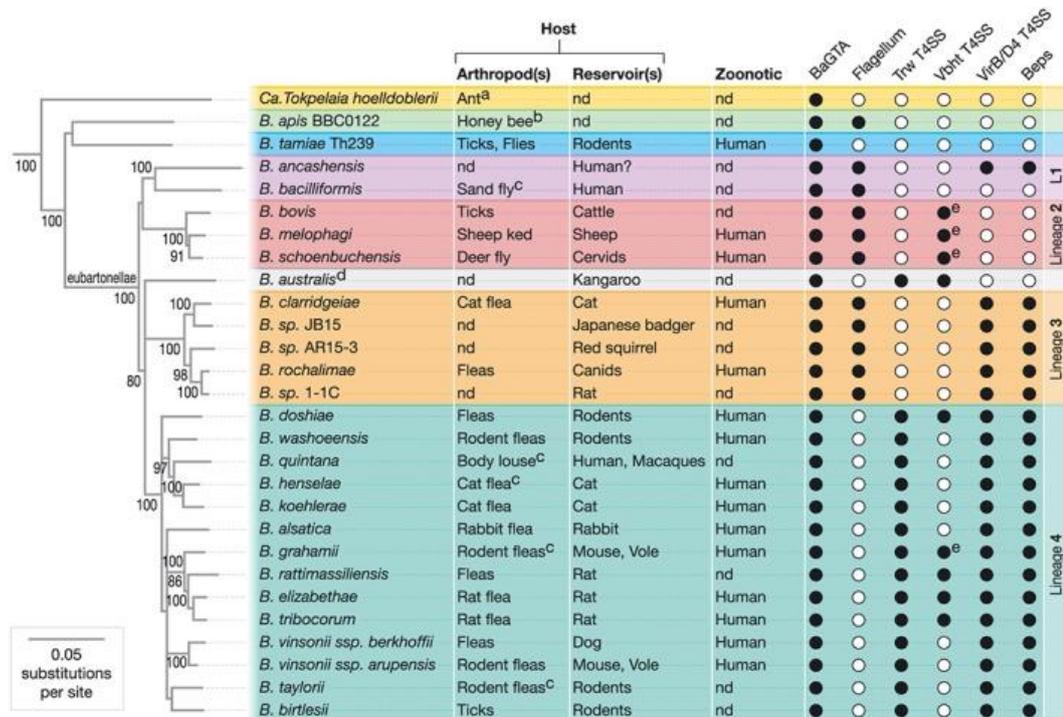
*Bartonella bacilliformis* es una bacteria gram negativa y facultativa intracelular. Pertenece a las alfa-proteobacterias junto a los géneros *Rickettsia*, *Ehrlichia*,

*Brucella* y *Agrobacterium*. Esta bacteria posee de 2 a 16 flagelos en un polo (Figura 2) y su morfología es pleomórfica (puede tomar forma de coco, bacilo o cocobacilo) (22). La cantidad de flagelos es especialmente importante puesto que podría ser un factor importante al momento de colonizar a su hospedero (11).



**Figura 2.** Flagelos de *B. bacilliformis* ATCC35686 en microscopía electrónica realizada por Garcia-Quintanilla et. al. Las flechas señalan la adhesina BadA que parecería tener un rol en el proceso de infección. Barra de escala: 100 nm (21).

El género *Bartonella* puede dividirse en tres clados: i) *Bartonella apis*, ii) *Bartonella tamiae* y iii) Eubartonellae. *B. apis* es una especie simbiote de la abeja y la única especie no patogénica del género. Es la más cercana al ancestro común del género *Bartonella* y parte de su genoma accesorio ha sido obtenido por herencia vertical desde este ancestro. *B. tamiae* comparte características ancestrales pero no tiene varios de los factores de virulencia expuestos por el grupo Eubartonellae, por lo que se asume que su adquisición de patogenicidad es distinta a la vista en dicho grupo (2). Por último, está el grupo de Eubartonellae, que puede dividirse de acuerdo al hospedero de cada una de las especies. Son cuatro los linajes dentro del clado Eubartonellae: L1-L4. L1 está conformado por *B. bacilliformis* y *B. ancashensis*, siendo estas las dos únicas especies que infectan exclusivamente a humanos. El linaje L2 está restringido a rumiantes, mientras que los linajes L3 y L4 infectan otra variedad de hospederos. *B. henselae* y *B. quintana* pertenecen al linaje L4 (Figura 3). *B. bacilliformis* y *B. ancashensis* pertenecen al mismo clado y son las especies que, según la evidencia, infectan humanos exclusivamente (23).



**Figura 3.** Filogenia del género *Bartonella* según Wagner et. al. Se aprecia la separación de *B. apis* y *B. tamiiae* del resto de la taxa. Se aprecia también los cuatro linajes descritos en el grupo eubartonellea. *B. ancashensis* y *B. bacilliformis* pertenecen al linaje L1 (23).

### 1) Proceso de infección de *Bartonella bacilliformis*

Tras su inoculación en el torrente sanguíneo, *Bartonella* spp. permanece indetectable por aproximadamente cuatro días. Al quinto día, *Bartonella* causa aglutinación, iniciando la formación de biofilms en la vía sanguínea. Se desconocen los factores responsables de esta replicación. Durante este periodo, *Bartonella* spp. aún no ingresa a los eritrocitos, permaneciendo en el entorno extracelular y, por ende, siendo susceptible al sistema inmune (24). Se ha comprobado que, como en otras especies de alfa y gamma proteobacteria, *B. bacilliformis* posee moléculas de flagelina que no son reconocidas por TLR5, lo que evita la activación de NF-κB

(25). Se ha reportado también la inusual estructura de los lipopolisacáridos en el género *Bartonella* spp., lo cual también pertenece a un mecanismo de evasión por parte de la bacteria, ya que estas estructuras impiden el reconocimiento por parte de los receptores tipo Toll (TLR) (26).

Para poder iniciar exitosamente la infección, es necesaria la adherencia del patógeno a células del hospedero, en el caso de *Bartonella* spp., los eritrocitos. En *Bartonella tribocorum*, se ha visto el uso del Sistema de Secreción Tipo IV (T4SS), transportadores multicomponentes que codifican para subunidades de pili en la superficie (27). Sin embargo, T4SS no se encuentra presente en *B. bacilliformis*. En estudios filogenéticos, se ha teorizado que la adquisición de este sistema de secreción por parte del Linaje 4 fue por transferencia horizontal de genes, luego de la divergencia con el ancestro en común con *B. bacilliformis*. Cabe resaltar que la ausencia de T4SS parece estar correlacionada con el comportamiento hematófago del artrópodo vector (23).

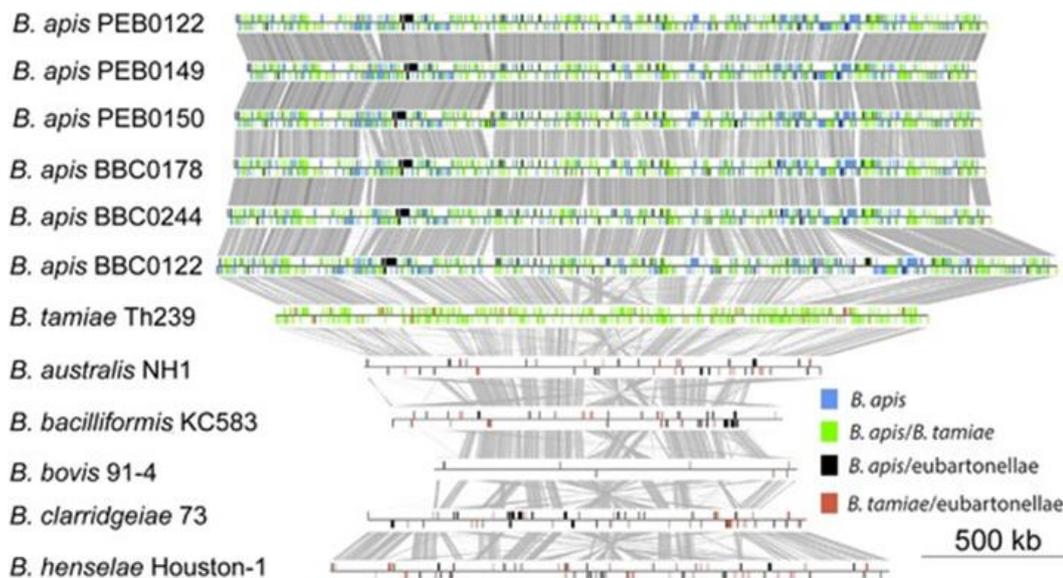
Aunque los mecanismos de entrada de *Bartonella* spp. a los eritrocitos aún no se comprenden completamente, se han identificado algunos factores clave. Uno de ellos es la proteína IalB, que contiene péptidos señal putativos y muestra alta homología con la proteína Ail de *Yersinia enterocolitica*, implicada en la invasión celular. En experimentos con *Escherichia coli* transformada con el gen *ialB*, esta adquirió la capacidad de invadir eritrocitos. Otro factor sería la deformina, que fue encontrada en cultivos de *B. bacilliformis* y *B. henselae*, sugiriendo que este mecanismo podría estar presente en el género *Bartonella*. En infecciones de *B.*

*bacilliformis*, se han observado invaginaciones cónicas, vacuolas internas entre otras alteraciones morfológicas. Por último, se ha encontrado que *B. bacilliformis* posee una hemolisina dependiente de contacto, que podría ser de utilidad al momento de escapar de una vacuola durante la invasión del eritrocito. Luego de entrar al eritrocito, se sabe que *Bartonella* spp. se replica dentro de una vacuola del eritrocito, pero no se conoce acerca de los factores que permiten esta supervivencia en *B. bacilliformis* (24).

### **G. Evolución del género *Bartonella***

La familia más cercana a *Bartonellaceae* es *Brucellaceae*. Sin embargo, incluso entre estas familias hay una distancia filogenética considerable (2). El género *Bartonella* ha tenido una historia evolutiva caracterizada por delecciones. Entre los tres grupos existentes en el género (*Bartonella apis*, *Bartonella tamiae* y Eubartonellae, clado al que pertenece *B. bacilliformis*), se comprobó que *B. apis* y *B. tamiae* son clados cercanos que divergieron en algún punto antes de la aparición del clado Eubartonellae, que es producto de un evento de radiación adaptativa.

En el género *Bartonella* aparecen dos eventos de adquisición de patogenicidad independientes. Uno por el lado de *B. tamiae* y otro por el lado de Eubartonellae. Realizando una comparación genómica de las especies pertenecientes al grupo Eubartonellae, se encontró que estas especies son muy distintas entre sí, siendo que la estructura genómica es menos consistente a lo largo de la historia evolutiva del grupo, mientras que *B. apis* y *B. tamiae*, fuera del grupo Eubartonellae, conservan dicha estructura (Figura 4) (2).



**Figura 4.** Comparación de genomas de especies del género *Bartonella* por Segers et. al. Se aprecia la conservación de la estructura genómica de *B. apis* y *B. tamiae*. Se observa también cómo las especies pertenecientes al grupo Eubartonellae presentan una estructura genómica menos conservada, explicándose con la adquisición y delección observadas en estos genomas (2).

### 1) Genómica en el estudio de agentes patógenos

Desde la aparición de la tecnología de secuenciación, la investigación en enfermedades infecciosas ha utilizado la genómica para entender mejor a los distintos agentes patógenos. Esta tecnología puede ofrecer nuevas perspectivas en lo que se refiere a la ecología, evolución y transmisión de distintos agentes patógenos. En Estados Unidos, la genómica de patógenos es parte importante de casi todas las investigaciones en enfermedades infecciosas, ya sea en infecciones por alimentos, enfermedades como tuberculosis o influenza y enfermedades

parasitarias (28). Este tipo de información puede permitir ver una visión detallada del organismo, su evolución e inferir la función de genes de interés.

La genómica también ha sido vital para el seguimiento temporal y espacial de la resistencia a antibióticos en bacterias, ofreciendo conocimiento necesario para poder ejercer políticas públicas de manera oportuna (29). Desde una perspectiva enfocada en la habilidad de los microorganismos a adaptarse y sobrevivir a diferentes presiones de selección, la genómica ofrece un acercamiento adecuado para una perspectiva multifacética, ya que permite no solo monitorear la presencia de genes de interés, sino también la movilidad, persistencia y abundancia de los mismos (30).

A nivel de identificación de patógenos, la tecnología genómica proporciona también información de relevancia. En el caso de *Neisseria meningitidis*, por ejemplo, se reconoce a este patógeno como una de las bacterias más variables en la naturaleza. Esto debido a altas tasas de transferencia horizontal de genes. Sin embargo, con el surgimiento del secuenciamiento de próxima generación, se habilitó la posibilidad de tener grandes bases de datos de este microorganismo, así como mejores herramientas para identificar la diversidad en esta especie, su evolución y su estructura poblacional (31).

## **2) Pangenoma en estudios microbiológicos**

El año 2005 se introdujo el concepto del pangenoma, el cual se define como todo el repertorio genético de un grupo de estudio. Este puede dividirse en dos tipos de genomas: el genoma *core*, definido como el pool de genes en común de todos los

individuos que componen al grupo, y el genoma accesorio, definido como todos los genes que no están presentes en el genoma core. Los genes accesorios no tienen que ser necesariamente únicos de cada individuo.

Algunos autores clasifican a los pangenomas como abiertos o cerrados, de acuerdo a la capacidad de la especie de adquirir DNA exógeno, de tener la maquinaria necesaria para utilizarlo y poseer una cantidad considerable de rRNA. Estas categorías estarían íntimamente relacionadas al estilo de vida de la especie bacteriana (32). Sin embargo, algunos autores han argumentado en contra de esta clasificación debido a que consideran que es una visión binaria de los pangenomas, y es muy sensible a la cantidad de genomas disponibles al momento (33,34).

A partir del genoma *core* y accesorio, es posible sacar algunas conclusiones relacionadas a la evolución de un grupo taxonómico. Por ejemplo, en el caso del género *Pseudomonas*, se buscó detectar si el genoma accesorio se estructura y se mantiene mediante presión selectiva. Se encontró que 86.7% de los genes accesorios en común entre diferentes individuos está involucrado en una relación coincidente (coinciden a nivel de presencia). A partir de estas relaciones también se puede inferir la forma de adquisición de estos genes (35). Por lo tanto, la construcción del pangenoma puede ayudar en estudios evolutivos de un determinado grupo taxonómico.

Siendo que las definiciones taxonómicas son complejas en el dominio Bacteria debido a la alta variabilidad entre distintos individuos (36), la pangenómica ofrece

una óptica distinta, pudiendo aglomerar distintos individuos bajo una misma categoría taxonómica en función a los genes que comparten entre los mismos. Por ejemplo, en *E. coli*, se descubrió que para el año 2010, todas las cepas depositadas compartían el 40% de su composición genómica. Acordemente, conforme se unen más genomas al análisis, el genoma *core* de *E. coli* sigue siendo el 40% de la composición genómica de la especie (37). La pangenómica, entonces, ofrece una perspectiva compleja de la especie. Sin embargo, esto dependerá también de la consideración que se asuma al momento de definir un genoma *core*, ya que algunos estudios pueden cambiar el umbral inicialmente definido. En un estudio del 2021, se utilizaron más de 100 000 genomas de *E. coli* y *Shigella* para poder construir el pangenoma de la especie. En este caso, se definió el genoma *core* como aquellos genes homólogos compartidos por al menos 97% de los genomas analizados. En esta situación, solo 1.96% de los clusters fueron incluidos en el genoma *core* construido (38).

### 3) Genómica de *Bartonella bacilliformis*

El primer genoma de *Bartonella bacilliformis* fue reportado el año 2006, siendo una secuencia de *Bartonella bacilliformis* KC583, de número de acceso en Genbank: NC\_008783.1. El año 2013, se secuenció el genoma de *B. bacilliformis* cepa INS, aislada en Jaén-Cajamarca, Perú. Para este secuenciamiento se utilizó la tecnología SOLiD 3 Plus genome analyzer, con un *coverage* de 700. Se reportó que este genoma tiene una longitud de 1 444 107 pares de base (bp) con un porcentaje de contenido de GC de 38.3%. Se detectaron 1244 secuencias codificantes (CDSs), 4 rRNAs y 44 tRNAs. Un siguiente genoma sería reportado el año 2020, en donde se secuenció la cepa KC584, utilizando tecnología Illumina y PacBio para una mejor

calidad de contigs. A partir del ensamblaje híbrido usando UniCycler, se obtuvo una secuencia circular de 1 411 580 bp, con un porcentaje de contenido de GC de 38.2% (39).

Estudios comparativos con otras especies de *Bartonella* spp. han mostrado que *B. bacilliformis* tiene un genoma reducido a comparación de otras especies. En un estudio de genes patogénicos, se descubrió que 15 de los genes presentes en otras especies de *Bartonella* spp. estaban ausentes en el genoma de *B. bacilliformis* (40). Por ejemplo, el genoma de *B. bacilliformis* es de 1.45 Mb, mientras que el genoma de *B. tribocorum* es de 2.64 Mb (21).

Hasta 2011, eran escasos los estudios dedicados a explorar la diversidad genética de *B. bacilliformis*. Algunos de estos hicieron uso de técnicas como la comparación de variaciones de secuencia nucleotídica en loci tales como la citrato sintasa (*gltA*) o la región espaciadora intergénica 16S-23S (ISR) (41). Todos estos métodos delinearon genotipos en la especie según sus capacidades. Más recientemente, se discutió la incorporación de la tipificación multilocus de secuencias. Sin embargo, se ha reportado una debilidad en este método de identificación asociada con el número disponible de cepas registradas hasta el momento (42).

La genómica permite identificar nuevas variantes bacterianas y caracterizar factores relacionados con su patogenicidad. En el 2013, por ejemplo, se reporta la candidatura de *Bartonella ancashensis* (43). Su detección requirió secuenciamiento de un solo locus, en donde se detectó divergencia del gen de citrato sintasa con

respecto a *B. bacilliformis*. Múltiples análisis moleculares y microbiológicos concluyeron en una nueva especie de *Bartonella*. Análisis genómicos posteriores indicarían que esta especie es parecida a *B. bacilliformis* y co-circula con esta bacteria en la región de Ancash. *B. ancashensis* presenta flagelos como *B. bacilliformis*, pero además cuenta también con el sistema VirB T4SS, siendo esta una característica solo compartida con otros dos patógenos del género *Bartonella* spp. (44).

#### 4) Genes de virulencia asociados a *Bartonella bacilliformis*

La patogenicidad de un microorganismo es el resultado de la presencia de factores que codifican proteínas asociadas a la enfermedad producida. Es así como los factores de virulencia proporcionan al microorganismo la capacidad de evadir una respuesta inmune, colonizar y/o adherirse a su hospedero. Dentro de *B. bacilliformis*, son 3 las categorías de factores de virulencia indicados por los análisis comparativos realizados (Tabla 1): Factores de adherencia, de motilidad y de unión al grupo Heme. *B. bacilliformis* vendría a ser la única especie dentro del grupo Eubartonellae que no posee un sistema de secreción tipo VirB/D4 tipo IV o Trw tipo IV, como si es que lo poseen otras especies de este linaje o, incluso, *B. ancashensis*, que comparte el linaje L1 con *B. bacilliformis*. El proceso de infección en *B. bacilliformis* depende primariamente de la presencia del aparato flagelar. Durante diversos experimentos, este aparato ha sido removido, y se comprueba que el mutante no-móvil es incapaz de invadir el eritrocito. Aún es poco entendida la invasión mediada por el flagelo, aunque se especula que el movimiento giratorio podría ser clave en el proceso (45).

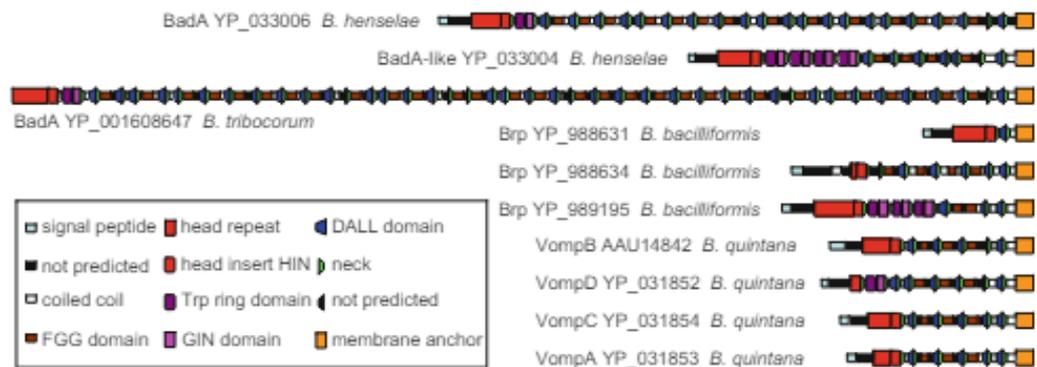
El locus asociado a la invasión (*Invasion associated locus*, IalB) está localizado en la membrana interna. Curiosamente, posee una secuencia secretoria y una secuencia aminoacídica similar a proteínas de la membrana externa involucradas en la virulencia, que darían una primera impresión de ser una proteína de la membrana externa. Los mutantes  $\Delta$ IalB han mostrado una capacidad disminuida para unirse al eritrocito, reduciendo su capacidad de unión entre 47 a 53% relativo al control (46).

Las proteínas de unión al grupo Heme (Hbp) son una familia de proteínas. Todas las *Bartonella* spp. poseen uno o más genes que codifican proteínas pertenecientes a esta familia. Su rol es proveer la proteína Heme requerida para el crecimiento de la bacteria. Diferentes especies de *Bartonella* expresan distintas combinaciones de estas proteínas. Estas combinaciones han sido asociadas al estilo de vida de la especie. HbpA/Pap31 es una proteína de la membrana exterior de *B. bacilliformis*. Es la proteína mejor caracterizada en la especie. En análisis previos de HbpA/Pap31 en *B. bacilliformis* se ha encontrado la presencia de un número variable de 6 repeticiones de aminoácidos en tándems. Previamente se han asociado estas repeticiones en tándems al desarrollo de anticuerpos poco funcionales en infecciones de otros microorganismos, por lo que se especula una asociación de esta proteína con la evasión inmune de *B. bacilliformis* (47).

HbpB y HbpC han sido menos descritos en este patógeno. Sin embargo, hay algunas caracterizaciones realizadas en otras especies. En *Bartonella quintana*, que posee las proteínas HbpA, HbpB, HbpC, HbpD y HbpE, se observa que los genes *hbpB* y *hbpC* son expresados predominantemente en condiciones ambientales con altas

concentraciones de hemina. En la misma especie se observa también que los individuos infectados expresan anticuerpos anti-HbpB y anti-HbpC, pero no HbpA, HbpC o HbpE, especulando que habría un rol de estas proteínas en las evasiones inmunológicas. En *Bartonella henselae*, se ha demostrado que la expresión de HbpC solo incrementa a temperaturas parecidas a las temperaturas corporales de artrópodos, relacionando a esta proteína con la adaptabilidad del patógeno a su vector artrópodo. Se observó también que los cambios en la expresión de *hbpC* afectan los niveles de mRNA de otros genes *hbp*, especialmente *hbpA* (48). Si bien no son observaciones directamente hechas en *B. bacilliformis*, se podría especular que estas funciones de adaptabilidad para la toma de las proteínas Heme tendría un rol similar en la especie, que tiene un vector artrópodo también.

Las adhesinas también juegan un rol como factores de virulencia en el género *Bartonella*. Son complejos proteicos con la habilidad de unirse con componentes específicos moleculares del hospedero. Dentro de estas adhesinas, se encuentra la familia de autotransportadores triméricos, o TAA, presentes en  $\alpha$ ,  $\beta$  y  $\gamma$ -proteobacteria. Las TAAs tienen una estructura de paleta trimérica, con un ancla a la membrana, un cuello y una cabeza (Figura 5). Todas las TAAs se caracterizan por su dominio ancla C-terminal. Dentro del género *Bartonella*, las TAA mejor caracterizadas son las *BadA* en *B. quintana*, caracterizadas por una considerable longitud (49).



**Figura 5.** Dominios presentes en los genes Brp identificados en distintas especies de *Bartonella* spp según O'Rourke et. al. (49).

Otros factores de virulencia han sido citados a lo largo de la literatura de *B. bacilliformis*. La data al respecto es escasa. Un estudio señala, por ejemplo, la presencia de una deformina extracelular como factor importante en la invasión del eritrocito (1,45). Sin embargo, si bien se ha reportado su aislamiento, su secuenciamiento, intentado en el año 1994, no pudo ser posible. Los autores lo atribuyeron a la baja cantidad obtenida, que hizo el secuenciamiento imposible (50). En el año 2004 se reportaría que la deformina es una molécula pequeña de 1400 Da (51). Asimismo, se cita la proteína conservada GroEL como un factor que favorece la proliferación vascular de la célula (1). En algunas bases de datos, se ha observado que esta proteína no aparece como factor de virulencia asociado a *B. bartonella*, si bien se caracteriza en otras especies bacterianas.

**Tabla 1.** Factores de virulencia descritos en *B. bacilliformis* en la literatura. Se citan los factores, su localización y su rol durante el proceso patogénico.

<b>Factores de virulencia</b>	<b>Localización</b>	<b>Función en patogénesis</b>
Adhesinas tipo BadA/Vomp	Membrana, con ancla en la membrana.	Unión a componentes moleculares específicos del hospedero.
IalB	Membrana interna	Capacidad de unión al eritrocito.
Flagelo	Membrana	Capacidad mecánica en la invasión del eritrocito
Proteínas de unión al grupo Heme	Membrana exterior	Adquisición de Heme para el crecimiento de la bacteria
Deformina	Extracelular	Invasión del eritrocito
GroEL	Factor de secreción	Proliferación vascular de la célula

#### **H. Determinación de la selección positiva en estudios evolutivos**

Entender los roles de las fuerzas evolutivas en el curso de la evolución es un problema fundamental dentro de la biología. Estas fuerzas se definen como mecanismos que conducen el cambio en la frecuencia de alelos entre las poblaciones en el tiempo. La mutación, la deriva génica, el flujo génico y la selección natural (52). Esta última fue la propuesta por Darwin y Wallace en 1858, y establece que ciertos rasgos en una población se vuelven menos o más comunes

de acuerdo a la supervivencia y heredabilidad de los mismos. Varios tipos de selección se agrupan en esta categoría, tales como la selección negativa o purificadora y la selección positiva.

La selección requiere de dos características para ser considerada como tal: El rasgo debe de ser beneficioso para la supervivencia y reproducción del organismo, y debe de ser heredable, de modo tal que el organismo pueda pasar este mismo rasgo a su descendencia (53). En estudios evolutivos, detectar la selección positiva es de gran importancia. Con este fin, se suele recurrir a tests estadísticos que usan data de divergencia entre especies, algunos incluso añadiendo el polimorfismo a la ecuación (54). En estos tests estadísticos contra una hipótesis nula son comunes en este tipo de estudios, en dónde se requieren asumir varios factores para simplificar las matemáticas involucradas (55).

Muchos modelos han ido siendo creados a lo largo del tiempo, siendo uno de ellos el test de rama-sitio (*branch-site*). Su objetivo es detectar la selección en ramas pre especificadas en un árbol, analizando cómo esta selección afecta algunos codones de un gen codificante. La selección positiva se define de la siguiente manera (Eq. 1):

$$\omega = \frac{d_n}{d_s} \text{ (Eq. 1)}$$

En dónde:

$\omega$ : Fuerza selectiva

$d_n$ : Tasa de mutaciones no-sinónimas

$d_s$ : Tasa de mutaciones sinónimas

Se considera un resultado positivo cuando  $\omega > 1$ , mientras que  $\omega \leq 1$  niega la posibilidad de un evento de selección positiva, pero no la posibilidad de que otro tipo de selección pueda estar actuando en un determinado gen (56).

Que una región de un gen esté bajo selección positiva indica la posibilidad de la selección natural de un rasgo fenotípico o funcional siendo beneficiado. Para ilustrar este caso, se puede citar por ejemplo el trabajo realizado por Nandi et. al., en dónde se evaluaron todos los genes de *Burkholderia pseudomallei* distribuidos en 2 cromosomas. Los autores reportaron la presencia de genes bajo selección positiva en factores relacionados con la virulencia de este patógeno tales como algunos genes involucrados en la formación del pili tipo IV (TFP4). Para poder comprobar la influencia de esta selección funcional, se generaron mutantes sin el locus de TFP4. Se observó que estos mutantes TFP4 tuvieron una virulencia reducida a comparación del *wildtype*, concluyendo que no es coincidencia que se exhiba selección positiva en factores que contribuyen al proceso patogénico de la bacteria en cuestión (57).

En el contexto de este trabajo, las señales moleculares se definen como patrones distintivos en la variación secuencial del ADN que pueden detectarse haciendo uso de estadística (58). Estas señales reflejan mutaciones fijadas en la población en genes codificadores de proteínas, y como resultado, establecer cambios funcionales en las proteínas finales (59).

#### IV. PLANTEAMIENTO DEL PROBLEMA

*Bartonella bacilliformis* presenta una característica evolutiva notable: comparte un ancestro común con *Bartonella apis*, una bacteria perteneciente a la microbiota de las abejas. Esta relación plantea interrogantes respecto a la evolución de la patogenicidad en el género *Bartonella*, particularmente en cómo ocurrió la introducción de la patogenicidad en su historia evolutiva, además de cómo *B. bacilliformis* y el Linaje 1 se diferenciaron del resto de especies al tener al humano como hospedero final. Esto plantea un caso de estudio común en salud pública, en donde hay un interés en el proceso que conlleva a bacterias a adquirir patogenicidad con la finalidad de prevenir eventos similares en el futuro (60). Es en este proceso de *spillover* en donde estudiar la evolución del patógeno a su hospedero mediante sus adaptaciones moleculares es de importancia (61).

La genómica comparativa ofrece herramientas para explorar esta transición. Mediante esta metodología, es posible determinar diferencias entre individuos del mismo taxón, recopilando diferencias entre estas, además de determinar regiones con mayor dinámica evolutiva, lo que podría contribuir con potenciales aplicaciones prácticas en el manejo clínico de las infecciones.

Sin embargo, la investigación en *Bartonella bacilliformis* se ha centrado principalmente en sus impactos en la salud pública y epidemiología, aspectos que están bien documentados en la literatura científica local. Exploraciones genómicas y evolutivas específicas para *B. bacilliformis* que planteen cómo se adaptan a su hospedero final, particularmente durante la invasión de eritrocitos, no han sido llevadas a cabo previamente. Esto debido a que por mucho tiempo los estudios

genómicos de la especie se limitaron por la poca cantidad de genomas disponibles. Algunos de estos trabajos de análisis bioinformático fueron realizados de manera descriptiva, pero no necesariamente desde una perspectiva evolutiva (47,62). El trabajo realizado por Paul et. al. tomó una perspectiva desde la evolución, en dónde se enfocó en las mutaciones como principal conductor de la diversificación en la especie (91). Sin embargo, el estudio realizó una comparación únicamente a nivel de la especie y no a nivel del género.

### **i) Pregunta de investigación**

¿Qué cambios moleculares permitieron a *Bartonella bacilliformis* adaptarse específicamente a la invasión de eritrocitos humanos, en contraste con otras especies del género *Bartonella* que no presentan el mismo hospedero?

### **ii) Justificación**

*Bartonella bacilliformis* es el agente causal de la enfermedad de Carrión o bartonelosis humana, una enfermedad endémica en regiones andinas del Perú. Es por esto que es clasificada como una enfermedad desatendida y poco explorada con enfoques biológicos más modernos como el secuenciamiento de próxima generación. Este estudio aborda una brecha en el conocimiento de esta bacteria al investigar los mecanismos moleculares y evolutivos mediante los cuales un microorganismo no patógeno puede transformarse en un patógeno humano. El enfoque presentado integra análisis genómico comparativo y evolutivo, y busca comprender de esta manera el rol de los factores de virulencia en el ciclo de vida de *B. bacilliformis*.

Gracias a la disponibilidad de datos genómicos en repositorios públicos, además de un mejor acceso a tecnología de secuenciación genómica, es posible obtener una mayor cantidad de genomas del patógeno en cuestión. Esto permitirá llenar un vacío en el conocimiento de cómo la presión selectiva influye en la patogenicidad de *B. bacilliformis*, contribuyendo así con el entendimiento de los procesos que conducen a la emergencia de patógenos humanos, con potenciales aplicaciones en la prevención y control de enfermedades infecciosas.

## **V. METODOLOGÍA**

### **A. Cultivo, extracción de ADN y generación de librerías genómicas**

Los aislados de *Bartonella* se cultivaron en medios bifásicos, cuya preparación consta de una fase sólida y una líquida. En primer lugar, la fase sólida se preparó mezclando 10 gr. de bacto agar, 2,5 gr. de cloruro sódico, 0,5 gr. de dextrosa, 10 gr. de peptona proteosa con 500 mL de agua destilada. Una vez preparado, los cultivos fueron incubados a 37°C durante 24 horas. Tras este proceso, se cogió el matraz con el medio sólido y se añadieron 15 mL de la fase líquida (RPMI con 10% de SFB). A continuación, se realizó un segundo proceso de esterilidad durante una semana y se incubó el medio bifásico a 28°C. Una vez finalizado el proceso, los matraces se almacenaron en refrigeración a 4°C.

Se extrajo ADN genómico de los aislados utilizando el kit de purificación de ADN GeneJET (Thermo Fisher Scientific, EE.UU) siguiendo las instrucciones del fabricante. El gADN extraído se cuantificó con un fluorómetro Qubit 4.0, y se construyeron bibliotecas genómicas utilizando el kit de bibliotecas genómicas Nextera XT (Illumina, EE.UU). Las bibliotecas se secuenciaron en la plataforma MiSeq de Illumina, generando lecturas de 250 pb por pares con una cobertura media de 80X.

### **B. Ensamblaje e identificación de nuevos genomas de *Bartonella***

Se obtuvieron secuencias para los 13 aislados. Se procedió a realizar un ensamblaje utilizando SPAdes ver. 3.15.5 (63) y se realizó la identificación de estos genomas

utilizando el repositorio MLST ver. 2.22.1 (64). Se identificó 1 cepa como *B. ancashensis*, mientras que 12 fueron identificadas como *B. bacilliformis*.

### **C. Genomas de repositorios públicos**

Se descargaron secuencias de todas las especies conocidas de *Bartonella* disponibles en la base de datos abiertos Genome del National Center for Biotechnology Information (NCBI), con el objetivo de obtener una amplia representación del género. Posteriormente, se pasó a la descarga de todas las cepas disponibles de *Bartonella bacilliformis*, la especie de interés en este estudio. Se realizó la descarga de genomas ya ensamblados. Este proceso de recopilación de secuencias permitió obtener un conjunto de datos completo y diverso para realizar análisis genéticos y comparativos en el estudio de la virulencia dentro del género *Bartonella*. 17 genomas fueron obtenidos para *B. bacilliformis*, mientras que 39 genomas fueron obtenidos de todas las especies de *Bartonella* descritas hasta el momento.

### **D. Control de calidad**

Una vez completada la descarga de secuencias genómicas, se procedió a realizar una evaluación exhaustiva de todos los genomas obtenidos utilizando la herramienta QUAST (65). QUAST es una herramienta ampliamente utilizada en bioinformática para evaluar la calidad de los genomas ensamblados y realizar comparaciones entre ellos. Mediante esta evaluación, se obtuvieron métricas clave como la longitud total del genoma, el número de contigs, la cobertura promedio y otros indicadores que permitieron una caracterización precisa de los genomas descargados. Debido a que los genomas fueron de especies distintas con longitudes

de genomas distintas, no se utilizó una única métrica, y se consideraron factores tales como el grado de ensamblaje y el número de contigs (si es que aplicaba). Se prefirió el mayor grado de ensamblaje posible y el menor número de contigs en los ensamblados escogidos.

### **E. Anotación de genomas y construcción del pangenoma**

Los genomas descargados fueron anotados utilizando la herramienta BAKTA (66). Esta herramienta presenta la ventaja de contar con distintas bases de datos incorporadas. Posteriormente, se llevaron a cabo dos análisis pangenómicos: uno considerando todos los genomas disponibles de la especie *Bartonella bacilliformis* y otro enfocado en el género *Bartonella*.

En el primer caso, se utilizaron todos los genomas de cepas de *Bartonella bacilliformis* disponibles. En este análisis, se aplicó un criterio más estricto utilizando una identidad de BLASTP de 90 debido a que la comparación es intraespecie. Este enfoque permitió realizar un análisis pangenómico específico para la especie *Bartonella bacilliformis*.

En el segundo caso, se seleccionaron los genomas más completos de cada especie de *Bartonella* obtenidas, utilizando una identidad de BLASTP del 70 como criterio de selección debido a que se trataban de especies distintas (35). Estos genomas fueron utilizados en el análisis pangenómico del género *Bartonella*.

Para llevar a cabo la construcción del pangenoma, se empleó la herramienta Panaroo (67), que facilita la construcción de pangenomas a partir de genomas bacterianos. Como resultado de estos análisis, se obtuvieron tablas de presencia y ausencia de genes, que indican qué genes están presentes o ausentes en cada genoma evaluado. Además, se generó un alineamiento de los genes core, que son aquellos genes compartidos por la mayoría de los genomas y que proporcionan información sobre las relaciones evolutivas dentro del grupo a comparar.

Para graficar el pangenoma, se utilizó un gráfico de tipo UpSet, que permite una mejor representación de conjuntos complejos que interseccionan entre sí. Esto se hizo mediante una librería pyUpSet v. 0.1 (68) implementada mediante scripts realizados localmente que permitieron la conversión rápida de tablas para su posterior graficación.

Algunas regiones en ciertos genomas fueron clasificadas como profágicas durante el proceso de anotación. Para evaluar mejor el origen y clasificación de estas regiones, se usó la plataforma web phastest.ca (69), que realiza una identificación y anotación rápida de secuencias profágicas a partir de genomas bacterianos. Los genomas con regiones profágicas relevantes fueron comparados con el genoma de la cepa de referencia, KC584. Se utilizó la opción *lite*, que utiliza de referencia la base de datos Swissprot (70). El resto de parámetros fueron estándar. Asimismo, se compararon qué proteínas son únicas para ciertos grupos de interés. Se extrajeron del *output* de Panaroo y fueron caracterizadas manualmente utilizando la base de datos InterPro (71), que incorpora a varias bases de datos de proteínas. En caso no

podiera ser encontrada información en base a esta base de datos, se utilizó UniProt (72)

Se realizó, asimismo, una comparación de categorías COG en dos genomas de interés (KC583 y Ver097), para obtener un panorama general del tipo de genes que estos aislados poseen. Para ello se utilizó el servidor web EggNOG Mapper (73) y, a partir de los resultados obtenidos, se realizaron gráficas de barras circulares comparando los datos de ambos genomas.

#### **F. Filogenia de género y especie**

Se realizaron dos tipos de filogenias: Una del género *Bartonella* y otra del linaje L1 del género *Bartonella* (*B. bacilliformis* y *B. ancashensis*). Para ambos casos, se utilizó un *output* producido por Panaroo: un alineamiento de los genes core. Para obtener el archivo filogenético tipo Newick, se utilizó IQ-TREE (74). Una ventaja de este software es que evalúa como parte de su flujo de trabajo el modelo más apto para la filogenia a realizar. La edición del árbol se realizó en la herramienta web iTOL (75).

#### **G. Genómica comparativa**

Para comparar y evaluar las similitudes genómicas entre los aislados del patógeno y las cepas de referencia se utilizó el Índice de Identidad de Alineación (ANI, por sus siglas en inglés). El ANI se calculó utilizando el paquete pyani (76). Este compara las secuencias nucleotídicas de los genomas completos para determinar el porcentaje de identidad promedio entre las secuencias comparadas. Este enfoque proporciona una medida cuantitativa de la similitud genómica y permite una

comparación precisa entre las distintas cepas. Para los análisis realizados en este estudio, se utilizó la opción de alineamiento MAFFT.

## **H. Extracción de genes de virulencia**

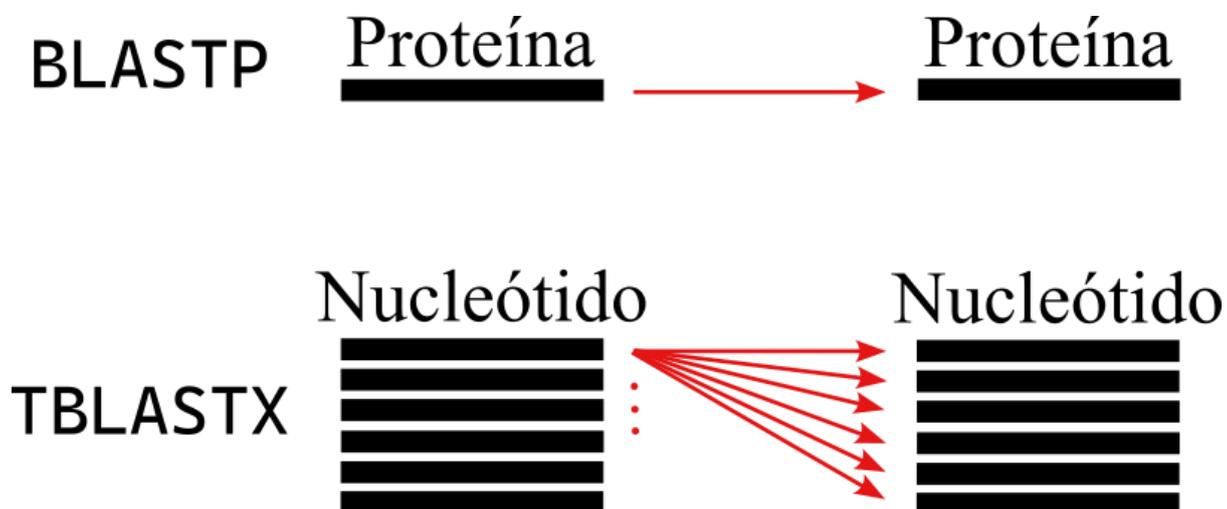
La detección de genes de virulencia se realizó en la plataforma ofrecida por VFDB online (77) en su versión más reciente, en dónde se pueden subir genomas en versión completa o draft, y se obtiene una lista de los genes de virulencia encontrados y registrados en VFDB. Estos resultados fueron luego también comparados con lo obtenido en la anotación realizada en BAKTA. Se escogieron sólo aquellos genes que figuraron en ambos outputs. Dos genes citados en la literatura no aparecieron en esta evaluación. Se decidió utilizar únicamente aquellos genes que pudieran ser evaluados por las herramientas bioinformáticas utilizadas.

## **I. Alineamiento de genes de virulencia**

Para detectar los cambios en las secuencias de aminoácidos, se llevaron a cabo alineamientos múltiples que permitieron identificar adiciones, deleciones y diferencias entre las secuencias obtenidas de todas las cepas de *Bartonella*. Los alineamientos se llevaron a cabo utilizando el algoritmo MAFFT basado en aminoácidos, haciendo uso de un script de Perl doméstico: translatorx.pl. A continuación, estos fueron visualizados gráficamente utilizando el paquete de R ggmsa (78). Esto con el fin de visualizar las variaciones presentes en las secuencias proteicas y su implicancia en la diversidad del género *Bartonella*.

## J. Filogenia de genes de virulencia

Se realizaron alineamientos de las secuencias aminoacídicas correspondientes utilizando el algoritmo MAFFT. Estos alineamientos se sometieron luego a un análisis filogenético utilizando IQ TREE (74). Para detectar diferencias entre los operones bacterianos en diferentes especies, se utilizó TBLASTX (Figura 6), que permite la comparación entre los 6 marcos de lecturas posibles de un genoma con una base de datos compuesta de todos los marcos de lectura posibles. Esto obtuvo como resultado una sintenia, que fue graficada utilizando el paquete de R gggenomes (79).



**Figura 6.** (Arriba) Alineamiento realizado por el algoritmo de BLASTP, en dónde se alinea una secuencia proteica contra otra secuencia proteica. (Abajo) Alineamiento realizado por el algoritmo de TBLASTX, en dónde se traducen las secuencias de nucleótidos en los 6 marcos de lectura posibles para ser comparados contra otra secuencia nucleotídica traducida a los 6 marcos de lectura posibles.

## **K. Prueba de selección positiva para genes de virulencia**

Para detectar la influencia de selección positiva en los genes de virulencia analizados, se utilizó el paquete PAML (80). Este paquete emplea métodos estadísticos y modelos evolutivos para identificar regiones que podrían estar bajo selección positiva en secuencias nucleotídicas de genes.

En el caso de los genes de adherencia y unión al grupo Heme, se detectaron secuencias duplicadas y/o posibles pseudogenes, que fueron agrupados como parte de estos grupos de virulencia. Para poder determinar cuál secuencia utilizar y cuál descartar, se realizó un filtro utilizando Orthofinder (81), paquete especializado en la detección de genes ortólogos. Ello permitió agrupar los genes de todas las especies de acuerdo a ortogrupos. Se procedió a descartar duplicados que pertenecieran al mismo ortogrupo, y se trató en lo posible de someter a la comparación a todos los ortogrupos posibles detectados por la herramienta.

El proceso de evaluación de selección positiva comenzó con la alineación de las secuencias nucleotídicas de genes de virulencia obtenidas previamente mediante el algoritmo MAFFT el cuál fue ejecutado mediante el script translatorx.pl. Este permite realizar un alineamiento en base a codones aplicado a la secuencia genética. Luego, se aplicó Gblocks v. 0.91b (82) en un loop de bash. Gblocks se ejecutó con sus comandos en default, especificando que el alineamiento era de codones. Este paquete permitió obtener un alineamiento limpio sin gaps que pudieran ser aplicados en PAML.

PAML calcula estadísticas de selección positiva, como las razones de tasas de sustitución no sinónimas a sinónimas (dN/dS), y realiza pruebas de hipótesis para determinar si estas estadísticas son significativamente diferentes. Usando un código escrito en Bash para este estudio, se utilizó el modelo rama-nodo y se aplicaron 2 hipótesis en los genes de virulencia, diferenciada por los sitios marcados en el árbol para cada una. Se aplicó también 2 hipótesis nulas para estas pruebas. Con el paquete, se obtuvieron distintos outputs, entre ellos los archivos *mlc*, que contienen el LnL de la tasa de sustitución no sinónima a sinónima ( $\omega$ ). Estos LnL fueron comparados en una tabla aplicando la prueba de Chi cuadrado ( $\chi^2$ ), obteniendo así la significancia de cada hipótesis para cada gen. Aquellos genes que tuvieran una significancia de  $p=0.05$  serían luego evaluados usando el output de PAML, para ubicar la región exacta que estaría siendo sometida a esta presión evolutiva.

Para determinar la influencia de los sitios bajo influencia de selección positiva, se realizó un alineamiento estructural usando ColabFold (83), una implementación de AlphaFold 2 junto con MMSeqs en un Google Colab de libre acceso. Se editaron las estructuras resultantes utilizando PyMol (84). Por último, para poder determinar alguna diferencia entre estructuras, se realizó una alineación estructural con este mismo paquete.

## VI. RESULTADOS

### A. Filogenia, pangenómica y virulencia de *B. bacilliformis*

Se obtuvieron 12 genomas nuevos de *Bartonella bacilliformis* (Bb) y 1 genoma nuevo de *Bartonella ancashensis* para este estudio (Tabla 2). Asimismo, estos genomas se complementaron con 17 genomas que se encuentran en repositorios públicos (Tabla 3). En el caso del genoma de referencia, KC583, este fue secuenciado a partir de la cepa ATCC 35686. Su versión más actualizada es KC584. Para el análisis de la especie Bb en este trabajo, se decidió utilizar ambos genomas a modo de poder obtener la mayor cantidad de información posible. Sin embargo, a lo largo del estudio se tomó a KC584 como el genoma de referencia.

**Tabla 2.** Estadísticas básicas de los ensamblajes obtenidos en el laboratorio. Se identifican 12 genomas como *B. bacilliformis* y 1 genoma como *B. ancashensis*.

	Nombre	Nº de pares de base	Nivel de ensamblaje	Nº de contigs
<i>B. bacilliformis</i>				
1	UPCH_1473_B12	1 367 912	Draft	19
2	UPCH_1460_A18	1 367 325	Draft	14
3	UPCH_1461_A19	1 394 834	Draft	17
4	UPCH_1464_B01	1 367 771	Draft	16
5	UPCH_1456_A02	1 367 846	Draft	17

6	UPCH_1469_B06	1 437 717	Draft	18
7	UPCH_1472_B10	1 430 882	Draft	15
8	UPCH_1466_B03	1 437 432	Draft	15
9	UPCH_1474_B14	1 434 768	Draft	19
10	UPCH_1477_B17	1 400 678	Draft	16
11	UPCH_1461_A19	1 394 834	Draft	17
12	UPCH_1478_B19	1 368 156	Draft	19
<i>B. ancashensis</i>				
1	UPCH_1458_A07	1 442 729	Draft	34

**Tabla 3.** Estadísticas básicas de los genomas de la especie *Bartonella bacilliformis* disponibles en GenBank. A partir de aquí se realizó la selección de genomas. El año representa el año en que fueron subidas a la base de datos.

N° de acceso	Nombre del organismo	Cepa	Tamaño kb	Año
GCA_001559035.1	<i>B. bacilliformis</i>	ATCC:35685D-5	1 443 404	2016
GCA_001624625.1	<i>B. bacilliformis</i>	USM-LMMB 07	1 405 613	2016
GCA_001625425.1	<i>B. bacilliformis</i>	USM-LMMB 06	1 401 011	2016
GCA_000015445.1	<i>B. bacilliformis</i>	KC583	1 445 021	2007

GCA_000311905.1	<i>B. bacilliformis</i>	INS	1 444 107	2012
GCA_000607205.1	<i>B. bacilliformis</i>	San Pedro 600-02	1 403 817	2014
GCA_000607225.1	<i>B. bacilliformis</i>	Heidi Mejía	1 406 538	2014
GCA_000607245.1	<i>B. bacilliformis</i>	Perú-18	1 390 379	2014
GCA_000709735.1	<i>B. bacilliformis</i>	Hosp800-02	1 429 421	2014
GCA_000709755.1	<i>B. bacilliformis</i>	CUSCO5	1 401 647	2014
GCA_000709775.1	<i>B. bacilliformis</i>	Cond044	1 439 147	2014
GCA_000709795.1	<i>B. bacilliformis</i>	Ver097	1 394 706	2014
GCA_000709815.1	<i>B. bacilliformis</i>	Perú38	1 421 778	2014
GCA_000709835.1	<i>B. bacilliformis</i>	VAB9028	1 413 459	2014
GCA_000709855.1	<i>B. bacilliformis</i>	Ver075	1 413 194	2014
GCA_000709875.1	<i>B. bacilliformis</i>	CAR600-02	1 417 443	2014
GCA_009498695.1	<i>B. bacilliformis</i>	KC584	1 411 655	2019

Los nuevos genomas de Bb presentaron un tamaño aproximado de 1.4 megabases de acuerdo a lo esperado. De todos los genomas obtenidos entre los nuevos secuenciados y los provenientes de repositorios públicos, el más pequeño corresponde a la cepa Peru-18 (pb = 1390379), mientras que el genoma de mayor tamaño es el correspondiente a la cepa KC583 (pb = 1445021). Al analizar los

valores de *Average Nucleotide Identity* (ANI, por sus siglas en inglés) entre todos los genomas obtenidos además de otras especies del mismo género (*B. henselae*, *B. schoenbuchensis*, *B. melophagi* y *B. quintana*), se observó que se formaron 3 clusters distintos (Figura 7): El primero vendría a ser el clado que junta a las especies que no son Bb, en dónde se encuentran *B. ancashensis*, UPCH\_1458\_A07, además de *B. henselae*, *B. schoenbuchensis*, *B. melophagi* y *B. quintana*. En la Figura 6 este clado se distingue por poseer un color distinto. Esto se debe al bajo valor ANI que obtuvieron las comparaciones de estas especies versus los genomas de Bb. Es importante resaltar que, entre *B. ancashensis* y UPCH\_1458\_A07 se observa un ANI superior a 0.99, confirmando que UPCH\_1458\_A07 pertenece a la especie *B. ancashensis* (Tabla S1).

Hay un segundo clado observable en la matriz ANI referente a los genomas UPCH\_1456\_A13, UPCH\_1459\_A13, UPCH\_1473\_B12, UPCH\_1478\_B19, UPCH\_1464\_B01, UPCH\_1460\_A18 y Ver097. Entre estas cepas el índice ANI es de 0.99, pero, comparando cada una de estas cepas versus el resto de genomas de Bb, se aprecia que el índice ANI es más bien de 0.94. Esto revela una diferenciación relevante a nivel de secuencia entre este cluster y el resto de genomas de Bb.

Por último, el tercer cluster corresponde al resto de genomas de Bb. Entre estos genomas, el valor suele oscilar entre 0.98 y 0.99, pero se observa la formación de grupos dentro de este cluster también, representados por un color más intenso en la matriz de ANI. Esto ocurre entre los genomas UPCH\_1469\_B06, Cond044,



**Figura 7.** Matriz ANI entre las cepas de *B. bacilliformis* y *B. ancashensis*, siendo que la escala va desde color verde (valores superiores a 0.90) hacia color morado (valores inferiores a 0.9).

En el análisis del pangenoma de Bb, se encontró que Ver097 tiene menos genes en comparación con el resto de las cepas analizadas. Esta diferencia en el número de genes puede indicar una reducción en el tamaño del genoma de Ver097 en relación con otras cepas.

En cuanto a la composición del pangenoma, se identificó que la mayoría de los genes pertenecen a la categoría de genes core. Esto significa que estos genes están presentes en la mayoría de las cepas analizadas, lo que indica una alta conservación en términos de secuencia y función.

Encontramos que 1027 genes coinciden entre todos los aislados analizados, según la Figura 8. Asimismo, encontramos que 51 genes están presentes en todos los genomas exceptuando los que pertenecen al previamente denominado cluster Ver097. De manera interesante, este cluster contiene 57 genes que pertenecen exclusivamente al mismo, y que no se hallan en el resto de genomas evaluados (Tabla 4).

**Tabla 4.** Genes únicamente presentes en el cluster Ver097 (n=56). Se presentan únicamente las proteínas únicas dentro de las que pudieron ser identificadas (n=21). Los genes restantes (n=35) fueron anotados como proteínas hipotéticas. Las

funciones encontradas fueron halladas en la base de datos InterPro a menos de que se señale lo contrario. En algunos casos, ninguna función relevante descrita ha sido hallada.

Nombre de proteína	Función	InterPro/Ref
Sintasa de Lisilfosfatidilglicerol TM	Esta proteína participa en la síntesis de lisilfosfatidilglicerol (Lys-PG), un componente crucial en la membrana bacteriana, y que le da su carga positiva.	IPR022791
Helicasa Dependiente de ATP	Las helicasas dependientes de ATP desenrollan ADN o ARN de doble hebra utilizando energía de la hidrólisis de ATP. Son cruciales para la replicación y reparación.	Múltiples proteínas tales como: IPR004589, IPR014152
SecD	SecD es parte del sistema Sec translocasa en bacterias, esencial para la translocación de proteínas a través de membranas.	IPR005791
Proteína Integral de Membrana Familia TerC	Las proteínas de la familia TerC están involucradas en la resistencia al telurio y podría estar involucrada en la expulsión de iones de telurio,	IPR005496

	contribuyendo a la resistencia antibiótica.	
Proteína con Dominio BTP	El dominio BTP tiene un pliegue parecido a la histona, y se predice que se une al ADN.	IPR006565
Deshidrogenasa L-lactato [Citocromo]	NA	NA
Proteína con Dominio DUF1561	En <i>Leptospira</i> se han relacionado proteínas que contienen este dominio con una capacidad de modificación de la virulencia.	(85)
Proteína con Dominio Asparaginasa	Las asparaginasa hidrolizan la asparagina a aspartato y amoníaco, desempeñando roles importantes en el metabolismo del nitrógeno. Sin embargo, mayor información de esta proteína no ha sido encontrada.	NA
Proteína con Dominio SH3b	Los dominios SH3b están asociados a dominios de unión a ligandos. Aparecen en varias bacterias y fagos. En particular, estos dominios son conocidos por unirse a motivos en la pared celular.	(86)

L-serina Amonioliasa	Este enzima cataliza la conversión de L-serina a piruvato y amoníaco.	IPR004642
Proteína con Dominio DsbC	DsbC está involucrada en el plegamiento de proteínas y la formación de enlaces disulfuro en bacterias.	IPR018950
Proteína de Recombinación de ADN RmuC	Aún no se conoce específicamente el rol de RmuC pero parece estar involucrada en la protección del ADN contra nucleasas o parece estar involucrada en la escisión de estructuras secundarias del ADN.	IPR003798
Proteína con Dominio Autotransportador	Los autotransportadores facilitan el transporte de proteínas a través de membranas bacterianas. Respecto a esta proteína en particular, no hay suficiente caracterización.	NA

Otro de los cluster previamente identificados en la evaluación ANI presenta también genes únicos a comparación del resto de genomas: UPCH\_1469\_B06, Cond044, UPCH\_1466\_B03 y UPCH\_1474\_B14. De manera interesante, el aislado UPCH\_1472\_B10, el cuál no figura en el cluster mostrado por ANI, comparte varios genes con este grupo (Tabla 5). Por último, un genoma único a comparación del resto de genomas evaluados fue el genoma 1416, el cuál presenta 11 genes

únicos a comparación del repositorio genético de la especie (Tabla 6). Otros genomas han presentado 1 o 2 genes únicos a comparación del resto, pero no han sido incluidos en el gráfico por ser un número bajo de incidencias.

**Tabla 5.** (Arriba) Genes únicamente presentes en el grupo conformado por UPCH\_1469\_B06, Cond044, UPCH\_1466\_B03, UPCH\_1474\_B14 y UPCH\_1472\_B10 (n=42). (Abajo) Genes únicamente presentes en el grupo conformado por UPCH\_1469\_B06, Cond044, UPCH\_1466\_B03 y UPCH\_1474\_B14 que no se hallaron en UPCH\_1472\_B10. Se presentan únicamente los que pudieron ser identificados (n=35). Los genes restantes (n=16) fueron anotados como proteínas hipotéticas. Las funciones encontradas fueron halladas en la base de datos InterPro a menos de que se señale lo contrario. En algunos casos, ninguna función relevante descrita ha sido hallada.

Proteína	Función	InterPro/Ref.
UPCH_1469_B06, Cond044, UPCH_1466_B03, UPCH_1474_B14 y UPCH_1472_B10		
Proteína de cola X	Es una proteína de cola de fago. Aparentemente tiene una función de unión al peptidoglicano.	IPR008861
Proteína que contiene dominio helix-turn-helix	Este dominio es común en muchas proteínas, particularmente aquellas involucradas en la transcripción y unión al ADN.	IPR018060

Proteína que contiene dominio DUF2730	En algunas bacterias está relacionada al bacteriófago Mu.	A0A009ZC78A 0A022PD31 (UniProt)
Proteína que contiene dominio DUF2190	En algunas bacterias, está relacionada con un bacteriófago.	A0A0N7ACR5 (UniProt)
Proteína relacionada con fagos	Las proteínas relacionadas con fagos a menudo incluyen varios componentes estructurales y enzimáticos cruciales para el ciclo de vida del fago.	NA
Proteína de ensamblaje de la baseplate	Esta proteína está involucrada en el ensamblaje de la estructura de la baseplate de los bacteriófagos, lo que es crucial para su capacidad de inyectar material genético en las células huésped.	IPR013046
Proteína de morfogénesis del virion del fago	Estas proteínas son esenciales para el ensamblaje de nuevos viriones durante el ciclo de vida del fago, contribuyendo a la integridad estructural y funcionalidad del fago.	IPR006522

Proteína de cola del fago	Las proteínas de la cola son críticas para la unión de los bacteriófagos a las células huésped e inyectando su material genético en dichas células.	NA
Proteína que contiene dominio DUF3486	NA	NA
Proteína que se une al peptidoglicano	Estas proteínas interactúan con el peptidoglicano, un componente principal de las paredes celulares bacterianas.	IPR002477
Proteína de ensamblaje de la baseplate del fago V	Esta proteína es parte del complejo de ensamblaje de la baseplate en los bacteriófagos, crucial para su capacidad de unirse y penetrar las células bacterianas.	IPR013046
Proteína reguladora GemA	GemA está involucrada en la regulación de varios procesos dentro de los bacteriófagos, incluyendo la expresión génica durante la infección.	A0A1C6ZDJ0 (UniProt)

Proteína que contiene dominio DUF3164	NA	NA
Proteína de la familia Baseplate J/gp47	Caracterizada en el bacteriófago P2.	IPR006949
Proteína de partición de la familia ParB/RepB/Spo0J	Se encuentran en proteínas relacionadas a la partición de plásmidos y cromosomas.	IPR004437
Proteína de la cápside	Las proteínas de la cápside forman la envoltura protectora alrededor del material genético viral en fagos.	NA
Proteína prophage I tipo Mu	Asociadas con bacteriófagos.	NA
Proteína del sistema de inyección contractil VgrG/Pvc8	Forman parte de los sistemas de inyección contractiles usados por algunos bacteriófagos para la infección de sus huéspedes.	IPR045362
ATPasa de la familia AAA	Se encuentran en todos los organismos y su función es variada, pero están más que todo relacionadas a la regulación celular, procesos de proteólisis y transporte intracelular.	IPR003959

Proteína menor de la cabeza del fago	Las proteínas menores de la cabeza contribuyen al ensamblaje para formar una cápside icosaédrica.	P19727 (UniProt)
Proteína que contiene dominio Subtilisina en la cola del fago	Una proteína de la cola de un fago.	IPR035089
Proteína que contiene dominio DUF935	Hipotética proteína de bacteriófago.	B3G130 (UniProt)
Proteína que contiene dominio C-terminal de la transposasa Mu	Este dominio está asociado con la actividad de la transposasa.	IPR009004
Proteína reguladora de la longitud de la cola del fago	Esta proteína regula la longitud de la cola durante el ensamblaje del fago.	IPR010090
Proteína que contiene dominio DUF4815	De función desconocida pero ha sido encontrada en bacterias y virus.	IPR032096
UPCH_1469_B06, Cond044, UPCH_1466_B03 y UPCH_1474_B14		
Subunidad de la familia Translocasa Twin-Arginina TatA/TatE	Esta proteína es parte del sistema de translocasa Tat, que es responsable de la translocación de proteínas plegadas a través de la membrana interna en las	IPR006312

	bacterias. Las subunidades TatA y TatE desempeñan papeles cruciales en la formación del complejo translocón.	
Antitoxina de Muerte Celular Programada PemI	PemI es una antitoxina que contrarresta los efectos de las toxinas producidas durante la muerte celular programada en las bacterias. Está involucrada en sistemas de mantenimiento de plásmidos.	IPR007159
Proteína de Unión de Sustrato del Transportador ABC	Las proteínas de unión de sustrato (SBP) son componentes críticos de los transportadores ABC, facilitando la captación de alta afinidad de varios sustratos, incluyendo aminoácidos y azúcares. Determinan la especificidad del sustrato.	IPR023920
Interferasa de mRNA	Las interferasas de mRNA son enzimas que degradan moléculas de mRNA, desempeñando un	IPR031451 IPR012933

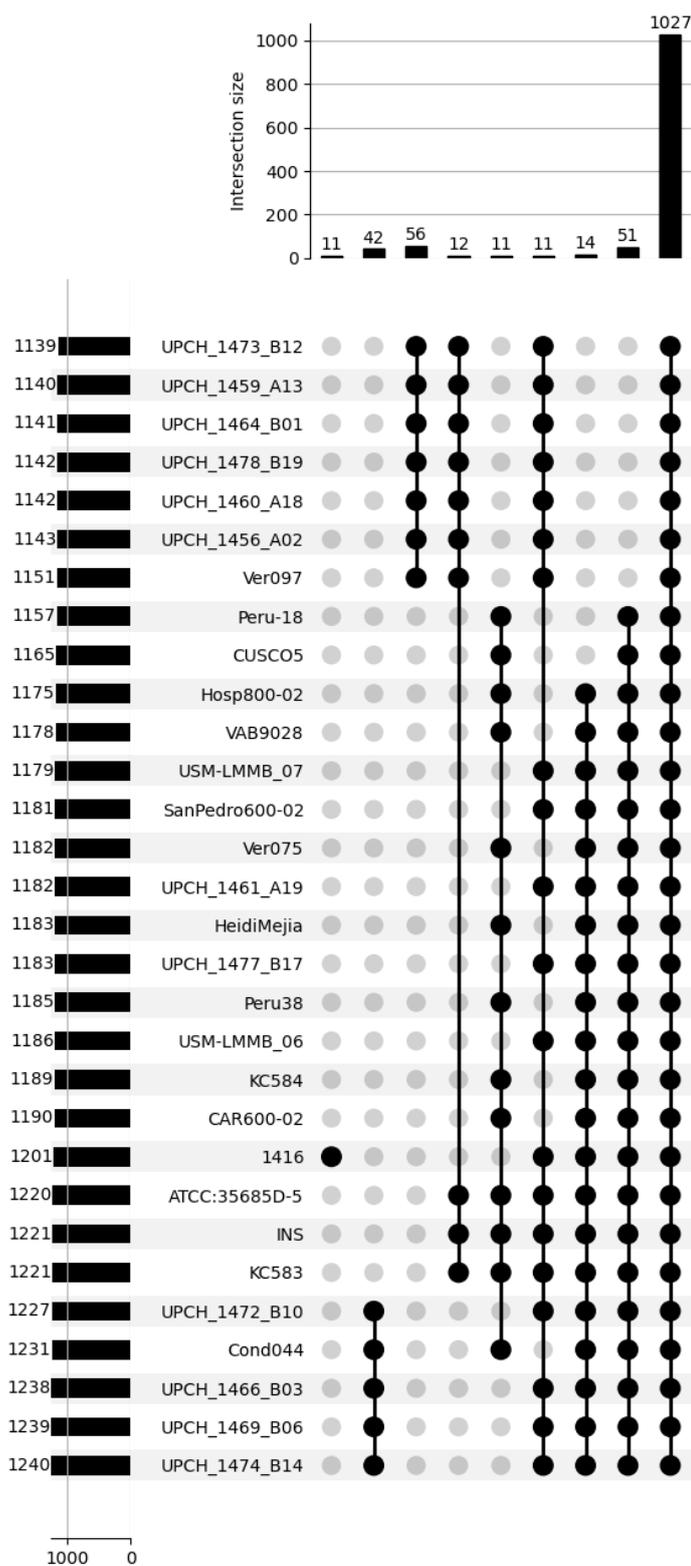
	papel en la regulación de la expresión génica.	
Proteína similar a TrbI de la conjugación bacteriana	Las proteínas similares a TrbI están involucradas en la conjugación bacteriana, facilitando la transferencia de material genético entre bacterias mediante contacto directo. Pertenecientes a los sistemas tipo IV.	IPR005498

Por otro lado, se observó que los genomas UPCH\_1469\_B06, Cond044, UPCH\_1466\_B03 y UPCH\_1474\_B14 presentaron un número mayor de regiones codificantes en comparación con el resto de genomas. Según la matriz de presencia y ausencia obtenida de la reconstrucción pangénómica, UPCH\_1474\_B14 es el genoma de Bb con mayor número de regiones codificantes, con 1240 regiones de este tipo (Figura 8). Los 3 genomas que le siguen pertenecen todos al grupo previamente mencionado.

**Tabla 6.** Genes únicamente presentes en el genoma 1416 (n=11). Se presentan únicamente los que pudieron ser identificados (n=9). Los genes restantes (n=2) fueron anotados como proteínas hipotéticas.

Proteína	Función	InterPro/Ref.
Proteína con Dominio DUF1561	En <i>Leptospira</i> se han relacionado proteínas que contienen este dominio con una capacidad de modificación de la virulencia.	(85)
Proteína Ribosomal 50S L7/L12-Serina Acetiltransferasa	NA	A0A482PEM3 UniProt
Helicasa de ADN Dependiente de ATP RecG	RecG es una helicasa dependiente de ATP que juega un papel crítico en la reparación y replicación del ADN. Ayuda a procesar uniones Holliday y es un intermedio hacia productos más maduros.	IPR004609
Glutamina--Fructosa-6-Fosfato Transaminasa (Isomerizante)	Esta enzima cataliza la formación de glucosamin-6-fosfato.	IPR005855
Enzima Bifuncional UDP-N-Acetilglucosamina Difosforilasa/Glucosamina-1-Fosfato N-Acetiltransferasa GlmU	GlmU es una enzima trimérica bifuncional que cataliza las 2 últimas reacciones en secuencia de la vía de UDP-GlcNAc.	IPR005882
Proteína de Biogénesis de Citocromo C	Esta proteína podría estar involucrada en la guía de los apocitocromos hacia su	IPR003567

	unión covalente realizada por la liaza citocromo-c-haem.	
Proteína Accesorio para el Ensamblaje de Clústeres Hierro-Azufre	Los clusters de Hierro-Azufre son cofactores importantes para varias proteínas involucradas en la transferencia de electrones, y catálisis de tipo redox y no-redox.	IPR011339
Trifosfohidrolasa de Deoxiguanosina Trifosfato	Algunos miembros de esta familia actúan como factores de restricción del hospedero, defendiendo al organismo contra infecciones virales. Otras están involucradas en la reparación del ADN.	IPR050135
Biotina Carboxilasa	La biotina carboxilasa cataliza la carboxilación de enzimas dependientes de biotina, desempeñando un papel en la síntesis de ácidos grasos y la gluconeogénesis.	IPR011764



**Figura 8.** Representación gráfica que muestra conjuntos de genes y su intersección entre los genomas. Consiste en una matriz binaria donde cada conjunto se

representa como una columna y las intersecciones entre conjuntos se muestran como puntos de conexión. La altura de los puntos indica el número de elementos compartidos. Se observa que la mayoría de intersecciones son elementos (genes) compartidos por todas las cepas.

Por otra parte, el programa identificó en total 1412 genes totales, clasificando estos en las 4 categorías que utiliza este paquete en función de la presencia en un porcentaje de las cepas. De acuerdo a estas estadísticas, Bb tiene un pangenoma *core* grande, que ocupa el 72.73% del pangenoma. La segunda categoría más abundante es la de los genes *shell*, pertenecientes al 22.17% del pangenoma. Curiosamente, la categoría menos abundante es la categoría *soft core*, en dónde solo 12 genes pertenecen al 95% o más de las cepas evaluadas (Tabla 7).

**Tabla 7.** Estadísticas del pangenoma de *Bartonella bacilliformis* en dónde Ver097 se incluye. 1027 genes son compartidos por el 100% de cepas evaluadas, siendo el mayor porcentaje de genes en el pangenoma.

Categoría	Presente en...	Cantidad de genes
Genes <i>core</i>	(99% <= cepas <= 100%)	1027
Genes <i>soft core</i>	(95% <= cepas < 99%)	12
Genes <i>shell</i>	(15% <= cepas < 95%)	313
Genes <i>cloud</i>	(0% <= cepas < 15%)	60

<b>Genes totales</b>	<b>(0% &lt;= cepas &lt;= 100%)</b>	<b>1412</b>
----------------------	------------------------------------	-------------

Resalta el hecho de que la mayoría de estos genes que comparte este grupo a diferencia del resto pertenecen todos a proteínas asociadas a fagos. Para comprender mejor la naturaleza de estos genes adicionales en este grupo, se procedió a identificar estas regiones profágicas (Tabla 8).

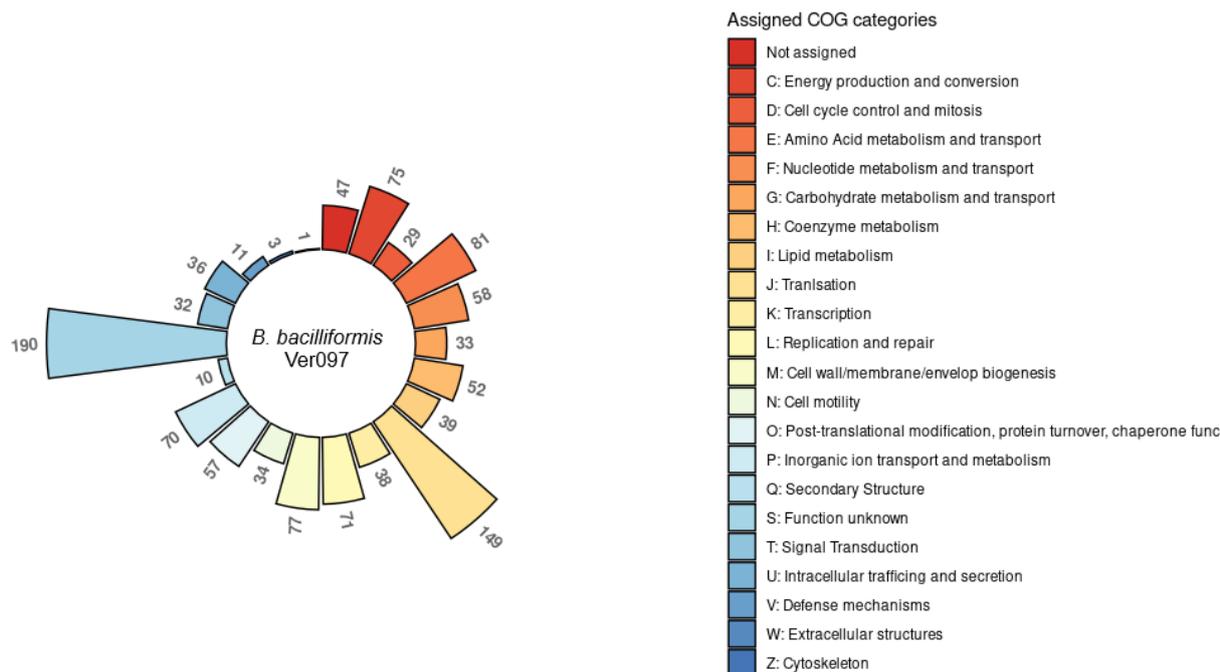
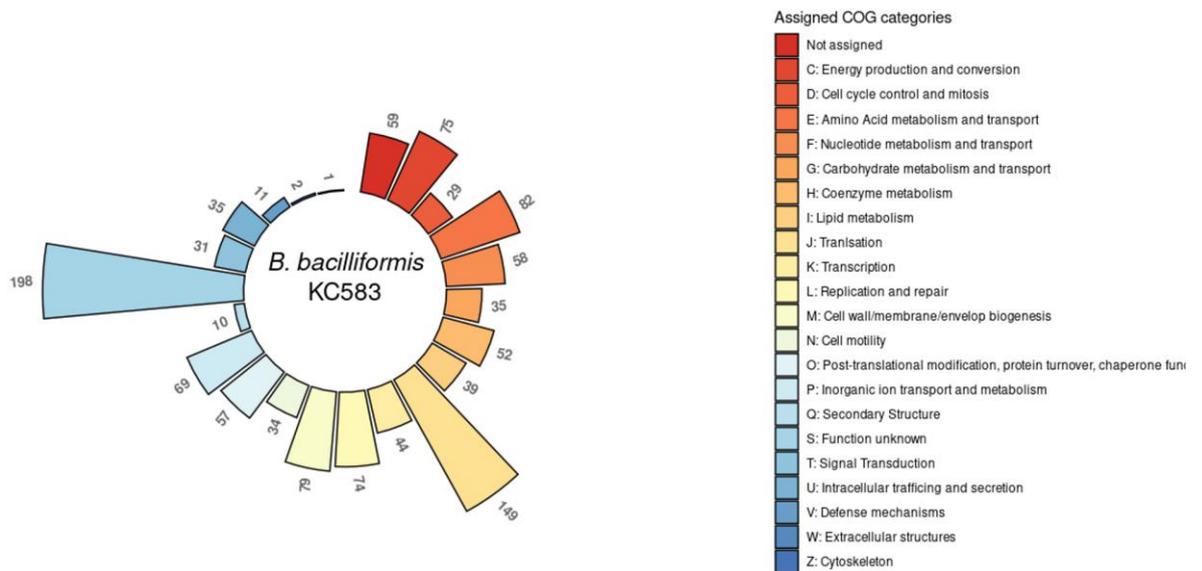
**Tabla 8.** Regiones profágicas encontradas por el servidor web Phastest. Para poder calificar una región como “intacta”, el programa utiliza un puntaje basándose en las secuencias codificantes que se encuentran en la región, comparándolas con una base de datos. A una mayor cantidad de proteínas fágicas de un solo taxón, mayor puntaje. Hay 3 categorías: Incompleto, cuestionable e intacto, siendo esta última la categoría más alta.

<b>Genoma</b>	<b>Nivel</b>	<b>Kb</b>	<b>#ORFs</b>	<b>Posición</b>	<b>Cont.</b>	<b>Fago más común</b>	<b>GC%</b>
Cond044	Intacto	28.2	40	787425 - 815698	2	Vibrio 12B12	41.9
1469	Intacto	28.2	37	44425 - 72698	3	Vibrio 12B12	41.9
1466	Intacto	28.2	37	39807 - 68080	3	Vibrio 12B12	42
1474	Intacto	28.2	37	39807 - 68080	3	Vibrio 12B12	42

Todas las regiones identificadas como profágicas en estos genomas fueron descritas como intactas por el servidor web. Asimismo, todas presentaron la misma longitud

de 28.2 Kb. Resalta también que el taxón más prevalente fue el fago Vibrio 12B12, al cuál estaban asociados la mayoría de ORFs identificados en cada uno de estos genomas.

Para poder realizar una mejor comparación con el cluster Ver097, se realizó una anotación a base de la asignación de categorías COG (*Clusters of Orthologous Genes*) presentes en los genomas. En KC583, se identificaron 198 genes con función desconocida, 149 genes relacionados con la categoría de traducción, 59 genes no asignados a ninguna categoría específica y 34 genes asociados con la categoría de motilidad celular. En el caso del genoma Ver097, varias categorías tuvieron un número reducido de genes. Se observa, por ejemplo, que Ver097 tiene menos genes en las categorías transcripción (n=38), replicación (n=71) y pared celular (n=77). Curiosamente, Ver097 presenta también 1 gen más en la categoría de secreción (n=36), así como 1 gen más también en la categoría de traducción de señales (n=32) (Figura 9).



**Figura 9.** Categorías COG asignadas en el servidor web eggNOG Mapper. (Arriba) Categorías COG asignadas al genoma KC583. (Abajo) Categorías COG asignadas al genoma Ver097.

Se construyó un árbol filogenético basado en genes *core* para el linaje L1 (Figura 10). El clado Ver097 formó su propio grupo con una longitud de rama considerable respecto al resto de Bb. Se demuestra que estas cepas no pertenecen al clado de *B. ancashensis*. Las distancias entre los 4 clados formados en el grupo de Bb es reducida, y con menor longitud de la escala mostrada por el árbol. Esto demuestra la poca diferencia que hay internamente entre los diferentes clados de Bb. Por otra parte, *B. ancashensis* se muestra como un clado marcadamente diferenciado y apropiadamente separado de Bb.



**Figura 10.** Árbol filogenético del Linaje 1 del género *Bartonella* con una escala de 0.1. Se distingue la separación entre Bb y *B. ancashensis* aunque dentro del propio clado que ocupan todos los genomas pertenecientes a Bb, se distingue también una marcada diferencia entre el clado Ver097 y el resto de genomas correspondientes a Bb.

Los resultados obtenidos a partir de los análisis de genómica comparativa determinan una marcada diferencia del clado Ver097 al respecto del resto de genomas de Bb. La definición y filogenética de este grupo es un estudio en curso. Es por ello que, a partir de los resultados obtenidos, y con el objetivo de poder obtener una mejor resolución de la historia evolutiva de Bb, se considera a Ver097 como una especie distinta a Bb.

### **B. Filogenia y pangenómica en el género *Bartonella***

Se obtuvieron 39 genomas de repositorios públicos (Tabla 9). Se seleccionaron aquellos genomas que tuvieran condiciones óptimas de ensamblaje, tales como haber llegado al nivel de cromosoma o, de lo contrario, que tuvieran la menor cantidad de contigs o scaffolds posibles. Debido a que durante la selección de genomas, se hallaron nuevas especies recién incorporadas al género. Entre estas nuevas especies incluidas en este árbol se encuentra a *B. rattaaustraliani*, descrita el año 2009 (87), *B. florencae*, descrita el año 2014, *B. senegalensis*, descrita el año 2013 (88), *B. kosoyi* y *B. krasnovii*, descritas el año 2020 (89) o *B. fuyuanensis*, descrita el año 2015 (90).

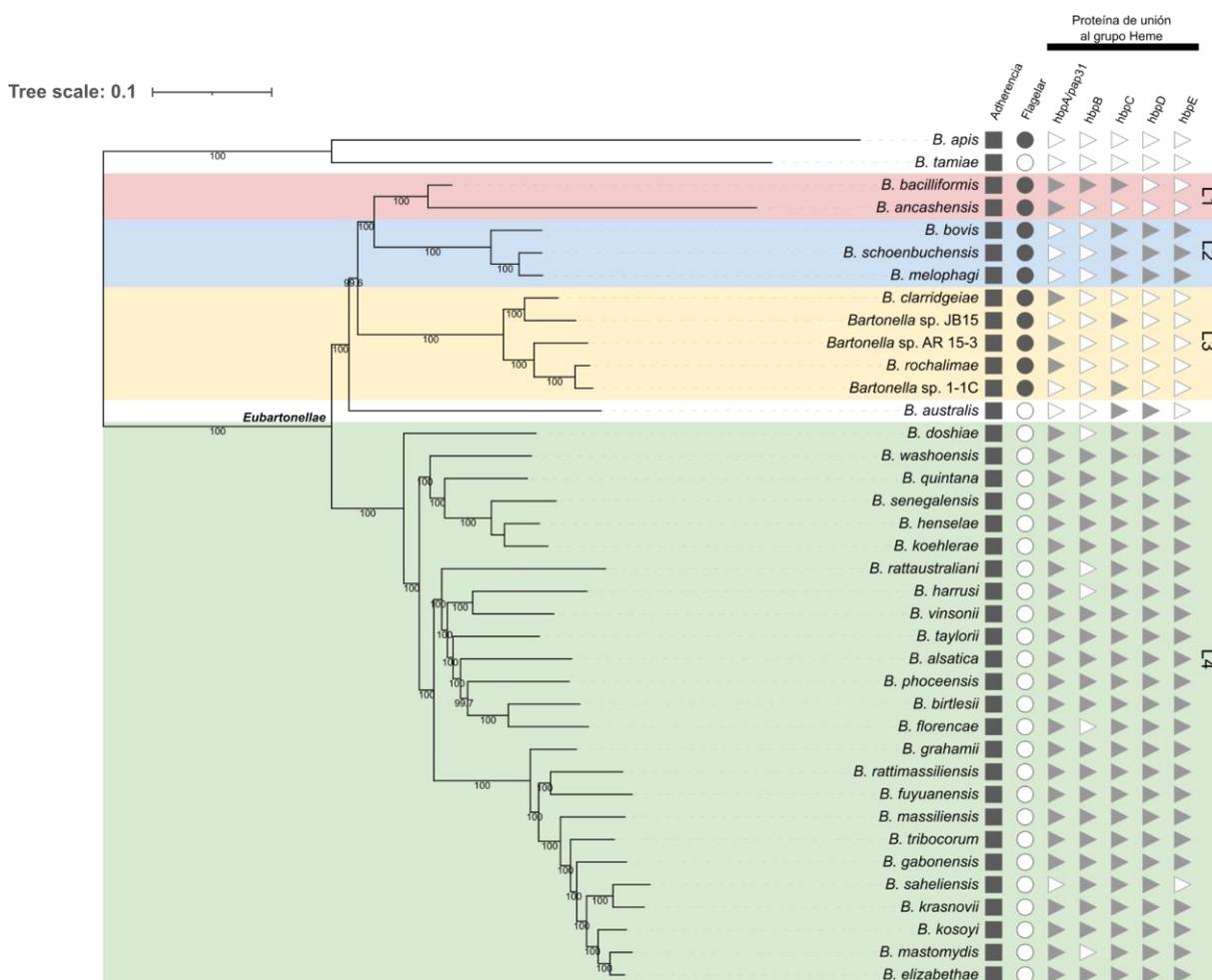
**Tabla 9.** Estadísticas básicas de los genomas del género *Bartonella* disponibles en GenBank. A partir de aquí se realizó la selección de genomas. El año representa el año en que fueron subidas a la base de datos.

N° de acceso	Nombre del organismo	Cepa	Tamaño kb	Año
GCF_002007565.1	<i>Bartonella apis</i>	BBC0122	2 907 212	2017
GCA_000341355.1	<i>Bartonella australis</i>	Aust/NH1	1 596 490	2014
GCF_000384965.1	<i>Bartonella bovis</i>	91-4	1 624 667	2016
GCA_000612965.1	<i>Bartonella henselae</i>	BM1374163	1 905 383	2014
GCA_000022725.1	<i>Bartonella grahamii</i>	as4aup	2 369 520	2014
GCA_000706625.1	<i>Bartonella koehlerae</i>	C-29	1 740 241	2014
GCA_000278255.1	<i>Bartonella melophagi</i>	K-2C	1 518 819	2014
GCA_000312565.2	<i>Bartonella rattaaustraliani</i>	AUST/NH4	2 158 445	2014
GCA_000278215.1	<i>Bartonella rattimassiliensis</i>	15908	2 076 034	2014
GCA_000706645.1	<i>Bartonella rochalimae</i>	ATCC BAA-1498	1 506 733	2014
GCF_002022685.1	<i>Bartonella schoenbuchensis</i>	R1	1 728 487	2017
GCA_000278275.1	<i>Bartonella tamiae</i>	Th239	2 252 122	2014
GCA_000196435.1	<i>Bartonella tribocorum</i>	CIP 105476	2 642 404	2014
GCA_000312525.1	<i>Bartonella florencae</i>	R4	2 053 511	2012
GCA_000312545.1	<i>Bartonella senegalensis</i>	OS02	2 002 317	2012
GCA_000400165.1	<i>Bartonella birtlesii</i>	E11	1 840 919	2013
GCA_001281405.1	<i>Bartonella ancashensis</i>	20.00	1 467 695	2015

GCA_002022445.1	<i>Bartonella sp.</i>	AR 15-3	1 630 082	2017
GCA_002022605.1	<i>Bartonella sp.</i>	JB15	1 494 018	2017
GCA_002810325.1	<i>Bartonella sp.</i>	1-1c	1 600 621	2017
GCA_003606325.3	<i>Bartonella kosoyi</i>	Tel Aviv	2 276 164	2019
GCA_003606345.3	<i>Bartonella krasnovii</i>	OE 1-1	2 186 621	2019
GCA_009936175.1	<i>Bartonella quintana</i>	MF1-1	1 588 683	2019
GCA_013388295.1	<i>Bartonella alsatica</i>	CIP 105477	1 659 117	2020
GCA_014197255.1	<i>Bartonella fuyuanensis</i>	DSM 100694	1 944 744	2020
GCA_900185775.1	<i>Bartonella mastomydis</i>	8	2 045 026	2019
GCA_900445535.1	<i>Bartonella doshiae</i>	NCTC12862	1 844 489	2018
GCA_900638615.1	<i>Bartonella elizabethae</i>	NCTC12898	2 018 538	2018
GCA_900638635.1	<i>Bartonella vinsonii</i>	NCTC12905	1 995 085	2018
GCA_902150025.1	<i>Bartonella massiliensis</i>	OS09	2 277 694	2019
GCA_902162175.1	<i>Bartonella sahelensis</i>	77	2 258 762	2019
GCA_902825145.1	<i>Bartonella phoceensis</i>	CIP107707	1 828 078	2021
GCA_903679515.1	<i>Bartonella gabonensis</i>	669	1 971 183	2020
GCF_000253015.1	<i>Bartonella clarridgeiae</i>	73	1 522 743	2011
GCF_000278135.1	<i>Bartonella washoensis</i>	Sb944nv	1 970 822	2012
GCF_009498695.1	<i>Bartonella bacilliformis</i>	KC584	1 411 655	2019
GCF_023920085.1	<i>Bartonella taylorii</i>	IBS296	1 948 309	2022

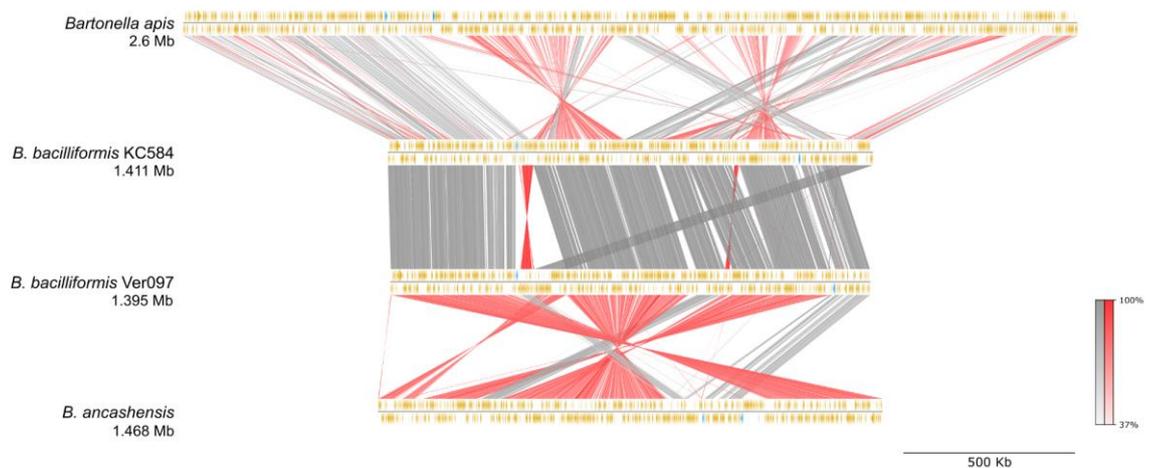
GCF_024297065.1	<i>Bartonella harrusi</i>	117A	2 235 184	2022
-----------------	---------------------------	------	-----------	------

Estas especies no han sido incluidas en algunas filogenias realizadas previamente. Se optó por realizar una filogenia nueva del género, en dónde se obtuvieron los linajes anteriormente descritos en la bibliografía. Se halló que todas las especies añadidas a esta filogenia pertenecen al Linaje 4 (Figura 11).



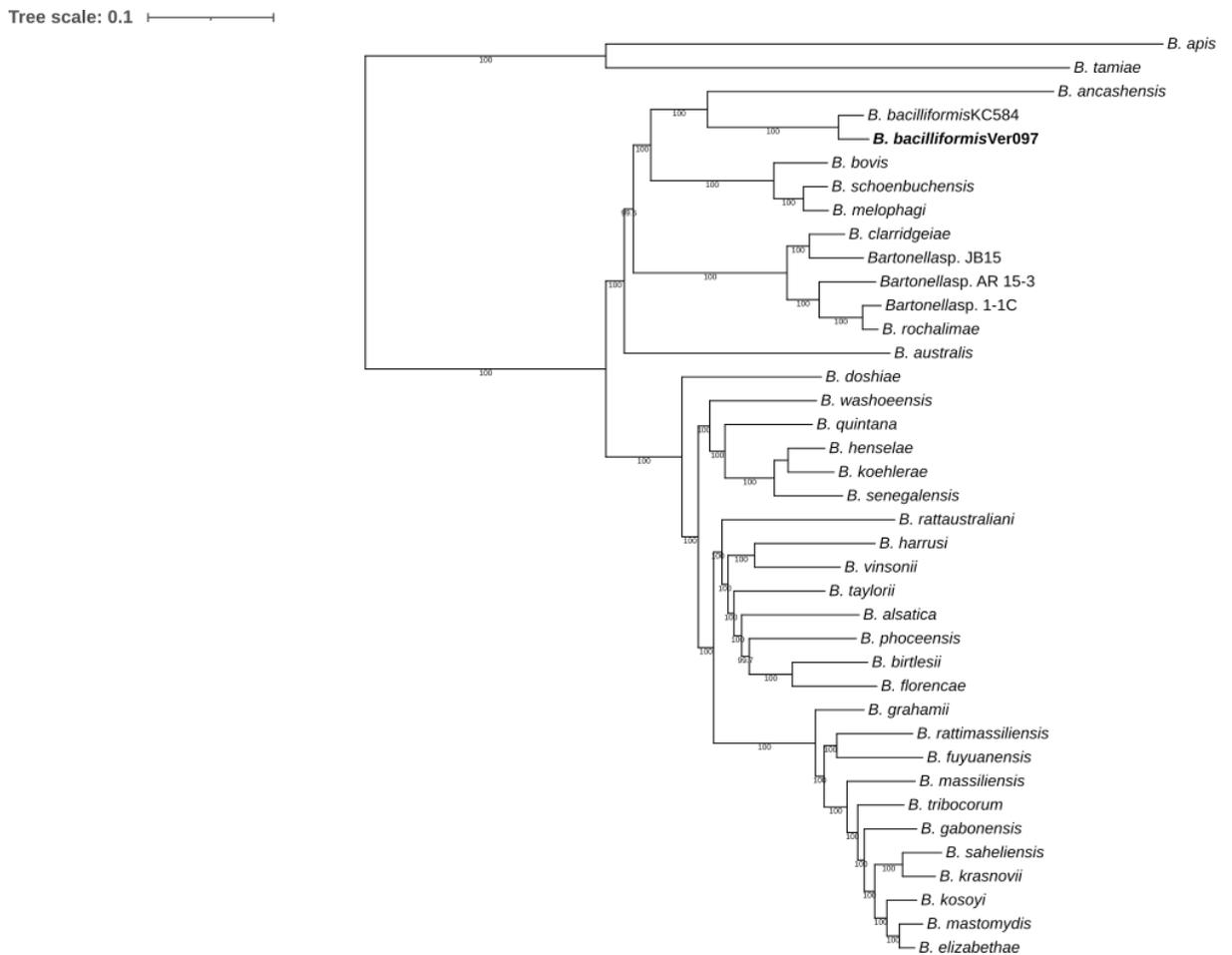
**Figura 11.** Árbol del clado Eubartonellae basado en el alineamiento de genes core del género *Bartonella* (identidad = 0.70). A la derecha figuran las categorías de virulencia proveídas por VFDB. Se presenta un esquema de presencia/ausencia para cada una de las categorías de genes de interés.

Se procedió con una alineación genómica, que reveló que la cepa Ver097 de *B. bacilliformis* posee un genoma más pequeño que la cepa KC584, con una diferencia de aproximadamente 0.016 Mb. Esta reducción de tamaño se atribuye a varias eliminaciones puntuales observadas entre la cepa estándar y el aislado Ver097. La organización del genoma de la cepa Ver097 diverge de *B. ancashensis* de la misma línea, lo que sugiere una estructura genómica más única y posiblemente un camino evolutivo distinto. Ver097 no es similar a otras especies del género, como *B. ancashensis*, *B. schoenbuchensis* y *B. henselae* (Figura 12).



**Figura 12.** Alineación genómica entre especies del género *Bartonella*. Los colores de las líneas dibujadas entre genomas se refieren a la orientación que toman estas secciones en los genomas evaluados. El rojo indica inversión, mientras que el gris indica la misma orientación que la sección comparada. Se aprecia que, entre *B. bacilliformis* KC584 y *B. bacilliformis* Ver097 los bloques de identidad no son continuos.

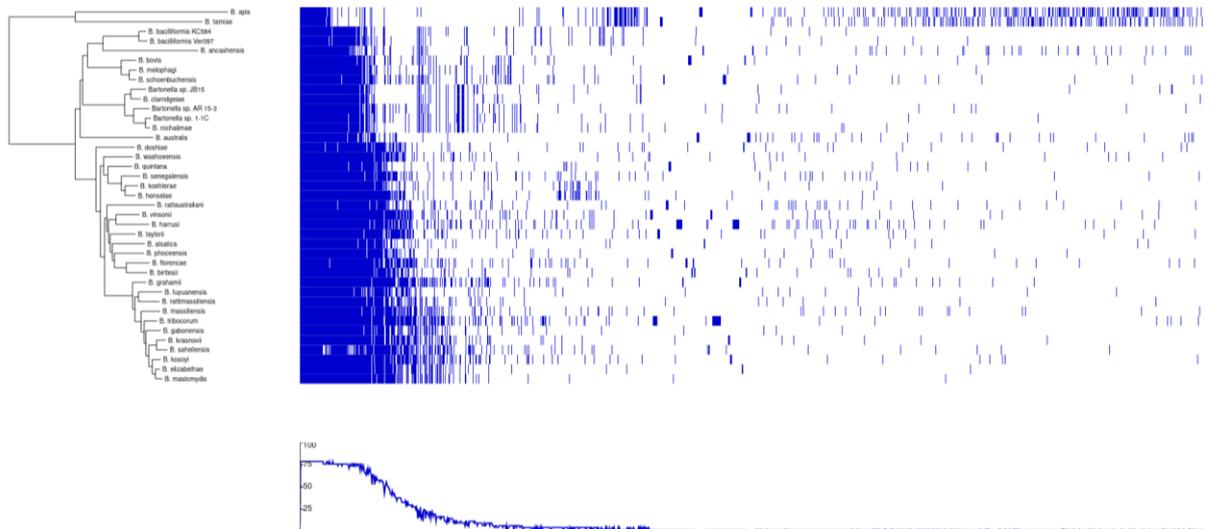
Para asegurar que Ver097 no era un aislado mal identificado como *B. bacilliformis*, se realizó un árbol filogenético de todo el género y con Ver097 incorporado en el análisis. Se comprueba que Ver097 sigue ubicándose en el clado L1 y próximo a *B. bacilliformis* KC584 (Figura 13).



**Figura 13.** Filogenia del género *Bartonella* en dónde Ver097 (en negrita) se sigue acoplando al linaje L1 y junto a *B. bacilliformis* KC584, comprobando que *B. bacilliformis* Ver097 no ha sido incorrectamente identificada como de otra especie.

Por último se realizó un análisis pangenómico de todo el género *Bartonella* incluyendo a Ver097 dentro del análisis pangenómico (Figura 14). 294 genes pertenecen al genoma *core* del género, de un universo de 11476 genes (Tabla 10).

La mayor cantidad de genes en el género *Bartonella* caen en la categoría de genes *cloud* (n = 9524), indicando como la mayoría de genes en el pool genético del género *Bartonella* pertenecen a menos del 15% de cepas evaluadas en el análisis.



**Figura 14.** Pangenoma del género *Bartonella* en dónde Ver097 se incluye. A nivel de género, se observa el limitado genoma core, mientras que los genes accesorios son los que ocupan la mayor parte de la gráfica.

A partir de la tabla de presencia y ausencia de genes, se determinaron 98 genes únicos para *B. bacilliformis* y *B. bacilliformis* Ver097, 18 genes que únicamente pertenecen a *B. bacilliformis* y 11 genes que únicamente pertenecen a *B. bacilliformis* Ver097. Además, 121 proteínas hipotéticas también fueron detectadas como únicas para estos genomas a comparación del género *Bartonella*.

Dentro del grupo de genes únicos para Bb y Bb Ver097, los genes *fliE*, *fliY*, *fliO*, *motB* y *motC*, además de un gen putativo relacionado al motor flagelar fueron reconocidos como únicos para ambas cepas. Esto se debe a que el punto de corte

para ser agrupados con otros genes fue 70 de identidad en BLASTP. Si bien otras especies de *Bartonella* cuentan con estos genes, Bb y Bb Ver097 exhibieron una identidad menor de 70 al compararlos con estos.

Otros genes de interés en este grupo único fueron genes relacionados a la conformación de la membrana, tales como la proteína CvpA, involucrada en la homeostasis de la membrana, proteínas conteniendo el dominio SH3b, que puede estar involucrada en la invasión de células por parte del patógeno, y proteínas relacionadas con la toma de heme o hierro dependientes de TonB. En el grupo de genes exclusivo de Ver097, algunas proteínas pudieron ser identificadas, como la proteína GpU, relacionada al ensamble de la cola del fago P2 (Tabla S2).

**Tabla 10.** Estadísticas del pangenoma del género *Bartonella* en donde Ver097 se incluye. 294 genes son compartidos por el 100% de cepas evaluadas.

<b>Categoría</b>	<b>Presente en...</b>	<b>Cantidad de genes</b>
Genes <i>core</i>	(99% <= cepas <= 100%)	294
Genes <i>soft core</i>	(95% <= cepas < 99%)	99
Genes <i>shell</i>	(15% <= cepas < 95%)	1559
Genes <i>cloud</i>	(0% <= cepas < 15%)	9524
<b>Genes totales</b>	<b>(0% &lt;= cepas &lt;= 100%)</b>	<b>11476</b>

**C. Relaciones filogenéticas entre genes de virulencia en el Linaje 1 de Eubartonella**

A partir de las evaluaciones realizadas a cada genoma utilizando la base de datos VFDB, se obtuvieron los siguientes genes de virulencia en *Bartonella bacilliformis* (Tabla 11). Estos fueron los genes seleccionados para construir filogenias genéticas y también para pasar por pruebas de selección positiva, dependiendo si las otras especies poseían o no estos genes.

**Tabla 11.** Genes de virulencia hallados en los genomas de Bb KC584 y Bb Ver097. 3 son las categorías de virulencia identificadas en las bases de datos: Adherencia, Motilidad y otras, en dónde figuran las proteínas de unión al grupo heme.

Genes de virulencia (VFDB)	Genes de virulencia (BAKTA)	Clase VFDB	Factor VFDB	¿Son genes core?
badA/vomp/brp	group_268, group_764, group_751	Adherencia	Bartonella adhesin A	Sí/No/Sí
flaA	flaA	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flaF	flaF	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flbT	flbT	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgA	flgA	Motilidad y mejora de	Flagella	Sí

		unión al eritrocito		
flgB	flgB	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgC	flgC	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgD	flgD	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgE	flgE	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgF	flgF	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgG	flgG	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgH	flgH	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgI	flgI	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flgK	flgK	Motilidad y mejora de unión al eritrocito	Flagella	Sí

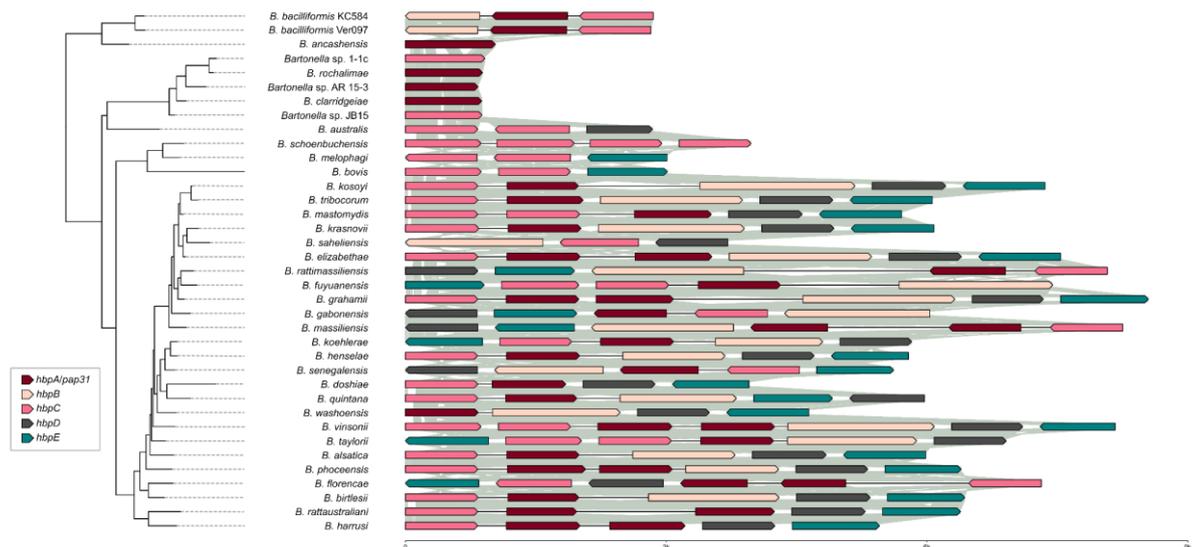
flgL	flgL	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flhA	flhA	Motilidad y mejora de unión al eritrocito	Flagella	Sí
flhB	flhB	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliE	fliE	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliF	fliF	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliG	fliG	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliI	fliI	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliL	fliL	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliM	fliM	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliP	fliP	Motilidad y mejora de	Flagella	Sí

		unión al eritrocito		
fliQ	fliQ	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliR	fliR	Motilidad y mejora de unión al eritrocito	Flagella	Sí
fliY	fliY	Motilidad y mejora de unión al eritrocito	Flagella	Sí
hbpA/pap31	group_815	Otros	Heme protein binding	Sí
hbpB	group_816	Otros	Heme protein binding	Sí
hbpC	group_812	Otros	Heme protein binding	Sí
ialB	ialB	Adherencia	Invasion-associated locus B	Sí
motA	motA	Motilidad y mejora de unión al eritrocito	Flagella	Sí
motB	motB	Motilidad y mejora de unión al eritrocito	Flagella	Sí
omp43	group_500	Adherencia	Omp43	Sí
omp89	bamA	Adherencia	Omp89	Sí

### 1. Hbp (Heme binding protein)

En el análisis filogenético realizado a los genes concatenados relacionados a la familia Hbp, se puede observar que el linaje L1 conserva su conformación. Lo

mismo sucede con el resto de clados en este árbol. Sin embargo, el orden no es el mismo, y el linaje L2 aparece más cercano al linaje L4 en el caso de los genes Hbp (Figura 15). A nivel de organización genética, se observa que no hay diferencia entre la distribución y tamaño de los genes Hbp para el caso de Bb KC584 y Bb Ver097. En general, se observa que la distribución y tamaño de estos genes es variada en el género, y que no necesariamente presentan una estructura conservada. Resulta interesante observar que el gen *hbpB* se encuentra presente en Bb y en la mayoría de especies pertenecientes al linaje L4, aunque muestran una diferencia de tamaño considerable, siendo que el gen *hbpB* en el linaje L4 es más largo. Por último, respecto al alineamiento realizado entre estas secuencias, se observa que los genes *hbp* en Bb KC584 y Bb Ver097 comparten estructura, orientación e identidad. Mientras tanto, al comparar el gen *hbpA/pap31* con el mismo gen en *B. ancashensis*, se observa que la identidad sólo es compartida con algunas regiones en esta especie.

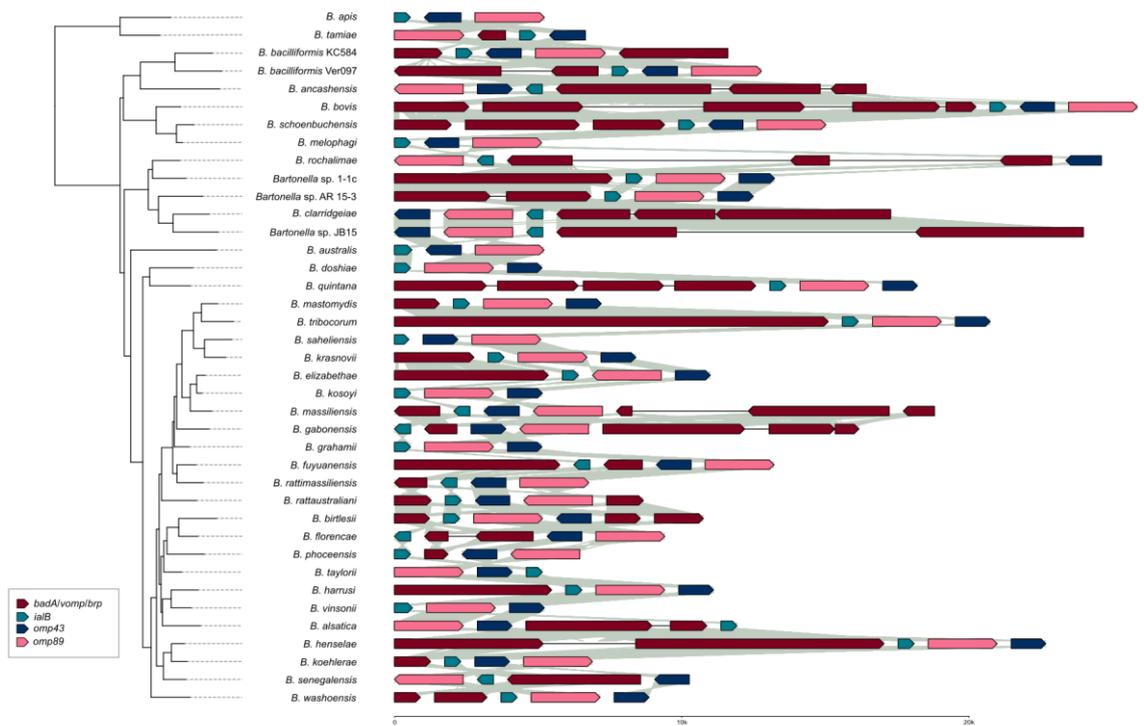


**Figura 15.** Árbol de genes concatenados pertenecientes a la familia Hbp identificados por VFDB en el género *Bartonella*. El alineamiento, realizado mediante TBLASTX, tuvo un valor mínimo de  $1e^{-4}$ .

## 2. Adherencia

En el caso de los genes de adherencia, a primera vista se destaca la organización poco conservada de estos genes en las distintas especies evaluadas. Todos los genes anotados como *badA/vomp/brp* son muy disímiles en tamaño y se pueden presentar en diferentes números en los genomas evaluados. Por otra parte, *ialB*, *omp43* y *omp89* se presentan de una manera más conservada, siempre con un tamaño y distribución similar en todos los genomas. Estas secuencias en particular parecen alinearse en bloque y exhiben una conservación a nivel de identidad también. Esto probablemente pueda estar relacionado a una función conservada para estos genes.

Dentro del linaje L1, la estructura no cambia, y la única diferencia notable reside en *B. ancashensis*, que tiene un gen *badA/vomp/brp* más que las otras dos cepas. Resulta interesante que un gen *badA/vomp/brp* no tenga mucha relación a nivel de identidad ni con Bb KC584 ni con *B. ancashensis*, siendo un gen particularmente único en este aspecto y con solo unas pocas regiones similares a las de las otras 2 especies (Figura 16).

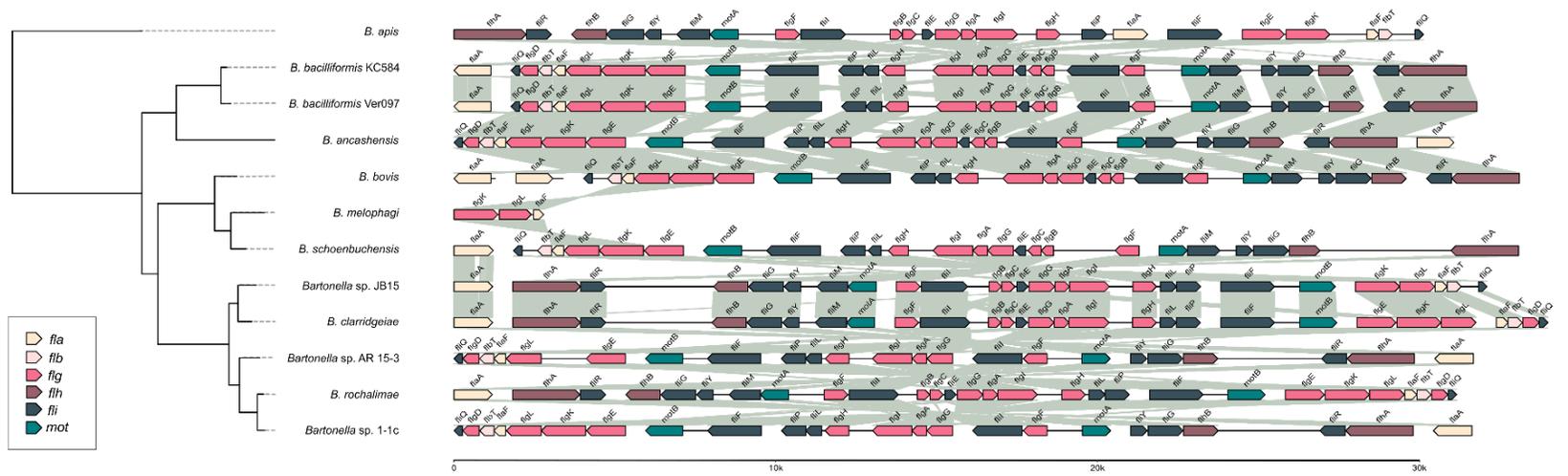


**Figura 16.** Árbol de genes concatenados pertenecientes a los genes de adherencia identificados por VFDB en el género *Bartonella*. El alineamiento, realizado mediante TBLASTX, tuvo un valor mínimo de  $1e^{-4}$ .

### 3. Motilidad

Se evaluó la motilidad en 3 linajes del género *Bartonella*: L1, L2 y L3. Se observó que la motilidad no se presenta en el linaje L4. En este sentido, los linajes evaluados se conservan a este nivel, y no se presenta un cambio en la topología del árbol de genes concatenados. No se observa una variación notable en la organización y conservación de los genes involucrados. A nivel de identidad, los genes de motilidad muestran un considerable grado de conservación, alineándose en bloques entre los distintos genomas pertenecientes al grupo Eubartonella. Una diferencia significativa se encuentra cuando se evalúan los genes móviles de la especie *B. apis*

con los de Bb KC584, en dónde se observa una inversión de los elementos genéticos. Curiosamente se observa que no sería el único evento de inversión en estas especies, puesto que nuevamente se presenta este reordenamiento de genes en *Bartonella* sp. JB-15, *Bartonella* sp. AR 15-3, *B. rochalimae* y *Bartonella* sp. AR 1-1c. Asimismo, la identidad entre los genes de *B. apis* y Bb KC584 no presenta una diferenciación importante a este nivel, con una identidad mínima de  $1e^{-4}$  para todos los genes asociados con la motilidad (Figura 17).



**Figura 17.** Árbol de genes concatenados pertenecientes a los genes de motilidad identificados por VFDB en el género *Bartonella*. El alineamiento, realizado mediante TBLASTX, tuvo un valor mínimo de  $1e^{-4}$ .

#### **D. Selección positiva en genes de virulencia de *B. bacilliformis***

Se sometió al género *Bartonella* a 2 hipótesis de selección positiva usando el modelo rama-nodo: La primera buscó comprobar la presencia de esta fuerza evolutiva en la divergencia entre *B. bacilliformis* KC584 y Ver097 del resto de especies de *Bartonella* (H1). Y la segunda buscó comprobar esto mismo pero en la divergencia del linaje L1 (H2). Los resultados son indicados en la Tabla 12:

**Tabla 12:** Valores de la prueba de selección positiva, en dónde se pusieron a prueba 2 hipótesis. H1: Clado de *B. bacilliformis* KC584 y *B. bacilliformis* Ver097 versus las especies del género *Bartonella*. H2: Linaje 1 compuesto por *B. ancashensis*, *B. bacilliformis* KC584 y *B. bacilliformis* Ver097 versus las especies del género *Bartonella*. (\*): En estos genes, a H1 consistió en Ver097 versus el resto de especies del género *Bartonella*, mientras que H2 se refiere a KC584 y Ver097 versus el resto de especies del género *Bartonella*. Esto debido a que *B. ancashensis* no tenía el gen ortólogo en estos casos.

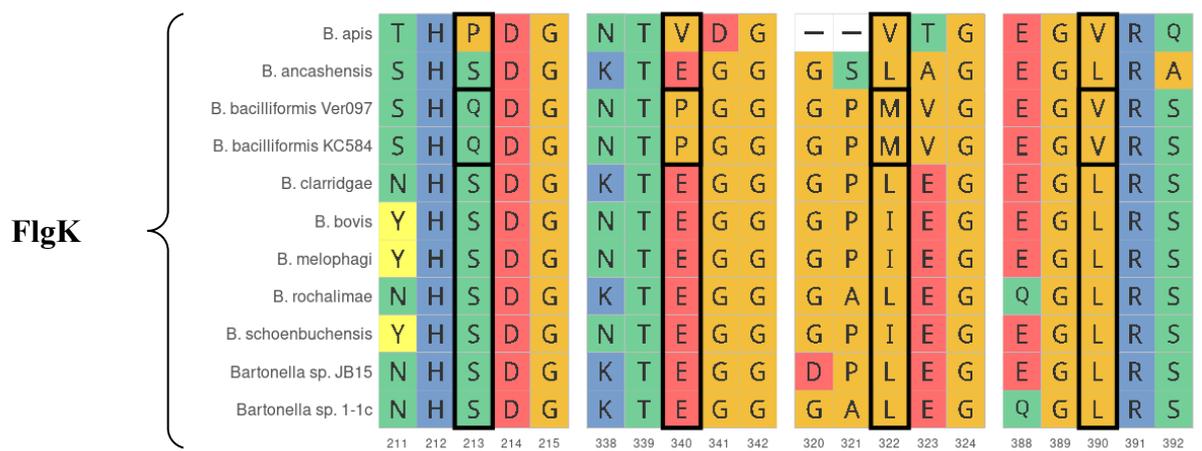
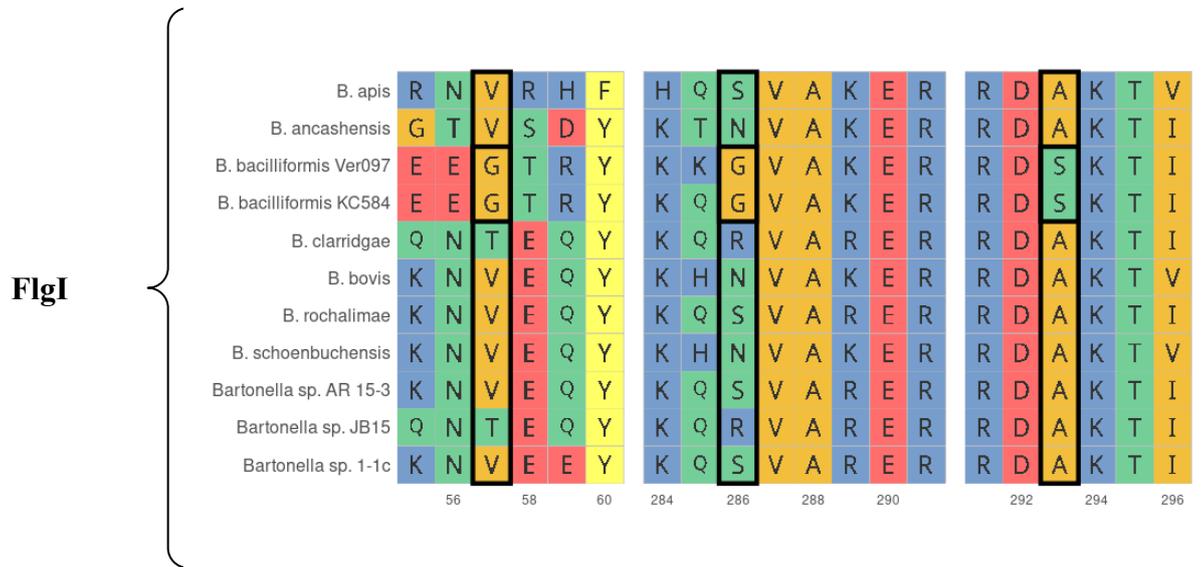
Gen	Hipótesis 1						Hipótesis 2					
	H0	H1	LRT	LRT absoluto	valor p	valor p corregido	H0	H2	LRT	LRT absoluto	valor p	valor p corregido
flaA	-6210,965	-6210,965	0,000	0,000	1,000	1,000	-6208,151	-6206,205	3,892	3,892	0,049	0,424
flaF	-1488,896	-1488,896	0,000	0,000	1,000	1,000	-1488,810	-1488,782	0,056	0,056	0,813	1,000
flbT	-1970,630	-1973,342	-5,424	5,424	1,000	1,000	-1969,965	-1969,951	0,028	0,028	0,868	1,000
flgA	-2387,831	-2387,734	0,193	0,193	0,661	1,000	-2386,820	-2386,507	0,627	0,627	0,428	0,882

flgB	-1946,803	-1948,794	-3,982	3,982	0,046	0,207	-1941,840	-1940,457	2,766	2,766	0,096	0,562
flgC	-1821,902	-1821,902	0,000	0,000	0,995	1,000	-1819,340	-1818,649	1,382	1,382	0,240	0,696
flgD	-1596,557	-1596,758	-0,402	0,402	0,526	1,000	-1592,929	-1586,509	12,841	12,841	0,000	0,008
flgE	-5948,986	-5948,875	0,221	0,221	0,638	1,000	-5946,776	-5946,649	0,255	0,255	0,614	0,976
flgF	-3575,836	-3575,152	1,369	1,369	0,242	0,770	-3578,071	-3578,071	0,000	0,000	1,000	1,000
flgG	-3312,716	-3312,275	0,882	0,882	0,348	0,909	-3312,758	-3312,758	0,000	0,000	1,000	1,000
flgH	-3259,436	-3259,573	-0,273	0,273	0,601	1,000	-3258,892	-3258,365	1,053	1,053	0,305	0,762
flgI	-5458,503	-5453,917	9,171	9,171	0,002	0,030	-5460,435	-5460,435	0,000	0,000	1,000	1,000
flgK	-7123,494	-7119,221	8,547	8,547	0,003	0,030	-7123,858	-7123,577	0,560	0,560	0,454	0,883
flgL	-4733,055	-4730,721	4,669	4,669	0,031	0,179	-4733,176	-4733,048	0,257	0,257	0,612	0,976
flhA	-9086,374	-9084,433	3,882	3,882	0,049	0,207	-9087,389	-9087,389	0,000	0,000	1,000	1,000
flhB	-4868,329	-4868,329	0,000	0,000	1,000	1,000	-4865,222	-4863,589	3,267	3,267	0,071	0,495
fliE	-1912,614	-1912,244	0,741	0,741	0,389	0,909	-1914,451	-1913,382	2,137	2,137	0,144	0,696
fliF	-7841,625	-7841,426	0,398	0,398	0,528	1,000	-7839,874	-7838,913	1,921	1,921	0,166	0,696
fliG	-5598,402	-5593,499	9,806	9,806	0,002	0,030	-5602,561	-5599,563	5,995	5,995	0,014	0,167

fliI	-5605,611	-5605,611	0,000	0,000	1,000	1,000	-5605,611	-5605,611	0,000	0,000	0,998	1,000
fliL	-2105,003	-2102,515	4,975	4,975	0,026	0,179	-2103,792	-2102,889	1,805	1,805	0,179	0,696
fliM	-5022,483	-5022,107	0,752	0,752	0,386	0,909	-5024,317	-5024,307	0,020	0,020	0,888	1,000
fliP	-3553,745	-3553,745	0,000	0,000	1,000	1,000	-3555,568	-3555,568	0,000	0,000	1,000	1,000
fliQ	-1050,357	-1050,357	0,000	0,000	1,000	1,000	-1050,357	-1051,104	-1,494	1,494	0,222	0,696
fliR	-4036,555	-4036,353	0,405	0,405	0,525	1,000	-4036,645	-4036,172	0,947	0,947	0,330	0,771
fliY	-2255,225	-2256,261	-2,072	2,072	0,150	0,525	-2256,180	-2256,261	-0,161	0,161	0,689	1,000
motA	-4043,156	-4043,166	-0,018	0,018	0,892	1,000	-4042,563	-4042,489	0,149	0,149	0,699	1,000
motB	-5281,812	-5277,528	8,567	8,567	0,003	0,030	-5281,812	-5281,468	0,687	0,687	0,407	0,882
badA/vomp/brp (*)	-1044,596	-1042,726	3,739	3,739	0,053	0,207	-1047,290	-1047,290	0,000	0,000	1,000	1,000
ialB	-6218,356	-6218,356	0,000	0,000	1,000	1,000	-6220,218	-6220,400	-0,364	0,364	0,547	0,976
omp43	-17660,900	-17660,470	0,860	0,860	0,354	0,909	-17667,160	-17666,491	1,338	1,338	0,247	0,696
omp89	-30382,801	-30382,801	0,000	0,000	1,000	1,000	-30384,928	-30378,797	12,263	12,263	0,000	0,008
hbpA/pap31	-11918,748	-11918,748	0,000	0,000	1,000	1,000	-11916,907	-11916,962	-0,110	0,110	0,741	1,000
hbpB (*)	-6547,749	-6547,749	0,000	0,000	1,000	1,000	-6544,866	-6544,727	0,278	0,278	0,598	0,976

hbpC (*)	-4240,591	-4240,591	0,000	0,000	1,000	1,000	-4243,398	-4242,760	1,276	1,276	0,259	0,696
----------	-----------	-----------	-------	-------	-------	-------	-----------	-----------	-------	-------	-------	-------

La selección positiva se determina de acuerdo a regiones específicas obtenidas como parte del cálculo final, siendo que las regiones que fueron determinadas como bajo la influencia de la selección se muestran en la Figura 18:



### Flg

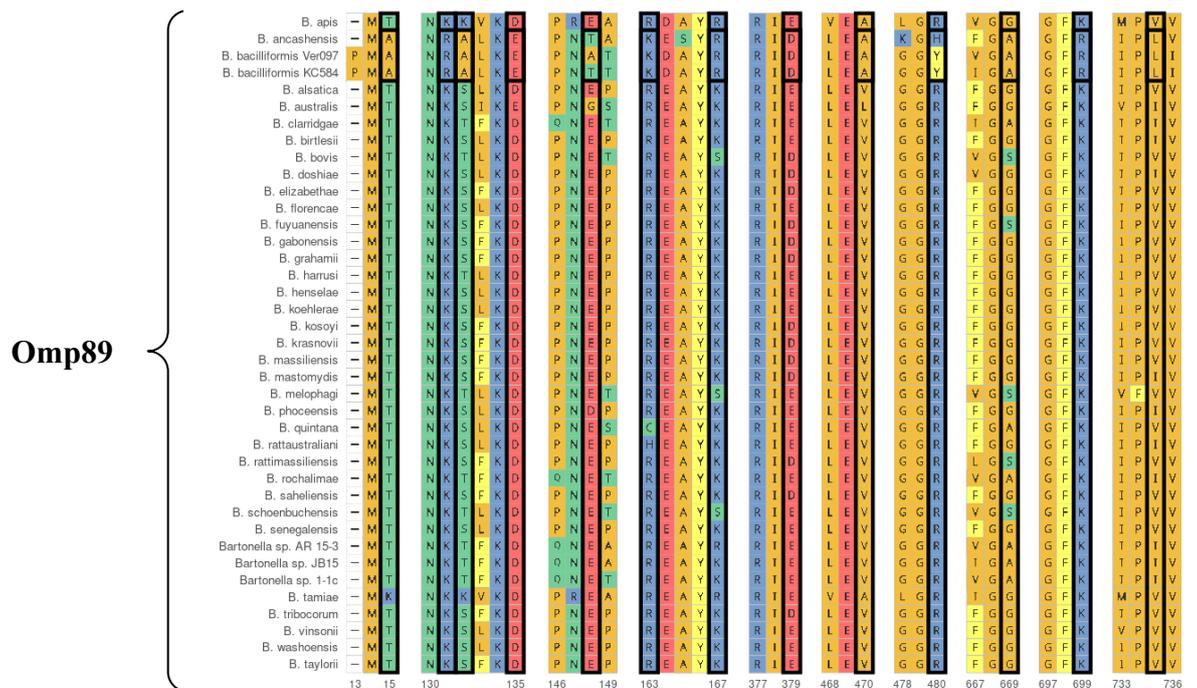
B. apis	A	H	T	T	P	A	E	E	L	Q	Q	T	M	A	M	Q	K	G	L	D	P	S	D	V	I	L	G	S	E	T	A
B. ancashensis	A	H	T	A	P	P	D	S	L	Q	Q	N	M	A	L	E	N	D	L	N	P	A	E	T	I	L	N	S	E	G	A
B. bacilliformis Ver097	A	H	S	A	P	K	D	V	T	Q	Q	S	A	T	L	Q	E	D	L	E	P	A	D	A	I	L	A	S	E	G	K
B. bacilliformis KC584	A	H	S	A	P	K	D	V	T	Q	Q	S	A	T	L	Q	E	D	L	E	P	A	D	A	I	L	A	S	E	G	K
B. claridgae	A	H	T	A	P	A	D	V	L	Q	Q	T	M	S	L	K	N	N	L	N	P	A	D	T	I	L	N	S	E	G	T
B. bovis	A	H	T	A	P	A	D	V	L	Q	Q	S	M	I	S	Q	D	D	L	S	P	A	D	T	I	L	N	S	E	G	T
B. rochalimae	A	H	T	A	P	T	D	T	L	Q	Q	T	M	S	L	K	N	N	L	N	P	A	D	T	I	L	N	S	E	G	T
B. schoenbuchensis	A	H	T	A	P	A	D	V	L	Q	Q	P	M	A	S	Q	D	D	L	S	S	A	D	T	V	L	N	S	E	G	T
Bartonella sp. AR 15-3	A	H	T	A	P	T	D	V	L	Q	Q	P	M	S	L	K	N	N	L	N	P	A	D	T	I	L	N	S	E	G	T
Bartonella sp. JB15	A	H	T	A	P	A	D	I	L	Q	Q	T	M	S	L	K	N	N	L	N	P	A	D	T	I	L	N	S	E	G	T
Bartonella sp. 1-1c	A	H	T	A	P	T	D	A	L	Q	Q	T	M	S	L	K	N	N	L	N	P	A	D	T	I	L	N	S	E	G	T

### MotB

B. ancashensis	D	T	S	V	N	Q	I	I	Q	K	S	E	T	D	A	N	Q	Q
B. bacilliformis Ver097	E	T	T	V	D	P	I	V	D	E	N	S	T	D	S	N	T	A
B. bacilliformis KC584	E	T	T	V	D	P	I	V	D	E	N	S	T	D	S	N	K	A
B. claridgae	Q	T	S	L	D	Q	I	V	Q	E	Q	G	T	D	A	N	K	K
B. bovis	Q	A	S	V	G	Q	I	V	Q	N	H	E	T	D	A	N	K	Q
B. rochalimae	Q	T	S	L	D	Q	I	V	Q	E	K	G	T	D	A	N	K	K
B. schoenbuchensis	Q	T	S	V	G	Q	I	V	Q	K	H	E	T	D	A	N	K	Q
Bartonella sp. AR 15-3	Q	T	S	L	D	Q	I	V	Q	K	Q	G	T	D	A	N	K	K
Bartonella sp. JB15	Q	T	S	L	D	Q	I	V	Q	E	K	G	T	D	A	N	K	K
Bartonella sp. 1-1c	Q	T	S	L	D	Q	I	V	Q	E	K	G	T	D	A	N	K	K

### FlgD

B. ancashensis	M	T	V	P	G	F	L	R	D	F	S	M	N	E	K	V	L	A	N	F	K	D	K	D	Y	V	K	S	L	I	G	P	G	V	T	V	-	E	T	D
B. bacilliformis Ver097	M	T	V	P	G	F	L	R	G	F	S	M	N	D	K	V	L	A	N	F	E	D	K	D	Y	V	K	S	L	I	G	P	G	V	T	V	-	P	D	D
B. bacilliformis KC584	M	T	V	P	G	F	L	R	G	F	S	M	N	D	K	V	L	A	N	F	E	D	K	D	Y	V	K	S	L	I	G	P	G	V	T	V	-	P	D	D
B. claridgae	M	V	L	P	R	I	Y	S	Q	F	S	S	N	E	T	V	L	A	N	F	T	D	E	K	Y	V	V	S	L	I	G	P	G	V	T	V	-	E	G	Y
Bartonella sp. AR 15-3	M	A	L	L	R	S	Y	S	Q	F	S	S	N	E	T	M	L	S	N	F	T	D	E	K	Y	V	V	S	L	I	G	P	G	V	T	V	-	E	G	Y
Bartonella sp. 1-1c	M	A	L	P	R	S	Y	S	Q	F	S	S	N	E	T	M	L	S	N	F	T	D	E	K	Y	V	V	S	L	I	G	P	G	V	T	V	-	E	G	Y



**Figura 18.** Alineamiento de las secuencias aminoacídicas de las proteínas que figuran como bajo selección positiva utilizando las secuencias de *B. apis* como referencia. En el recuadro negro se muestran las posiciones específicas que están siendo sometidas a esta fuerza evolutiva. Los colores son relativos a la química de la cadena lateral de los aminoácidos, siendo que naranja = hidrofóbico alifático, amarillo = hidrofóbico aromático, verde = hidrofílico y polar no cargado, rojo = hidrofílico ácido y azul = hidrofílico básico.

En la proteína FlgI, los aminoácidos en las siguientes posiciones son identificados como bajo la influencia de la selección natural: V57G, S286G y A293S. En el caso de S286G, Bb y Bb Ver097 se diferencian en, en vez de contar con un aminoácido hidrofóbico alifático, cuentan con un hidrofílico no cargado. Mientras que en el caso de A293S, se pasa de un aminoácido hidrofílico no cargado a uno hidrofóbico alifático. En la posición V57G no ocurre una diferenciación química.

En el caso de la proteína FlgK, las posiciones P213Q, V340P y V322M están bajo selección. En el caso de la valina en la posición 390, la referencia *B. apis* también tiene el mismo aminoácido en dicha posición. Pero es importante recalcar que, en el resto de las especies, el aminoácido en esa posición es la leucina. Solo en el caso de P213Q ocurre un cambio químico, en donde la prolina no está en Bb y Bb Ver097 y, en su lugar, se encuentra un aminoácido de carácter hidrofílico y no cargado, como es el caso de la glutamina.

En la proteína FliG, las posiciones bajo la influencia de selección son T77S, A134K, L152T, M185A, K220E, D255E, V291A, G309A y A363K. Se dieron diferencias relevantes en el caso de A134K y A363K, en donde Bb y Bb Ver097 tienen un aminoácido hidrofílico básico en estas posiciones. Asimismo, también ocurre una diferencia relevante en L152T, en donde se observa que Bb y Bb Ver097 tiene un aminoácido hidrofílico y polar no cargado en dicha posición. En la sustitución K220E, Bb y Bb Ver097 tienen un aminoácido hidrofílico ácido en vez del hidrofílico básico que se encuentra en la referencia.

En la proteína MotB, los sitios bajo selección son S79T, Q99P, Q133D, E152S, A300S y Q350A. Solo en el caso de el primer alineamiento no ocurre un cambio químico significativo. En el caso de Q99P, Bb y Bb Ver097 tienen un aminoácido hidrofóbico alifático en dicha posición, en vez del aminoácido hidrofílico y polar no cargado que exhiben las otras especies. En el caso de Q133D, se observa un cambio de un aminoácido polar y no cargado a uno cargado negativamente. Para

E152S, el cambio es de un aminoácido cargado negativamente a uno polar y no cargado, disminuyendo potencialmente la capacidad de formar interacciones iónicas. En A300S, ocurre un cambio de un aminoácido hidrofóbico a uno polar y no cargado, lo que podría introducir capacidad de formar puentes de hidrógeno. Finalmente, en Q350A, el cambio es inverso: de un aminoácido polar y no cargado a uno hidrofóbico, reduciendo la interacción con moléculas polares.

En las proteínas FlgD y Omp89, la selección resultó ser positiva para la segunda hipótesis, en donde se evaluaron las proteínas pertenecientes a todo el linaje L1, que contiene a Bb, Bb Ver097 y *B. ancashensis*. Para la primera proteína, en donde *B. clarridgeae* fue la referencia, las siguientes posiciones fueron positivas para la prueba de selección: L3V, I11F, S13R, Q14G (Q14D en el caso de *B. ancashensis*), S69M, T79K, S85A, T103E (T103K en el caso de *B. ancashensis*), K106D, V115K, V135L, K141T y G155T. En el caso de Omp89, las posiciones que resultaron positivas a la prueba de selección fueron: T15A, K131R, K132A, D135E, E148T (siendo el único caso en donde Bb Ver097 exhibe un aminoácido distinto a Bb, E148A), R163K, R167R, E379D, V470A (siendo que la referencia, *B. apis*, también exhibe alanina en esta posición, la valina señala es el aminoácido del resto de especies), R480Y (R480 para *B. ancashensis*), G669A, K699R y V735L.

Son 48 los sitios que fueron determinados como bajo selección positiva. De estos, se observa que en 9 casos, las especies evaluadas prefieren aminoácidos hidrofílicos a comparación de los aminoácidos hidrofóbicos exhibidos por las especies de referencia. Mientras que en 7 sitios, las especies de interés exhiben preferencia por

aminoácidos hidrofóbicos a comparación de los hidrofílicos en las especies de referencia. Por último, en 13 casos, se observa que hay una preferencia por aminoácidos hidrofílicos en las especies de interés y en las de referencia, pero que en el caso de las primeras, se prefieren otras propiedades, como aminoácidos hidrofílicos básicos (4 sitios), aminoácidos hidrofílicos ácidos (2 sitios) o aminoácidos polares no cargados (2 sitios) (Tabla 13).

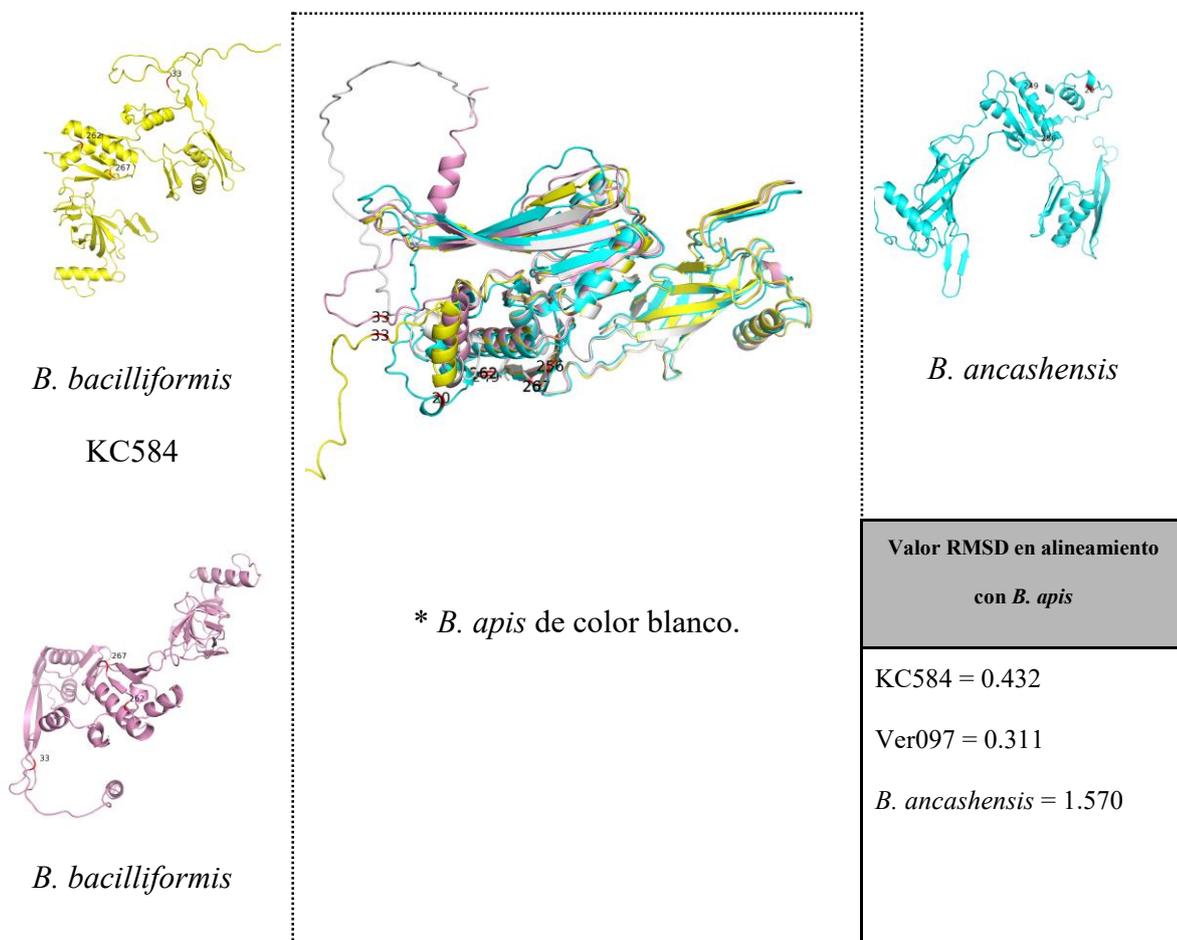
**Tabla 13:** Resumen de posiciones bajo selección positiva que exhibieron una diferenciación química al ser comparadas con la secuencia de referencia.

Categoría	N° incidencias	Posiciones
No cambio químico	18	V57G (FlgI), S286G (FlgI), A293S (FlgI), P213Q (FlgK), V340P (FlgK), V322M (FlgK), T77S (FliG), A134K (FliG), L152T (FliG), M185A (FliG), K220E (FliG), D255E (FliG), V291A (FliG), G309A (FliG), A363K (FliG), S79T (MotB), Q99P (MotB), Q133D (MotB), E152S (MotB)
Hidrofílico → Hidrofóbico	7	Q99P (MotB), V291A (FliG), G309A (FliG), A300S (MotB), Q350A (MotB), V322M (FlgK), T79K (FlgD)
Hidrofóbico → Hidrofílico (polar no cargado)	6	S286G (FlgI), T103E (FlgD), T79K (FlgD), S85A (FlgD), E148T (Omp89), V470A (Omp89)
Hidrofóbico → Hidrofílico (básico)	3	V340P (FlgK), A134K (FliG), K220E (FliG)
Hidrofílico (polar no cargado) → Hidrofílico (básico)	4	S13R (FlgD), Q14G (FlgD), E152S (MotB), T103E (FlgD)
Hidrofílico (básico) → Hidrofílico (ácido)	2	Q133D (MotB), K220E (FliG)

Hidrofílico → Hidrofílico (mismo grupo)	5	I11F (FlgD), V115K (FlgD), K141T (FlgD), G155T (FlgD), S69M (FlgD)
Hidrofílico (ácido) → Hidrofílico (polar no cargado)	2	E152S (MotB), E148T (Omp89)

Finalmente, de los genes sometidos a la prueba, 6 exhibieron la posibilidad de estar bajo selección natural. Curiosamente, 5 de estos pertenecen a la categoría “Motilidad” (*flgI*, *flgK*, *flgC*, *fliG* y *motB*), mientras que 1 pertenece a la categoría “Adherencia” (*omp89*). Estos fueron los elegidos para ser sometidos a modelamiento y alineamiento estructural (Figura 19).

### A) FlgI



Ver097

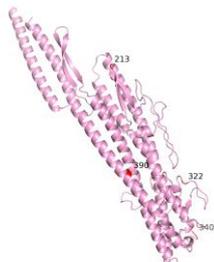


**B) FlgK**



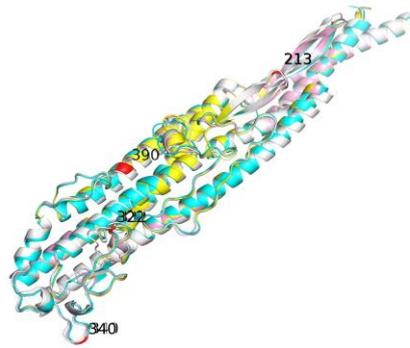
*B. bacilliformis*

KC584

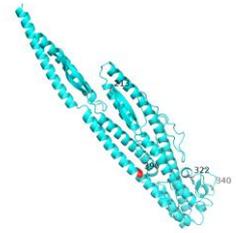


*B. bacilliformis*

Ver097



\* *B. apis* de color blanco.



*B. ancashensis*

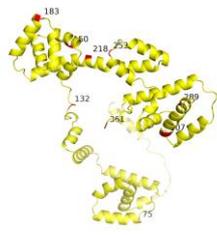
Valor RMSD en alineamiento  
con *B. apis*

KC584 = 0.578

Ver097 = 0.529

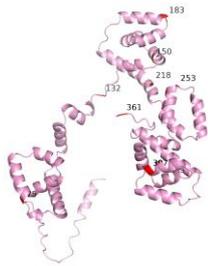
*B. ancashensis* = 0.648

**C) FlgG**



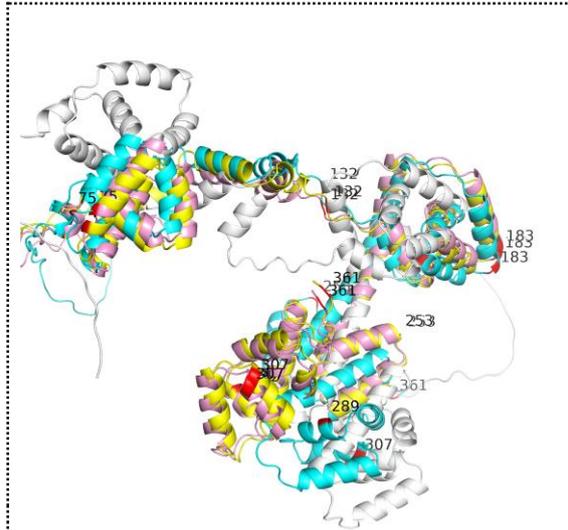
*B. bacilliformis*

KC584

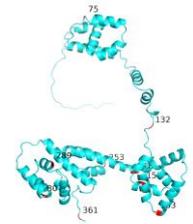


*B. bacilliformis*

Ver097



\* *B. apis* de color blanco.



*B. ancashensis*

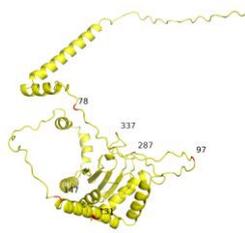
Valor RMSD en alineamiento  
con *B. apis*

KC584 = 21.414

Ver097 = 20.560

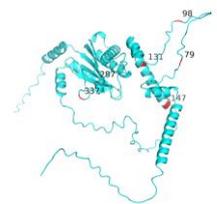
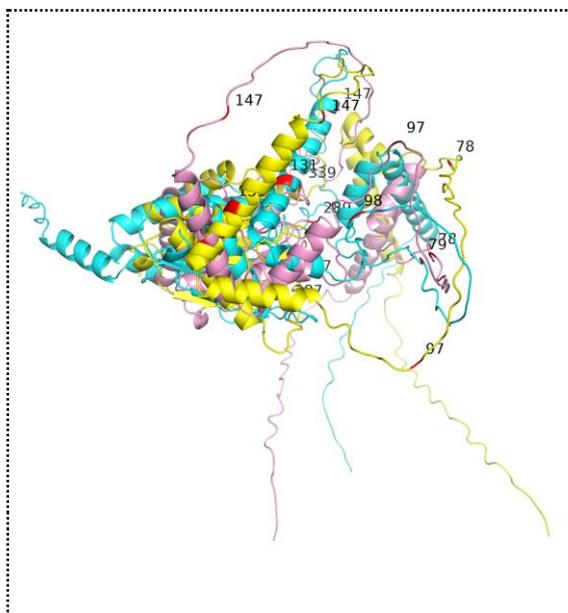
*B. ancashensis* = 15.223

#### D) MotB



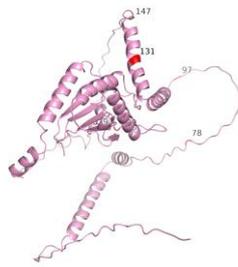
*B. bacilliformis*

KC584



*B. ancashensis*

Valor RMSD en alineamiento



*B. bacilliformis*

Ver097

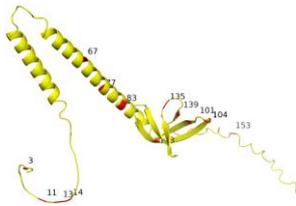
\* *B. apis* no tiene esta proteína.

con *B. ancashensis*

KC584 = 17.621

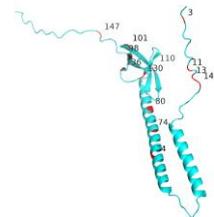
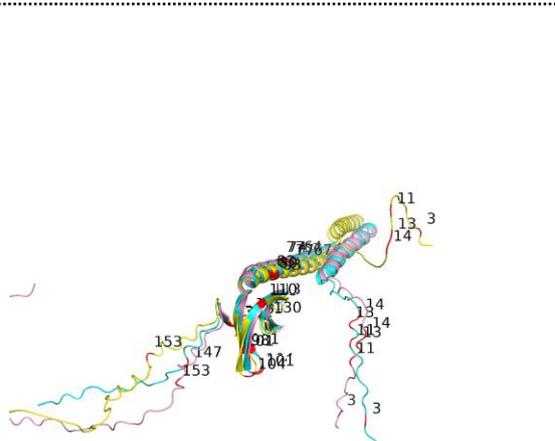
Ver097 = 18.987

### E) FlgD

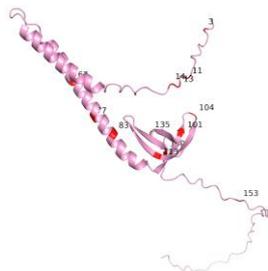


*B. bacilliformis*

KC584



*B. ancashensis*



*B. bacilliformis*

\* *B. apis* no tiene esta proteína.

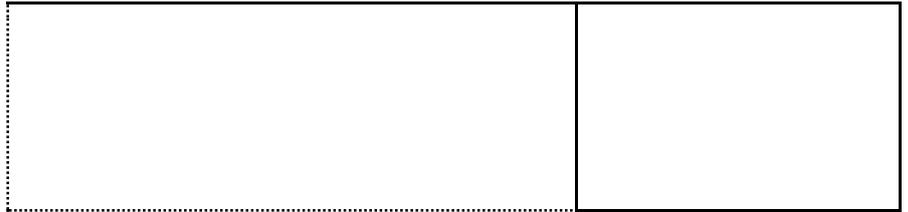
Valor RMSD en alineamiento

con *B. ancashensis*

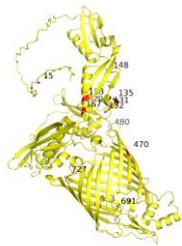
KC584 = 3.041

Ver097 = 1.162

Ver097

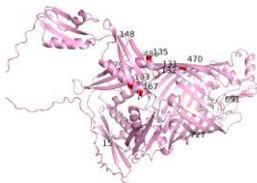


### F) Omp89



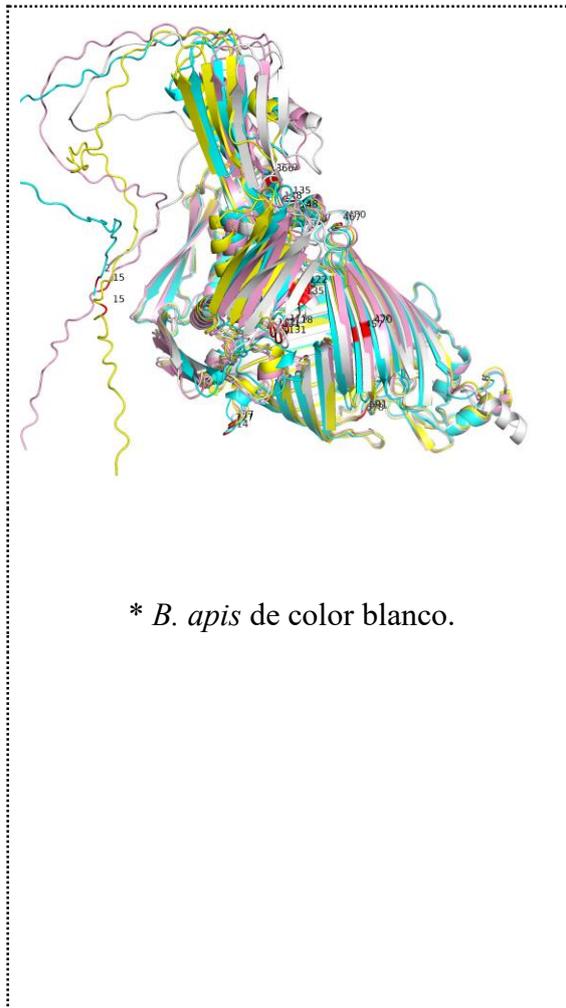
*B. bacilliformis*

KC584

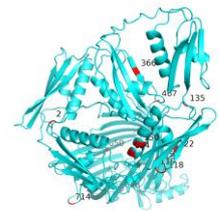


*B. bacilliformis*

Ver097



\* *B. apis* de color blanco.



*B. ancashensis*

Valor RMSD en alineamiento  
con *B. apis*

KC584 = 1.348

Ver097 = 1.489

*B. ancashensis* = 1.764

**Figura 19.** Modelamiento y alineamiento estructural de las 6 proteínas bajo selección positiva: A) FlgI, B) FlgK, C) FliG, D) MotB, E) FlgD y F) Omp89. En el caso de FlgI, FlgK, FliG y MotB, la selección positiva fue comprobada usando la primer hipótesis, H1 (Bb y Bb Ver097 versus el género *Bartonella*), mientras que

en el caso de FlgD y Omp89, la selección positiva fue comprobada usando la segunda hipótesis, H2 (Bb, Bb Ver097 y *B. ancashensis* versus el género *Bartonella*). En el cuadro derecho inferior, figuran los valores de distancia RMSD del alineamiento estructural realizado.

En el alineamiento estructural, el valor RMSD (error raíz media cuadrática en español) cercano a 0 indica estructuras más parecidas entre sí, mientras que valores más lejanos indican estructuras que son más disímiles entre sí. Para este alineamiento, se realizó la comparación con estructuras de *B. apis*, siendo esta la única especie no patogénica del género *Bartonella*.

En estas comparaciones, las proteínas MotB y FliG de Bb y Bb Ver097 exhibieron un RMSD considerable al compararlas con proteínas de otras especies, con valores de 17.621 y 18.987 respectivamente para MotB al comparar estas proteínas con la proteína MotB de *B. ancashensis*. Asimismo, valores de 21.414 y 20.560 respectivamente fueron obtenidos para el alineamiento estructural de FliG al compararlo con la proteína FliG de *B. apis*. Es importante resaltar que, a comparación de las proteínas de Bb y Bb Ver097, *B. ancashensis* exhibió un RMSD de 15.223 al compararlo con la proteína de *B. apis*.

## VII. DISCUSIÓN

En este estudio se buscó trazar la evolución de los genes de virulencia dentro del linaje L1 del género *Bartonella* comparándolas con el resto de especies de este género. Primero, se realizó una reconstrucción filogenética de todas las especies actualmente consideradas dentro de este género. Como indica la literatura, las especies *B. bacilliformis* y *B. ancashensis* pertenecen al linaje L1 (23). Además, se secuenciaron 14 nuevos genomas pertenecientes al género *Bartonella*, en donde 13 fueron identificados como *B. bacilliformis* y 1 como *B. ancashensis*. Dentro de la especie *B. bacilliformis*, 6 fueron identificadas como pertenecientes a la cepa Ver097, una cepa que ha estado largamente involucrada en una discusión taxonómica (91,92). Los genomas publicados, entonces, colaborarán con futuros esfuerzos para una mejor caracterización de esta especie.

### A. Divergencia de Ver097 del resto de *B. bacilliformis*

La clasificación taxonómica en el dominio Bacteria es un campo complejo que se enfrenta a varios retos. En particular, el impacto de sucesos como la transferencia horizontal de genes (THG) en la evolución bacteriana ha planteado muchas interrogantes sobre cómo es que conceptualizamos las especies bacterianas. En el pasado, las observaciones fenotípicas constituían la base de la categorización taxonómica. Sin embargo, el auge de las tecnologías actuales de secuenciación ha permitido agregar una nueva dimensión a la interrogante. Por ejemplo, se puede argumentar que la existencia de un genoma central ha apoyado la idea de la delimitación de las especies en bacterias (93). En el año 2006 el papel de la genómica en la definición de una especie bacteriana seguía siendo discutido,

argumentando que en ciertos grupos taxonómicos no se podían encontrar clusters definitivos mediante filogenómica, mientras que otros (especialmente patógenos) sí podían formar perfectamente un cluster. Así pues, se argumentó que los nichos ecológicos tenían una influencia significativa en cómo se agrupaban determinadas bacterias (94).

Thompson et. al. argumentaron que la clasificación bacteriana podría funcionar utilizando únicamente datos genómicos, y sugirieron cinco indicadores genómicos para discernir entre diferentes especies. Estos indicadores son los siguientes: i) índice de Karlin (<10 sugiere que ambos genomas pertenecen a la misma especie), ii) *Average Aminoacidic Identity* o AAI (>95% de AAI sugiere que ambos genomas pertenecen a la misma especie), iii) *Average Nucleotidic Identity* o ANI (>95% de ANI para un criterio de especie), iv) >95% de identidad basada en genes de alineación múltiple y v) 70% en el índice *Genome-to-Genome Hybridization similarity* (GGDH *in silico*) (95).

Durante los análisis de genómica comparativa realizados en este estudio a los genomas de Bb obtenidos, resaltó lo divergente que era la cepa Ver097 a comparación del resto. Esta cepa presentó un ANI inferior al esperado para cepas que se consideran de la misma especie. Asimismo, en el resto de indicadores de identidad genómica discutidos por Thompson et. al., como AAI (< 95,38 entre Ver097 y KC584) y GGDH (< 70 entre Ver097 y KC584), Ver097 mostró valores fuera de lo esperado para cepas agrupadas dentro de la especie Bb (Figura S1, data no publicada). Se ha evaluado la longitud del genoma de Ver097 y se distingue un

genoma menor en longitud a comparación de la cepa de referencia, además de exhibir el menor número de secuencias codificantes a comparación de otras cepas de Bb.

Estos hallazgos indicarían que Ver097 pertenece a un clado distinto, ya sea dentro de la misma especie *B. bacilliformis* o como otra especie. Tomando en cuenta que los objetivos del presente trabajo requerían de genómica comparativa, se consideró a esta cepa como una especie distinta basándonos en los criterios indicados por Thompson et. al.. Sin embargo, es necesario una mayor caracterización de esta cepa a otros niveles para poder determinar su categoría taxonómica. Estos resultados abren nuevas vías de investigación para entender una posible divergencia dentro de la especie *B. bacilliformis*.

En esta discusión, es importante considerar el caso de *B. schoenbuchensis* y *B. melophagi*, dos especies de *Bartonella* que fueron divididas en especies distintas, pero que podrían pertenecer a un mismo complejo taxonómico (96). Este podría ser el caso de *B. bacilliformis*.

### **B. Los genes de virulencia de la especie *B. bacilliformis* son parte del genoma *core* de la especie más no del género *Bartonella***

En el análisis pangenómico, a grandes rasgos, los genes se dividen en dos categorías principales: los genes *core* y los genes accesorios. Algunas herramientas dividen esta última categoría en genes *soft core*, *shell* y *cloud*, dependiendo de la incidencia de genes en el número de cepas evaluadas. Si un gen se encuentra en todos los genomas analizados, entonces se considera que pertenece al genoma *core*. Por otra

parte, los genes que se encuentran esporádicamente en uno o varios genomas pero que no llegan a estar en todos, se consideran genes accesorios.

Los genes *core* se consideran como fundamentales para el estilo de vida del grupo taxonómico evaluado. En este estudio, se determina que el genoma *core* de la especie Bb es el mayor componente del pangenoma, siendo que 1027 genes lo componen. Son pocos los genes que pertenecen a menos del 15% de cepas evaluadas (n = 60). Los genes de virulencia evaluados a nivel de la especie Bb se encuentran en el genoma *core*. Es necesario considerar en la discusión que Ver097 fue incluido en este análisis pangenómico, debido a su estatus taxonómico actual. Es así como se comprobó que 57 genes solo pertenecen a los genomas de Ver097. Se resalta que ninguno de estos fue identificado como factor de virulencia, indicando que la diferencia entre las diferentes cepas de Bb no se encuentra en sus factores de virulencia.

Este tipo de estructura pangenómica, en donde el genoma *core* es de gran tamaño a comparación del genoma accesorio, es común en patógenos intracelulares. Es el caso de *Legionella pneumophila*, bacteria gram-negativa perteneciente a las  $\gamma$ -Proteobacterias y patógeno intracelular facultativo que invade habitualmente protozoarios. Esta especie tiene un pangenoma compuesto en un 66.9% por su genoma *core* (n = 1979 de un total de 2957 genes) (97). También es el caso de *Corynebacterium pseudotuberculosis*, patógeno intracelular facultativo perteneciente a la clase Actinomycetia. Su genoma *core* se compone de 1504 genes, correspondientes al 54% del pangenoma de toda la especie (98). Por lo tanto, la

estructura pangenómica mostrada por Bb no es inusual, y corresponde a lo esperado en patógenos intracelulares. Esto se debe al nicho ecológico que ocupa. Es conocido que los patógenos intracelulares pasan por un proceso de pérdida de genes, producto probablemente de la deriva génica a la que se someten. Esto impacta en su pangenoma, en dónde la presencia de genes accesorios es inusual, mientras que el componente *core* es más relevante (99).

A nivel de género, sin embargo, el peso del genoma *core* es más reducido, mientras que el genoma accesorio pasa a ser el de mayor tamaño, siendo que 9524 genes del pangenoma en total están presentes en menos del 15% de cepas evaluadas. Es importante resaltar que el resto de patógenos en el género *Bartonella* son intracelulares también, exceptuando uno. Sin embargo, considerando el evento de radiación adaptativa que derivó en las especies del grupo Eubartonellae, además de la diversidad de nichos ecológicos presentes en el género, representado por la diversidad de hospederos, el tamaño del genoma *core* resultaría en una evidencia de las diversas adaptaciones experimentadas en el género *Bartonella*, y en cómo se ha adaptado a diversos hospederos, que vendrían a significar nichos distintos entre sí (2).

Resulta interesante que los genes de virulencia no estén presentes en el genoma *core* del género pero que presenten una estructura conservada. Las tres categorías de virulencia presentan una estructura y distribución muy similar entre sí, como es en el caso del aparato móvil (Figura 17). Y, cuando no se observa una conservación en la distribución como en las proteínas de adherencia o de unión al grupo heme, de

todos modos, el alineamiento mediante TBLASTX muestra que hay una significancia en estos alineamientos mediante el valor E superior a  $1e^{-4}$  (Figura 15 y Figura 16).

Es importante recalcar que, para que un gen pertenezca al genoma *core*, este tiene que superar un porcentaje de identidad de BLASTP, que es el algoritmo utilizado por el paquete Panaroo para refinar la primera clusterización del proceso (100). Sin embargo, en el alineamiento genético realizado individualmente entre los factores de virulencia presentados, se utilizó TBLASTX, un algoritmo que traduce las secuencias de nucleótidos a secuencias de aminoácidos. Esto permite una mayor amplitud al momento de evaluar una probable homología entre 2 secuencias.

Por lo tanto, se podría decir que las secuencias aminoacídicas entre las diferentes especies en el género *Bartonella* divergen entre sí debido a eventos que pueden alterar su identidad, tales como las mutaciones o deleciones. Sin embargo, al momento de comparar estas secuencias mediante el algoritmo TBLASTX, que puede tomar en cuenta varios inicios de lectura, se observa que aún hay relación entre estos genes, y que su conservación a este nivel probablemente ha sido crucial para la ocupación de nichos ecológicos del género *Bartonella*.

**C. *B. ancashensis* es una especie muy diferenciada de *B. bacilliformis* a partir de distintos índices**

*B. ancashensis* es la otra especie que acompaña a Bb en el Linaje 1 del grupo Eubartonellae. Esta especie se caracteriza por contar con un sistema de excreción

*virB/virD4* que es un factor de virulencia clave que se utiliza para translocar las proteínas efectoras de *Bartonella* (Beps) en las células huésped con el fin de subvertir sus funciones (101). A lo largo de los análisis realizados en este estudio, también exhibe otras particularidades, como un ANI de 0.8325 (Tabla S1) al compararlo con la cepa de referencia KC584, evidenciando lo distante que es su genoma. A nivel filogenético, su rama se muestra distante de la cepa de referencia de Bb. Asimismo, dos proteínas asociadas a la virulencia de *B. ancashensis* exhibieron valores significativos de presión selectiva, junto con las proteínas de Bb y Bb Ver097 (FlgD y Omp89). Si bien este estudio no se centró en *B. ancashensis*, algunas de sus características genómicas fueron resaltantes, y demuestra que aún hay por explorar en esta especie recientemente descrita e incorporada en el género *Bartonella*.

#### **D. Varios genes exclusivos a *B. bacilliformis* y *B. bacilliformis* Ver097 están relacionados a funciones de membrana y adherencia**

Dentro de los genes exclusivos hallados en la reconstrucción pangénomica, es de interés notar que varios de los genes señalados como únicamente presentes en Bb y Bb Ver097 pertenecen a genes expresados a nivel de la membrana. Se puede nombrar, por ejemplo, a la proteína CvpA, implicada en la homeostasis de la membrana, algunos genes flagelares como *motB* y porinas. Estos genes pertenecen a la categoría de genes *cloud* debido a que los análisis pangénómicos toman en cuenta la identidad en un BLASTP al agrupar genes. Esto quiere decir que los genes previamente nombrados han sido agrupados entre ellos y no en el resto de especies

de *Bartonella* debido a que superaron un umbral de identidad. Como ejemplo, *motB* puede estar presente en el resto de especies flageladas de *Bartonella*. Sin embargo, al momento de comparar las identidades de estas secuencias, el paquete informático Panaroo agrupó a los genes *motB* de Bb y Bb Ver097 en un grupo y a los demás genes *motB* en otro grupo.

El énfasis en proteínas de membranas podría ser resultado de una adaptación particular de Bb a su rol patogénico intracelular teniendo a humanos como hospederos únicos. En el género *Bartonella*, el rol de ciertas proteínas de membrana como la adhesina Bartonella adhesin A (BadA) en *B. henselae* o Vombs en *B. quintana* son esenciales para la adhesión a las proteínas de la matriz celular de los eritrocitos. Por otra parte, el escape inmunológico de *Bartonella* se da por componentes liposacáridos no reconocidos por receptores Toll tipo 4 (TLR4) debido a su inusual composición. BadA también puede participar de este escape (102). En el caso de *B. bacilliformis*, se ha descrito que parte de su invasión inmunológica a receptores Toll tipo 5 (TLR5) se debe a un motivo encontrado en la secuencia de flagelina (*flaA*) que posee (25). Por lo tanto, estas proteínas podrían tener un rol similar en la patogénesis de Bb.

Se requeriría un análisis con mayor detalle para poder determinar si estas divergencias son significativas a nivel funcional y si la adaptación al hospedero es influyente. Se sabe que secuencias aminoacídicas con más del 50% de identidad tienden a exhibir una mayor probabilidad a presentar funciones disímiles entre sí. Mientras que con una identidad menor al 40%, hay una mayor probabilidad de que

la funcionalidad no pueda ser determinada (103). En este estudio, se utilizó una identidad del 70% para la clusterización de genes durante el análisis pangénomico. Esto siguiendo un protocolo propuesto previamente para evaluar pangenomas a nivel de género. Sin embargo, una clusterización a partir del 50% podría ser útil para agrupar en base a funcionalidad, y poder determinar si efectivamente las divergencias a nivel de secuencia podrían traducirse a divergencias a nivel funcional.

#### **E. Existe una presión selectiva sobre genes relacionados a la adherencia y la motilidad**

Se sometió a los genes de virulencia de Bb a una prueba de selección positiva. De los genes de virulencia obtenidos mediante una identificación usando la base de datos VFDB, seis exhibieron la posibilidad de estar bajo selección natural. Curiosamente, cinco de estos pertenecen a la categoría “Motilidad” (*flgI*, *flgK*, *flgC*, *fliG* y *motB*), mientras que uno pertenece a la categoría “Adherencia” (*omp89*). Sin embargo, se debe tomar en cuenta que, en el caso de Bb y Bb Ver097, los genes asociados a la motilidad cumplen una función de adhesión. Observando que son estos los genes bajo presión selectiva positiva, se podría resolver la hipótesis del estudio, y plantear que estos genes están bajo esta presión debido a su adaptación a sus hospederos humanos.

El eritrocito humano presenta algunas particularidades a comparación de los eritrocitos de otros animales: Sharma et. al realizaron un análisis comparativo de las membranas eritrocíticas de varios mamíferos, determinando que hay una

diferencia en el perfil de sialoglicoprotéico dependiendo de la especie analizada (104). Las sialoglicoproteínas son glicoproteínas que contienen también un ácido siálico. Las glicoforinas son proteínas transmembrana de los eritrocitos, altamente glicosiladas en la cadena externa a la superficie. Su tamaño va desde 20 hasta 30 kDa. Hay cuatro tipos de glicoforinas: A, B, C y D. Todas estas se encuentran presentes en los eritrocitos humanos. Si bien no tienen función desconocida, contribuyen en la carga negativa observada en la superficie eritrocítica. Se ha observado que distintos patógenos interactúan con estas glicoproteínas, tales como virus o parásitos (105). En el caso de *Plasmodium falciparum*, por ejemplo, en el que su ligando se une preferentemente a la glicoforina A. También puede unirse a la glicoforina B si la primera no está disponible, ya que es estructuralmente similar a la glicoforina A. El ligando EBA-175, una de las proteínas funcionales de unión de *P. falciparum*, requiere interactuar con los residuos de ácido siálico, además de interactuar también con la cadena polipeptídica de la glicoforina (106).

En el caso de las bacterias, la adhesión a una superficie celular ocurre con una primera interacción hidrofóbica, producto de que las membranas de la célula hospedera y el patógeno están cargadas negativamente. Este impedimento inicial debe de ser traspasado por la bacteria para lograr su cometido, por lo que depende de las interacciones hidrofóbicas para ello. Luego de esto, la bacteria requerirá poder unirse a su hospedero de una manera más fija. Para ello, hará uso de adhesinas específicas, que le permiten unirse a su huésped de manera complementaria. En bacterias gram-negativas, las fimbrias o pili son las principales estructuras

adhesivas, utilizadas por microorganismos tales como la *E. coli* uropatogénica o *Salmonella* (107).

En el caso de Bb y Bb Ver097, se ha comprobado mediante experimentos de interacción de proteínas que Bb KC584 se une a  $\alpha$  y  $\beta$  espectrina, además de a las glicoforinas A y B. Removiendo el ácido siálico cargado negativamente utilizando neuraminidasa, las interacciones con Bb KC584 se incrementaron, sugiriendo que el ácido siálico actúa ocultando los sitios de unión. Se comprobó también la especificidad de las interacciones de Bb KC584 con los eritrocitos, debido a que se observó que las proteínas de *E. coli* no se unían con estas proteínas eritrocíticas (108).

Por lo tanto, se podría especular que las condiciones de la membrana eritrocítica humana ejercen una presión selectiva sobre algunos genes de adherencia de Bb y Bb Ver097. Según los resultados observados, varios de los sitios bajo presión selectiva exhiben un cambio de propiedades químicas, varios de ellos hacia características hidrofílicas de distintos tipos (sin carga, ácidas o básicas). Por otro lado, los sitios que pasan de ser hidrofílicos a hidrofóbicos más bien son menos comunes en lo evaluado. Habría, entonces, una tendencia hacia la presencia de aminoácidos con características hidrofílicas, favoreciendo quizá la segunda parte de la adhesión de la bacteria con la membrana, luego de las interacciones hidrofóbicas iniciales. Esto considerando las características propias de la membrana eritrocítica humana, y conociendo que la unión a las glicoforinas A y B es un momento fundamental en la infección de Bb y Bb Ver097.

La gran mayoría de sitios bajo presión selectiva, sin embargo, no exhibieron un cambio químico significativo ( $n = 18$ ). Pero esto no quiere decir que no haya ventajas considerables producto del cambio de aminoácidos: Se sabe, por ejemplo, que la glicina puede alterar la flexibilidad de una proteína (ej. V57G en FlgI). Asimismo, la prolina pone rigidez en una estructura, al tener un anillo. Al ser sustituida, esto puede otorgar más flexibilidad a la cadena aminoacídica (ej. Q99P en MotB). Entonces, si bien en estos 18 sitios no se observó un cambio químico significativo, si es posible que a nivel estructural y dinámico exista un cambio en cómo las proteínas cumplen su función. A nivel de alineamiento estructural, de hecho, algunas proteínas de Bb exhibieron un gran contraste al ser comparadas con su homólogo en la especie no patogénica, *B. apis*. Es el caso de FliG, en dónde ambas estructuras de Bb y Bb Ver097 exhibieron un RMSD mayor a 20. Interesantemente, el otro patógeno perteneciente a este linaje L1, *B. ancashensis*, exhibió una mayor cercanía estructural con la proteína de *B. apis*. Quizá, entonces, a este nivel se están apreciando adaptaciones en la conformación de estas proteínas producto del rol en la infección de Bb y Bb Ver097. Es el caso también de MotB, en dónde las pertenecientes a Bb y Bb Ver097 exhiben una diferenciación al compararlas con su homólogo en *B. ancashensis*.

Se requeriría una dinámica molecular, probablemente, y experimentos de interacción entre proteínas para poder determinar, finalmente, si las proteínas y, específicamente, los sitios que se han descrito en este trabajo, efectivamente son relevantes en el aún poco comprendido proceso de infección de *B. bacilliformis*.

## VIII. CONCLUSIÓN

En este estudio se analizaron los genes asociados a la virulencia de *Bartonella bacilliformis* identificando señales de selección natural en genes previamente asociados con la invasión del eritrocito humano.

La reconstrucción del pangenoma de *B. bacilliformis* determinó un genoma *core* amplio compuesto por 1027 genes esenciales para esta especie. Los factores de virulencia identificados se encuentran dentro de este genoma *core* confirmando la importancia funcional de los factores de virulencia.

La comparación entre los factores de virulencia identificados utilizando una base de datos y la literatura reveló discrepancias entre ambas fuentes pues la base de datos de virulencia carece de factores reportados en estudios anteriores para la especie. Esto pone de manifiesto la necesidad de integrar ambos datos para futuros trabajos genómicos en *B. bacilliformis*.

Los genomas de alta calidad obtenidos a partir de nuevos aislados demostraron que la cepa Ver097 presenta una divergencia genómica ANI de 94% con respecto a las demás cepas de *B. bacilliformis*. Esta divergencia está por debajo del umbral de 95% establecido para individuos de la misma especie, indicando que Ver097 representa una especie distinta dentro del Linaje 1 bajo criterios de genómica comparativa. La confirmación de este estatus taxonómico requiere análisis fisiológicos y morfológicos complementarios.

La filogenia del género *Bartonella*, reconstruida a partir de los genes *core*, confirmó la adaptación del género a diversos hospederos. Además, las pruebas de presión selectiva sobre genes asociados a la virulencia de *B. bacilliformis* identificaron regiones bajo selección positiva, siendo la mayoría genes de adherencia. Estos hallazgos bioinformáticos sugieren que la adherencia constituye un mecanismo clave para la virulencia de *B. bacilliformis* en humanos, amplían la comprensión teórica del arsenal patogénico de *B. bacilliformis* y proporcionan candidatos moleculares para estudiar el proceso de invasión al eritrocito.

## IX. LIMITACIONES

El bajo número de genomas disponibles y la data asociada a ellos vendría a ser una limitante importante. Una de las teorías de la divergencia de Ver097 vendría a ser que sería un ecotipo o biotipo que es distinto debido al nicho ecológico que ocupa. Sin embargo, con la escasa información existente respecto a periodos de colecta o detalles de la misma, no es posible explorar mejor esta teoría.

Encontramos otra limitación a nivel de la información disponible en las bases de datos respecto a la virulencia en *B. bacilliformis*. La base de datos VFDB, acaso de las bases de datos más activas en estos años, ha tenido su última versión en el año 2022. Y, sin embargo, la literatura reporta otros genes de virulencia que no aparecen listados en la misma. Se observa una barrera respecto al conocimiento que fue caracterizado con métodos microbiológicos y moleculares y a los criterios actuales que se consideran en genómica. Para poder hacer a este estudio más reproducible, se eligieron únicamente los factores de virulencia que pudieran ser reconocidos a través de métodos genómicos.

En este estudio se hicieron pruebas de selección positiva únicamente en regiones sin gaps o deleciones. Es por ello que nada más se comenta los resultados de estas pruebas. Otros estudios señalan la importancia de las deleciones en la evolución de *B. bacilliformis*, pero si bien esto ha sido mencionado a lo largo del documento, no ha sido tomado en cuenta al momento de realizar las pruebas.

Para poder comprobar los supuestos obtenidos en este estudio mediante métodos *in silico*, se requiere primero pasar por una dinámica molecular que permita observar los efectos de estas regiones bajo selección positiva. Asimismo, se requiere realizar pruebas experimentales en medios controlados que permitan observar la importancia de estos factores en la infección de *B. bacilliformis*.

## **X. RECOMENDACIONES**

En relación a estos hallazgos, la primera recomendación sería aumentar el secuenciamiento genómico de *B. bacilliformis* y *B. ancashensis*, como las dos especies del Linaje 1 que actualmente tienen sus zonas endémicas en el Perú. Este secuenciamiento tiene que estar acompañado de un seguimiento detallado del muestreo y las condiciones del mismo. Solo de esta forma se podría obtener mayor información respecto a estos taxones.

Por otra parte, la caracterización molecular de una infección por *B. bacilliformis* será de mucha importancia en un futuro, porque podrá ayudar a determinar algunos elementos involucrados que probablemente por el momento no se estén tomando en cuenta. A partir de ello, se podrán actualizar también las bases de datos de virulencia, de modo tal que pueda haber una mejor caracterización genómica de *B. bacilliformis*. Asimismo, las conclusiones observadas en este estudio podrían corroborarse con este tipo de análisis.

## XI. REFERENCIAS BIBLIOGRÁFICAS

1. Ruiz J. JMM Profile: Bartonella bacilliformis: a forgotten killer. *J Med Microbiol.* 2022;71(12):001614.
2. Segers FH, Kešnerová L, Kosoy M, Engel P. Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *ISME J.* mayo de 2017;11(5):1232-44.
3. Gomes C, Ruiz J. Carrion's Disease: the Sound of Silence. *Clin Microbiol Rev.* 29 de noviembre de 2017;31(1):e00056-17.
4. Allison MJ, Pezzia A, Gerszten E, Mendoza D. A case of Carrion's disease associated with human sacrifice from the Huari culture of Southern Peru. *Am J Phys Anthropol.* 1974;41(2):295-300.
5. Ministerio de Salud. Enfermedad de Carrión (Bartonelosis) en el Perú. Ministerio de Salud, OGE, INS; 2001. 88 p. (Módulos Técnicos. Serie de Documentos Monográficos; vol. 13).
6. Salinas D. Daniel Alcides Carrión: la teoría unicista. *Rev Fac Med.* enero de 2016;64(1):93-7.
7. Enfermedad de Carrión : Inoculaciones en Seres humanos [Internet]. [citado 24 de diciembre de 2024]. Disponible en:  
[https://sisbib.unmsm.edu.pe/bvrevistas/anales/v61\\_n4/enfer\\_carr.htm](https://sisbib.unmsm.edu.pe/bvrevistas/anales/v61_n4/enfer_carr.htm)
8. Hugo Vizcarra E. La Bartonella de Alberto Barton T. y las nuevas

Bartonellas. Folia Dermatológica Peru [Internet]. diciembre de 1997 [citado 24 de diciembre de 2024];8(4). Disponible en:

[https://sisbib.unmsm.edu.pe/bvrevistas/fofia/vol8\\_n4\\_dic97/bartonella.htm](https://sisbib.unmsm.edu.pe/bvrevistas/fofia/vol8_n4_dic97/bartonella.htm)

9. Saettone-León A. Enfermedad de Carrión: verruga peruana. DERMATOL PERU. 2018;28(1):70-84.
10. Centro Nacional de Epidemiología P y C de E. Número de casos de enfermedad de Carrión, Perú 2015-2020. MINSA. 2020;
11. Minnick MF, Anderson BE, Lima A, Battisti JM, Lawyer PG, Birtles RJ. Oroya Fever and Verruga Peruana: Bartonelloses Unique to South America. PLoS Negl Trop Dis. 17 de julio de 2014;8(7):e2919.
12. Huarcaya E, Maguiña C, Torres R, Rupay J, Fuentes L. Bartonelosis (Carrion's Disease) in the pediatric population of Peru: an overview and update. Braz J Infect Dis. octubre de 2004;8:331-9.
13. Gladis Ventura Egúsquiza. Diagnóstico Bacteriológico de la Bartonelosis humana o Enfermedad de Carrión [Internet]. Disponible en: [https://bvs.minsa.gob.pe/local/MINSA/1285\\_INS79.pdf](https://bvs.minsa.gob.pe/local/MINSA/1285_INS79.pdf)
14. Anaya E, Mendoza G, García-Uscamayta L, Fernández Y. Prueba de Elisa indirecta del lisado total de Bartonella bacilliformis para el diagnóstico rápido de la enfermedad de Carrión. Rev Peru Med Exp Salud Publica. abril de 2008;25(2):243-6.
15. Mallqui V, Spielmon EC, Verástegui M, Maguiña-Vargas C, Pinell-Salles P,

- Lavarello R, et al. Sonicated Diagnostic Immunoblot for Bartonellosis. Clin Diagn Lab Immunol. enero de 2000;7(1):1-5.
16. Ready PD. Biology of phlebotomine sand flies as vectors of disease agents. Annu Rev Entomol. 2013;58:227-50.
  17. Cohnstaedt LW, Caceres AG, Beati L, Munstermann LE. The Population Structure of *Lutzomyia verrucarum* (Diptera: Psychodidae), a *Bartonella bacilliformis* and *Leishmania peruviana* Vector in Peru. J Med Entomol. 1 de enero de 2012;49(1):77-84.
  18. Moo-Llanes DA, Arque-Chunga W, Carmona-Castro O, Yañez-Arenas C, Yañez-Trujillano HH, Cheverría-Pacheco L, et al. Shifts in the ecological niche of *Lutzomyia peruensis* under climate change scenarios in Peru. Med Vet Entomol. 2017;31(2):123-31.
  19. Ulloa GM, Vásquez-Achaya F, Gomes C, del Valle LJ, Ruiz J, Pons MJ, et al. Molecular Detection of *Bartonella bacilliformis* in *Lutzomyia maranonensis* in Cajamarca, Peru: A New Potential Vector of Carrion's Disease in Peru? Am J Trop Med Hyg. noviembre de 2018;99(5):1229-33.
  20. Holden N, Pritchard L, Toth I. Colonization outwith the colon: plants as an alternative environmental reservoir for human pathogenic enterobacteria. FEMS Microbiol Rev. 1 de julio de 2009;33(4):689-703.
  21. Garcia-Quintanilla M, Dichter AA, Guerra H, Kempf VAJ. Carrion's disease: more than a neglected disease. Parasit Vectors. 26 de marzo de 2019;12:141.

22. Maguiña Vargas C, Ugarte-Gil C, Breña Chávez P, Ordaya Espinoza E, Ventosilla López P, Huarcaya Castilla E, et al. Actualización de la enfermedad de Carrión. *Rev Medica Hered.* enero de 2008;19(1):36-41.
23. Wagner A, Dehio C. Role of distinct type-IV-secretion systems and secreted effector sets in host adaptation by pathogenic *Bartonella* species. *Cell Microbiol.* marzo de 2019;21(3):e13004.
24. Deng H, Pang Q, Zhao B, Vayssier-Taussat M. Molecular Mechanisms of *Bartonella* and Mammalian Erythrocyte Interactions: A Review. *Front Cell Infect Microbiol.* 12 de diciembre de 2018;8:431.
25. Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, et al. Evasion of Toll-like receptor 5 by flagellated bacteria. *Proc Natl Acad Sci.* 28 de junio de 2005;102(26):9247-52.
26. Focà A, Liberto MC, Quirino A, Matera G. Lipopolysaccharides: From Erinyes to Charites. *Mediators Inflamm.* 2012;2012:684274.
27. Seubert A, Hiestand R, De La Cruz F, Dehio C. A bacterial conjugation machinery recruited for pathogenesis. *Mol Microbiol.* 2003;49(5):1253-66.
28. Armstrong GL, MacCannell DR, Taylor J, Carleton HA, Neuhaus EB, Bradbury RS, et al. Pathogen Genomics in Public Health. *N Engl J Med.* 26 de diciembre de 2019;381(26):2569-80.
29. Baker S, Thomson N, Weill FX, Holt KE. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens.

- Science. 18 de mayo de 2018;360(6390):733-8.
30. Inzaule SC, Tessema SK, Kebede Y, Ogwel Ouma AE, Nkengasong JN. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. *Lancet Infect Dis.* septiembre de 2021;21(9):e281-9.
  31. Caugant DA, Brynildsrud OB. *Neisseria meningitidis*: using genomics to understand diversity, evolution and pathogenesis. *Nat Rev Microbiol.* febrero de 2020;18(2):84-96.
  32. Rouli L, Merhej V, Fournier PE, Raoult D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 26 de junio de 2015;7:72-85.
  33. Parmigiani L, Wittler R, Stoye J. Revisiting pangenome openness with *k*-mers. *Peer Community J* [Internet]. 2024 [citado 25 de diciembre de 2024];4. Disponible en:  
<https://peercommunityjournal.org/articles/10.24072/pcjournal.415/>
  34. Tonkin-Hill G, Gladstone RA, Pöntinen AK, Arredondo-Alonso S, Bentley SD, Corander J. Robust analysis of prokaryotic pangenome gene gain and loss rates with Panstripe. *Genome Res.* 1 de enero de 2023;33(1):129-40.
  35. Whelan FJ, Hall RJ, McInerney JO. Evidence for Selection in the Abundant Accessory Gene Content of a Prokaryote Pangenome. *Mol Biol Evol.* 1 de septiembre de 2021;38(9):3697-708.
  36. Caputo A, Fournier PE, Raoult D. Genome and pan-genome analysis to

- classify emerging bacteria. *Biol Direct*. 26 de febrero de 2019;14(1):5.
37. Mira A, Martín-Cuadrado AB, D'Auria G, Rodríguez-Valera F. The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol*. 20 de septiembre de 2010;13(2):45-57.
  38. Abram K, Udaondo Z, Bleker C, Wanchai V, Wassenaar TM, Robeson MS, et al. Mash-based analyses of *Escherichia coli* genomes reveal 14 distinct phylogroups. *Commun Biol*. 26 de enero de 2021;4(1):1-12.
  39. Dichter AA, Schultze TG, Becker SA, Tsukayama P, Kempf VAJ. Complete Genome Sequence of *Bartonella bacilliformis* Strain KC584 (ATCC 35686). *Microbiol Resour Announc*. 2 de enero de 2020;9(1):e01377-19.
  40. Saenz HL, Engel P, Stoeckli MC, Lanz C, Raddatz G, Vayssier-Taussat M, et al. Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. *Nat Genet*. diciembre de 2007;39(12):1469-76.
  41. Chaloner GL, Palmira Ventosilla, Birtles RJ. Multi-Locus Sequence Analysis Reveals Profound Genetic Diversity among Isolates of the Human Pathogen *Bartonella bacilliformis*. *PLoS Negl Trop Dis*. 19 de julio de 2011;5(7):e1248.
  42. Ruiz J, Pons MJ. Revisiting *Bartonella bacilliformis* MLST. *Infect Genet Evol*. 1 de septiembre de 2018;63:231-5.
  43. Mullins KE, Hang J, Jiang J, Leguia M, Kasper MR, Maguiña C, et al. Molecular Typing of "Candidatus *Bartonella ancashi*," a New Human

- Pathogen Causing Verruga Peruana. *J Clin Microbiol.* noviembre de 2013;51(11):3865-8.
44. Mullins KE, Hang J, Clifford RJ, Onmus-Leone F, Yang Y, Jiang J, et al. Whole-Genome Analysis of *Bartonella ancashensis*, a Novel Pathogen Causing Verruga Peruana, Rural Ancash Region, Peru. *Emerg Infect Dis.* marzo de 2017;23(3):430-8.
  45. Soni T. Endothelial Cell Factors Involved in *Bartonella Bacilliformis* Pathogenesis. *Biol Diss [Internet]*. 30 de abril de 2009; Disponible en: [https://scholarworks.gsu.edu/biology\\_diss/39](https://scholarworks.gsu.edu/biology_diss/39)
  46. Coleman SA, Minnick MF. Establishing a Direct Role for the *Bartonella bacilliformis* Invasion-Associated Locus B (IalB) Protein in Human Erythrocyte Parasitism. *Infect Immun.* julio de 2001;69(7):4373-81.
  47. Ruiz J, Gomes C. In silico analysis of Pap31 from *Bartonella bacilliformis* and other *Bartonella* spp. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis.* octubre de 2020;84:104482.
  48. Roden JA, Wells DH, Chomel BB, Kasten RW, Koehler JE. Hemin Binding Protein C Is Found in Outer Membrane Vesicles and Protects *Bartonella henselae* against Toxic Concentrations of Hemin. *Infect Immun.* marzo de 2012;80(3):929-42.
  49. O'Rourke F, Schmidgen T, Kaiser PO, Linke D, Kempf VAJ. Adhesins of *Bartonella* spp. En: Linke D, Goldman A, editores. *Bacterial Adhesion: Chemistry, Biology and Physics [Internet]*. Dordrecht: Springer Netherlands;

2011 [citado 25 de diciembre de 2024]. p. 51-70. Disponible en:

[https://doi.org/10.1007/978-94-007-0940-9\\_4](https://doi.org/10.1007/978-94-007-0940-9_4)

50. Xu YH, Lu ZY, Ihler GM. Purification of deformin, an extracellular protein synthesized by *Bartonella bacilliformis* which causes deformation of erythrocyte membranes. *Biochim Biophys Acta*. 22 de marzo de 1995;1234(2):173-83.
51. Derrick SC, Ihler GM. Deformin, a substance found in *Bartonella bacilliformis* culture supernatants, is a small, hydrophobic molecule with an affinity for albumin. *Blood Cells Mol Dis*. 2001;27(6):1013-9.
52. Shook B, Nelson K, Aguilera K, Braff L. Forces of Evolution. 9 de diciembre de 2019 [citado 25 de diciembre de 2024]; Disponible en: [https://pressbooks-dev.oer.hawaii.edu/explorationsbioanth/chapter/\\_\\_unknown\\_\\_-3/](https://pressbooks-dev.oer.hawaii.edu/explorationsbioanth/chapter/__unknown__-3/)
53. Stephen F. Schaffner, Pardis C. Sabeti. Scitable by Nature Education. 2008 [citado 25 de diciembre de 2024]. Evolutionary Adaptation and Positive Selection in Humans | Learn Science at Scitable. Disponible en: <https://www.nature.com/scitable/topicpage/evolutionary-adaptation-in-the-human-lineage-12397/>
54. Zeng K, Fu YX, Shi S, Wu CI. Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants. *Genetics*. noviembre de 2006;174(3):1431-9.
55. Zeng K, Shi S, Wu CI. Compound Tests for the Detection of Hitchhiking

- Under Positive Selection. *Mol Biol Evol.* 1 de agosto de 2007;24(8):1898-908.
56. Yang Z, dos Reis M. Statistical Properties of the Branch-Site Test of Positive Selection. *Mol Biol Evol.* 1 de marzo de 2011;28(3):1217-28.
57. Nandi T, Ong C, Singh AP, Boddey J, Atkins T, Sarkar-Tyson M, et al. A Genomic Survey of Positive Selection in *Burkholderia pseudomallei* Provides Insights into the Evolution of Accidental Virulence. *PLOS Pathog.* 1 de abril de 2010;6(4):e1000845.
58. Nielsen R. Molecular Signatures of Natural Selection. *Annu Rev Genet.* 1 de diciembre de 2005;39(1):197-218.
59. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol.* 1 de diciembre de 2000;15(12):496-503.
60. Plowright RK, Ahmed AN, Coulson T, Crowther TW, Ejotre I, Faust CL, et al. Ecological countermeasures to prevent pathogen spillover and subsequent pandemics. *Nat Commun.* 26 de marzo de 2024;15(1):2577.
61. El salto de la barrera de especie en la emergencia creciente de zoonosis - Anales RANM [Internet]. *Anales de la Real Academia Nacional de Medicina de España.* 2019 [citado 30 de abril de 2025]. Disponible en: [https://analesranm.es/revista/2018/135\\_02/13502rev04](https://analesranm.es/revista/2018/135_02/13502rev04)
62. Henriquez-Camacho C, Ventosilla P, Minnick MF, Ruiz J, Magaña C. Proteins of *Bartonella bacilliformis*: Candidates for Vaccine Development.

Int J Pept. 2015;2015(1):702784.

63. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* mayo de 2012;19(5):455-77.
64. Seemann T. mlst [Internet]. Disponible en: <https://github.com/tseemann/mlst>
65. Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. Icarus: visualizer for de novo assembly evaluation. *Bioinformatics.* 1 de noviembre de 2016;32(21):3321-3.
66. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genomics.* 5 de noviembre de 2021;7(11):000685.
67. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 22 de julio de 2020;21(1):180.
68. Kyle. pyUpSet [Internet]. 2015. Disponible en: <https://github.com/ImSoErgodic/py-upset>
69. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res.* 5 de julio de 2023;51(W1):W443-50.
70. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A.

- UniProtKB/Swiss-Prot. *Methods Mol Biol* Clifton NJ. 2007;406:89-112.
71. Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, et al. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res.* 20 de noviembre de 2024;gkae1082.
  72. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* 18 de noviembre de 2024;gkae1010.
  73. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol.* 1 de diciembre de 2021;38(12):5825-9.
  74. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol.* enero de 2015;32(1):268-74.
  75. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2 de julio de 2021;49(W1):W293-6.
  76. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 17 de diciembre de 2015;8(1):12-24.
  77. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 1 de enero de

2005;33(Database Issue):D325-8.

78. Zhou L, Feng T, Xu S, Gao F, Lam TT, Wang Q, et al. ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Brief Bioinform.* 1 de julio de 2022;23(4):bbac222.
79. Hackl T, Ankenbrand MJ, Adrichem B van. gggenomes: A Grammar of Graphics for Comparative Genomics [Internet]. 2024. Disponible en: <https://github.com/thackl/gggenomes>
80. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol.* 1 de agosto de 2007;24(8):1586-91.
81. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 14 de noviembre de 2019;20(1):238.
82. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol Biol Evol.* 1 de abril de 2000;17(4):540-52.
83. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods.* junio de 2022;19(6):679-82.
84. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.
85. Chaurasia R, Vinetz JM. In silico prediction of molecular mechanisms of toxicity mediated by the leptospiral PF07598 gene family-encoded virulence-

modifying proteins. *Front Mol Biosci.* 23 de enero de 2023;9:1092197.

86. Vázquez R, Gutiérrez D, Grimon D, Fernández L, García P, Rodríguez A, et al. The New SH3b\_T Domain Increases the Structural and Functional Variability Among SH3b-Like CBDs from Staphylococcal Phage Endolysins. *Probiotics Antimicrob Proteins* [Internet]. 30 de julio de 2024 [citado 26 de diciembre de 2024]; Disponible en: <https://doi.org/10.1007/s12602-024-10309-0>
87. Gundi VAKB, Taylor C, Raoult D, La Scola B. *Bartonella rattaaustraliani* sp. nov., *Bartonella queenslandensis* sp. nov. and *Bartonella coopersplainsensis* sp. nov., identified in Australian rats. *Int J Syst Evol Microbiol.* 2009;59(12):2956-61.
88. Mediannikov O, El Karkouri K, Robert C, Fournier PE, Raoult D. Non-contiguous finished genome sequence and description of *Bartonella florenciae* sp. nov. *Stand Genomic Sci.* 10 de octubre de 2013;9(1):185-96.
89. Gutiérrez R, Shalit T, Markus B, Yuan C, Nachum-Biala Y, Elad D, et al. *Bartonella kosoyi* sp. nov. and *Bartonella krasnovii* sp. nov., two novel species closely related to the zoonotic *Bartonella elizabethae*, isolated from black rats and wild desert rodent-fleas. *Int J Syst Evol Microbiol.* 2020;70(3):1656-65.
90. Li DM, Hou Y, Song XP, Fu YQ, Li GC, Li M, et al. High Prevalence and Genetic Heterogeneity of Rodent-Borne *Bartonella* Species on Heixiazi Island, China. *Appl Environ Microbiol.* diciembre de 2015;81(23):7981-92.

91. Paul S, Minnick MF, Chattopadhyay S. Mutation-Driven Divergence and Convergence Indicate Adaptive Evolution of the Intracellular Human-Restricted Pathogen, *Bartonella bacilliformis*. *PLoS Negl Trop Dis*. 11 de mayo de 2016;10(5):e0004712.
92. Mendoza-Mujica G, Flores-León D, Ruiz J. Molecular Characterization of Fluoroquinolone-Resistant *Bartonella bacilliformis*. *Pathogens*. 10 de julio de 2021;10(7):876.
93. Riley MA, Lizotte-Waniewski M. Population Genomics and the Bacterial Species Concept. En: Gogarten MB, Gogarten JP, Olendzenski LC, editores. *Horizontal Gene Transfer: Genomes in Flux* [Internet]. Totowa, NJ: Humana Press; 2009 [citado 26 de diciembre de 2024]. p. 367-77. Disponible en: [https://doi.org/10.1007/978-1-60327-853-9\\_21](https://doi.org/10.1007/978-1-60327-853-9_21)
94. Konstantinidis KT, Ramette A, Tiedje JM. The bacterial species definition in the genomic era. *Philos Trans R Soc B Biol Sci*. 11 de octubre de 2006;361(1475):1929-40.
95. Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, Thompson FL. Microbial genomic taxonomy. *BMC Genomics*. 23 de diciembre de 2013;14(1):913.
96. do Amaral RB, Cardozo MV, Varani A de M, Furquim MEC, Dias CM, de Assis WO, et al. First Report of *Bartonella* spp. in Marsupials from Brazil, with a Description of *Bartonella harrusi* sp. nov. and a New Proposal for the Taxonomic Reclassification of Species of the Genus *Bartonella*.

- Microorganisms. 9 de agosto de 2022;10(8):1609.
97. D'Auria G, Jiménez-Hernández N, Peris-Bondia F, Moya A, Latorre A. Legionella pneumophila pangenome reveals strain-specific virulence factors. BMC Genomics. 17 de marzo de 2010;11(1):181.
98. Soares SC, Silva A, Trost E, Blom J, Ramos R, Carneiro A, et al. The Pan-Genome of the Animal Pathogen *Corynebacterium pseudotuberculosis* Reveals Differences in Genome Plasticity between the Biovar ovis and equi Strains. PLoS ONE. 14 de enero de 2013;8(1):e53818.
99. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. The Ecology and Evolution of Pangenomes. Curr Biol CB. 21 de octubre de 2019;29(20):R1094-103.
100. Le DQ, Nguyen TA, Nguyen SH, Nguyen TT, Nguyen CH, Phung HT, et al. Efficient inference of large prokaryotic pangenomes with PanTA. Genome Biol. 6 de agosto de 2024;25(1):209.
101. Fromm K, Dehio C. The Impact of Bartonella VirB/VirD4 Type IV Secretion System Effectors on Eukaryotic Host Cells. Front Microbiol [Internet]. 15 de diciembre de 2021 [citado 26 de diciembre de 2024];12. Disponible en: <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.762582/full>
102. Jin X, Gou Y, Xin Y, Li J, Sun J, Li T, et al. Corrigendum: Advancements in understanding the molecular and immune mechanisms of Bartonella

pathogenicity. *Front Microbiol* [Internet]. 27 de julio de 2023 [citado 26 de diciembre de 2024];14. Disponible en:  
<https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2023.1260035/full>

103. Sangar V, Blankenberg DJ, Altman N, Lesk AM. Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinformatics*. 8 de agosto de 2007;8:294.
104. Sharma S, Punjabi V, Zingde SM, Gokhale SM. A Comparative Protein Profile of Mammalian Erythrocyte Membranes Identified by Mass Spectrometry. *J Membr Biol*. 1 de noviembre de 2014;247(11):1181-9.
105. Hollox EJ, Louzada S. Genetic variation of glycoporphins and infectious disease. *Immunogenetics*. 1 de junio de 2023;75(3):201-6.
106. Jaskiewicz E, Jodłowska M, Kaczmarek R, Zerka A. Erythrocyte glycoporphins as receptors for Plasmodium merozoites. *Parasit Vectors*. 24 de junio de 2019;12(1):317.
107. Ofek I, Bayer EA, Abraham SN. Bacterial Adhesion. En: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editores. *The Prokaryotes: Human Microbiology* [Internet]. Berlin, Heidelberg: Springer; 2013 [citado 26 de diciembre de 2024]. p. 107-23. Disponible en:  
[https://doi.org/10.1007/978-3-642-30144-5\\_50](https://doi.org/10.1007/978-3-642-30144-5_50)
108. Buckles EL, McGinnis Hill E. Interaction of Bartonella bacilliformis with human erythrocyte membrane proteins. *Microb Pathog*. 1 de septiembre de

2000;29(3):165-74.

## XII. ANEXOS

**Tabla S1.** Tabla de valores del ANI según alineamiento MAFFT. El límite inferior de especie en el índice ANI es de 0,95. Debajo de este valor, se considera que se están comparando 2 genomas de especies bacterianas distintas. De color verde claro, todos los valores que son igual o superiores a las 0,95. De color verde oscuro, todos los valores que son menores a 0,95.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	
1	USM-LMMB_07	1,0000	0,9898	0,9900	0,9910	0,9901	0,9440	0,9441	0,9909	0,9962	0,9951	0,9909	0,9441	0,9952	0,8340	0,9951	0,9439	0,9951	0,9899	0,8318	0,9441	0,9951	0,9950	0,9989	0,9987	0,9908	0,9952	0,9951	0,9952	0,9910	0,9917	0,9439	0,9441	
2	UPCH_1474_B14	0,9898	1,0000	0,9995	0,9896	0,9994	0,9439	0,9439	0,9891	0,9900	0,9899	0,9895	0,9439	0,9899	0,8371	0,9900	0,9438	0,9896	1,0000	0,8350	0,9438	0,9898	0,9898	0,9901	0,9898	0,9894	0,9899	0,9898	0,9900	0,9896	0,9898	0,9439	0,9439	
3	UPCH_1469_B06	0,9900	0,9995	1,0000	0,9894	0,9998	0,9437	0,9439	0,9891	0,9901	0,9901	0,9893	0,9438	0,9901	0,8371	0,9902	0,9437	0,9897	0,9995	0,8351	0,9438	0,9901	0,9901	0,9902	0,9900	0,9892	0,9902	0,9901	0,9903	0,9893	0,9899	0,9437	0,9438	
4	UPCH_1477_B17	0,9910	0,9896	0,9894	1,0000	0,9894	0,9436	0,9437	0,9977	0,9913	0,9926	0,9998	0,9437	0,9926	0,8347	0,9926	0,9434	0,9914	0,9896	0,8324	0,9436	0,9926	0,9924	0,9912	0,9911	0,9998	0,9927	0,9926	0,9927	0,9999	0,9910	0,9435	0,9437	
5	Cond044	0,9901	0,9994	0,9998	0,9894	1,0000	0,9437	0,9437	0,9890	0,9901	0,9900	0,9892	0,9436	0,9900	0,8371	0,9901	0,9436	0,9898	0,9994	0,8351	0,9436	0,9900	0,9899	0,9900	0,9900	0,9892	0,9900	0,9900	0,9900	0,9900	0,9893	0,9900	0,9437	0,9437
6	UPCH_1459_A13	0,9440	0,9439	0,9437	0,9436	0,9437	1,0000	0,9999	0,9433	0,9440	0,9436	0,9434	0,9999	0,9437	0,8323	0,9439	0,9999	0,9436	0,9436	0,8300	0,9999	0,9436	0,9435	0,9440	0,9436	0,9434	0,9438	0,9437	0,9436	0,9434	0,9433	0,9999	0,9999	
7	UPCH_1473_B12	0,9441	0,9439	0,9439	0,9437	0,9437	0,9999	1,0000	0,9433	0,9441	0,9437	0,9434	0,9999	0,9438	0,8324	0,9439	0,9999	0,9437	0,9437	0,8300	0,9999	0,9438	0,9436	0,9439	0,9438	0,9434	0,9439	0,9438	0,9437	0,9434	0,9432	0,9999	0,9999	
8	VAB9028	0,9909	0,9891	0,9891	0,9977	0,9890	0,9433	0,9433	1,0000	0,9908	0,9925	0,9978	0,9434	0,9925	0,8348	0,9926	0,9432	0,9912	0,9891	0,8326	0,9435	0,9926	0,9925	0,9910	0,9910	0,9978	0,9925	0,9925	0,9926	0,9978	0,9911	0,9435	0,9435	
9	Ver075	0,9962	0,9900	0,9901	0,9913	0,9901	0,9440	0,9441	0,9908	1,0000	0,9958	0,9910	0,9442	0,9959	0,8339	0,9959	0,9439	0,9957	0,9899	0,8318	0,9441	0,9958	0,9957	0,9964	0,9961	0,9909	0,9959	0,9958	0,9958	0,9911	0,9918	0,9441	0,9442	
10	KC584	0,9951	0,9899	0,9901	0,9926	0,9900	0,9436	0,9437	0,9925	0,9958	1,0000	0,9924	0,9437	0,9998	0,8345	0,9998	0,9435	0,9963	0,9898	0,8326	0,9437	0,9998	0,9999	0,9951	0,9952	0,9926	0,9998	0,9998	0,9998	0,9925	0,9919	0,9436	0,9439	
11	Hosp800-02	0,9909	0,9895	0,9893	0,9998	0,9892	0,9434	0,9434	0,9978	0,9910	0,9924	1,0000	0,9437	0,9924	0,8350	0,9925	0,9436	0,9913	0,9894	0,8326	0,9438	0,9926	0,9924	0,9909	0,9910	0,9997	0,9925	0,9924	0,9925	0,9998	0,9910	0,9437	0,9437	
12	UPCH_1478_B19	0,9441	0,9439	0,9438	0,9437	0,9436	0,9999	0,9999	0,9434	0,9442	0,9437	0,9437	1,0000	0,9438	0,8324	0,9440	0,9999	0,9438	0,9436	0,8300	0,9999	0,9438	0,9436	0,9439	0,9437	0,9435	0,9439	0,9438	0,9437	0,9435	0,9434	0,9999	0,9999	
13	KC583	0,9952	0,9899	0,9901	0,9926	0,9900	0,9437	0,9438	0,9925	0,9959	0,9998	0,9924	0,9438	1,0000	0,8345	0,9999	0,9436	0,9963	0,9898	0,8324	0,9438	0,9998	0,9998	0,9951	0,9952	0,9926	0,9999	0,9999	0,9998	0,9925	0,9920	0,9437	0,9440	

14	UPCH_1458_A07	0,8340	0,8371	0,8371	0,8347	0,8371	0,8323	0,8324	0,8348	0,8339	0,8345	0,8350	0,8324	0,8345	1,0000	0,8345	0,8324	0,8342	0,8370	0,9996	0,8325	0,8345	0,8344	0,8306	0,8306	0,8347	0,8348	0,8345	0,8344	0,8344	0,8344	0,8342	0,8324	0,8324
15	INS	0,9951	0,9900	0,9902	0,9926	0,9901	0,9439	0,9439	0,9926	0,9959	0,9998	0,9925	0,9440	0,9999	0,8345	1,0000	0,9438	0,9963	0,9898	0,8323	0,9440	0,9998	0,9998	0,9952	0,9952	0,9926	0,9999	0,9999	0,9998	0,9926	0,9921	0,9439	0,9441	
16	Ver097	0,9439	0,9438	0,9437	0,9434	0,9436	0,9999	0,9999	0,9432	0,9439	0,9435	0,9436	0,9999	0,9436	0,8324	0,9438	1,0000	0,9438	0,9438	0,8304	0,9999	0,9438	0,9437	0,9439	0,9439	0,9435	0,9440	0,9437	0,9438	0,9435	0,9433	0,9999	0,9999	
17	Peru38	0,9951	0,9896	0,9897	0,9914	0,9898	0,9436	0,9437	0,9912	0,9957	0,9963	0,9913	0,9438	0,9963	0,8342	0,9963	0,9438	1,0000	0,9898	0,8323	0,9439	0,9963	0,9962	0,9955	0,9952	0,9915	0,9963	0,9962	0,9964	0,9914	0,9921	0,9438	0,9439	
18	UPCH_1466_B03	0,9899	1,0000	0,9995	0,9896	0,9994	0,9436	0,9437	0,9891	0,9899	0,9898	0,9894	0,9436	0,9898	0,8370	0,9898	0,9438	0,9898	1,0000	0,8352	0,9437	0,9899	0,9898	0,9900	0,9899	0,9893	0,9899	0,9899	0,9901	0,9895	0,9898	0,9437	0,9437	
19	B. ancashensis	0,8318	0,8350	0,8351	0,8324	0,8351	0,8300	0,8300	0,8326	0,8318	0,8326	0,8326	0,8300	0,8324	0,9996	0,8323	0,8304	0,8323	0,8352	1,0000	0,8300	0,8322	0,8322	0,8285	0,8285	0,8326	0,8325	0,8323	0,8325	0,8324	0,8317	0,8299	0,8299	
20	UPCH_1464_B01	0,9441	0,9438	0,9438	0,9436	0,9436	0,9999	0,9999	0,9435	0,9441	0,9437	0,9438	0,9999	0,9438	0,8325	0,9440	0,9999	0,9439	0,9437	0,8300	1,0000	0,9436	0,9435	0,9439	0,9436	0,9433	0,9438	0,9437	0,9436	0,9434	0,9432	0,9999	0,9999	
21	UPCH_1461_A19	0,9951	0,9898	0,9901	0,9926	0,9900	0,9436	0,9438	0,9926	0,9958	0,9998	0,9926	0,9438	0,9998	0,8345	0,9998	0,9438	0,9963	0,9899	0,8322	0,9436	1,0000	0,9997	0,9952	0,9952	0,9927	0,9998	0,9998	0,9998	0,9927	0,9920	0,9436	0,9438	
22	USM-LMMB_06	0,9950	0,9898	0,9901	0,9924	0,9899	0,9435	0,9436	0,9925	0,9957	0,9999	0,9924	0,9436	0,9998	0,8344	0,9998	0,9437	0,9962	0,9898	0,8322	0,9435	0,9997	1,0000	0,9952	0,9952	0,9926	0,9997	0,9997	0,9998	0,9926	0,9919	0,9437	0,9438	
23	Peru-18	0,9989	0,9901	0,9902	0,9912	0,9900	0,9440	0,9439	0,9910	0,9964	0,9951	0,9909	0,9439	0,9951	0,8306	0,9952	0,9439	0,9955	0,9900	0,8285	0,9439	0,9952	0,9952	1,0000	0,9998	0,9908	0,9953	0,9950	0,9954	0,9909	0,9918	0,9440	0,9440	
24	CUSCO5	0,9987	0,9898	0,9900	0,9911	0,9900	0,9436	0,9438	0,9910	0,9961	0,9952	0,9910	0,9437	0,9952	0,8306	0,9952	0,9439	0,9952	0,9899	0,8285	0,9436	0,9952	0,9952	0,9998	1,0000	0,9909	0,9954	0,9951	0,9954	0,9910	0,9917	0,9438	0,9439	
25	CAR600-02	0,9908	0,9894	0,9892	0,9998	0,9892	0,9434	0,9434	0,9978	0,9909	0,9926	0,9997	0,9435	0,9926	0,8347	0,9926	0,9435	0,9915	0,9893	0,8326	0,9433	0,9927	0,9926	0,9908	0,9909	1,0000	0,9926	0,9926	0,9926	0,9999	0,9909	0,9434	0,9434	
26	1416	0,9952	0,9899	0,9902	0,9927	0,9900	0,9438	0,9439	0,9925	0,9959	0,9998	0,9925	0,9439	0,9999	0,8348	0,9999	0,9440	0,9963	0,9899	0,8325	0,9438	0,9998	0,9997	0,9953	0,9954	0,9926	1,0000	0,9998	0,9998	0,9926	0,9921	0,9439	0,9440	
27	ATCC:35685D-5	0,9951	0,9898	0,9901	0,9926	0,9900	0,9437	0,9438	0,9925	0,9958	0,9998	0,9924	0,9438	0,9999	0,8345	0,9999	0,9437	0,9962	0,9899	0,8323	0,9437	0,9998	0,9997	0,9950	0,9951	0,9926	0,9998	1,0000	0,9998	0,9925	0,9919	0,9437	0,9439	
28	SanPedro600-02	0,9952	0,9900	0,9903	0,9927	0,9900	0,9436	0,9437	0,9926	0,9958	0,9998	0,9925	0,9437	0,9998	0,8344	0,9998	0,9438	0,9964	0,9901	0,8325	0,9436	0,9998	0,9998	0,9954	0,9954	0,9926	0,9998	0,9998	1,0000	0,9927	0,9922	0,9437	0,9439	
29	HeidiMejia	0,9910	0,9896	0,9893	0,9999	0,9893	0,9434	0,9434	0,9978	0,9911	0,9925	0,9998	0,9435	0,9925	0,8344	0,9926	0,9435	0,9914	0,9895	0,8324	0,9434	0,9927	0,9926	0,9909	0,9910	0,9999	0,9926	0,9925	0,9927	1,0000	0,9910	0,9434	0,9434	
30	UPCH_1472_B10	0,9917	0,9898	0,9899	0,9910	0,9900	0,9433	0,9432	0,9911	0,9918	0,9919	0,9910	0,9434	0,9920	0,8342	0,9921	0,9433	0,9921	0,9898	0,8317	0,9432	0,9920	0,9919	0,9918	0,9917	0,9909	0,9921	0,9919	0,9922	0,9910	1,0000	0,9432	0,9434	
31	UPCH_1460_A18	0,9439	0,9439	0,9437	0,9435	0,9437	0,9999	0,9999	0,9435	0,9441	0,9436	0,9437	0,9999	0,9437	0,8324	0,9439	0,9999	0,9438	0,9437	0,8299	0,9999	0,9436	0,9437	0,9440	0,9438	0,9434	0,9439	0,9437	0,9437	0,9434	0,9432	1,0000	0,9999	
32	UPCH_1456_A02	0,9441	0,9439	0,9438	0,9437	0,9437	0,9999	0,9999	0,9435	0,9442	0,9439	0,9437	0,9999	0,9440	0,8324	0,9441	0,9999	0,9439	0,9437	0,8299	0,9999	0,9438	0,9438	0,9440	0,9439	0,9434	0,9440	0,9439	0,9439	0,9434	0,9434	0,9434	0,9999	1,0000



**Tabla S2 - Genes únicos de *B. bacilliformis* KC584 y *B. bacilliformis* Ver097**

Gene name	Annotation	Ver097	KC584
	DUF1561 domain-containing protein	BDMONH_03810	KOKJPD_03945
	DUF1561 domain-containing protein	BDMONH_05605	KOKJPD_04840
	DUF1561 domain-containing protein	BDMONH_02170	KOKJPD_05325
	UPF0434 protein BARBAKC583_1098	4_refound_150_stop	KOKJPD_04945
	Immunodominant surface antigen Brp family	4_refound_139	KOKJPD_01395
	Thioredoxin domain-containing protein	BDMONH_04205	KOKJPD_03400
	Type I secretion outer membrane protein TolC family	BDMONH_03775	KOKJPD_03015
	L-lactate dehydrogenase [cytochrome];L-lactate dehydrogenase cytochrome	BDMONH_00215	KOKJPD_05405
	ATP-dependent RNA helicase SUPV3L1/SUV3	BDMONH_01675	KOKJPD_00615
	metalloprotease;Peptidase M48 family	BDMONH_01685	KOKJPD_00625
	BrpA	BDMONH_04855	KOKJPD_04085
	UPF0235 protein NCTC12862_00242	BDMONH_02240	KOKJPD_01465
	PG-binding-1 domain-containing protein;peptidoglycan-binding protein	BDMONH_02705	KOKJPD_01935
	tRNA(Ile)-lysidine synthase	BDMONH_01845	KOKJPD_00780
	OpgC protein	BDMONH_01795	KOKJPD_00730
	SPOR domain-containing protein;Sporulation and cell division repeat protein	BDMONH_04135	KOKJPD_03330
	hypothetical protein;Tetratricopeptide repeat protein	BDMONH_01960	KOKJPD_00920

	DUF6101 domain-containing protein	BDMONH_02540	KOKJPD_01770
	hypothetical protein;Mitochondrial inner membrane protein	BDMONH_01375	KOKJPD_00315
	Autotransporter domain-containing protein;autotransporter outer membrane beta-barrel domain-containing protein	BDMONH_03110	KOKJPD_02335
	DUF1561 domain-containing protein	BDMONH_03755	KOKJPD_02995
	protein-L-isoaspartate O-methyltransferase	BDMONH_03770	KOKJPD_03010
	DUF1561 domain-containing protein	BDMONH_05560	KOKJPD_04835
	Putative membrane protein	BDMONH_02515	KOKJPD_01745
	Putative membrane protein	BDMONH_03100	KOKJPD_02330
	DUF6460 domain-containing protein	BDMONH_05425	KOKJPD_04650
	Nitrogen regulatory protein P-II	BDMONH_04065	KOKJPD_03260
	TIGR01244 family phosphatase	BDMONH_05985	KOKJPD_05255
fliE	flagellar hook-basal body complex protein FliE	BDMONH_05890	KOKJPD_05155
	DUF5330 domain-containing protein	BDMONH_02700	KOKJPD_01930
	SH3b domain-containing protein	BDMONH_05195	5_refound_240_stop
	Glycine zipper 2TM domain-containing protein	BDMONH_01695	KOKJPD_00635
sdhD	succinate dehydrogenase hydrophobic membrane anchor protein	BDMONH_01515	KOKJPD_00450
	MliC domain-containing protein	BDMONH_01245	KOKJPD_00185
mnhG	monovalent cation/H(+) antiporter subunit G	BDMONH_01240	KOKJPD_00180
atpC	ATP synthase F1 subunit epsilon	BDMONH_01630	KOKJPD_00565
	ATP synthase subunit	BDMONH_02460	KOKJPD_01690

	PRC domain-containing protein	BDMONH_05275	KOKJPD_04500
	TonB system transport protein ExbD	BDMONH_01865	KOKJPD_00800
	CvpA protein	BDMONH_02985	KOKJPD_02215
	outer membrane protein assembly factor BamE	BDMONH_03700	KOKJPD_02930
	Flagellar motor switch protein FliN	BDMONH_05935	KOKJPD_05200
	Copper chaperone PCu(A)C	BDMONH_00845	KOKJPD_06060
	Flagellar biosynthesis protein FliO	BDMONH_04705	KOKJPD_03855
	Lectin-like protein BA14k	BDMONH_05205	KOKJPD_04430
pal	peptidoglycan-associated lipoprotein Pal	BDMONH_01840	KOKJPD_00775
	DUF1036 domain-containing protein	BDMONH_01710	KOKJPD_00650
dut	dUTP diphosphatase	BDMONH_01870	KOKJPD_00815
	nuclease	BDMONH_04680	KOKJPD_03825
	cytochrome C	BDMONH_00395	KOKJPD_05610
	disulfide bond formation protein B	BDMONH_01880	KOKJPD_00825
	Lipopolysaccharide-assembly	BDMONH_01590	KOKJPD_00525
	Rod binding protein	BDMONH_05965	KOKJPD_05230
	DNA-binding protein	BDMONH_01380	KOKJPD_00320
	LPS ABC transporter substrate-binding protein LptA	BDMONH_01040	KOKJPD_06245
	DUF218 domain-containing protein	BDMONH_01560	KOKJPD_00495
	DedA family protein	BDMONH_01885	KOKJPD_00830
	5-formyltetrahydrofolate cyclo-ligase	BDMONH_01950	KOKJPD_00910
	shikimate kinase	BDMONH_01260	KOKJPD_00200
	Regulator of sigma E protease	BDMONH_03470	KOKJPD_02700

	DUF4167 domain-containing protein	BDMONH_02085	KOKJPD_01105
lptC	LPS export ABC transporter periplasmic protein LptC	BDMONH_01045	KOKJPD_06250
	septation protein A	BDMONH_01145	KOKJPD_00080
	thiamine phosphate synthase	BDMONH_01800	KOKJPD_00735
tsaB	tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex dimerization subunit type 1 TsaB	BDMONH_00425	KOKJPD_05640
mtgA	monofunctional biosynthetic peptidoglycan transglycosylase	BDMONH_01655	KOKJPD_00595
	YkuD domain-containing protein	BDMONH_01275	KOKJPD_00215
	energy transducer TonB	BDMONH_02855	KOKJPD_02085
	Putative membrane protein	BDMONH_01875	KOKJPD_00820
	SURF1 family	BDMONH_01450	KOKJPD_00385
	EipB-like	BDMONH_03440	KOKJPD_02665
	OMP-b-brl domain-containing protein	BDMONH_00295	5_refound_226
	MgtE-N domain-containing protein	BDMONH_05870	KOKJPD_05135
	Outer membrane protein omp25/ropB family	BDMONH_00305	KOKJPD_05505
	septal ring lytic transglycosylase RlpA family lipoprotein	BDMONH_04555	KOKJPD_03700
	OMP-b-brl domain-containing protein	BDMONH_00300	KOKJPD_05500
	ABC transporter permease	BDMONH_01565	KOKJPD_00500
motB	flagellar motor protein MotB	BDMONH_05825	KOKJPD_05080
purK	5-(carboxyamino)imidazole ribonucleotide synthase	BDMONH_01965	KOKJPD_00925

	peptidase M23	BDMONH_03165	KOKJPD_02375
	MFS transporter	BDMONH_04290	KOKJPD_04020
	Porin	BDMONH_02805	KOKJPD_02035
	DUF2087 domain-containing protein	BDMONH_04310	KOKJPD_04000
	Putative flagellar motor protein	BDMONH_05815	KOKJPD_05070
motC	chemotaxis protein MotC	BDMONH_05820	KOKJPD_05075
	MFS transporter	BDMONH_04105	KOKJPD_03300
	Autotransporter domain-containing protein	BDMONH_05845	5_refound_242
	50S ribosomal protein L7/L12-serine acetyltransferase	BDMONH_04180	KOKJPD_03375
	DUF1561 domain-containing protein	BDMONH_04315	KOKJPD_03995
	TonB-dependent heme/hemoglobin receptor family protein	BDMONH_02850	KOKJPD_02080
	Putative adhesin/invasin	BDMONH_03115	KOKJPD_02340
	Inh domain-containing protein	BDMONH_01165	KOKJPD_00100
	Outer membrane autotransporter	BDMONH_05840	KOKJPD_05105
	DUF1376 domain-containing protein	BDMONH_01905	KOKJPD_00855
maf	septum formation inhibitor Maf	BDMONH_01985	KOKJPD_00940
	hypothetical protein;invasion-associated locus B family protein	BDMONH_02175	KOKJPD_01400
	TamB inner membrane protein subunit of TAM complex	BDMONH_00905	KOKJPD_06120
	autotransporter outer membrane beta-barrel domain-containing protein	4_refound_151	KOKJPD_04990
	Phage P2 GpU	BDMONH_05380	

	GNAT family N-acetyltransferase	BDMONH_05310	
	transcriptional repressor	BDMONH_02135	
	DUF1561 domain-containing protein		KOKJPD_05365
	BTP domain-containing protein	BDMONH_01020	
	DUF1561 domain-containing protein	BDMONH_05530	
	DUF1561 domain-containing protein	BDMONH_00315	
	Asparaginase domain-containing protein	BDMONH_00535	
	Conserved domain protein		KOKJPD_04780
	DUF1561 domain-containing protein	BDMONH_05535	
	ERF family protein		KOKJPD_04810
secG	preprotein translocase subunit SecG	BDMONH_03195	
	CSD domain-containing protein	BDMONH_04615	
	DUF1561 domain-containing protein	BDMONH_05540	
	L-serine ammonia-lyase	BDMONH_03405	
	DUF2059 domain-containing protein		KOKJPD_02775
	Smr domain protein		KOKJPD_05890
atpH	F0F1 ATP synthase subunit delta	BDMONH_01610	
	DUF2062 domain-containing protein	BDMONH_02895	
	queuosine precursor transporter	BDMONH_01460	
	Ancillary SecYEG translocon subunit		KOKJPD_01680
	TPM-phosphatase domain-containing protein	BDMONH_04280	
	Type V secretory pathway adhesin AidA		KOKJPD_05100
	Cell division and transport-associated protein TolA	BDMONH_01755	

	TolA domain protein		KOKJPD_00690
	DNA polymerase III subunit gamma/tau	BDMONH_00410	
	DUF1561 domain-containing protein		KOKJPD_05515
	DUF1561 domain-containing protein		KOKJPD_03380
	DUF2497 domain-containing protein		KOKJPD_03020

**Figura S1.** (Arriba) Eje X: Porcentaje de identidad aminoacídica promedio.

Umbral de corte del AAI% a nivel de especie: > 95,38 (línea azul) Eje Y:

Porcentaje de identidad entre secuencias de 16S rRNA. Punto de corte a nivel de especie en %16S rRNA > 98.7% (línea roja).

A la izquierda del umbral de AAI%, se ubican los puntos correspondientes a Ver097. (Abajo) Índice de hibridación genoma-genoma, con punto de corte del 70% para especies. Se observan dos grupos con índice mayor al 70%.

