Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California

Doctorado en Ciencias en Ecología Marina

Genetic characterization of "Candidatus Xenohaliotis californiensis" associated with blue (Haliotis fulgens) and yellow abalone (Haliotis corrugata) on the Pacific coast of Baja California

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de

Doctor en Ciencias

Presenta:

Francesco Cicala

Ensenada, Baja California, México

2017

Tesis defendida por

Francesco Cicala

y aprobada por el siguiente Comité

_____

Dr. Axayácatl Rocha Olivares
Director de tesis

Miembros del Comité

Dr. Jorge Abelardo Cáceres Martínez

Dr. Miguel Ángel del Río Portilla

Dra. Mónica Hernández Rodríguez

Dr. James Douglas Moore

Dr. Helmut Maske Rubach

Dra. María Lucila del Carmen Lares Reyes
Coordinadora del Posgrado de Ecología Marina

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **Francesco Cicala** como requisito parcial para la obtención del grado de Doctor en Ciencias en Ecología marina.

Genetic characterization of "Candidatus Xenohaliotis californiensis" associated with blue (Haliotis fulgens) and yellow abalone (Haliotis corrugata) on the Pacific coast of Baja California.

Resumen aprobado por:

_____

Dr. Axayácatl Rocha Olivares
Director de tesis

Las poblaciones de abulón que viven en la costa del Pacífico de la península de Baja California (México), han estado sujetas a eventos de mortandad masiva. Dichos eventos pueden relacionarse con diferentes causas incluyendo el síndrome de deshidratación (o WS por sus siglas en inglés), que es una enfermedad de debilitamiento crónica que afecta a la mayoría de las especies de abulón de Norte América. "*Candidatus* Xenohaliotis californiensis" (o *C*Xc) ha sido descrito como el agente etiológico de este síndrome. Sin embargo, su papel en el desencadenamiento de esta enfermedad no está totalmente entendido. Al respecto, la ausencia (o presencia) de *C*Xc en abulones enfermos (o sanos) así como diferentes tasas de susceptibilidad y mortalidad, han sido ya reportadas entre las diferentes especies de abulón. Posiblemente, estas observaciones se pueden relacionar con: (*i*) la existencia de diferentes cepas genéticas de *C*Xc, la cuales podrían diferir por su patogenicidad y con (*ii*) la presencia de microbiotas alternativos los cuales podrían prevenir la infección por parte de *C*Xc. El día de hoy, ambas consideraciones han sido escasamente investigadas. En relación a esto, *C*Xc es todavía una bacteria no cultivable y el gen *16S rRNA* ha sido el único secuenciado hasta la fecha. Este limitado conocimiento genético representa uno de los principales obstáculos en la comprensión de las características ecológicas y biológicas de esta bacteria. Además, no existe información sobre las microbiotas que albergan las especies Mexicanas de abulón. Como consecuencia, los objetivos principales de esta tesis doctoral son ampliar el conocimiento genético de *C*Xc y evaluar la comunidad bacteriana en la cual *C*Xc vive. En el capítulo 3, se propone con éxito un acercamiento independiente de cultivo con el cual fue posible amplificar y secuenciar seis loci codificantes y no codificantes anteriormente desconocidos de este procariota. Esta nueva información genética, fue inicialmente utilizada para aclarar la sistemática y las relaciones evolutivas entre *C*Xc y otras bacterias genéticamente cercanas. Las reconstrucciones filogenéticas más confiables colocan a *C*Xc como un taxón-hermano de *Neorickettsia* lo cuales forman un linaje basal recíprocamente monofilético a las otras *Anaplasmataceae*. Además, de acuerdo con los resultados de la línea temporal evolutiva, *C*Xc puede ser considerado como el taxón más basal dentro de las *Anaplasmataceae*, lo cual apoya la hipótesis de un origen marino de esta familia bacteriana. En el capítulo 4, se utilizaron dos métodos de tipificación de secuencias para analizar la diversidad genética de *C*Xc a diferentes niveles, incluyendo el geográfico, es decir, a lo largo de las costas de Baja California Sur; el de especies hospederas, que incluye a los abulones azul, amarillo y rojo; y el relacionado con el WS, es decir entre abulones sanos y con signos externos de WS. Como resultado, no se encontró polimorfismo genético en 358 secuencias de ADN analizadas con los acercamientos de "*Multilocus Sequence Typing*" y "*Multi Spacer Typing*" ni en 7,117 secuencias del gen ribosomal 16S, obtenidas por medio de un análisis metagenómico. La ausencia de variabilidad genética puede ser el resultado de presiones selectivas, las cuales modelaron el genoma de esta *Rickettsia*. Por lo tanto, es posible que los abulones de California y Baja California están siendo afectados por un único linaje de *C*Xc. En el capítulo 5, se describe el análisis metagenómico, para evaluar las comunidades bacterianas presentes en el tracto gastrointestinal (GI) de los abulones azul (HF) y amarillo (HC). Los resultados revelaron cambios estructurales significativos en los microbiomas de ambas especies; a pesar de esto, no se identificaron cambios funcionales inferidos por medio del análisis PICRUSt. Sin embargo, la microbiota de HC presentó una composición más diversa y compleja, lo que se reflejó en las funciones metabólicas inferidas.

Palabras clave: *Candidatus* Xenohaliotis californiensis, Caracterización genética, Microbiota de *Haliotis fulgens* y *Haliotis corrugata*

Abstract of the thesis presented by Francesco Cicala as a partial requirement to obtain the degree of Doctor in Science in Marine Ecology.

Genetic characterization of "Candidatus Xenohaliotis californiensis" associated with blue (Haliotis fulgens) and yellow abalone (Haliotis corrugata) on the Pacific coast of Baja California.

Abstract approved by:

_____

Dr. Axayácatl Rocha Olivares

Thesis Director

Abalone populations inhabiting the Pacific coast of Baja California peninsula (Mexico) have experienced unusual mortality events. Such events may have been associated with different causes including the withering syndrome (WS), a chronic wasting disease affecting most if not all North American abalone species. The etiological agent of this syndrome has been described as "*Candidatus* Xenohaliotis californiensis" (or *CXc*,). Nevertheless, its role in triggering this syndrome is not completely understood. Indeed, the absence (or presence) of *C*Xc in diseased (or healthy) abalone as well as different susceptibility and mortality rates among abalone species have been previously reported. These observations may be related to: (*i*) the existence of different *C*Xc genetic strains differing in their pathogenicity and/or; (*ii*) the presence of alternative microbiota arrangements which may prevent infections by *C*Xc. Thus far, both considerations have been poorly investigated. In this context, *C*Xc remains an uncultured bacterium and the *16S rRNA* gene has been the only gene sequenced to date. This limited genetic knowledge represents one of the main obstacles in understanding several ecological and biological features of this bacterium. Also, no information exists about the microbiota composition harbored by Mexican abalone species. Consequently, the main goal of this doctoral dissertation is to extend the genetic knowledge of *C*Xc and assess the bacterial community in which *C*Xc lives. In Chapter 3, I successfully applied a culture-independent approach in order to amplify and sequence six previously unknown coding and non-coding loci of this prokaryote. This new genetic information, was initially used to clarify the systematic positions and the evolutionary relationship between *C*Xc and other genetically close bacteria. Our most reliable phylogenetic reconstructions place *C*Xc as a sister-taxon of *Neorickettsia* forming a basal lineage reciprocally monophyletic to the rest of *Anaplasmataceae*. Moreover, according to our evolutionary timeline, *C*Xc may be considered as the most basal *Anaplasmataceae*, which supports a marine origin of this bacterial family. In Chapter 4, I used two sequence typing approaches to analyze the genetic diversity of *C*Xc at different levels, namely geographically among fishing cooperatives from Baja California Sur in the blue and yellow abalone, among host species (blue, yellow, and red abalone), and between healthy blue and yellow abalone and those with WS. I found no genetic polymorphisms in the 358 DNA sequences from Multilocus Sequence Typing and Multi Spacer Typing approaches nor in the 7,117 DNA *16S rDNA* reads from metagenomic analyses. The absence of genetic variability may result from selective pressures shaping the genome of this rickettsia. Hence, the evidence suggests that a single lineage of CXc is infecting different abalone species off the coast of California and Baja California. In Chapter 5, I used a metagenome analysis to assess the bacterial communities of the gastro-intestinal (GI) tract of blue (HF) and yellow (HC) abalone. Our results revealed significant differences in the microbiomes of both species but no concomitant functional shifts based on PICRUST inferences. HC presented a more diverse and complex microbiota composition, which was reflected in more diversified metabolic functions in the microbiome.

Keywords: *Candidatus* Xenohaliotis californiensis, Genetic characterization, Microbiota of *Haliotis fulgens* y *Haliotis corrugate*

Dedicatoria

In ricordo di mio padre, alla forza di mia madre, al coraggio di mio fratello.

Agradecimientos

Al CICESE, y UC Davis por darme la posibilidad de desarrollar mi proyecto en sus estructuras.

Al CONACyT y UC Mexus por las becas otorgadas para mis estudios de Doctorado y Estancia.

Al Dr. Axayácatl Rocha Olivares por permitirme participar en este proyecto y formar parte de su equipo de trabajo.

Al Dr. Jorge A. Cáceres Martínez por su apoyo y valiosas críticas durante el desarrollo de este proyecto.

Al Dr. Miguel Ángel del Río Portilla, por sus valiosos comentarios durante la realización de esta tesis.

A la Dra. Mónica Hernández Rodríguez por sus comentarios y dedicación en la revisión de esta tesis.

Al Dr. Helmut Maske Rubach por todo su apoyo, asesoría y paciencia a lo largo de todos estos años.

Al Dr. James Douglas Moore todo su apoyo, la valiosa asesoría y por aceptarme en su equipo de trabajo.

A los departamentos de personal y estudiantes de Ecología Marina, Acuicultura y Microbiología en particular a Nallely Cano Domínguez por ayudarme en el desarrollo de mi trabajo.

A mis compañeros y amigos del laboratorio de Ecología Molecular en particular modo a Pablo Suarez Moo y Alejandro Cisterna por haber compartito conmigo la experiencia de trabajar con *Candidatus* Xenohaliotis californiensis y otras bacterias.

A mis amigos Andrea Lievana, Roberto Cruz Flores y Ashley Vater por sus valiosos comentarios y sugerencias en las primeras ediciones de mi tesis.

A los “Cataneros” por haberme regalado entre los momentos más lindo de mi vida.

A todas las personas que en un modo o en el otro me están acompañando en mi "segunda vida Mexicana".

Tabla de Contenido

List of figures

List of tables

Chapter 1: Introduction

1.1 Generalities of marine mollusk diseases

Recent research has highlighted that marine mollusks are constantly threatened by a great variety of pathogenic agents such as bacteria, viruses and protozoa, among others (Burge *et al.*, 2014). These microorganisms pose a hazard being responsible for epidemiological outbreaks with mortality rates that may reach 100% in both wild and farmed mollusk populations. As a result, the occurrence of infectious diseases may produce unpredictable ecological shifts and/or economic losses (Gestal *et al.*, 2008; Sweet & Bateman, 2015). Gestal *et al.* (2008) pointed out that the occurrence and development of epidemiological outbreaks may be a response to:

1. increase in host population density;
2. animal movements and/or local or international transfer;
3. habitat alteration (including both natural and anthropogenic changes);
4. diversity in immune competency in the host population.
5. diversity in pathogen virulence.

The first three factors may be prevented using appropriate management strategies, whereas the fourth and fifth factors need an exhaustive knowledge of hosts and pathogens to propose correct diagnostic tools in order to develop prevention or recovery plans. Moreover, with regard to the 4th factor, only limited control strategies may be applied (Gestal *et al.*, 2008). Indeed, mollusks rely only on the innate immune system to overcome disease as no true adaptive immune system has been described in these animals (Mydlarz *et al.*, 2006). Therefore, vaccination cannot be used to prevent infectious diseases and drugs would be difficult to apply particularly in wild populations (Gestal *et al.*, 2008).

Accordingly, the identification, isolation and characterization of epidemiological agents is a fundamental task to completely understand infectious diseases (Harvell *et al.*, 1999). However, for the majority of mollusk pathogens, the current diagnostic methods are restricted to histological, physiological and ultra-structural analysis (Harvell *et al.*, 1999; Figueras & Novoa, 2004; Gestal *et al.*, 2008). With these methods, the characterization of some pathogens may be particularly challenging due to inter-specific similarities and/or intra-specific morphological plasticity and the occurrence of inconsistent results, among others (Luna *et al.*, 2007; Batista *et al.*, 2007; Justice *et al.*, 2008).

In recent decades, genetic analyses have provided several advantages in the study of pathogens (Harvell *et al.*, 1999; Figueras & Novoa, 2004; Luna *et al.*, 2007; Gestal *et al.*, 2008) and since 2008, these studies, have been included in the Manual of Diagnostic Tests for Aquatic Animals as an essential diagnostic tool to identify microorganisms (OIE, 2008).

1.2 Biology of abalone

Abalone belong to the class Gastropoda, order Vetigastropoda, family Haliotidae and genus *Haliotis* (Cox, 1962). Abalone live in groups; the gregarious life-strategy is needed to increase the probability of fertilization. Indeed, all *Haliotis* spp. present separate sexes and during spawning the gametes are released into the water column where fertilization occurs (Cox, 1962; Leighton, 2000). Early stages of all abalone are represented by pelagic free-swimming lecithotrophic larvae. Generally, after 2 weeks larvae settle on the substrate and after about 3 years they reach adult size (Leighton, 2000). Adult abalone are benthic and may reach species-specific depths (3m for shallow *H. fulgens* up to 30m for the deep *H. sorenseni*) where they graze on algae and/or seagrasses (Cox, 1962; Leighton, 2000; Guzman del Próo *et al.*, 2003; Flores, 2014).

Since the 16th and 17th centuries, abalone have played an important role in the development of the Baja California peninsula (Revollo & Sáenz-Arroyo, 2012). Moreover, in the second part of the 19th century, due the high local and international demand, abalone fishery became a lucrative trade (Cariño Olvera & Monteforte, 2008). The first industrial fisheries were established by Chinese that exported large quantities of abalone to different Asiatic and European countries (Revollo & Sáenz-Arroyo, 2012). After the Second World War the abalone fishery was passed under the control of Mexican authorities (Cox, 1962). In recent years, the Federation of Fishing Cooperatives (FEDECOOP acronym in Spanish) have generated more than 20,000 jobs making the abalone fishery the economic base of many families (Revollo & Sáenz-Arroyo, 2012). The Baja California peninsula has seven exploitable abalone species (Morales-Bojórquez *et al.*, 2008) and catches have been focused according to the abundance of abalone species (Flores, 2014). At present, only the blue (*Haliotis fulgens* or HF) and yellow (*H. corrugata* or HC) abalone are exploited since the density of the other species is insufficient (Morales-Bojórquez *et al.*, 2008; SAGARPA, 2009). Since the second half of the 20th century, abalone catches have decreased from 5,900 (in 1950) to 102 t (in 2012). The causes of this decline are not entirely clear (Cáceres-Martínez *et al.*, 2011) and can be related to

different factors such as overfishing, environmental changes (e.g. El Niño, poor water conditions, etc.) and the presence of disease outbreaks (Cáceres Martínez, 2002; Crosson *et al.*, 2014).

1.3 Abalone pathogens and diseases

Several ecological and biological factors make abalone susceptible to infectious diseases, such as: (*i*) forming high density patches naturally promoting pathogen transmission; (*ii*) the anthropogenic introduction of infected abalone contributing to the disease dispersion; (*iii*) the many environmental changes occurring in the intertidal coastal zone that may seriously debilitate abalone populations; and (*iv*) the grazing feeding behavior that may promote the gathering of a large number of microorganism including those that can represent an immune challenge (Harvell *et al.*, 1999; Gestal *et al.*, 2008; Burge *et al.*, 2014). Additionally, some pathogenic microorganisms may affect abalone in any stage of their ontogenetic development (Bateman *et al.*, 2011; Sweet & Bateman, 2015).

The main mollusk diseases are listed in the Manual of Diagnostic Tests for Aquatic Animals (http://www.oie.int/international-standard-setting/aquatic-manual/access-online/).

1.4 Early studies on Withering syndrome (WS)

Withering syndrome is a chronic fatal disease affecting all tested abalone species (Friedman, 2012). Even though WS was believed to be endemic of southern California (USA) and Baja California (Mexico) (Friedman, 2012; Flores, 2014). The commercial trade of infected abalones has increased the geographic distribution of the withering syndrome which is presently recognized as a worldwide distributed disease (Crosson *et al.*, 2014).

Among abalone populations in the U. S., anomalous high mortalities were observed along the Pacific coast of California since the second half of 20th century (Cox, 1962). However, the first well-documented report was proposed by Haaker *et al.* in 1992. These authors described the mass mortality experienced by black abalone (*Haliotis cracherodii*) population of Santa Cruz Island in 1986. Remarkably, dying abalone were characterized by several morphological changes including color alteration, pedal muscle and mantle retraction, among others. To collectively describe these morphological disorders and the resulting

shrunken appearance of diseased abalone, these authors proposed the name of withering syndrome (WS). Thenceforth, successive WS outbreaks where reported in different areas of southern California and often associated with populations decreases (Davis *et al.,* 1992; Haaker *et al.,* 1992.; Steinbeck *et al.,* 1992; Tissot, 1995). A different number of studies, pointed out that many factors could explain the abalone die-off including, pollutants, increase in water temperature, food shortages and the presence of etiological agents (Lafferty & Kuris, 1993; Friedman *et al.*, 1997). Although, the relation between these factors and the expression of WS is still a topic of debate, disease outbreaks in U. S. abalone populations were principally associated with the presence of a Rickettsiales-like prokaryotes (RLP) (Friedman *et al.*, 1997).

Significant decrease in Mexican abalone populations have also been observed, such as the collapse of black abalones population in different areas of Baja California, such as Cedros and San Benito islands (Cáceres-Martínez, 2002). Additionally, the first black abalone die-off, was observed in Bahía Tortugas (Valles-Ríos, 2000). These organisms were morphological characterized by clinical sings of WS; also histological analysis revealed the presence of RLP inclusions in the gastrointestinal cells of dying abalone (Valles-Ríos, 2000). These results suggested, a strong correlation between WS and a RLP (Valles-Ríos, 2000). In contrast, successive investigations conducted on red (*H. rufescens*), yellow (*H. corrugata*) and blue (*H. fulgens*) abalones weakened such correlation, because histological analysis revealed the presence of RLP inclusions in both sick and healthy abalone with similar prevalence ranges (Caceres-Martinez & Tinoco-Orta, 2001; Álvarez Tinajero *et al.*, 2002). Accordingly, Cáceres-Martínez (2002) proposed that Mexican abalone reduction could be due to overfishing, environmental changes, among others.

The main evidence in supporting the WS-RLP cause relation was based on the amplification of *16S rRNA* of the RLP affecting abalone whit the proposed name "*Candidatus* Xenohaliotis californiensis" (or *C*Xc) (Friedman *et al.*, 2000). Moreover, they designed specific *in situ* hybridization probes to physically associate the amplified DNA sequence with histological inclusions. Finally, the *C*Xc role in the development of the WS was established by treating infected abalones with external signs of WS with oxytetracycline; the abalone regained their normal healthy condition after antibiotic treatment (Friedman *et al.*, 2000).

Recently *in situ* hybridization analysis confirmed that inclusions associated with RLP observed in Mexican black (*H. cracherodii*), red (*H. rufescens*), blue and yellow abalone, correspond to "*Candidatus* Xenohaliotis californiensis" (Cáceres-Martínez *et al.*, 2011).

1.5 "*Candidatus* Xenohaliotis californiensis" and differences in disease susceptibility

"Candidatus Xenohaliotis californiensis" (*C*Xc) is a pleomorphic, gram-negative coccobacillus that inhabits abalone gastrointestinal epithelia (Friedman *et al.*, 2000). The cycle of infection of *C*Xc is thought to occur entirely inside the host cells and may be conceptualized in the following steps (Cáceres Martínez, 2002): (*i*) Once inside the host, the rickettsial bacteria reaches the esophagus where they adhere and penetrate into the columnar epithelial cells; (*ii*) *C*Xc multiply inside intracytoplasmic vacuoles normally apically located in host cells (Friedman *et al.*, 2000). This activity, may result in host cells hypertrophy (Valles-Ríos, 2000; Cáceres Martínez, 2002) and depending on parasite load (e.g. number of vacuoles present into a cells) infections may provoke intestinal cells necrosis; hence physiological starvation, pedal muscle catabolism, metaplasia and finally host dead (Friedman *et al.*, 2000; Moore *et al.*, 2001; Braid *et al.*, 2005; Crosson *et al.*, 2014). (*iii*) The infected cells containing *C*Xc colonies may break down or peel off to release this bacterium with fecal material and restart a new cycle of infection (Cáceres Martínez, 2002). *C*Xc is considered an obligate endoparasite (Friedman, 2012; Crosson *et al.*, 2014) and its transmission is assumed to be horizontal, likely via fecal-oral route (Crosson *et al.*, 2014); however, recent PCR analysis have shown that *C*Xc may survive outside its host for undetermined time, which in turn suggests that a water-borne transmission may be not discounted (Friedman *et al.*, 2014a; Cruz-Flores *et al.*, 2015).

Nevertheless, *C*Xc has been observed in all tested abalone species, losses due to WS vary among species (Wetchateng *et al.*, 2010; Crosson *et al.*, 2014) with mortality rates varying from up to 100% as in black and white abalone (Altstatt *et al.*, 1996; Raimondi *et al.*, 2002; Friedman *et al.*, 2007), through moderate mortality in red abalone (Moore *et al.*, 2000, 2001; Caceres-Martinez & Tinoco-Orta, 2001), to little effects and low mortalities in yellow and blue abalone (Álvarez Tinajero *et al.*, 2002; Moore *et al.*, 2009). Furthermore, long-term thermal studies induced clinical WS signs in red abalone while blue abalone appeared more resistant (Vilchis *et al.*, 2005; Moore *et al.*, 2009). Nevertheless, temperature remains a key WS trigger factor. In this context blue abalone held at 25 °C experienced more morphological changes and higher mortality those held at 20 °C (García-Esquivel *et al.*, 2007).

In addition to differences in disease susceptibilities among abalone species, populations of a single species from different geographic areas may respond differently to *C*Xc (Cáceres-Martínez *et al.*, 2011; Crosson *et al.*, 2014; Friedman *et al.*, 2014b). In this context, significant differences in survival were observed among the black abalone populations from San Nicolas Island and Carmel Point, with the latter presenting higher mortality rates (Friedman *et al.*, 2014b). Moreover, since 2009 differences in mortalities were

observed in both yellow and blue abalone, among closely localities of Baja California Sur (Cáceres-Martínez *et al.*, 2011).

This geographical and/or species-specific mortality rates may relate to the existence of different genetic *C*Xc strains, some pathogenic and others non-pathogenic (Cáceres-Martínez *et al.*, 2011). Additionally, recent research has suggested that the abalone gut microbiota composition may help prevent infection by pathogenic bacteria (Hyun *et al.*, 2013; Iehata *et al.*, 2014) including *C*Xc. Nevertheless, no efforts have been made to explore the microbiome organization in which *C*Xc lives.

1.6 Genetic knowledge of "*Candidatus* Xenohaliotis californiensis"

The main hazard of microorganisms is their ability to modify their genome composition in order to face environmental change. Moreover, obligate endocellular bacteria may shape their genome via local point mutations (e.g. according to sympatric evolutionary model) and/or acquiring (or removing) genes via horizontal transfer (e.g. following a sympatric evolutionary model) (Merhej & Raoult, 2011). This mutational history may shed light on the bacterial phylogenetic relationships and provide useful information to understand disease outbreaks; hence, the possibility to propose more adequate management or recovery plans. Additionally, DNA sequencing analysis has become an essential approach to detect and characterize these microorganisms (Maiden *et al.*, 1998; Fournier *et al.*, 2003, 2004; Fournier & Raoult, 2007).

The genetic study of *C*Xc has been hampered by the inability to grow it axenically and although new methodology has been proposed to simplify analysis of this pathogen outside of it hosts cells (Cruz-Flores *et al.*, 2015), *16S rRNA* remains the only known gene (Friedman *et al.*, 2000). Nevertheless, this genetic information has been used to evaluate the systematic position of *C*Xc inside order Rickettsiales and to assess its genetic variability. Friedman *et al.* (2000) proposed the first phylogenetic analysis of this bacterium, reporting a genetic similarity between *C*Xc and some *Anaplasmataceae* species, and placing *C*Xc inside the family *Rickettsiaceae* (Friedman *et al.*, 2000). Even though subsequent analysis have replaced *C*Xc inside the *Anaplasmataceae* family (Garrity *et al.*, 2004; Friedman, 2012; Martijn *et al.*, 2015), the new systematic position of *C*Xc may been considered still unsettled. In this context, the bootstrap values defining the taxonomic boundary of *C*Xc in different phylogenetic reconstructions, are generally low (around 70%) (Friedman *et al.*, 2000; Lee *et al.*, 2005; Vannini *et al.*, 2005).

The *16S rRNA* gene was also used to assess intra-specific genetic variability of *C*Xc. Despite the growing number of infected abalone species and of *16S rRNA* sequences from *C*Xc, no genetic variability useful to evaluate the existence of different strains has been found (Balseiro *et al.*, 2006; Friedman, 2012; Kiryu *et al.*, 2013).

Finally, as proposed for other Rickettsiales (Merhej & Raoult, 2011), *C*Xc may reorganize its genome composition exchanging genetic information with neighboring bacteria. Accordingly, the genome organization of *C*Xc and even its presence/absence inside abalone microbiota, may depend on the microbiome composition in which *C*Xc lives. However, no data is available on the microbiome organization of *Haliotis fulgens* and *Haliotis corrugata.*

Chapter 2. Justification, Hypothesis and Objective

2.1 Justification:

The genetic knowledge of *C*Xc is needed to clarify several aspects of this bacterium, including its role in triggering the WS. Moreover, the identification of polymorphic loci will allow us to assess the possible existence of different linages and strains of *C*Xc. This will allow understanding how the genetic variation is related to the geographic origin of the samples, to host species, and to the presence of WS. Finally, the assessment the gut microbiome will help to shed light on the microbial ecology of *C*Xc.

2.2 Working hypotheses:

1. According to the Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology (Stackebrandt *et al.*, 2002), a multi-genetic information will help to improve the estimation of evolutionary relationships of *C*Xc.

2. An extended genetic knowledge of *C*Xc will help to evaluate the possible existence of different *C*Xc strains. Moreover, the detected genetic variability will be correlated with different habitats (geography and/or abalone species), and with the presence of the WS.

3. Different environmental adaptations among abalone species may be traced to their gut microbiota composition. Consequently, assessing the microbiota will provide ecological information of predominant bacterial groups.

2.3 Objectives:

2.3.1 General objective:

Genetically characterize "*Candidatus* Xenohaliotis californiensis" associated with the yellow (*H. corrugata*) and blue (*H. fulgens*) abalones in the west coast of Baja California Sur.

2.3.2 Specific objectives:

1. Establish the systematic position of "*Candidatus* Xenohaliotis californiensis" using multi-genetic analyses.

2. Evaluate the genetic variability of "*Candidatus* Xenohaliotis californiensis" and its relationship with: (*i*) different geographic areas; (*ii*) different *Haliotis* spp.; (iii) the presence of the withering syndrome.

3. Characterize the structure and function of the gut microbiome of the blue and yellow abalone by a metagenomic analysis.

Chapter 3: Multigenetic characterization of "*Candidatus* Xenohaliotis californiensis"

3.1 Abstract

"*Candidatus* Xenohaliotis californiensis" (or *CXc*) is the etiological agent of the withering syndrome (WS), a chronic wasting disease affecting most if not all North American species of abalone, and has been described as a *Rickettsiales*-like prokaryote. Genetic data about this species is limited to the *16S rRNA* gene. The inability to grow it axenically has hindered its genetic and genomic characterization and, in consequence, a thorough analysis of its systematics. Here, we amplified and sequenced five genes (*16S rRNA, 23S rRNA*, *ftsZ*, *virD4*, and *virB11*) of *CXc* from infected abalone to analyze its phylogenetic position. Phylogenies from concatenated DNA and amino acid sequences with representative genera of most *Rickettsiales* unequivocally place *CXc* in the family *Anaplasmataceae*. Furthermore, the family has two reciprocally monophyletic lineages: one leading to (*Neorickettsia*, *CXc*) and the other to ((*Ehrlichia*, *Anaplasma*), *Wolbachia*). A molecular-clock Bayesian reconstruction places *CXc* as the most basal lineage in *Anaplasmataceae*. These phylogenetic hypotheses shed light on patterns of host evolution and of ecological transitions. Specifically, *Neorickettsia* and *CXc* inhabit aquatic hosts whereas the rest of *Anaplasmataceae* are found in terrestrial hosts. Additionally, our evolutionary timeline places the directly transmitted marine *CXc* as the basal *Anaplasmataceae*, ancestral to both fresh water and terrestrial species with adaptations leading to more complex life-cycles involving intermediate vectors or reservoir species; thus supporting the hypothesis of a marine origin for this bacterial family.

3.2 Introduction

Among *Alphaproteobacteria*, the order *Rickettsiales* comprises a diversity of obligate endocellular gram-negative bacteria infecting a wide variety of metazoans. Best known are the pathogenic bacteria of vertebrates transmitted by ticks (*Anaplasma*, *Ehrlichia* and *Rickettsia*) or by trematodes and/or fishes (*Neorickettsia*) as well as arthropod endosymbionts (*Wolbachia*) (Rikihisa, 2006; Weinert *et al.*, 2009; Rar & Golovljova, 2011). Infection of blood cells in vertebrates, including humans, causes clinical diseases collectively called rickettsioses, ehrlichiosis and anaplasmosis; a febrile illness often followed by lymphadenopathy and hematological abnormalities among other medical disorders (Rikihisa, 2006; Darby

et al., 2007; Weinert *et al.*, 2009; Rar & Golovljova, 2011; Kang *et al.*, 2014). Recently, technological advances have increased our ability to detect and describe *Rickettsiales*-like prokaryotes (RLPs) using an expanded set of molecular and morphological analyses. In this context, a multi-genetic understanding has been essential in shedding light on the phylogenetic relationships of different bacterial groups including *Alphaproteobacteria* and those in the order *Rickettsiales*. For instance, the taxonomy of this class has been recently revised into new subclasses according to genetic information from both the small (*16S rRNA*) and large (*23S rRNA*) subunit ribosomal RNA genes (Ferla *et al.*, 2013), and the systematic position and boundaries of several *Rickettsiales* species, have been resolved using DNA sequences from *16S rRNA*, *groESL*, and surface protein genes (Dumler *et al.*, 2001). The order *Rickettsiales* includes three recognized families (*Rickettsiaceae*, *Anaplasmataceae* and *Holosporaceae*) and two more have been recently described (*Candidatus* Paracaedibacteraceae and *Candidatus* Midichloriaceae) (Montagna *et al.*, 2013; Hess *et al.*, 2016). Additionally, an increasing number of poorly characterized species are provisionally labeled with the generic name of "Rickettsia" (Merhej & Raoult, 2011) or RLP (Friedman *et al.*, 2000), and require additional efforts to determine their phylogenetic and systematic positions (Kang *et al.*, 2014). In this context, *CXc* has been described as a new obligate RLP endoparasite (Friedman *et al.*, 2000; Moore *et al.*, 2002; Friedman, 2012). However, the study of its genetic organization has been hampered by the inability to grow it axenically. Even with the advent of new methods to study it outside of its host (Cruz-Flores *et al.*, 2015), the *16S rRNA* remains the only known gene from this bacterium. Moreover, using *16S rRNA* gene sequences, previous phylogenetic reconstructions placed *CXc* as an independent linage inside *Anaplasmataceae* (Garrity *et al.*, 2004; Lee *et al.*, 2005; Vannini *et al.*, 2005; Friedman, 2012; Martijn *et al.*, 2015). However, the systematic position and evolutionary relationship of *CXc* remains unsettled due the low bootstrap values (typically around 70% bootstrap support) (Friedman *et al.*, 2000; Lee *et al.*, 2005; Vannini *et al.*, 2005; Martijn *et al.*, 2015). Consequently, here we analyze the systematic position and evolutionary relationships of *CXc* within the order *Rickettsiales*, with special interest in its relationship with other members of the family *Anaplasmataceae*, using DNA sequences from five genes and the corresponding amino acid sequences from three of them.

3.3 Materials and methods

3.3.1 Sample collection and "*Candidatus* Xenohaliotis californiensis" detection.

The source of *CXc* DNA included gastrointestinal tissue of wild-caught blue (*Haliotis fulgens*) and yellow (*H. corrugata*) abalone sampled from eight localities during commercial fishing operations on the Pacific coast of southern Baja California (Mexico). Approximately 30 mg of postesophageal tissue were collected and immediately transferred to sterile 1.5 ml microcentrifuge tubes containing molecular grade ethanol and DNA was extracted using DNeasy tissue kits (Qiagen, Valencia, CA, USA). In addition, we analyzed fecal material from farmed red abalone (*H. rufescens*) maintained in UC Davis Bodega Marine Laboratory, California, USA. Approximately 500 mg of feces from infected specimens were collected using sterile pipettes and processed with the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA, USA). To evaluate the presence of *CXc*, diagnostic PCRs were performed using *16S* primers specific for *CXc* (RA5-1 and RA3-6, Andree *et al.,* 2000). However, inconsistent results using the standard OIE protocol (Friedman, 2012) required the design of a new set of diagnostic primers (ss16S-F: GCCTCAGTTTGGCTGGGTTCTTCA and ss16S-R: GAATTGCCACTTTAAAGTATGGACGG). For both primer sets, PCR reactions (15 µl final volume) contained: 90 ng of tissue DNA, 1X PCR Buffer (Kapa Biosystems, Woburn, MA, USA), 0.2 mM of dNTPs (New England Biolabs, Beverly, MA, USA), 0.3 µM of each primer, and 1U of *Taq* polymerase (Kapa Biosystems, Woburn, MA, USA). Thermal cycling consisted of 4 min at 94 °C, 40 cycles of 1 min at 94 °C; 30 s at 62°C (RA-set; Andree *et al.,* 2000) and 66°C (16Sss-set; new proposed diagnostic primers), and 30 s at 72 °C, followed by 8 min at 72°C. Amplification was verified with 1.5% agarose gel electrophoresis. At least seven amplicons for each primer set were purified enzymatically (ExoSAP-IT, USB-Affymetrix, Cleveland, OH, USA) before cycle sequencing using an ABI 3730XL automatic sequencer (Macrogen Inc., Seoul, Korea). Resulting *16S rRNA* sequences were trimmed and verified with Codon Code Aligner v3.7.1 (Codon Code Corporation, Dedham, MA, USA) and aligned using ClustalW within the program Geneious R9 (Kearse *et al.*, 2012). Online BLAST searches (Altschul *et al.*, 1990) were used to confirm the taxonomic identity of the sequences.

3.3.2 Gene selection and primer design

A set of 16 genes was originally selected (products shown in parentheses): *16S rRNA* (small ribosomal subunit RNA), *23S rRNA* (large ribosomal subunit RNA), *coxA* (cytochrome oxidase A subunit),

ftsZ (a cell division protein), *virD4* (a type IV secretion system –T4SS – protein), *virB11* (a T4SS protein), *tpiA* (triose phosphate isomerase), *gap* (glyceraldehyde-3- phosphate dehydrogenase), *groEL* (a chaperone protein), *atpD* (ATP synthase F1 beta subunit), *tkt* (transketolase, a pentose phosphate pathway protein), *recA* (recombinase A), *omp* (outer membrane protein), *rpoB*, *rpoC* and *rpoD* (RNA polymerase subunits). These loci were selected because (*i*) they have conserved regions allowing primer design across species of *Rickettsiales*, (*ii*) they are present in most *Rickettsiales* species, and (*iii*) they have proven useful in molecular systematic analyses of this group (Dumler *et al.*, 2001; Casiraghi *et al.*, 2005; Baldo *et al.*, 2006; Min *et al.*, 2008; Gillespie *et al.*, 2010). DNA sequences from the available target genes of *Rickettsiales* species were downloaded from GenBank and aligned using the Multiple Expectation-Maximization for Motif Elicitation (MEME) algorithm. This algorithm allows multiple alignment of DNA or amino acid sequences and facilitates searches for conserved uninterrupted motifs that can be used to design primers (Bailey *et al.*, 2006). Primers were designed using the MEME suite with a maximum degeneracy of 25-35% of bases (http://meme suite.org/)

3.3.3 Optimization PCR, cloning, and sequencing.

Amplification of all loci was carried out in 20 µl reactions containing 100 ng of fecal or 90 ng of tissue DNA, 1X PCR Buffer, 1.5 mM MgCl2 (both Kapa Biosystems, Woburn, MA, USA), 0.2 mM dNTPs (New England Biolabs, Beverly, MA, USA), 0.5 µM of each primer, 0.4 mM bovine serum albumin (BSA, New England Biolabs, Beverly, MA, USA) and 1U of *Taq* polymerase (Kapa Biosystems, Woburn, MA, USA). Thermal cycling conditions were 94 °C for 4 min, followed by 40 cycles of 94 °C for 1 min; a gradient (40–60 °C) as annealing temperature for 30 s, 72 °C for 30 s, and a final extension for 8 min at 72 °C. After verification using 1.5% agarose gel electrophoresis, amplicons of expected sizes were cleaned with QIAquick PCR Purification (Qiagen, Valencia, CA, USA) and cloned using a PCR Cloning Kit (New England Biolabs, Beverly, MA, USA). Forty transformed colonies for each locus were used as DNA template in 15 µl colony PCR reactions consisting of 1X PCR Buffer (Kapa Biosystems, Woburn, MA, USA), 0.2 mM of dNTPs (New England Biolabs, Beverly, MA, USA), 0.3 µM of each vector primer (provided in the cloning kit) and 1U of *Taq* polymerase (Kapa Biosystems, Woburn, MA, USA). Thermal cycling conditions were as follows: one cycle of 4 min at 94 °C, 30 cycles of 1 min at 94 °C, 30 s at 57°C, and 30 s at 72 °C, followed by a final 8 min incubation at 72 °C. Amplicon size and quality were verified by 1.5% agarose gel electrophoresis before enzymatic purification (ExoSAP-IT, USB-Affymetrix, Cleveland, OH, USA) and cycle sequenced using an ABI

3730XL automatic DNA sequencer (Macrogen Inc., Seoul, Korea). DNA sequences were verified using Codon Code Aligner v3.7.1 (Codon Code Corporation, Dedham, MA, USA).

3.3.4 DNA sequence control tests

DNA sequences obtained from colony PCR experiments were subject to several controls in order to maximize the probability of their being from *CXc*. Initially, an identity control was assessed by comparison with the GenBank database using BLAST (Altschul *et al.*, 1990). For the confirmed rickettsial sequences, a second control consisted of *in situ* hybridization (ISH) assays to verify their physical association with bacterial inclusions in infected tissues. Digoxygenin-dUTP (DIG) labeled probes were prepared using the PCR DIG Probe Synthesis Kit (Roche Applied Science, Mannheim, Germany). Each DIG probe was obtained by nested PCR using locus generic primers and purified amplicons from colony PCRs as templates. Subsequently, ISH was carried out on histological slides prepared with postesophagus, digestive gland, kidney and gill tissues from infected red abalone following Antonio *et al.* (2000). A third control consisted of comparing haplotype-specific PCRs co-amplification patterns tests. For this, we designed specific primers for the DNA sequences (haplotypes) being tested as originating from *CXc* for each locus. These haplotype-specific primers were then used in conjunction with the improved diagnostic *16S rRNA* primers (ss16S set) to assess co-amplification patterns in infected and uninfected abalone (n=16 yellow, n= 15 blue, and n= 8 red abalone). Amplification of both markers in the same organisms (i.e., co-amplification) was interpreted as corroborating evidence that the source of the sequence was *CXc*, whereas lack of co-amplification was interpreted as evidence that the source was another RLP. For final corroboration, all PCR products from these experiments were sequenced as described above.

3.3.5 Phylogenetic analyses

Phylogenetic analyses were conducted separately for each gene and using concatenated DNA and amino-acid (AA) sequences (for protein-coding genes), the latter to improve the resolution of deeper nodes. In order to assess the phylogenetic position of *CXc* in the order *Rickettsiales*, homologous sequences were obtained from representative species of as many lineages as possible for which genomic data were available; including the two well-characterized families *Anaplasmataceae* (*Ehrlichia, Anaplasma, Wolbachia*, and *Neorickettsia*), and *Rickettsiaceae* (*Rickettsia* and *Orientia*); also including *Holosporaceae*

(*Holospora*), *Ca.* Paracaedibacteraceae (*Paracaedibacter* and *Odysella*), *Ca.* Midichloriaceae (*Ca.* Midichloria and an undescribed endosymbiont of *Acanthamoeba* str. UWC8), as well as unclassified RLP *Caedibacter* and *Ca.* Hepatobacter (Annex 1). Phylogenetic reconstructions were outgroup rooted with sequences from Alphaproteobacteria (SAR11 clade) for the *16S rRNA*, *23S rRNA* and *ftsZ* genes (Annex 1). Given that the *rvh* T4 secretion system evolved after the split of "*Candidatus* Pelagibacter" from the *Rickettsiales* ancestor (Gillespie *et al.*, 2010), *virD4* and *virB11* phylogenetic reconstructions were rooted with other outgroup taxa. Furthermore, since the *rvh* P-T4SSs of *Rickettsiales* evolved via horizontally gene transfer from ϒ-*proteobacteria* (Gillespie *et al.* (2010), we rooted these reconstructions with their closest *rvh* xenolog: *trw* P-T4SS sequences from *Xanthomonas* and *Lysobacter* (Annex 1). *Caedibacter varicaedens*, *Ca.* Caedibacter acanthamoebae, *Ca.* Hepatobacter penaei, *Ca.* Odyssella thessalonicensis, *Ca.* Paracaedibacter acanthamoebae, *Holospora obtuse*, and *Holospora undulata* lack T4SS genes hence they were excluded in *virD4* and *virB11* as well as from amino acid based reconstructions (Annex 1). In concatenated DNA and amino-acid analyses, outgroup sequences consisted of assembled chimeras from α- (*16S rRNA*, *23S rRNA* and *ftsZ*) and ϒ- (*virD4* and *virB11*) *proteobacteria*. Alignments were obtained with ClustalW as implemented in Geneious R9 (Kearse *et al.*, 2012). We used maximum likelihood (ML) and Bayesian inference (BI) as methods of phylogenetic reconstruction. The best-fit model of sequence evolution (Annex 2) was estimated with jModeltest v2.1.3 for nucleotide and ProtTest v2.4 for amino-acid sequences (Abascal *et al.*, 2005; Darriba *et al.*, 2012). ML heuristic searches were conducted using 100 random taxon addition replicates with tree bisection and reconnection (TBR) branch swapping with RAxML v7.2.8 (Stamatakis, 2006), whereas BI was conducted using Mr Bayes version 3.2 (Huelsenbeck & Ronquist, 2001). For each data set, the default number of generations (20,000) was increased until achieving a standard deviation smaller than 0.01. Topological congruence was assessed with the Shimodaira-Hasegawa test (SH-test; Shimodaira & Hasegawa, 1999) as implemented in PAUP* version 4 (Swofford, 2002). The SH-test null distribution was estimated by non-parametric resampling generated by a log-likelihood RELL method in PAUP version 4 (Swofford, 2002).

To investigate the chronology of evolutionary changes in hosts in the order *Rickettsiales* leading to *C*Xc, we performed a Bayesian molecular clock-constraint phylogeny with *16S rRNA* gene DNA sequences using the program BEAST version 1.8 (Drummond *et al.*, 2012). The MCMC consisted of 1,000,000 generations with sampling every 200 generations. We used the GTR+Γ+I substitution model and a rate of nucleotide substitution estimated for endosymbiont bacteria (Moran *et al.*, 1993). In this analysis we included data from additional *Rickettsiales* in order to widen the taxonomic scope of the reconstruction.

3.4 Results

3.4.1 *Candidatus* Xenohaliotis californiensis detection.

Inconsistent results and a high frequency of false negative assays with the OIE detection primers (Andree *et al.*, 2000), led us to design a new set of diagnostic primers for *CXc*. DNA sequences of diagnostic amplicons confirmed the presence of *CXc* . Consequently, all diagnostic tests were performed with the proposed ss16S primer set (Table 1).

Table 1: Haplotype-specific primers for *Candidatus* Xenohaliotis californiensis used in control tests.

Locus	ID Primers	5' - Sequence - 3'	Length	Annealing temperature °C	Expected fragment size (bp)
16S rRNA	ss16S.F	GCCTCAGTTTGGCTGGGTTCTTCA	24	66	426
	ss16S.R	GAATTGCCACTTTAAAGTATGGACGG	26		
16S rRNA	ss16SV1.F	GCGGTAGGCTTAATACATGCAAGTTG	26	63	1412
	ss16S2.R	GTTGTCGCCCCCACTGTGAGTGGTTATC	28		
23S rRNA	23S.s1F	CAGTTCGGTTTCTATCCTCCGCAA	24	66	376
	5S.s1R	GGGTGTTTCATCTCCCCTATCAAT	24		
virD4	VirD4.s1F	CTCAGGGTCATGCTGTGTATTGC	23	66	289
	VirD4.s2R	CAACACAACTGCCAAATTGTAGGT	24		
virB11	VirB11.ss2F	GCTCAATCTACTGAGCAGAAAGTTAG	26	60	508
	VirB11.ss1R	CGCAAGCAGGCTTCTATCAATTCTTG	26		
ftsZ	ftsZ.ss1F	CCTGTAATTGCAAGAATTTCTCGTG	25	61	300
	ftsZ.ss1R	CTCTACCTTTGCCTTGCATAACAGAT	26		

3.4.2 Gene amplification, sequencing, and control tests

Fifteen pairs of generic rickettsial primers were designed using the MEME algorithm, and most amplified products of the expected size (Annex 3). Only those designed for *recA*, *rpoB* and *rpoC* failed to yield positive amplifications. Cloning of PCR products revealed sequences from more than one bacterial species for all but the *virB11* gene, for which a single haplotype was obtained. BLAST analyses revealed that only *23S rRNA*, *ftsZ*, *virD4*, *virB11*, *tpiA*, *coxA*, *gap*, and *tkt* genes contained at least one rickettsial haplotype (72-85% similar to the closest rickettsial sequence from GenBank, Annex 4). Successive analyses were carried out on these rickettsial haplotypes. Physical confirmation of *CXc* origin through ISH was observed for the ribosomal gene probes (*16S rRNA* and *23S rRNA*) in the post-esophagus of infected red abalone

(Annex 5). Other probes did not produce detectable signal in the histological slides. However, co-amplification patterns provided reliable evidence that the source of rickettsial gene sequences was *CXc*. Thirty two of the thirty nine tested abalone were carriers of *CXc* and positive co-amplification of the *16S rRNA* gene (Annex 6) with the haplotype-specific primers was confirmed for the genes *23S rRNA*, *ftsZ*, *virD4* and *virB11* (Table 1). Co-amplification mismatches with the genes *coxA*, *tpiA*, *gap* and *tkt* suggests a different RLP source for these DNA sequences. Accordingly, phylogenetic analyses were conducted on *16S rRNA*, *23S rRNA*, *ftsZ*, *virD4* and *virB11* DNA sequences.

3.4.3 Nucleotide and amino acid phylogenetic analyses

Phylogenetic reconstructions were based on 1,304 bp for *16S rRNA*, 312 bp for *23S rRNA*, 429 bp for *ftsZ,* 800 bp for *virB11*, 1,111 bp for *virD4*, for a total of 3,956 bp for the concatenated analysis. For each gene ML and 253 BI reconstructions were identical or not significantly different (SH tests: *16S rRNA* p = 0.052, *23s rRNA* p = 0.076, *ftsZ* p = 0.081, *virB11* p = 0.065, *virD4* p = 1.0), hence we focus on the BI reconstruction. Furthermore, no significant differences were found between the BI individual-gene trees and the concatenated phylogeny in three of the five genes (SH tests: *16S rRNA* p= 0.36, *23s rRNA* p = 0.44, *ftsZ* p = 0.25, *virB11* p = 0.002, *virD4* p = 0.015). Topological differences between the concatenated tree and *virB11* and *virD4* trees, computed on the subset of taxa possessing these genes, related to insufficient resolution of deep nodes (Annex 7-8), hence we focus on the results of the concatenated tree as the most reliable reconstruction. Salient features of this tree included the monpophyly of three well-supported monophyletic families *Anaplasmataceae*, *Ca.* Midichloriaceae, and *Rickettsiaceae*; with the latter as the most ancestral. This lineage also included the unclassified *Rickettsiales* bacterium strain Ac37b (Genbank CP009217) as sister to *Rickettsiaceae*. Of the rest of *Rickettsiales* bacteria, a second well-supported group, sister to the one grouping *Anaplasmataceae*, *Ca.* Midichloraceae, and *Rickettsiaceae*, included the family *Holosporaceae*, *Ca.* Hepatobacter and *Caedibacter. Candidatus* Paracaedibacteraceae was the most ancestral lineage among *Rickettsiales*. *Candidatus* Xenohaliotis californiensis was included in the family *Anaplasmataceae* (posterior probability = 1.0), appearing as sister to *Neorickettsia* and with it in reciprocal monophyly to the monophyletic lineage grouping *Wolbachia*, *Ehrlichia* and *Anaplasma* (topology A, Fig. 1). In order to improve deep resolution, we carried out phylogenetic reconstructions based on AA sequences of the protein coding genes (136 codons for the *ftsZ*, 273 codons for the *virB11* and 383 codons for the *virD4*) separately and concatenated (792 codons). Protein sequences satisfactorily resolved most of the shallow branches of the phylogeny encompassing the monophyly of several genera and the clade

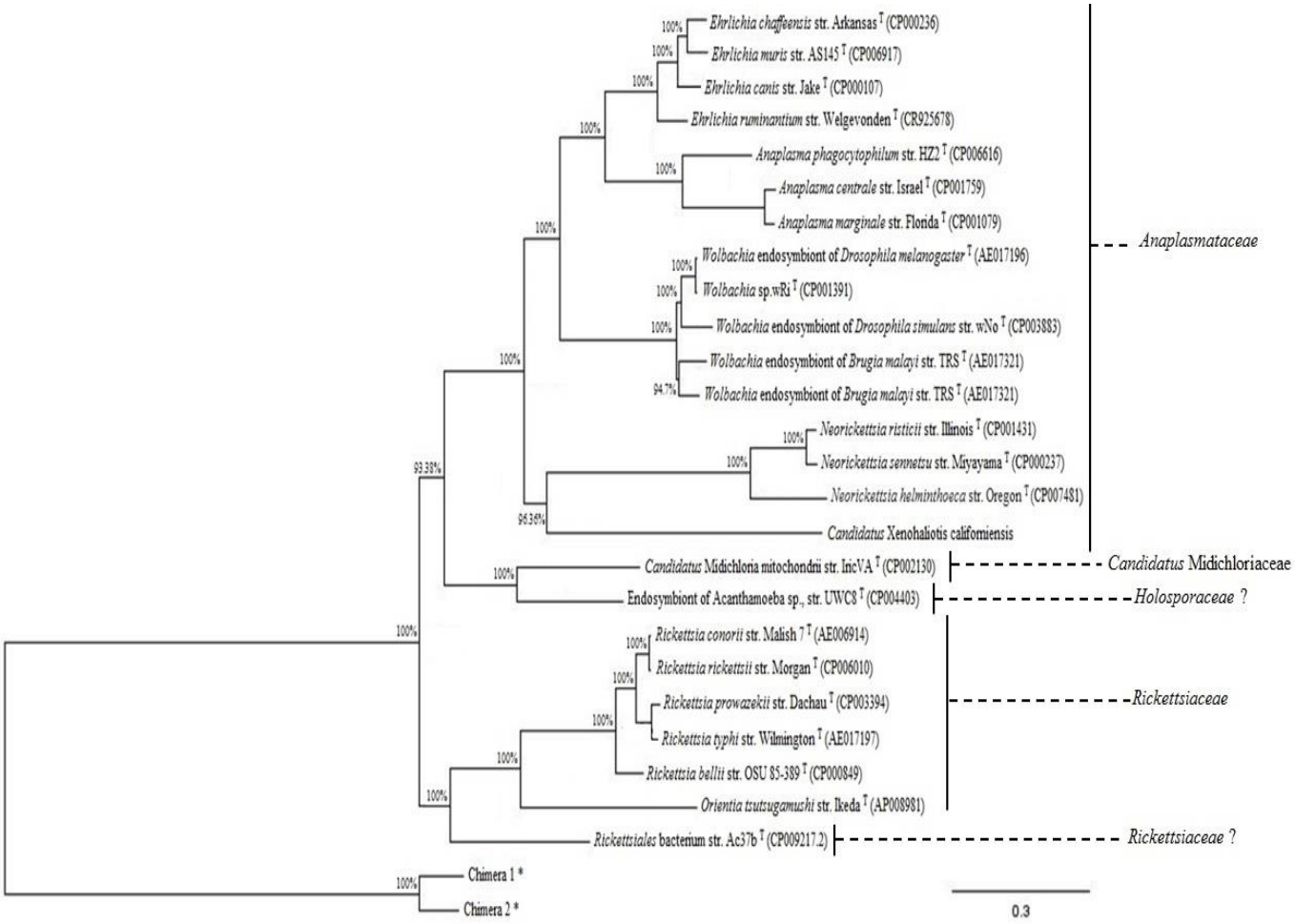


Figure 1. Bayesian phylogenetic tree (Mr Bayes version 3.2) obtained with the concatenated nucleotide sequences of five genes. Nodes labels show posterior probabilities. *= Outgroups were obtained using DNA sequences of different bacteria species and are reported as Chimera.

Ehrlichia, *Anaplasma* and *Wolbachia*. The concatenated AA phylogeny was well-resolved and placed *CXc* as sister to *Neorickettsia*, in the same topology as the DNA based tree (topology A, Fig. 2).

The clock constraint tree places the transition to Rickettsial intracellular habitation between 1070 and 1030 million years ago (mya). The most recent common ancestor (MRCA) of the families *Anaplasmataceae*, *Ca.* Midichloriaceae, and *Rickettsiaceae* lineages lived ca. 870 mya. In *Anaplasmataceae*, the MRCA of bacteria infecting terrestrial hosts (*Ehrlichia, Anaplasma*, and *Wolbachia*) lived 455 mya, which diverged from bacteria infecting aquatic hosts (*Neorickettsia* and *Ca.* Xc) after 568 mya.

Notably, in this 284 reconstruction the marine *CXc* RLP appears as the most basal lineage of the family *Anaplasmataceae* (topology B, Fig. 3).

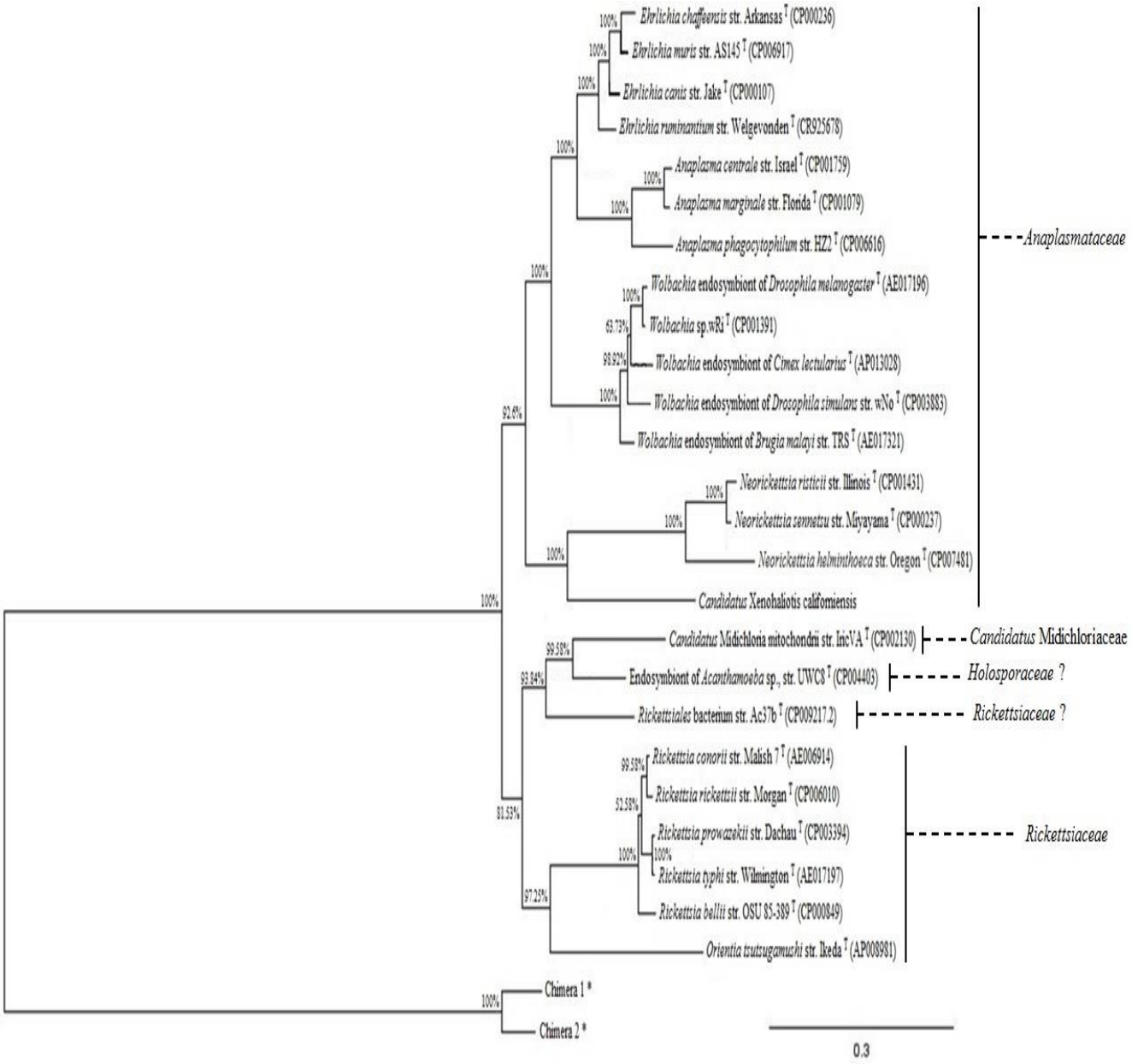


Figure 2. Bayesian phylogenetic tree (Mr Bayes version 3.2) obtained with the concatenated amino acid sequences of tree genes. Nodes labels show posterior probabilities. *= Outgroups were obtained using DNA sequences of different bacteria species and are reported as Chimera.

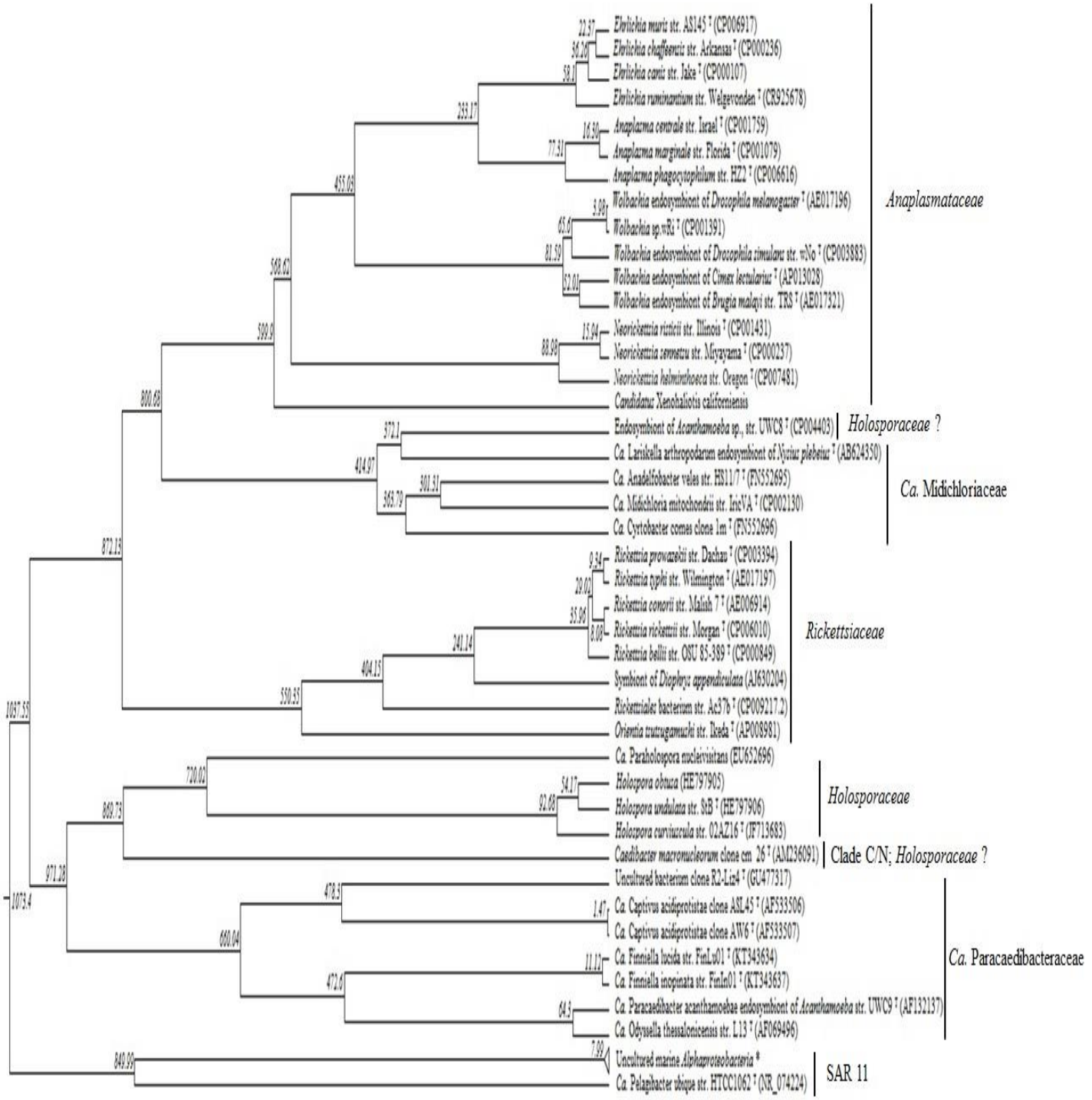


Figure 3. Molecular clock-constraint phylogeny based on Bayesian inference of the order Rickettsiales. Numbers at nodes indicate divergence time in millions of years ago. The collapsed note labelled as "Uncultured marine Alphaproteobacteria" include: Alphaproteobacterium strain HIMB59 (CP003801); Uncultured marine microorganism clone NB062806_218 (KC425551) and Uncultured Alphaproteobacterium clone ARTE1_103 (GU230260).

3.5 Discussion

3.5.1 Sequencing of *Candidatus* Xenohaliotis californiensis DNA

Our experimental approach successfully allowed us to sequence the target *CXc* DNA without the need of isolation or axenic culture. The physical association of the amplified DNA with bacterial inclusions in infected tissues was confirmed through ISH for the two ribosomal genes. We speculate that the lack of detectable ISH signal in the non-ribosomal genes, using the same experimental approach, is likely due to a number of target gene sequences below the detection level of ISH (Speel *et al.*, 1999). To our knowledge, the structural genes used in this study are single copy in rickettsial bacteria and may provide an insufficient number copies for ISH detection. In contrast, ribosomal RNA can accumulate in the cytoplasm reaching millions of copies and up 33% of the total dry cell weight (El-Sharoud, 2008). Nevertheless, the adoption of a third verification protocol based on co-amplification of target genes with *C*Xc-diagnostic *16S rRNA* allowed us to differentiate target and non-target sequences.

3.5.2 *Candidatus* Xenohaliotis californiensis phylogenetic affinities

Even though we found topological differences in individual gene trees, they were only significantly different from the concatenated analysis in the T4SS gene trees. These differences may relate to the distinct evolutionary dynamics of these genes, which are well documented (Gillespie et al., 2010). Nevertheless, the reconstruction carried out with the five concatenated genes (3,956 bp) represents the most extensive analysis of *CXc* evolutionary relationships and provided improved resolution surrounding its phylogenetic position among Rickettsiales (most posterior probabilities >95% in *Anaplamataceae*) compared to published reconstructions based on the *16S rRNA* gene (Friedman *et al.*, 2000; Lee *et al.*, 2005; Vannini *et al.*, 2005; Martijn *et al.*, 2015).

Sometimes the use of DNA sequences may lead to alignments that do not reflect the appropriate mutational history and thus incorrect phylogenetic trees (Philippe *et al.*, 2011). The difficulties in reconstructing accurate phylogenetic trees may relate to speciation events closely spaced in time that do not allow for the evolution of synapomorphic characters or, in the case of ancient nodes and relationships, to long terminal branches abounding of multiple substitutions at the same sequence position, erasing the original phylogenetic signal and producing homoplasies. In consequence, resolving some deep divergences may be challenging even with the use of long DNA sequences (Mossel & Steel, 2005; Philippe

et al., 2011). The use of proteins or expressed sequence tags may help to resolve phylogenetic incongruences and is becoming a standard approach to investigate evolutionary relationships (Schreiber *et al.*, 2009; Philippe *et al.*, 2011). Hence, we investigated phylogenetic reconstructions based on AA sequences. In contrast to recent studies that use only the *16S rRNA* gene DNA sequences (Friedman *et al.*, 2000; Lee *et al.*, 2005; Vannini *et al.*, 2005; Martijn *et al.*, 2015), our most reliable reconstructions (DNA and AA concatenated phylogenetic trees) placed *CXc* unequivocally as a member of the family *Anaplasmataceae* and as sister of *Neorickettsia*, in reciprocal monophyly to the rest of the family members (topology A). Additionally, our molecular-clock phylogeny suggests a basal position of *CXc* within the family *Anaplasmataceae* (topology B). Given the aquatic habits of the *C*Xc-*Neorickettsia* clade, our finding supports the hypothesis of an aquatic origin of *Anaplasmataceae* and the subsequent transition to the terrestrial environment by the lineages giving rise to *Wolbachia*, *Anaplasma* and *Ehrlichia* (Kang *et al.*, 2014).

3.5.3 Evolutionary timeline and habitat/host transitions

In agreement with previous publications, the origin of species belonging to the families *Rickettsiaceae*, *Anaplasmataceae* and *Candidatus* Midichloriaceae evolved around 870 million years ago (Luo *et al.*, 2013). At that time, genome changes may have involved genetic reorganization which in turn may have modified the metabolic pathways used by these bacterial families (Fuxelius *et al.*, 2007; Merhej & Raoult, 2011; Luo *et al.*, 2013). The position of *CXc* as the most basal *Anaplamataceae* lineage (topology B) in the molecular-clock constraint tree is noteworthy since in the concatenated analysis it is not (topology A). Given the wider taxonomic scope of the molecular clock tree, this could result from improved character polarization provided by increased taxonomic sampling. Also, the basal position of *CXc* may have resulted from enforcing a uniform substitution rate in the ultrametric tree, placing the longest branch among *Anaplasmataceae* sequences (i.e., the one leading to *CXc*) at the most basal position in the family.

In light of the above and given that "*Candidatus* Xenohaliotis californiensis" has been detected only in marine invertebrate hosts, it is possible to reevaluate the evolution of some biological characteristics within the *Anaplasmataceae* family consistent with our phylogenetic insights: (*i*) the ability to live in the marine environment should be considered an ancestral characteristic of this bacterial family; (*ii*) the ability to infect invertebrate hosts is a plesiomorphic character; and (*iii*) since transmission of *CXc* is direct and presumed to be fecal-oral (Moore *et al.*, 2001), the absence of known intermediate vectors or reservoir

organisms mediating transmission is an ancestral character (Cruz-Flores *et al.*, 2015). Consequently, vector mediated infection evolved as a derived character in lineages infecting land animals. Moreover, we speculate that the use of fish and trematode hosts used by *Neorickettsia* in aquatic environments (Headley *et al.*, 2011) likely preceded the use of arthropods used by *Wolbachia*, *Anaplasma* and *Ehrlichia* in terrestrial environments.

3.5.4 Diversity of rickettsial endosymbionts in abalone gastrointestinal tract

Given the approach of subsequent refinement of the primers used to target *CXc* DNA, we detected new bacterial DNA sequences assignable to the order *Rickettsiales* but not to the newly described *C*Xc. This suggests that abalone species may harbor still uncharacterized RLP organisms, as previous proposed by Crosson *et al.* (2014). However, further research is needed to determine if the detected sequences belong to potentially pathogenic species.

NOTE

This chapter has been published as:

Cicala, F., J. D. Moore, J. Cáceres-Martinez, M. A. Del Río-Portilla, M. Hernández-Rodríguez, R. Vásquez-Yeomans and A. Rocha-Olivares (2017). "Multigenetic characterization of '*Candidatus* Xenohaliotis californiensis'." *International Journal of Systematic and Evolutionary Microbiology* **67**(1): 42-49.

Chapter 4: Monomorphic pathogens: the case of *Candidatus* Xenohaliotis californiensis from abalone in California, USA and Baja California, Mexico

4.1 Abstract

Withering syndrome (or WS) is a chronic wasting disease affecting abalone species attributed to the pathogen "*Candidatus* Xenohaliotis californiensis" (*C*Xc). Wild populations of blue (*Haliotis fulgens*) and yellow (*H. corrugata*) abalone have experienced unusual mortality rates since 2009 off the Baja California peninsula and WS has been hypothesized as a possible cause. Currently, little information is available about the genetic diversity of *CXc* and particularly the possible existence of strains differing in pathogenicity. In a recent phylogenetic analysis, we characterized five coding genes from this rickettsial pathogen. Here, we analyze those genes and two additional intergenic non-coding regions following multi-locus sequence typing (MLST) and multi-spacer typing (MST) approaches to assess the genetic variability of *CXc* and to clarify its relationships with blue, yellow and red (*H. rufescens*) abalone. Moreover, we used *16S rRNA* sequences from gut microbiomes of blue and yellow abalone to complete the genetic characterization of *CXc*. Overall, more than 150 abalone of the three species were analyzed and a total of 358 DNA sequences and 7,117 *16S rRNA* reads from *CXc* were used to evaluate its population genetic structure. Our findings suggest the absence of polymorphism in the DNA sequences of analyzed loci and the presence of a single lineage of *CXc* infecting abalone from California and Baja California. The absence of genetic variability may result from selective pressure shaping the genome of this species.

4.2 Introduction

The ability to detect genetic structure in pathogenic bacterial populations may provide useful information to understand epidemiological events and trace disease outbreaks (Maiden *et al.*, 1998; Fournier *et al.*, 2003, 2004; Fournier & Raoult, 2007). Accordingly, a variety of phenotypic and genotypic methods have been developed for bacterial inter- and intra-specific characterization (Fournier & Raoult, 2007). Among them, PCR-based DNA sequence-typing analyses are some of the most rapid, reliable and easiest methods to identify bacteria beyond the species hierarchical group (Maiden *et al.*, 1998; Fournier & Raoult, 2007).

Maiden *et al.* (1998) proposed Multi-locus Sequence Typing (or MLST) as a universal and portable method for characterizing bacteria strains. This method involves the sequencing of several housekeeping gene fragments and defines a bacterial strain in terms of its "sequence type" (or haplotype) consisting in the combined DNA sequences from the different loci (Maiden *et al.*, 1998). Moreover, Fournier *et al.* (2004) proposed a second approach named Multi Spacer Typing (or MST), which is based on the analysis of intergenic non-coding spacer sequences instead of structural genes. The assumption of MST is that non-coding regions are less subject to selective pressure than housekeeping genes and consequently may accumulate more genetic changes over time (Fournier & Raoult, 2007; Fournier *et al.*, 2004).

Both approaches have been successfully used for inter- and intra-specific genetic characterization of different bacterial species (Thompson *et al.*, 2005; Martino *et al.*, 2011; Santos *et al.*, 2012) including some in the order *Rickettsiales* (Baldo *et al.*, 2006; Fournier & Raoult, 2007; Fournier *et al.*, 2004; Zhu *et al.*, 2005). *Candidatus* Xenohaliotis californiensis (*CXc*) is an obligate Gram-negative intracellular bacterium in the family *Anaplasmataceae* from the order *Rickettsiales* (Garrity *et al.*, 2004; Friedman, 2012; Cicala *et al.*, 2017). It is an intracellular organism, inhabiting the cytoplasm of abalone (*Haliotis* spp.) digestive epithelial cells (Friedman *et al.*, 2000). *CXc* has been identified as the etiological agent of a chronic disease known as Withering syndrome (WS) (Haaker *et al.,* 1992; Gardner *et al.*, 1995) or abalone Rickettsiosis (Cáceres Martínez, 2002; Friedman, 2012). Advanced stages of WS include severe morphological and physiological anomalies resulting in physiological starvation, anorexia, catabolism of the food muscle and finally demise (Moore *et al.*, 2000; Friedman *et al.*, 2003; Crosson *et al.*, 2014).

On the Pacific coast of the Baja California peninsula (Mexico), the abalone fishery is among the most valuable (SAGARPA, 2009). The Mexican peninsula contains seven exploitable abalone species; however, the actual harvest is focused almost entirely on blue (*H. fulgens* or HF) and yellow (*H. corrugata* or HC) abalone (Morales-Bojórquez *et al.*, 2008; SAGARPA, 2009). Since 2009, unusual mortalities have been observed in wild populations of both species in different fishery areas of southern Baja California (Cáceres-Martínez *et al.*, 2011). Among the possible causes responsible for these unusual mortality events is the existence of *CXc* strains differing in pathogenicity (Cáceres-Martínez *et al.*, 2011). Thus far, *CXc* remains an uncultured bacterium and, until recently, the *16S rRNA* gene was the only known DNA sequence. The few sequences available from this ribosomal gene were insufficient to assess levels of intra-specific variability (Balseiro *et al.*, 2006; Friedman, 2012; Kiryu *et al.*, 2013). Thus, the limited genetic knowledge about *CXc* genetic variability has been the primary obstacle in understanding its influence on WS pathogenesis.

Recently, we characterized partial sequences from the *16S rDNA*, *23S rDNA*, *ftsZ*, *VirB11* and *VirD4* genes to resolve the phylogenetic position of *CXc* among *Rickettsiales* (Cicala *et al.*, 2017). Here, we use these genes under the framework of MLST, adding two intergenic regions, located between *23S-5S rDNA* (Genbank access number: KY882033) and *VirB11-VirD4* (Genbank access number: KY882034) genes, under a MST framework in order to characterize the genetic diversity of *CXc* at different levels: (*i*) geographically, across exploited areas in southern Baja California, (*ii*) taxonomically, between HC and HF hosts. Additionally, we also analyzed eight specimens of red abalone *H. rufescens* (or HR) from northern California, USA, to extend the geographic and taxonomic scope of the survey.

4.3 Materials and methods

4.3.1 Sampling, experimental design and DNA extraction.

Gastrointestinal tract (GI) tissues from wild HF and HC were sampled (April 2012 and May 2013) from eight localities during commercial fishing operations along the Pacific coast of southern Baja California, Mexico (Annex 9). We also collected GI tissue from HR abalone maintained in the UC Davis Bodega Marine Laboratory (California, USA).

The experimental design involved generating at least 3 sequences from each locus (n=7), species (n=3) and geographic locality (n=8 for HF and HC, and n=1 for HR). Additional sequences were subsequently obtained for the most variable loci, geographic areas, or both.

Approximately 30 mg of postesophageal tissue from each abalone was excised and immediately transferred to sterile 1.5 ml microcentrifuge tubes containing molecular grade ethanol (absolute grade).

Total DNA was extracted from abalone gut tissues using a DNeasy tissue kit (Qiagen, Valencia, CA, USA), following manufacturer's protocols. PCR fragments were produced using published (*23S rDNA*- ITS.1 and *ftsZ* loci; Cicala *et al.*, 2017) and newly designed primers (*16S rDNA*, *VirB11*- ITS.2-*VirD4,* Table 2).

4.3.2 Gene selection and PCR amplification

The MLST and MST approaches were performed using DNA sequences from five genes (products shown in parentheses): *16S rRNA* (small ribosomal sub-unit), *23S rRNA* (large ribosomal sub-unit), *ftsZ* (a cell division protein), *VirD4* (a type IV secretion system protein), *VirB11* (a type IV secretion system protein) and two intergenic non-coding spacer sequences located between the *23S-5S* rDNA (ITS.1) and *VirB11-VirD4* (ITS.2) genes.

Amplifications were carried out in 20 µl reactions containing: 120 ng of DNA, 1X PCR buffer, 1.5 mM $MgCl_2$ (both Kapa Biosystems, Woburn, MA, USA), 0.2 mM dNTPs (New England Biolabs, Beverly, MA, USA), 0.3 µM of each primer, and 1U of *Taq* polymerase (Kapa Biosystems, Woburn, MA, USA). Thermal cycling conditions were: 94 °C for 4 min, followed by 40 cycles of 94 °C for 1 min, with specific annealing temperatures for each locus (Table 2) of 30 s, 72 °C for 30 s, and a final extension for 8 min at 72 °C. For the *VirB11*- ITS.2 -*VirD4* region, nested PCRs (combining the primers VirB11ss3F and VirD4ss1R; Table 2) were used to increase the PCR yield or to reduce nonspecific amplification.

Table 2. Loci and primers designed for *Candidatus* Xenohaliotis californiensis amplification. L= primer Length; Tm = annealing temperature (°C); Size = approximate fragment size (base pairs).

Locus	ID Primers	5' - Sequence - 3'	L	Tm	Size	Reference
16S (for 454)	28F	GAGTTTGATCNTGGCTCAG	19	62	500	Ludwig *et al.,* 1993
	519R	GTNTTACNGCGGCKGCTG	18			Ruff-Roberts *et al.,* 1994
16S	ss16SV1.F	GCGGTAGGCTTAATACATGCAAGTTG	26	66	1400	Cicala *et al.,* 2017
	ss16S1.R	GAATTGCCACTTTAAAGTATGGACGG	26			this study
ftsZ	ftsZ.ss1F	CCTGTAATTGCAAGAATTTCTCGTG	25	61	300	Cicala *et al.,* 2017
	ftsZ.ss1R	CTCTACCTTTGCCTTGCATAACAGAT	26			Cicala *et al.,* 2017
23S - ITS.1	23S.s1F	CAGTTCGGTTTCTATCCTCCGCAA	24	66	370	Cicala *et al.,* 2017
	5S.s1R	GGGTGTTTCATCTCCCCTATCAAT	24			Cicala *et al.,* 2017
VirB11 - ITS.2 - VirD4	VirB11 ss3F	GAGCGTTTAATAACTGTTGAGGATGC	26	65	1400	this study
	VirD4 ss2R	CAACACAACTGCCAAATTGTAGGT	24			this study
	VirD4 ss1R	GCAATACACAGCATGACCCTGAG	23		1100	this study

Amplicon size and quality were verified by 1.5% agarose gel electrophoresis before enzymatic purification (ExoSAP-IT,USB-Affymetrix, Cleveland, OH, USA) and cycle sequencing using an ABI 3730XL automatic DNA sequencer (Macrogen Inc., Rockville, MD, USA).

4.3.3 Microbiome sequencing and bio-informatics pipeline

Additional *CXc* genetic diversity data was obtained from GI tract microbiomes. A fragment of approximately 400 bp of the *16S rRNA* gene was amplified using the universal eubacterial primers 28F (Ludwig *et al.*, 1993) and 519R (Ruff-Roberts *et al.*, 1994) (Table 1) from HF and HC GI tissue samples. The PCR master mix was prepared in a 20 µl reaction of 100 ng of DNA, 1X PCR Buffer (Kapa Biosystems, Woburn, MA, USA), 1.5 mM of magnesium chloride (Kapa Biosystems, Woburn, MA, USA), 0.2 mM of dNTPs (New England Biolabs), 0.5 µM of each primer, 0.4 mM of bovine serum albumin (New England Biolabs, Beverly, MA, USA) and 1U of Taq polymerase (Kapa Biosystems, Woburn, MA, USA). Thermal cycling conditions were as follows: an initial denaturation at 94 °C for 4 min, 40 cycles of 94 °C for 1 min; 62°C for 30 sec. and 72 °C for 30 sec., and a final extension step of 8 min at 72 °C. Confirmation of amplification was carried out by 1.5% agarose gel electrophoresis.

Amplicons were subject to massive parallel next generation sequencing using bacterial tag-encoded FLX Titanium amplicon pyrosequencing (bTEFAP) following Dowd *et al.* (Dowd *et al.*, 2008) (Research and Testing Laboratory, Lubbock, TX).

The *16S rRNA* raw reads were analyzed using the software package Quantitative Insights Into Microbial Ecology (QIIME) version 1.9.1 (Caporaso *et al.*, 2010b). Initial analyses included demultiplexing and primer deletion. Next, reads were filtered using quality scores generated during pyrosequencing according to the following quality control (QC) criteria: (*i*) minimum and maximum length settled at 250 and 550 bp, respectively; (*ii*) the default minimum quality Phred score was set at 25; (*iii*) no mismatches in primers and barcode sequences were allowed; (*iv*) maximum homopolymer length was set at 8 bp. After QC tests, the remaining sequences were clustered into operational taxonomic units (OTUs) at 99% sequence similarity using the UCLUST algorithm (Edgar, 2010). We chose the longest sequence as the OTU-representative sequence. These sequences were subsequently aligned with the Python Nearest Alignment Space Termination (PyNAST) algorithm (Caporaso *et al.*, 2010a). Chimeras and singletons were detected and removed with ChimeraSlayer (Haas *et al.*, 2011) as implemented in QIIME. Additionally, a second chimera

test control and taxonomic assignment was carried out using a BLAST search against the SILVA database (http://www.arb-silva.de/).

4.4 Results

4.4.1 Detection of *CXc* and DNA sequence data from tissues

The presence of *CXc* was detected using PCR assays with GI tissue from 69 abalone (32 HC, 29 HF and 8 HR; Annex 10). A total number of 375 sequences were successfully sequenced for MLST and MST analyses (Annex 11). The amplification of the *VirB11*- ITS.2 -*VirD4* region proved challenging due to the occurrence of false negatives, low PCR yield and occasional nonspecific amplification. This was solved by performing nested PCRs in five HC and four HF abalone (Annex 10). Finally, nine HC and seven HF were considered false negative organisms for this molecular marker (Annex 10).

Microbiomes were obtained from 86 additional abalone (37 HC and 49 HF). *C*Xc was detected in 50 of them (19 HC and 31 HF; Annex 12). The total number of reads (451,095 raw sequences) was reduced to 390,423 after QC, of which 7,117 belonged to *CXc*.

Table 3. Number of 16S rRNA gene reads of *Candidatus* Xenohaliotis californiensis (*C*Xc) found in microbiomes obtained by 454 pyrosequencing and frequency of ribotypes 1 (fR1) and 2 (fR2) in *H. corrugata* (HC) and *H. fulgens* (HF). N= number of analyzed abalone.

Species	N	*C*Xc Reads	fR1	fR2
HC	19	3861	0.99	0.01
HF	31	3256	0.98	0.02

4.4.2 Genetic diversity

Most loci analyzed through MLST and MST genotyping were monomorphic, showing a single DNA sequence in all analyzed abalone; hence, no genetic variation could be associated with (*i*) different geographic areas; (*ii*) host species (HC, HF, and HR).

However, a second *16S rDNA* haplotype was detected in one HC from the California fishery cooperative region (Annex 13a) and a second *ftsZ* haplotype was detected in one HF from Punta Abreojos (Annex 13b). The mutation types of these haplotypes consisted of a single nucleotide deletion for the *16S rDNA* and a substitution for the *ftsZ* genes (Table 4). These mutations produced no amino acidic changes in *ftsZ* protein and minor changes in the secondary structural in the *16S rRNA*.

According to the 454 pyrosequencing results, two *CXc* ribotypes (labelled as ribotype 1 and ribotype 2), differing in 3 nucleotide deletions (Table 4), were obtained from both HC and HF abalone. Nevertheless, ribotype 1 was dominant (> 98%) in both HF and HC (Table 3).

Table 4. Genetic variation in DNA sequences of loci analyzed from *Candidatus* Xenohaliotis californiensis in green, yellow and red abalone. GA = Genotyping approach, SS= Sequencing strategies; Size = average fragment length (base pairs), N= Number of 454 pyrosequencing reads or Sanger sequences analyzed; h= number of haplotypes; POL= polymorphism description (- monomorphic loci).

Locus	GA + SS	Size	N	h	POL
16S rDNA	Microbiome + 454	400	8060	2	3bp indel
16S rDNA	MLST + Sanger	900	52	2	1bp indel
23S rDNA	MLST + Sanger	250	54	1	-
ftsZ	MLST + Sanger	350	58	2	1 SNP
VirB11	MLST + Sanger	350	45	1	-
VirD4	MLST + Sanger	450	45	1	-
ITS.1	MST + Sanger	80	54	1	-
ITS.2	MST + Sanger	90	45	1	-

4.5 Discussion

The identification of intra-specific bacterial strains may play a pivotal role in tracing disease outbreaks and may improve understanding of the ecology and population structure of these microorganisms (Maiden *et al.*, 1998; Fournier *et al.*, 2003, 2004; Fournier & Raoult, 2007). With this in mind, we assessed the genetic diversity of *CXc* across space in California (USA) and Baja California Sur (Mexico), among host-species (HF, HC, and HR), using a total of seven molecular markers that included housekeeping genes and intergenic non-coding regions. Finally, to complete the genetic characterization of *CXc*, the ribosomal *16S rRNA* from HC and HF abalone was sequenced by both Sanger and 454 Pyrosequencing sequencing. The latter was included because we hypothesize that the physical amplicon separation that occurs during pyrosequencing may increase the sensitivity to detect polymorphisms compared to Sanger sequencing, given the much larger number of sequences (i.e., reads) generated.

All tested loci were monomorphic, except for the *16S rRNA* and *ftsZ* genes, for which a single alternative DNA sequence was observed. Nevertheless, the frequency of the second haplotype was too low ($f < 0.04$) to consider the gene as polymorphic. Similar results were obtained by 454 pyrosequencing.

The lack of genetic variability suggests the existence of a single and widespread lineage or strain of *CXc* in the samples analyzed. Furthermore, the simplest explanation for the near absence of genetic polymorphisms may lie in the highly conserved nature of the selected loci. However, the extreme absence of genetic polymorphisms in both coding and non-coding sequences, as well as the low inter and intra-specific genetic variability already reported in other species of order *Rickettsiales* (Roux & Raoult, 1995; Raoult & Roux, 1997; Andersson *et al.*, 1998) may suggest other possibilities. In this context, appreciable levels of genetic diversity are observed in most microbial taxa and the sequencing of housekeeping genes may provide a good resolution of their population genetic structure (Maiden, 2006; Linz *et al.*, 2007). On the other hand, a number of taxa, including many pathogenic bacteria as *Bacillus anthracis* and *Yersinia pestis*, display low or null levels of sequence diversity in these genes (Achtman, 2008). The simplest explanation for genetically monomorphic pathogens is that the ancestral population of all living descendants underwent a bottleneck (Achtman, 2008).

In the case of the order *Rickettsiales*, the transition from free-living to intracellular live strategy occurred around 1 billion of years ago (Luo *et al.*, 2013; Cicala *et al.*, 2017) and likely coincided with a profound genome size reduction that nowadays characterize most if not all *Rickettsiales* as well as other intracellular bacteria (Sällström & Andersson, 2005; Blanc *et al.*, 2007; Fournier *et al.*, 2009; Wolf & Koonin, 2013).

Moreover, a more recent loss of genetic diversity may have occurred in the last decades at the time in which *C*Xc first colonized California (USA) in the mid-1980s (Altstatt *et al.*, 1996; Moore *et al.*, 2002) and Baja California (Mexico) in the early 1990s (Altstatt *et al.*, 1996); since before this time *C*Xc was likely absent from this region (J.D. Moore, personal observations). Consequently, the low genetic variability of this bacterium may have resulted from a founder effect.

Accordingly, we speculate that the lack of genetic variability of *CXc* may be attributed to the synergistic contributions of these events involving a strong genomic reduction that presumably increased the levels selective pressure on the remaining genes and/or genome in conjunction with hypothetical changes in the effective size of the ancestral populations of extant *CXc*. Nevertheless, the obligate intracellular habitat of rickettsial bacteria has been proposed as a physical isolation that may lead to local genetic reorganizations following an allopatric evolutionary model (Darby *et al.*, 2007; Gillespie *et al.*, 2010; Merhej & Raoult, 2011). In other words, the possibility that different abalone species may harbor host-specific *CXc* strains cannot be ruled out until more variable genomic regions or SNPs are found and analyzed.

NOTE

We are preparing this chapter to be submitted for publication as:

Cicala, F., J. D. Moore, J. Cáceres-Martinez, M. A. Del Río-Portilla, M. Hernández-Rodríguez, R. Vásquez-Yeomans and A. Rocha-Olivares. " Monomorphic pathogens: the case of *Candidatus* Xenohaliotis californiensis from abalone in California, USA and Baja California, Mexico ". *Applied and Environmental Microbiology*.

Chapter 5: Structure, dynamics and predicted functional ecology of the gut microbiota of the blue (Haliotis fulgens) and yellow (H. corrugata) abalone from Baja California Sur, Mexico

5.1 Abstract

The GI microbiota of abalone contains a highly complex bacterial assemblage playing an essential role in the overall health of these gastropods. The gut bacterial communities characterized so far reveal considerable interspecific variability, likely resulting from bacterial interactions and constrained by the ecology of their host species; however, they remain poorly investigated. Additionally, the extent to which structural changes in the microbiota entail functional shifts in metabolic pathways of bacterial communities remains unexplored. In order to address these questions, we characterized the gut microbiota of the northeast Pacific blue (*Haliotis fulgens* or HF) and yellow (*Haliotis corrugata* or HC) abalone by *16S rRNA* 454 pyrosequencing to shed light on: (*i*) their gut microbiota structure; (*ii*) how bacteria may interact among them; and (*iii*) predicted shifts in bacterial metabolic functions associated with the observed structural changes. Our findings revealed that *Mycoplasma* dominated the GI microbiome in both species. However, the structure of the bacterial communities differed significantly in spite of considerable intra-specific variation. This resulted from differences of the species with most reads in each GI metagenome, suggesting host-specific adaptation of bacterial lineages to these sympatric abalone. We hypothesize that the presence of exclusive OTUs in each microbiota may relate to host-specific differences in competitive pressure. Significant differences in bacterial diversity were found for the explored metabolic pathways between species despite their functional overlap. A more diverse array of bacteria contributed to each function in HC, whereas a single or much fewer OTUs were generally observed in HF. The structural and functional analyses allowed us to describe a taxonomic and functional split between the microbiota of HF and HC abalone.

5.2 Introduction

The gastro-intestinal (GI) tract of metazoans may be considered a highly complex ecosystem inhabited by a large number of bacteria (Backhed, 2005). For instance, it is considered that, the commensal microbiota

harbored by the human GI tract far exceeds the total number of cells in the entire human body, and their collective genome (microbiome) is orders of magnitude larger than our own (Backhed, 2005; Bates *et al.*, 2006). Moreover, the GI microbiome has been associated with essential physiological activities such as food digestion, nutrient assimilation, and defense against invasion of foreign bacterial species; which in turn may prevent epidemiologic outbreaks (Blaut & Clavel, 2007; ten Doeschate & Coyne, 2008; Zhao *et al.*, 2012). Also, functional studies have revealed that the relationship between the gut microbiome and its host may be so close that bacteria may be directly involved in the maturation of the GI tract of the hosts species (Bano *et al.*, 2007; Bates *et al.*, 2006; Bry *et al.*, 1996).

As documented by cultured and uncultured approaches, the composition of the abalone gut microbiota may be influenced by a great variety of factors such as diet, environmental conditions and ontogenetic stages (Sawabe *et al.*, 2003; Tanaka *et al.*, 2003; Pang *et al.*, 2006; ten Doeschate & Coyne, 2008; Zhao *et al.*, 2012; Meryandini *et al.*, 2015). Also, the use of probiotics suggested that interspecific bacterial relationships may also shape the final gut microbiome composition residing in several marine invertebrates, including abalone (Macery & Coyne, 2004; Iehata *et al.*, 2014; Rungrassamee *et al.*, 2014). Overall, these factors may explain the consistent differences in the gut microbiome of abalone species studied so far. In this context, the most abundant bacteria in homogenate samples of the entire GI of *H. discus hannai* were fermenter γ- *proteobacteria*, such as *Vibrio halioticoli* as well as other *Vibrio* species, α- *proteobacteria*, *Mollicutes* and *Fusobacteria* (Tanaka *et al.*, 2003, 2004). Moreover, the intestinal microflora (from the stomach to the anus) of *Haliotis diversicolo*r was dominated by *Mollicutes*, *Flammeovirga*, as well as β- , a- , γ- and δ- *proteobacteria* (Huang *et al.*, 2010). In contrast, the bacterial composition of *H. gigantea* (from homogenate samples of the entire GI) appears less complex with a preponderance of γ- *proteobacteria* and *Mollicutes* (Iehata *et al.*, 2014).

The peninsula of Baja California harbors seven exploitable abalone species (Morales-Bojórquez *et al.*, 2008), two of which, the blue abalone *Haliotis fulgens* (HF, henceforth) and the yellow abalone *Haliotis corrugata* (HC, henceforth), sustain a high-valued fishery in the NW Mexican Pacific (Morales-Bojórquez *et al.*, 2008; SAGARPA, 2009). Despite the importance of the GI microbiomes for the survival of these abalone species, no efforts have been made to characterize them. Furthermore, it is equally uncertain which factors may shape their final composition as well as the functional roles played by the most representative bacterial groups.

Here, with the aim of addressing these questions, we analyze the structure of the GI microbiota of wild-caught specimens of HC and HF by means of *16S rRNA* amplification and 454 pyrosequencing (Roche).

Additionally, we analyze the functional shifts involved in structural changes using a predictive metagenomic analysis to focus on 86 genes involved in several metabolic pathways.

5.3 Materials and methods

5.3.1 Sample collection and genetic analyses.

Wild abalone (n = 31 HF, n = 35 HC) were sampled from the commercial harvested along the Pacific coast of central Baja California, Mexico. Approximately 30 mg of post esophageal tissue were dissected from visually healthy animals bearing no signs of the withering syndrome (Friedman, 2012), and immediately transferred to sterile 1.5 ml microcentrifuge tubes containing molecular grade ethanol, until further analysis. Abalone and bacterial DNA was extracted and purified from preserved tissues using DNeasy blood & tissue kit (Qiagen, Valencia, CA, USA) following manufacturer's protocols.

A fragment of the bacterial ribosomal *16S rRNA* spanning V1-V3 regions was PCR amplified using universal eubacterial primers 28F: 5' - GAGTTTGATCNTGGCTCAG - 3' (Ludwig *et al.*, 1993) and 519R: 5' - GTNTTACNGCGGCKGCTG - 3' (Ruff-Roberts *et al.*, 1994). PCR reactions (20 µl) contained: 1X PCR Buffer (Kapa Biosystems, Woburn, MA, USA), 1.5 mM magnesium chloride (Kapa Biosystems, Woburn, MA, USA), 0.2 mM dNTPs (New England Biolabs), 0.5 µM each primer, 0.4 mM bovine serum albumin (New England Biolabs, Beverly, MA, USA), 1U *Taq* polymerase (Kapa Biosystems, Woburn, MA, USA), and 100 ng purified DNA. Thermal cycling consisted of an initial incubation at 94 °C for 4 min, followed by 40 cycles of: 94 °C for 1 min; 62°C for 30 sec. and 72 °C for 30 sec., and a final incubation of 8 min at 72 °C. Confirmation of amplification was carried out by 1.5% agarose gel electrophoresis. Amplicons were subsequently tagged using Roche 454 adaptors and multiplex identifier (MID) tags for each organism, following the bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) approach of Dowd *et al.* (Dowd *et al.*, 2008). Following normalization, Roche 454 pyrosequencing was carried out in a GS FLX Titanium platform by Research and Testing Laboratory (Lubbock, TX).

5.3.2 Bioinformatic analyses

The *16S rRNA* reads were analyzed using Quantitative Insights Into Microbial Ecology (QIIME) software

version 1.9.1 (Caporaso *et al.*, 2010b). The first step consisted in demultiplexing and primer removal. Subsequently, reads were filtered according to Phred quality scores obtained from the 454 pyrosequencing. Acceptance quality criteria consisted of: (*i*) minimum and maximum lengths of 250 and 550 bp, respectively; (*ii*) default minimum quality Phred score of 25; (*iii*) no mismatch in primer sequences and MID-tag; (*iv*) maximum homopolymer length of 8 bp.

Sequences that met quality criteria were clustered in operational taxonomic units (OTUs) at 97% sequence similarity using the UCLUST algorithm (Edgar, 2010). The longest sequence from each OTU was selected as representative and these were subsequently aligned using Python Nearest Alignment Space Termination (also PyNAST) algorithm (Caporaso *et al.*, 2010a). ChimeraSlayer (Haas *et al.*, 2011) was used to detect and remove chimeras/singleton reads as implemented in QIIME. An additional chimera control test and taxonomic assignment were carried out by a BLAST search against the SILVA database (http://www.arb-silva.de/).

5.3.3 Ecological analysis

The number of reads of each OTU was used as an abundance proxy to estimate metagenomic diversity and structure of the abalone gut microbiota. Species richness (S), Shannon-Wiener (H) and Equitability (J) indices were calculated using Past V. 2.17c (Hammer *et al.*, 2001). Taxonomic differences between abalone species, were evaluate by both Student t-tests using R (Team, 2015) and by a linear discriminant analysis (LDA) effect size (LEfSe) (Segata *et al.*, 2011). To assess how exhaustively bacterial communities of both abalone species were sampled, rarefaction curves of discovered OTUs were generated for increasing numbers of sampled abalone. Also, OTU abundance was used to compute the non-parametric species richness estimator Chao 1 (Chao, 1984). Rarefaction curves were obtained using the EstimateS V.9.0.1 program (Colwell, 2013). Microbiome community structure was evaluated using non-paramteric multidimensional scaling (MDS) analyses using Bray-Curtis and Sorensen similarity indices based on read abundance and on presence/absence, respectively, as implemented in PRIMER V.6 (Clarke & Warwick, 2001). The statistical comparison of MDS results was performed with ANOSIM as implemented in Past V. 2.17c (Hammer *et al.*, 2001). To determine which OTUs were primarily responsible for the similarities within each species and dissimilarity between HF and HC, a SIMPER analysis with square root transformed data was performed using PRIMER V.6 (Clarke & Warwick, 2001).

Bacterial interactions in the microbiomes of both abalone species were estimated using Jaccard distance (J_d) as implemented in PRIMER V.6 (Clarke & Warwick, 2001). J_d is a measure of dissimilarity for all pairwise combinations of a data set and was calculated using OTUs presence/absence. J_d values close to 0 (from 0 to 0.33) were interpreted as co-occurrence (or putative mutualistic relationships) and values close to 1 (from 0.68 to 1) as interactions leading to exclusion (or putative competitive) (Rahel, 2000); whereas intermediate values were considered neutral relationships. We followed two approaches in the estimation of bacterial interactions. First, we computed J_d distances between all pairs of OTUs obtained from individual reads (putative species level). However, these mathematical results may not necessarily reflect a biological interaction. In order to improve the analysis, we posited that phylogenetically-related species share ecological functional attributes (Chaffron *et al.*, 2010); thus, we identified OTUs that could be grouped at the 90% similarity and recomputed mean Jaccard distances within groups only, in an attempt to cluster bacterial species with similar functions and explore ecological interactions within those groups. Distance matrices were computed with Geneious R9 (Kearse *et al.*, 2012). Finally, in order to test the hypothesis that abundant bacteria in the microbiomes are subject to less competition and more neutral or positive interactions, we correlated the mean Jaccard distance (or $\bar{J}_d$) of each OTU with its abundance (i.e. number of reads).

5.3.4 Functional prediction of metagenomes

A phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) (Langille *et al.*, 2013) was carried out to predict the functional attributes of metabolic genes from HF and HC GI microbiomes. Briefly, PICRUSt is a bioinformatic approach that uses information from a number of genetic markers, including the *16S rRNA*, to predict the metagenome functional content (Langille *et al.*, 2013; de Voogd *et al.*, 2015). These predictions were obtained by matching our *16S rRNA* gene sequences against the prearranged genomic KEGG database (Langille *et al.*, 2013; de Voogd *et al.*, 2015). The central result of PICRUSt consists of a table reporting the functional gene frequencies known as KEGG Orthologs (or KOs). KOs are hierarchically organized in sets of homologous sequences with known molecular function and assigned to biological pathways. We analyzed the data using the raw KOs counts as well as categorizing them by biological pathway. PICRUSt analyses use Greengenes (http://greengenes.lbl.gov/cgi-bin/nph-index.cgi) as taxonomic and functional reference database. To implement quality control, we computed weighted nearest sequenced taxon index (NSTI) values for each individual metagenome. NSTI was developed to evaluate the prediction accuracy of PICRUSt, since it

reflects the average genetic distance (measured as number of substitutions per site) between an OTU against a reference genome (Langille *et al.*, 2013; de Voogd *et al.*, 2015). Following suggested guidelines (Langille *et al.*, 2013), we eliminated observations with a NSTI higher than 0.17.

In order to assess the contribution of individual OTUs to predicted KO functions, first we focused our analyses on genes involved in metabolic pathways (KEGG IDs from EC:1.1.1 to EC:6.5.1). Next, we categorized the relative importance of KO genes by ranking them according to their raw counts. These counts were obtained with PICRUSt v.1.1.0 (Langille *et al.*, 2013) and were log normalized (Urbanová & Bárta, 2014). Finally, we focused our attention on a random subset (n =10) of KOs with the highest counts, as a first order analysis to characterize functional differences between these microbiomes. These analyses were carried out with the script *metagenome contributions.py*.

Finally, in order to compare the predicted ecological functions and the KOs abundance between microbiomes of both species of abalone, non-parametric MDS analysis based on Bray-Curtis similarity using log(x+1) transformation of data were performed in PRIMER V6 (Clarke & Warwick, 2001).

5.4 Results

5.4.1 Pyrosequencing and metagenome structure

Pyrosequencing yielded 451,095 raw *16S rRNA* reads of which 245,779 (54.5%) met quality criteria and were assigned to 281 OTUs, of which 87 had no match to databases (i.e., “No Hit”). The most abundant phyla (number of OTUs in parenthesis), included *Bacteroidetes* (13), *Fusobacteria* (14), *Proteobacteria* (56) and *Tenericutes* (96). Additionally, classified OTUs were assigned to 25 families and 47 genera. The five most abundant OTUs (n > 15876 reads) were assigned to class *Mollicutes* (order *Mycoplasmatales*). Notably, LEfSe analysis revealed that 63 of 96 *Mollicutes* OTUs were exclusive to either in HC or HF (Annex 14). Similar results were observed for other predominant bacterial families such as *Fusobacteriaceae* and *Vibrionaceae* (Annex 14).

The classes *Fusobacteria*, *Mollicutes*, α- and γ-*protobacteira* comprised 99% of the identifiable reads. Rarefaction curves suggest that the bacterial communities were sufficiently sampled in both abalone species, given their asymptotic shape and the proximity of the observed number of taxa in each species

to CHAO 1 estimates (Fig. 4). Species richness was significantly higher ($t = 6.07$; $p < 0.01$) in HF (mean ± S.D. HF: 56.45 ± 12.49; HC: 36.66 ± 13.16).

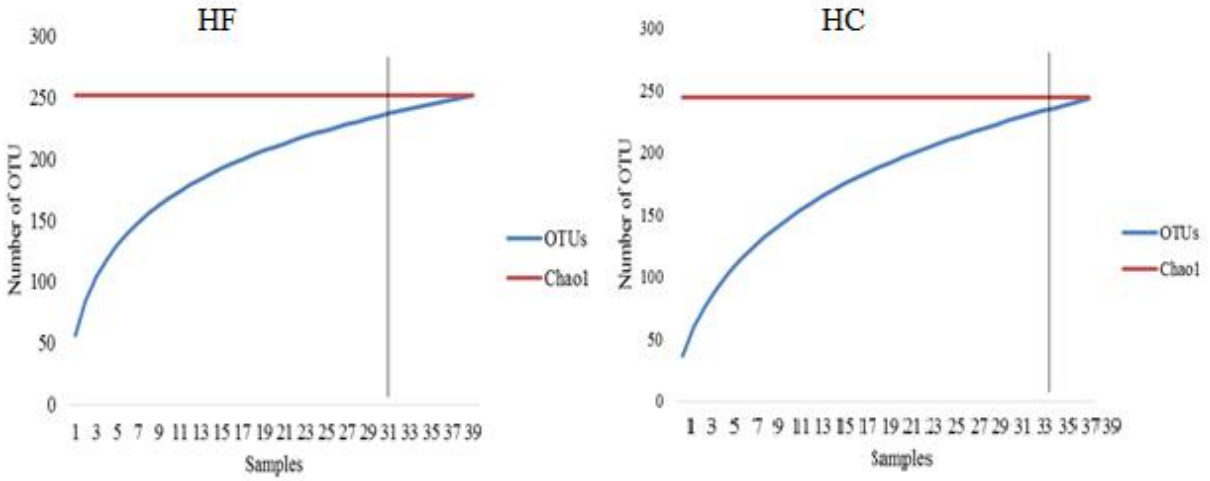


Figure 4. Rarefaction curves of the cumulative number of observed OTUs with increasing number of samples and Chao 1 estimation of total OTU richness in microbiomes of *Haliotis fulgens* (HF) and *H. corrugata* (HC). The vertical line indicates the actual sample size above which the curve is extrapolated to reach Chao 1 estimator.

Despite their similar composition at the class level (Fig. 5), MDS analyses at the highest taxonomic resolution revealed a clear-cut structural difference between the metagenomes of both species, in the presence of considerable intraspecific variation, using read numbers as proxy for abundance (Fig. 6) or solely on presence/absence data (Annex 15). Significant interspecific differentiation was corroborated by ANOSIM analyses in both cases ($R_{abundance} = 0.72$; $p < 0.001$; $R_{presence/absence} = 0.73$; $p < 0.001$).

Jaccard distances revealed that interspecific relationships among OTUs changed by an order magnitude with most involving competition (HC: 24,242 and HF: 24,118), followed by neutral (HC: 1321 and HF: 2123) and a smaller number of mutualistic interactions (HC: 315 and HF: 323). Furthermore, the $\bar{J}_d$ of the majority of OTUs decreased significantly with increasing read number in both species ($p < 0.001$; Fig. 7). The same trend was observed within genetic groups (Annex 16).

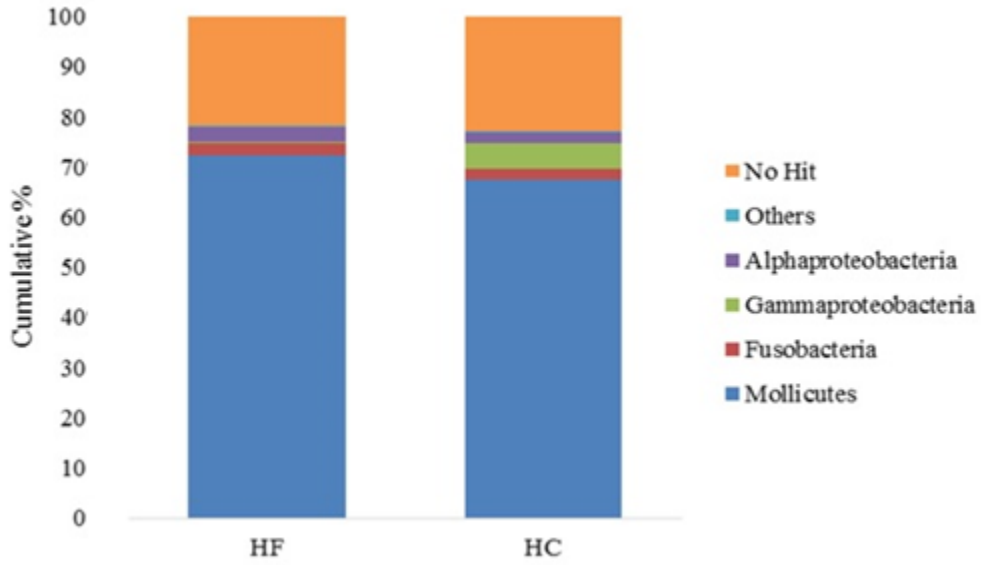


Figure 5. Major bacterial classes comprising the gut microbiota of *Haliotis fulgens* (HF) and *H. corrugata* (HC).

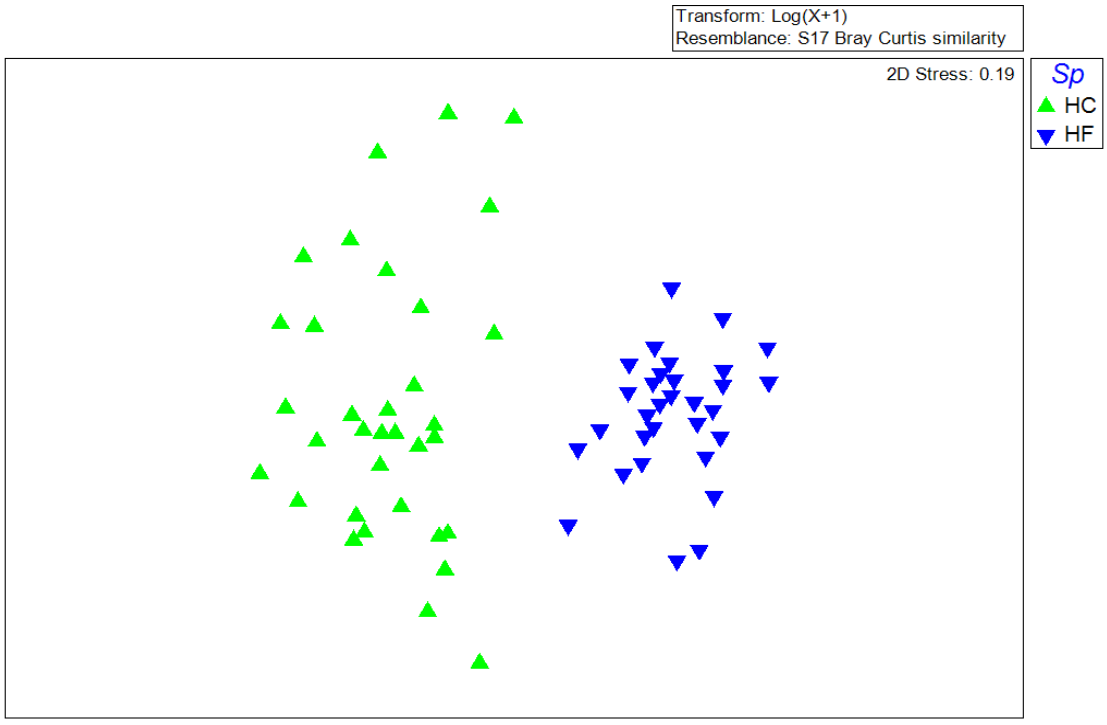


Figure 6. Non-metric multidimensional scaling (MDS) of Bray Curtis similarity index using log(x+1) transformed data based on read abundance of OTUs assembled at 97% similarity cut-off of the gut microbiota of *Haliotis fulgens* (HF) and *H. corrugata* (HC).

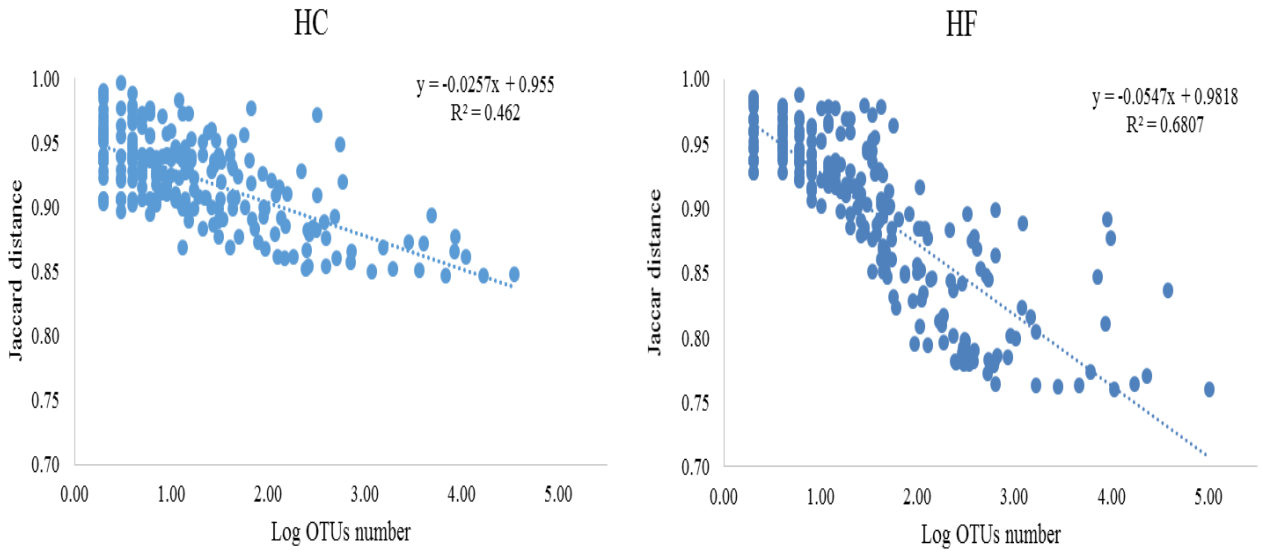


Figure 7. Regression analysis of average Jaccard distance based on all pairwise relationships as a function of the number of OTUs (log scale) in microbiomes of *Haliotis fulgens* (HF) and *H. corrugata* (HC).

5.4.2 Functional profiling

Most abalone possessed mean NSTI values from 0.07 to 0.17, except for seven HCs with higher values that made them unreliable; hence they were excluded from further analyses. PICRUSt identified 4,201 KOs genes involved in 262 metabolic functions. A one order of magnitude drop in log-normalized abundance was observed in the ranking of KOs (Annex 17); hence, *metagenome contribution*s analysis was carried out only on 10 KOs randomly selected from the 86 most abundant (Fig. 8).

According to PICRUSt, the metabolic functions in the HF microbiomes were generally enriched by one primary OTU, whereas many more OTUs contributed to the same function in HC (Fig.8).

MDS analyses performed using both KOs genes and ecological function counts revealed no clear functional separation of the GI microbiota of both species, even though the scatter of individual microbiomes is much larger in HC, which is consistent with its higher diversity (Fig. 6).

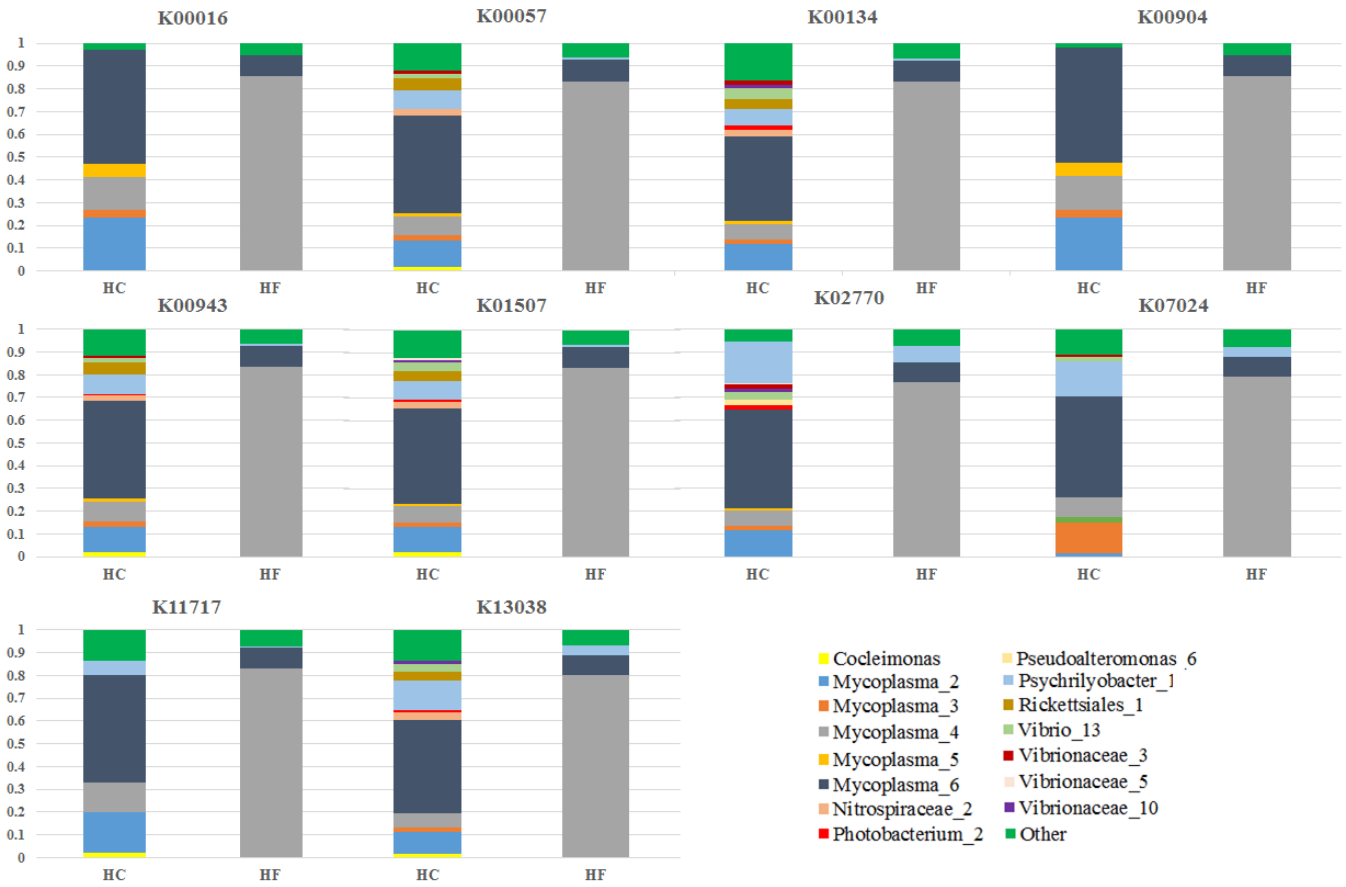


Figure 8. Bar plots of the relative bacterial gene count contributions of ten KOs with in the gut microbiomes of *Haliotis fulgens* (HF) and *H. corrugata* (HC) chosen at random among the 86 most abundant KOs (see text for details).

5.5 Discussion

5.5.1 Abalone microbiome composition and dynamics

Our data revealed that the post esophageal microflora of HC and HF are dominated by the same bacterial classes. *Mollicutes*, mostly represented by *Mycoplasma* spp., was by far the most abundant class, followed by *Fusobacteria*, α-*proteobacteria* and γ- *proteobacteria*, with the latter represented by the genera *Vibrio* and *Francisella*. These bacteria have also been found dominating in the GI microbiota of other abalone species (*H. discus hannai*, *H. diversicolor* and *H. gigantean* (Huang *et al.*, 2010; Iehata *et al.*, 2014; Tanaka *et al.*, 2003, 2004). Even though the taxonomic composition of HC and HF GI microbiotas bears

resemblance at high taxonomic levels, the species level composition showed significant differences. For instance, several *Mycoplasma* species predominant in one abalone species were either absent or at low abundance in the other. Furthermore, all *Vibrio* spp. presented a higher prevalence in HC (Annex 14).

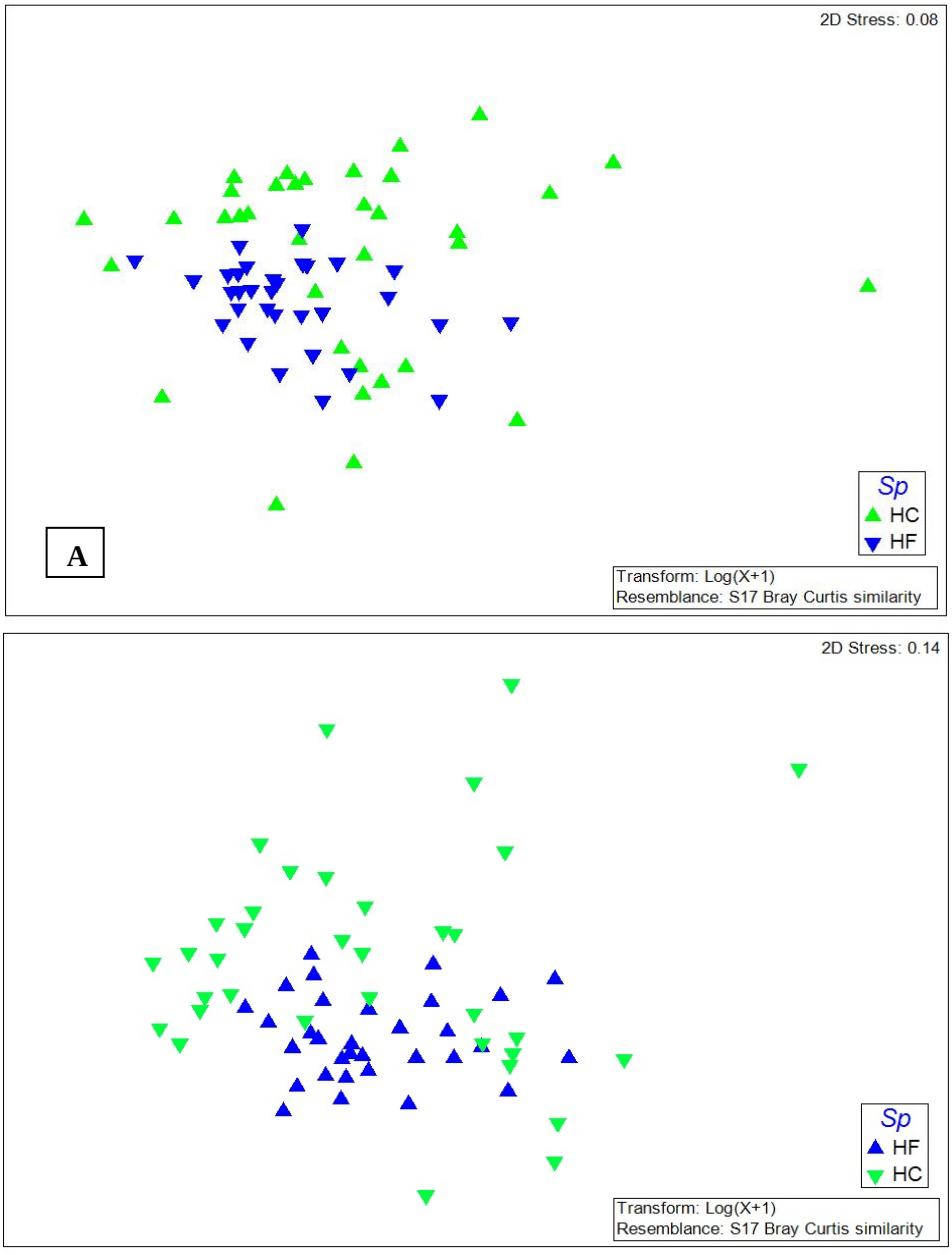


Figure 9. [A] Non-metric multidimensional scaling (MDS) of Bray Curtis similarity using log(x+1) transformed data of predicted ecological functions in the gut microbiomes of *Haliotis fulgens* (HF) and *H. corrugata* (HC). [B] MDS based on Bray Curtis similarity using log(x+1) transform data of KEGG orthologue counts in the gut microbiomes of *Haliotis fulgens* (HF) and *H. corrugata* (HC).

Notably, *Vibrio halioticoli* was absent in both HF and HC whereas it has been found at prevalence ranging from 40 to 65% in the GI of other marine invertebrates, including several abalone species (Sawabe et al., 2003). The discrepancy could be related to the anatomical source of the microbiomes, our samples come from post esophageal tissue only whereas other studies have analyzed the entire GI tracts (Tanaka *et al.*, 2004; Iehata *et al.*, 2014). Consequently, the presence of *V. halioticoli* in the rest of GI tissues of HC and HF should be explored.

Our findings suggest that for most bacterial species their abundance decreased with increasing competition (lower $\bar{J}_d$ values), as observed for *Mycoplasma*. Furthermore, $\bar{J}_d$ for a single bacterium changed according to its host, which suggests that the same bacterial species faces different competitive pressures depending on the microbiota composition. In other words, bacteria that may avoid competition in one microbiome, may be out competed in the other. However, this approach assumes untested hypothetical interspecific interactions (Faust & Raes, 2012). Consequently, the same analysis was conducted focusing on genetically similar (90% similarity) OTUs, assuming that genetically related bacteria may share similar ecological roles (Chaffron S *et al.*, 2010). This second approach produced similar results (Annex 16), which may support the analysis carried out using all pairs of OTUs.

Some pathogenic bacteria, such as *Candidatus* Xenohaliotis californiensis and *Francisella*, were observed in the microbiomes of healthy abalone. The former has been recognized as the etiological agent of the withering syndrome, a chronic wasting disease that possibly affects all North American abalone species (Friedman, 2012). This bacterium is a pleomorphic, gram-negative coccobacillus that inhabits abalone gastrointestinal epithelia and is considered an obligate endoparasite, like other *Rickettsiales* (Friedman *et al.*, 2000; Moore *et al.*, 2002; Crosson *et al.*, 2014;). We observed the presence of *Candidatus* Xenohaliotis californiensis in healthy blue and yellow abalone, which supports that the presence of this pathogen is not sufficient to trigger withering syndrome as previously proposed (Cáceres-Martínez *et al.*, 2011). Moreover, its absence and/or low intensity in abalone with morphological and histological signs of withering syndrome has already been reported (Balseiro *et al.*, 2006; Cáceres-Martínez *et al.*, 2011; Horwitz *et al.*, 2016). Accordingly, we posit that further investigations are needed to reveal all the factors involved in withering syndrome outbreaks.

Francisella is a γ- proteobacterium and has previously been found in abalone microbiota (Brevik *et al.*, 2011). It is a non-motile, pleomorphic gram-negative coccobacillus mainly known to be a facultative intracellular parasite of a wide range of hosts, including humans (Brevik *et al.*, 2011; Sjödin *et al.*, 2012). At the moment, this genus consists of five validated species (Brevik *et al.*, 2011). Additionally, a novel

species named *F. halioticida* (Brevik *et al.*, 2011) has been isolated from a die-off of farmed *H. gigantea* in Japan, for which it was identified as the etiological agent (Kamaishi *et al.*, 2010). The *Francisella* ribotype detected in *H. corrugata* from Mexico was only 89% similar to *F. halioticida 16S rRNA* sequence (Genbank accession number NR_118116), which suggests a different species. Moreover, the presence of this bacterium in only HC organisms and its low prevalence (n = 2), suggests that this strain should not be considered an established member of the HC abalone bacterial core. Also, our limited data do not support the pathogenicity of the detected strain.

Finally, many bacterial genera were detected in low read frequencies and prevalences. Some of these (e.g., *Acinetobacter*, *Alteromonas*, *Bacillus*, *Flavobacterium*) bear strong relationships with macroalgae (Goecke *et al.*, 2010) and may have been ingested while grazing.

5.5.2 Ecological functions of the abalone microbiome

According to the MDS of KOs gene counts and ecological functions, the inferred functionality of the GI microbiota of HF appears less variable than that of HC, which may reflect a higher degree of diet specialization. The diet diversity of wild HF in Baja California has been shown to be more limited and dominated by *Phyllospadix torreyi* (47%) and algae in the order *Gelidiales* (13%) (Guzman del Próo *et al.*, 2003). On the other hand, the diet of sympatric wild HC is more diverse consisting of different species of Phaeophyceae (10-20%), Rhodophyta (20%) and *Gelidiales* (20%) among others (Guzman del Próo *et al.*, 2003). Possibly, those dietary preferences may be the result of bathymetric adaptations; indeed, HC are generally found in relative deep waters (between 10-20m), whereas HF generally inhabit shallow regions (between 3-10m) (Guzman del Próo *et al.*, 2003).

In the microbiota of both HC and HF, *Mycoplasma* contributed to all predicted KEGG. However, a given KO was generally enriched by a single predominant *Mycoplasma* OTU in HF, whereas in HC it was generally enriched by two or more *Mycoplasma* as well as other bacterial-OTUs (Fig.8). In other words, our results not only support that different *Mycoplasma* species may be highly host-specific (Register *et al.*, 2015), but also that they may also be highly specific to ecological tasks and/or to specific steps along a metabolic route. This may also be reflected by the relative low J_d (interpreted as absence of competitive interactions) of *Mycoplasma* OTUs. Indeed, ecological specialization should translate into niche partitioning and a reduction of inter-specific competition.

To our knowledge, this is the first time that the ecological functions of *Mycoplasma* have been explored by a predictive analysis and our results may prove useful to reevaluate some of genetic features of these bacterial species. *Mycoplasma* possess the smallest genome size among bacteria (Bano *et al.*, 2007). Because of their small genome and number of genes, *Mycoplasma* species have been considered as obligate parasites and/or commensals species (Fraser *et al.*, 1995; Bano *et al.*, 2007). Furthermore, our results suggest that these bacteria should be considered members of the bacterial core in the GI microflora of abalone.

Vibrio was the second most common genus observed enriching the majority of predicted KEGG. These bacteria are Gram-negative γ- *proteobacteria* that include both pathogenic and non-pathogenic species (Colwell, 1996; Heidelberg *et al.*, 2000; Sawabe *et al.*, 2007). As it is the case in Asian abalone (Huang *et al.*, 2010; Iehata *et al.*, 2014; Sawabe *et al.*, 2003), our results confirm that *Vibrio* may be considered one of the main components of abalone GI microbiota in the eastern Pacific, at least in HC. Also, *Vibrio* has been found to play a pivotal role converting alginate to acetic acid (Sawabe *et al.*, 2003; Tanaka *et al.*, 2003). In *Vibrio cholera*, possibly the best studied *Vibrio* species, the genome has been found to be dynamic, continuously acquiring and losing genes over time (Heidelberg *et al.*, 2000; Okada *et al.*, 2005). Consequently, as suggested by the PICRUSt analysis, *Vibrio* may be involved in a large number of ecological functions in the GI of abalone. We also speculate that the dominance of *Vibrio* in HC may contribute to their ability to process a wider spectrum of food sources.

PICRUSt analysis also revealed that several other bacteria may play major roles in several metabolic and/or enzyme production pathways. *Fusobacteriaceae* (mainly represented by *Psychrilyobacter*) are obligate anaerobic gram-negative bacilli and are generally found in anoxic marine sediments. Species in this bacterial family have been detected in several marine invertebrate hosts such as shrimp (Chaiyapechara *et al.*, 2012), muricid snail (Ngangbam *et al.*, 2015), and abalone (Tanaka *et al.*, 2004) and have been shown to be involved in the degradation of organic compounds (Tanaka *et al.*, 2004; Chaiyapechara *et al.*, 2012; Hassenruck *et al.*, 2015;). Our findings suggest that these bacteria may be involved in several metabolic and/or enzyme production pathways such as glycero-phospholipid metabolism (K00057), starch and sucrose metabolism (K07024) and seleno-compound metabolism (K117170), among others, where they may take part in the production of dehydrogenases, lyases and other hydrolytic enzymes.

NOTE

This chapter has been submitted for publication as:

Cicala, F., J. D. Moore, and A. Rocha-Olivares (in review). "Structure, dynamics and predicted functional ecology of the gut microbiota of the blue (*Haliotis fulgens*) and yellow (*H. corrugata*) abalone from Baja California Sur, Mexico." Microbiome.

Chapter 6. Conclusions

6.1 Main results

The work presented in this thesis is the first attempt to extend the genetic knowledge of "*Candidatus* Xenohaliotis californiensis". We designed novel PCR primers allowing us to amplify and sequence six previously unknown loci of "*Candidatus* Xenohaliotis californiensis". These new DNA sequences, combined with the *16S rRNA*, were initially used in concatenated phylogenetic reconstructions to analyze the systematic and evolutionary relationships between *CXc* and other species belonging to the order Rickettsiales. According to our findings, we proposed a new reorganization inside Anaplasmataceae. In contrast with previous reports (Friedman *et al.*, 2000; Vannini *et al.*, 2005; Martijn *et al.*, 2015), we found that *CXc* is the sister group of *Neorickettsia* in reciprocal monophyly to the other species of Anaplasmataceae. Additionally, our molecular-clock phylogeny suggests a basal position of *CXc* which support the hypothesis of an aquatic origin and a successive terrestrial environment transition for this bacterial group (Kang *et al.*, 2014).

Considering that increased genetic knowledge of pathogenic bacteria is also required to: (*i*) understand the rise of outbreak events (Achtman, 2008), (*ii*) describe different pathogenic strains (Ellison *et al.*, 2008) and (*iii*) tracking their epidemic and/or pandemic spread (Achtman, 2008); we used the multi-locus sequence typing (MLST) and the multi spacer typing (MST) approaches to assess the genetic variability of *CXc*. Despite our efforts, low polymorphisms were observed among all molecular markers. Consequently, we conclude that a single virtually monomorphic strain of *C*Xc is infecting the three species of abalone in California and Baja California. Similar results were also observed analyzing more than 8,000 reads of the *16S rRNA* of *C*Xc generated by 454 pyrosequencing.

Several factors may explain the lack of genetic variability in *C*Xc such as; (*i*) a low intraspecific discrimination power for the used molecular markers; (*ii*) a strong genetic reduction and consequently a high selective pressure in the remaining genes or even on the entire genome of *C*Xc (*iii*) an early bottle neck event, in the ancestor population of *C*Xc after the transit to, the endo-cellular life strategy in abalone gut cells.

Our metagenome analysis suggests that HC and HF harbor significantly different species-specific microbiotas. Moreover, higher Jaccard distance values for *C*Xc were observed in HC than HF suggesting that this bacterium may face higher competitive pressures inside the microbiota of yellow abalone. Accordingly, we believe that these results may support the lower susceptibility rate previously reported for this abalone species (Álvarez Tinajero *et al.*, 2002; Friedman, 2012; Cruz Flores, 2013; Flores, 2014). In other words, the presence/absence of *C*Xc and/or differences in susceptibility and mortality rates among abalone, may be related with differences in microbiota composition of abalone species, and more effort in this direction should be made.

Using bTEFAP we were able to examine the microbiota of two Mexican abalone. *Mycoplasma* was by far the predominant genus detected in both species. However, the microbiomes were constituted by different bacterial species suggesting a host- specific adaptation for most of them. PICRUSt analysis supports that "host-specific" bacterial strains, may be responsible for the same molecular pathways inside different microbiota. Finally, the high number of unclassified OTUs as well as the relative high values of NSTI suggests that gut contents of proposed abalone species may be contain novel bacterial species.

6.2 Future tasks

In this work, I have used a cultivation independent approach to amplify and sequence new loci of *C*Xc. However such approach led to several challenges. Cloning results revealed the presence of several sequences assignable to the order Rickettsiales suggesting that both HF and HC abalone harbor uncharacterized RLPs, as observed for other abalone species (Crosson *et al.*, 2014; Horwitz *et al.*, 2016). Also, recent epidemiologic outbreak occurred in farmed *Haliotis midea* populations from South Africa, which were related to unknown RLPs. Accordingly, the pathogenicity of the bacteria harboring the uncharacterized RLP haplotypes from this work should be explored.

In situ hybridization (or ISH) was performed in order to directly test that the DNA fragment obtained belonged to *CXc*. Notably, a detectable signal was obtained only for the ribosomal genes (16S and 23S rRNA). I speculated that the lack of signal for non-ribosomal genes may be due to the fewer target regions for those genes. However, additional efforts are required to test this hypothesis using a higher number of DIG probes to increase the sensitivity of ISH. Additionally, more sensitive methods such as fluorescent in

situ hybridization (or FISH) could be used to test the physical location of the DNA sequences produced herein.

Even though I was able to amply and sequence new genes from *CXc*; the genome of *C*Xc remains mostly unknown. Recently, the Single-cell genome sequencing approach has attracted a great interest and was successfully used to sequence more than 800,000 bp in previously unknown RLPs (Martijn *et al.*, 2015) as well as other bacteria (Gawad *et al.*, 2016). This approach appears highly promising and may facilitate further genetic studies of *CXc*.

Also, the use of Single Nucleotide Polymorphisms (also SNPs) has been considered one of the most encouraging tool in discriminating bacterial strains (Sreevatsan *et al.*, 1997; Pearson *et al.*, 2004). Accordingly, the use of Single-cell genome sequencing to search for SNPs should represent a feasible way to search for *CXc* genetic variability and confirm or refute the lack of variation detected in this study.

As observed in other organisms, shifts of the bacterial community may result in epidemiologic outbreaks (Brown *et al.*, 2012; Koh *et al.*, 2015; Wardwell *et al.*, 2011). Accordingly, the occurrence of WS may be related to alterations of the microbiota balance of abalone. However this hypothesis has not been investigated and may provide a basis for developing future more extensive temporal and spatial sampling, and/or among different ontogenetic stages as well as physiological conditions. Also, additional effort should also be directed towards understanding the roles of environment variables or other factors that may also alter the gut microbiomes of abalone.

List of references

Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105.

Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* **62**, 53–70.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Altstatt, J. M., Ambrose, R. F., Engle, J. M., Haaker, P. L., Lafferty, K. D. & Raimondi, P. T. (1996). Recent declines of black abalone *Haliotis cracherodii* on the mainlandcoast of central California. *Mar Ecol Prog Ser* **142**, 185–192.

Alvarez-Martinez, C. E. & Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**, 775–808.

Álvarez Tinajero, M. D. C., Cáceres-Martínez, J., Gonzáles Avilés, J. G., Del Carmen Alvarez Tinajero, M., Caceres-Martinez, J. & Aviles, J. G. G. (2002). Histopathological evaluation of the yellow abalone *Haliotis corrugata* and the blue abalone *Haliotis fulgens* from Baja California, México. *J Shellfish Res* **21,** 825-830.

Anderson, C. L. & Karr, T. L. (2001). *Wolbachia*: evolutionary novelty in a rickettsial bacteria. *BMC Evol Biol* **1**, 10.

Andersson, S. G., Zomorodipour, a, Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C., Podowski, R. M., Näslund, a K., Eriksson, a S., Winkler, H. H. & Kurland, C. G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140.

Andree, karl B; Friedman, Carolyn S ;Moore, James D; Hedrick, R. P. (2000). A polymerase chain reaction assay for the detection of genomic DNA of a Rickettsiales-like prokaryote associated with withering syndrome in California abalone. *J Shellfish Res* **19**, 213–218.

Antonio, D. B., Andree, K. B., Moore, J. D., Friedman, C. S. & Hedrick, R. P. (2000). Detection of Rickettsiales-like prokaryotes by in situ hybridization in black abalone, *Haliotis cracherodii*, with withering syndrome. *J Invertebr Pathol* **75**, 180–182.

Backert, S. & Meyer, T. F. (2006). Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol* **9**, 207–217.

Backhed, F. (2005). Host-Bacterial Mutualism in the Human Intestine. *Science (80-)* **307**, 1915–1920.

Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* **34**, 369–373.

Baldo, L., Hotopp, J. C. D., Jolley, K. a., Bordenstein, S. R., Biber, S. a., Choudhury, R. R., Hayashi, C., Maiden, M. C. J., Tettelin, H. & Werren, J. H. (2006). Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. *Appl Environ Microbiol* **72**, 7098–7110.

Balseiro, P., Aranguren, R., Gestal, C., Novoa, B. & Figueras, A. (2006). *Candidatus* Xenohaliotis californiensis and *Haplosporidium montforti* associated with mortalities of abalone *Haliotis tuberculata* cultured in Europe. *Aquaculture* **258**, 63–72.

Bano, N., DeRae Smith, A., Bennett, W., Vasquez, L. & Hollibaugh, J. T. (2007). Dominance of *Mycoplasma* in the guts of the Long-Jawed Mudsucker, Gillichthys mirabilis, from five California salt marshes. *Environ Microbiol* **9**, 2636–2641.

Bateman, K. S., Hicks, R. J. & Stentiford, G. D. (2011). Disease profiles differ between non-fished and fished populations of edible crab (*Cancer pagurus*) from a major commercial fishery. *ICES J Mar Sci* **68**, 2044–2052.

Bates, J. M., Mittge, E., Kuhlman, J., Baden, K. N., Cheesman, S. E. & Guillemin, K. (2006). Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Dev Biol* **297**, 374–386.

Batista, F. M., Arzul, I., Ruano, F., Friedman, C. S., Boudry, P. & Renault, T. (2007). Detection of ostreid herpesvirus 1 DNA by PCR in bivalve molluscs : A critical review. *J Virol Methods* **139**, 1–11.

Blanc, G., Ogata, H., Robert, C., Audic, S., Suhre, K., Vestris, G., Claverie, J. M. & Raoult, D. (2007). Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet* **3**, 0103–0114.

Blaut, M. & Clavel, T. (2007). Metabolic diversity of the intestinal microbiota: implications for health and disease. *J Nutr* **137**, 751S–5S.

Braid, B. a., Moore, J. D., Robbins, T. T., Hedrick, R. P., Tjeerdema, R. S. & Friedman, C. S. (2005). Health and survival of red abalone, *Haliotis rufescens*, under varying temperature, food supply, and exposure to the agent of withering syndrome. *J Invertebr Pathol* **89**, 219–231.

Brevik, O. J., Ottem, K. F., Kamaishi, T., Watanabe, K. & Nylund, A. (2011). *Francisella halioticida* sp. nov., a pathogen of farmed giant abalone (Haliotis gigantea) in Japan. *J Appl Microbiol* **111**, 1044–1056.

Brown, K., DeCoffe, D., Molcan, E. & Gibson, D. L. (2012). Diet-induced dysbiosis of the intestinal microbiota and the effects on immunity and disease. *Nutrients* **4**, 1095–1119.

Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. (1996). A model of host-microbial interactions in an open mammalian ecosystem. *Science* **273**, 1380-83.

Burge, C. a, Mark Eakin, C., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., Petes, L. E., Prager, K. C., Weil, E. & other authors. (2014). Climate change influences on marine infectious diseases: implications for management and society. *Ann Rev Mar Sci* **6**, 249–77.

Cáceres-Martinez, J. & Tinoco-Orta, G. D. (2001). Symbionts of cultured red abalone *Haliotis rufescens* from Baja California, Mexico. *J Shellfish Res*. **20,** 875-881.

Cáceres-Martínez, J., Vásquez-Yeomans, R. & Flores-Saaib, R. D. (2011). Intracellular prokaryote *Xenohaliotis californiensis* in abalone *Haliotis* spp. from Baja California, México. *Cienc Pesq* **19**, 5–11.

Cáceres Martínez, J. (2002). Algunas Enfermedades Infecciosas en Moluscos. *Bol del programa Nac Sanid acuicoa y la red diagnostico* **2**, 1–10.

Caporaso, J. G., Bittinger, K., Bushman, F. D., Desantis, T. Z., Andersen, G. L. & Knight, R. (2010a). PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, K. & other authors. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336.

Cariño Olvera, M. & Monteforte, M. (2008). *Del saqueo a la conservación: Historia ambiental contemporánea de Baja California Sur, 1940-2003*.

Casiraghi, M., Bordenstein, S. R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J. J., Werren, J. H. & Bandi, C. (2005). Phylogeny of *Wolbachia pipientis* based on *gltA*, *groEL* and *ftsZ* gene sequences: Clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. *Microbiology* **151**, 4015–4022.

Chaffron S, Rehrauer H, Pernthaler J & Mering C. **(2010).** A global network of coexisting microbes from\nenvironmental and whole-genome sequence data. *Genome Res* **2010**, 947–959.

Chaiyapechara, S., Rungrassamee, W., Suriyachay, I., Kuncharin, Y., Klanchui, A., Karoonuthaisiri, N. & Jiravanichpaisal, P. (2012). Bacterial Community Associated with the Intestinal Tract of P. monodon in Commercial Farms. *Microb Ecol* **63**, 938–953.

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scand J Stat* **11**, 265–270.

Cicala, F., Moore, J. D., Cáceres-Martínez, J., del Río-Portilla, M. A., Hernández-Rodríguez, M., Vásquez Yeomans, R. & Rocha-olivares, A. (2017). Multigenetic characterization of '*Candidatus* Xenohaliotis californiensis'. *Int J Syst Evol Microbiol* **67**, 42–49.

Clarke, K. R. & Warwick, R. M. (2001). A further biodiversity index applicable to species lists: Variation in taxonomic distinctness. *Mar Ecol Prog Ser* **216**, 265–278.

Colwell, R. K. (2013). EstimateS Statistical estimation of species richness and shared species from samples. Version 9. User's Guide and application published at httppurl.oclc.orgestimates.pdf.

Colwell, R. R. (1996). Global climate and infectious disease: the cholera paradigm. *Science* **274**, 2025–2031.

Cox, K. W. (1962). California Abalones, Family Haliotidae. The resources ageny of California department of fish and Game. *Fish Bull No 118* **118**, 1–131.

Crosson, L. M., Wight, N., VanBlaricom, G. R., Kiryu, I., Moore, J. D. & Friedman, C. S. (2014). Abalone withering syndrome: Distribution, impacts, current diagnostic methods and new findings. *Dis Aquat Organ* **108**, 261–270.

Cruz-Flores, R., Cáceres-Martínez, J. & Vásquez-Yeomans, R. (2015). A novel method for separation of Rickettsiales-like organism '*Candidatus* Xenohaliotis californiensis' from host abalone tissue. *J Microbiol Methods* **115**, 79–82.

Cruz Flores, R. (2013). Distribución, prevalencia e intensidad de *Xenohaliotis californiensis* parásito del abulón *Haliotis fulgens* y *Haliotis corrugata*, en la península de Baja California, México. Tesis de Maestría en Ciencias. Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California. 80 pp.

Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J. & Andersson, S. G. E. (2007). Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet* **23**, 511–520.

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772–772.

Davis, G., Richards D. V., Haaker P. L., and Parker D. O. (1992). Abalone population declines and fishery management in southern California. Pages 237-249 in S. A. Guz- man del Proo, editor. Abalone of the world: biology, fisheries and culture. Fishing News Books, Cambridge, Massachusetts, USA.

ten Doeschate, K. I. & Coyne, V. E. (2008). Improved growth rate in farmed *Haliotis* midae through probiotic treatment. *Aquaculture* **284**, 174–179.

Dowd, S. E., Callaway, T. R., Wolcott, R. D., Sun, Y., McKeehan, T., Hagevoort, R. G. & Edrington, T. S. (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol* **8**, 125.

Drummond, A. J., Suchard, M. a., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973.

Dumler, J. S., Barbet, a F., Bekker, C. P., Dasch, G. a, Palmer, G. H., Ray, S. C., Rikihisa, Y. & Rurangirwa, F. R. (2001). Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combi. *Int J Syst Evol Microbiol* **51**, 2145–65.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461.

El-Sharoud, W. (2008). Bacterial physiology: A molecular approach. Springer.

Ellison, D. W., Clark, T. R., Sturdevant, D. E., Virtaneva, K., Porcella, S. F. & Hackstadt, T. (2008). Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa. *Infect Immun* **76**, 542–550.

Faust, K. & Raes, J. (2012). Microbial interactions: from networks to models. *Nat Rev Microbiol* **10**, 538–550.

Ferla, M. P., Thrash, J. C., Giovannoni, S. J. & Patrick, W. M. (2013). New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS One* **8**, 1–14.

Figueras, A. & Novoa, B. (2004). What has been going on in Europe in bivalve pathology? *Bull Eur Assoc Fish Pathol* **24**, 16–21.

Flores, M. M. (2014). Estudio epizootiológico de *Candidatus* Xenohaliotis californiensis (Prokaryota) en abulón azul (*Haliotis fulgens*) y amarillo (*H. corrugata*) de la península de Baja California. Tesis de Maestría en Ciencias. Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California. 59 pp.

Fournier, P. E., Dumler, J. S., Greub, G., Zhang, J., Wu, Y. & Raoult, D. (2003). Gene sequence-based criteria for identification of new *Rickettsia* isolates and description of *Rickettsia heilongjiangensis* sp. nov. *J Clin Microbiol* **41**, 5456–5465.

Fournier, P., Zhu, Y., Ogata, H. & Raoult, D. (2004). Use of highly variable intergenic spacer sequences for multispacer typing of *rickettsia conorii* strains use of highly variable intergenic spacer sequences for multispacer typing of *rickettsia conorii* strains. *Society* **42**, 5757–5766.

Fournier, P.-E. & Raoult, D. (2007). Identification of rickettsial isolates at the species level using multi-spacer typing. *BMC Microbiol* **7**, 72.

Fournier, P.-E., El Karkouri, K., Leroy, Q., Robert, C., Giumelli, B., Renesto, P., Socolovschi, C., Parola, P., Audic, S. & Raoult, D. (2009). Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. *BMC Genomics* **10**, 166.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G. & other authors. (1995). The minimal gene complement of *mycoplasma genitalium*. *Science (80-)* **270**, 397–403.

Friedman, C. S., Andree, K. B., Beauchamp, K. a, Moore, J. D., Robbins, T. T., Shields, J. D. & Hedrick, R. P. (2000). '*Candidatus* Xenohaliotis californiensis', a newly described pathogen of abalone, Haliotis spp., along the west coast of North America. *Int J Syst Evol Microbiol* **50**, 847–855.

Friedman, C. S., Scott, B. B., Strenge, R. E., Vadopalas, B. & Mccormick, T. B. (2007). Oxytetracycline as a tool to manage and prevent losses of the endangered white abalone , haliotis sorenseni , caused by withering syndrome. *J Shellfish Res* **26**, 877–885.

Friedman, C. S., Trevelyan, G., Robbins, T. T., Mulder, E. P. & Fields, R. (2003). Development of an oral administration of oxytetracycline to control losses due to withering syndrome in cultured red abalone *Haliotis rufescens*. *Aquaculture* **224**, 1–23.

Friedman, C. S., Wight, N., Crosson, L. M., White, S. J. & Strenge, R. M. (2014a). Validation of a quantitative PCR assay for detection and quantification of '*Candidatus* Xenohaliotis californiensis'. *Dis Aquat Organ* **108**, 251–259.

Friedman, C. S., Wight, N., Crosson, L. M., VanBlaricom, G. R. & Lafferty, K. D. (2014b). Reduced disease in black abalone following mass mortality: Phage therapy and natural selection. *Front Microbiol* **5**, 1–10.

Friedman, C. S. (2012). Infection with *Xenohaliotis californiensis*. *Man Diagnostic Tests Aquat Anim* 511–523.

Friedman, C. S., Thomson, M., Chun, C., Haaker, P. L. & Hedrick, R. P. (1997). Withering syndrome of the black abalone, *Haliotis cracherodii* (Leach): Water temperature, food availability, and parasites as possible causes. *J Shellfish Res* **16**, 403–411.

Friedman, C. S., Andree, K. B., Beauchamp, K. A., Moore, J. D., Robbins, T. T., Shields, J. D. & Hedrick, R. P. (2000b). '*Candidatus* Xenohaliotis californiensis', a newly described pathogen of abalone, Haliotis spp., along the west coast of North America. *Int J Syst Evol Microbiol* **50**, 847–855.

Fuxelius, H. H., Darby, A., Min, C. K., Cho, N. H. & Andersson, S. G. E. (2007). The genomic and metabolic diversity of *Rickettsia*. *Res Microbiol* **158**, 745–753.

García-Esquivel, Z., Montes-Magallón, S. & González-Gómez, M. A. (2007). Effect of temperature and photoperiod on the growth, feed consumption, and biochemical content of juvenile green abalone, Haliotis fulgens, fed on a balanced diet. *Aquaculture* **262**, 129–141.

Gardner, G. R., Harshbarger, J. C., Lake, J. L., Sawyer, T. K., Price, K. L., Stephenson, M. D., Haaker, P. L. & Togstad, H. a. (1995). Association of prokaryotes with symptomatic appearance of withering syndrome in black abalone Haliotis cracherodii. *J Invertebr Pathol*.

Garrity, G. M., Bell, J. a, Lilburn, T. G. & Lansing, E. (2004). Taxonomic Outline of the Prokaryotes. *Bergey's Man Syst Bacteriol* **2**, 1–399.

Gawad, C., Koh, W. & Quake, S. R. (2016). Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17**, 175–188. Nature Publishing Group.

Gestal, C., Roch, P., Renault, T., Pallavicini, A., Paillard, C., Novoa, B., Oubella, R., Venier, P. & Figueras, A. (2008). Study of Diseases and the Immune System of Bivalves Using Molecular Biology and Genomics. *Rev Fish Sci* **16**, 133–156.

Gillespie, J. J., Brayton, K. a., Williams, K. P., Quevedo Diaz, M. a., Brown, W. C., Azad, A. F. & Sobral, B. W. (2010). Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect Immun* **78**, 1809–1823.

Goecke, F., Labes, A., Wiese, J. & Imhoff, J. (2010). Chemical interactions between marine macroalgae and bacteria. *Mar Ecol Prog Ser* **409**, 267–299.

Guzman del Próo, S. a, Serviere-Zaragoza & Siqueiros-Beltrones, D. (2003). Natural diet of juvenile abalone *Haliotis fulgens* and *H. corrugata* (Mollusca: Gastropoda) in Bahia Tortugas, Mexico. *Pacific Sci* **57**, 319–324.

Haaker P.L., Parker D.O., Togstad H., Richards D.V., Davis G.E., Friedman C.S. (1992). Mass mortality and withering foot syndrome in black abalone, *Haliotis cracherodii*, in California. In: Shepard SA, Tegner MJ, Guzmán del Próo SA (eds.), Abalone of the World: Biology, Fisheries and Culture. Proc. 1st International Symposium on Abalone. Fishing News Books, Cambridge, pp. 214–224.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K. & other authors. **(2011).** Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**, 494–504.

Hammer, Ø., Harper, D. a. T. & Ryan, P. D. (2001). Paleontological statistics software package for education and data analysis. *Palaeontol Electron* **4**, 9–18.

Harvell, C. D., Kim, K., Burkholder, J. M., Colwell, R. R., Epstein, P. R., Grimes, D. J., Hofmann, E. E., Lipp, E. K., Osterhaus, a D. & other authors. **(1999).** Emerging marine diseases-climate links and anthropogenic factors. *Science.***285**, 1505–1510.

Hassenruck, C., Hofmann, L. C., Bischof, K. & Ramette, A. (2015). Seagrass biofilm communities at a naturally CO2-rich vent. *Environ Microbiol Rep* **7**, 516–525.

Headley, S. A., Scorpio, D. G., Vidotto, O. & Stephen Dumler, J. (2011). *Neorickettsia helminthoeca* and salmon poisoning disease: A review. *Vet J* **187**, 165–173.

Heidelberg, J. F., Eisen, J. a, Nelson, W. C., Clayton, R. a, Gwinn, M. L., Dodson, R. J., Haft, D. H., Tettelin, Â., Hickey, E. K. & other authors. **(2000).** DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483.

Horwitz, R., Mouton, A. & Coyne, V. E. (2016). Characterization of an intracellular bacterium infecting the digestive gland of the South African abalone *Haliotis midae. Aquaculture* **451**, 24–32.

Huang, Z.-B., Guo, F., Zhao, J., Li, W.-D. & Ke, C.-H. (2010). Molecular analysis of the intestinal bacterial flora in cage-cultured adult small abalone, *Haliotis diversicolor*. *Aquac Res* **41**, e760–e769.

Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Hyun, D. W., Kim, M. S., Shin, N. R., Kim, J. Y., Kim, P. S., Whon, T. W., Yun, J. H. & Bae, J. W. (2013). Shimia haliotis sp. nov., a bacterium isolated from the gut of an abalone, *Haliotis discus hannai*. *Int J Syst Evol Microbiol* **63**, 4248–4253.

Iehata, S., Inagaki, T., Okunishi, S., Nakano, M., Tanaka, R. & Maeda, H. (2010). Improved gut environment of abalone Haliotis gigantea through Pediococcus sp. Ab1 treatment. *Aquaculture* **305**, 59–65.

Iehata, S., Nakano, M., Tanaka, R. & Maeda, H. (2014). Modulation of gut microbiota associated with abalone *Haliotis gigantea* by dietary administration of host-derived Pediococcus sp. Ab1. *Fish Sci* **80**, 323–331.

Jaccard, P. (1908). Nouvelles Recherches Sur la Distribution Florale. *Bull Soc Vaudoise Sci Nat* **44**, 223–270.

Justice, S. S., Hunstad, D. a, Cegelski, L. & Hultgren, S. J. (2008). Morphological plasticity as a bacterial survival strategy. *Nat Rev Microbiol* **6**, 162–168.

Kamaishi, T., Miwa, S., Goto, E., Matsuyama, T. & Oseko, N. (2010). Mass mortality of giant abalone *Haliotis gigantea* caused by a *Francisella* sp. bacterium. *Dis Aquat Organ* **89**, 145–154.

Kang, Y.-J., Diao, X.-N., Zhao, G.-Y., Chen, M.-H., Xiong, Y., Shi, M., Fu, W.-M., Guo, Y.-J., Pan, B. & other authors. **(2014).** Extensive diversity of Rickettsiales bacteria in two species of ticks from China and the evolution of the Rickettsiales. *BMC Evol Biol* **14**, 167.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S. & other authors. **(2012).** Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

Kiryu, I., Kurita, J., Yuasa, K., Nishioka, T., Shimahara, Y., Kamaishi, T., Ototake, M., Oseko, N., Tange, N. & other authors. **(2013).** First detection of *Candidatus* Xenohaliotis californiensis the causative aget of withering syndrome, in Japanese Black abalone Haliotis discus discus in Japan. *Fish pathlogy* **2**, 35–41.

Koh, H.-W., Kim, M. S., Lee, J.-S., Kim, H. & Park, S.-J. (2015). Changes in the Swine Gut Microbiota in Response to Porcine Epidemic Diarrhea Infection. *Microbes Environ* **30**, 284–7.

Lafferty, K. D. & Kuris, a. M. (1993). Mass mortality of abalone *Haliotis cracherodii* on the California Channel Islands: tests of epidemiological hypotheses. *Mar Ecol Prog Ser* **96**, 239–248.

Langille, M., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J., Clemente, J., Burkepile, D., Vega Thurber, R. & other authors. **(2013).** Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**, 814–21.

Lee, K. B., Liu, C. Te, Anzai, Y., Kim, H., Aono, T. & Oyaizu, H. (2005). The hierarchical system of the 'Alphaproteobacteria': Description of Hyphomonadaceae fam. nov., Xanthobacteraceae fam. nov. and Erythrobacteraceae fam. nov. *Int J Syst Evol Microbiol* **55**, 1907–1919.

Leighton, D. L. (2000). The biology and culture of the California abalones. Dorrance Publishing Co., Pittsburgh, Pennsylvania.216 pp.

Lin, M., Zhang, C., Gibson, K. & Rikihisa, Y. (2009). Analysis of complete genome sequence of *Neorickettsia risticii*: causative agent of Potomac horse fever. *Nucleic Acids Res* **37**, 6076–91.

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F. & other authors. **(2007).** An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**, 915–918.

Ludwig, W., Mittenhuber, G. & Friedrich, C. G. (1993). Transfer of Thiosphaera pantotropha to *Paracoccus denitrificans*. *Int J Syst Bacteriol* **43**, 363–367.

Luna, R. A., Fasciano, L. R., Jones, S. C., Boyanton, B. L., Ton, T. T. & Versalovic, J. (2007). DNA pyrosequencing-based bacterial pathogen identification in a pediatric hospital setting. *J Clin Microbiol* **45**, 2985–2992.

Luo, H., Ros, M. C., Hughes, b A. L. & Morana, c M. A. (2013). Evolution of life history strategies in marine Alphaproteobacteria. *MBio* **4**, 1–8.

Macery, B. . & Coyne, V. E. (2004). Improved growth rate and disease resistance in farmed Haliotis midae through probiotic treatment. *Aquaculture* **245**, 249–261.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K. & other authors. **(1998).** Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–5.

Maiden, M. C. J. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* **60**, 561–588.

Martijn, J., Schulz, F., Zaremba-Niedzwiedzka, K., Viklund, J., Stepanauskas, R., Andersson, S. G., Horn, M., Guy, L. & Ettema, T. J. (2015). Single cell genomics of a rare environmental alphaproteobacterium provides unique insights into Rickettsiaceae evolution. *Isme j* **9**, 2373–2385.

Martino, M. E., Fasolato, L., Montemurro, F., Rosteghin, M., Manfrin, A., Patarnello, T., Novelli, E. & Cardazzo, B. (2011). Determination of microbial diversity of Aeromonas strains on the basis of multilocus sequence typing, phenotype, and presence of putative virulence genes. *Appl Environ Microbiol* **77**, 4986–5000.

Merhej, V. & Raoult, D. (2011). Rickettsial evolution in the light of comparative genomics. *Biol Rev* **86**, 379–405.

Meryandini, A., Junior, M. Z. & Rusmana, I. (2015). The Role of Agarolytic Bacteria in Enhancing Physiological Function for Digestive System of Abalone (Haliotis asinina). *j Appl EnvironBiol Sci* 5, 49–56.

Min, C. K., Yang, J. S., Kim, S., Choi, M. S., Kim, I. S. & Cho, N. H. (2008, January). Genome-based construction of the metabolic pathways of *Orientia tsutsugamush*i and comparative analysis within the *Rickettsiales* order. *Comp Funct Genomics* **2008,** 1-14

Moore, J. D., Robbins, T. T., Hedrick, R. P. & Friedman, C. S. (2001). Transmission of the Rickettsiales-like prokaryote '*Candidatus* xenohaliotis californiensis' and its role in Withering syndrome of California abalone, *Haliotis* spp. *J Shellfish Res*. **20,** 867-874.

Moore, J. D., Robbins, T. T. & Friedman, C. S. (2000). Withering syndrome in farmed red abalone *Haliotis rufescens*: Thermal induction and association with a gastrointestinal Rickettsiales-like prokaryote. *J Aquat Anim Health* **12**, 26–34.

Moore, J. D., Finley, C. a., Friedman, C. S. & Robbins, T. T. (2002). Withering syndrome and restoration of southern California abalone populations. *Calif Coop Ocean Fish Investig Reports* **43**, 112–117.

Moore, J. D., Juhasz, C. I., Robbins, T. T. & Ignacio Vilchis, L. (2009). Green abalone, *Haliotis fulgens* infected with the agent of withering syndrome do not express disease signs under a temperature regime permissive for red abalone, *Haliotis rufescens*. *Mar Biol* **156**, 2325–2330.

Morales-Bojórquez, E., Muciño-Díaz, M. O. & Vélez-Barajas, J. A. (2008). Analysis of the Decline of the Abalone Fishery (*Haliotis fulgens* and *H. corrugata*) along the Westcentral Coast of the Baja California Peninsula, Mexico. *J Shellfish Res* **27**, 865–870.

Moran, N. A., Munson, M. A., Baumann, P. & Ishikawa, H. (1993). A Molecular Clock in Endosymbiotic Bacteria is Calibrated Using the Insect Hosts. *Proc R Soc B Biol Sci*. **253,** 167-171.

Mossel, E. & Steel, M. (2005). How much can evolved characters tell us about the tree that generated them? *Math Evol Phylogeny*. Oxford University Press.

Mydlarz, L. D., Jones, L. E. & Harvell, C. D. (2006). Innate immunity, environmental drivers, and disease ecology of marine and freshwater invertebrates. *Annu Rev Ecol Syst* **37**, 251–288.

Ngangbam, A. K., Baten, A., Waters, D. L. E., Whalan, S. & Benkendorff, K. (2015). Characterization of Bacterial Communities Associated with the Tyrian Purple Producing Gland in a Marine Gastropod. *PLoS One* **10**, e0140725.

OIE. **(2008).** Aquatic Animal Health Code, Paris, (17th. ed.), *2008*.

Okada, K., Iida, T., Kita-Tsukamoto, K. & Honda, T. (2005). Vibrios commonly possess two chromosomes. *J Bacteriol* **187**, 752–757.

Pang, S. J., Xiao, T. & Bao, Y. (2006). Dynamic changes of total bacteria and *Vibrio* in an integrated seaweed–abalone culture system. *Aquaculture* **252,** 289–297.

Pearson, T., Busch, J. D., Ravel, J., Read, T. D., Rhoton, S. D., U'Ren, J. M., Simonson, T. S., Kachur, S. M., Leadem, R. R. & other authors. **(2004).** Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci U S A* **101**, 13536–13541.

Pettersoon, B., Leitner, T., Ronaghi, M., Bolske, G., Uhlen, M. & Johansson, K.-E. (1996). Phylogeny of the Mycoplasma mycoides Cluster as Determined by Sequence Analysis of the 16S rRNA Genes from the Two rRNA Operons. *J Bacteriol* **178**, 4131–4142.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G. & Baurain, D. (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol* **9**, e1000602.

Rahel, F. J. (2000). Homogenization of Fish Faunas Across the United States. *Science. 28***8,** 854–856.

Raimondi, P. T., Wilson, C. M., Ambrose, R. F., Engle, J. M. & Minchinton, T. E. (2002). Continued declines of black abalone along the coast of California: Are mass mortalities related to El Nino events? *Mar Ecol Prog Ser* **242**, 143–152.

Raoult, D. & Roux, V. (1997). Rickettsioses as paradigms of new or emerging infectious diseases. *Clin Microbiol Rev* **10**, 694–719.

Rar, V. & Golovljova, I. (2011). *Anaplasma*, *Ehrlichia*, and '*Candidatus* Neoehrlichia' bacteria: Pathogenicity, biodiversity, and molecular genetic characteristics, a review. *Infect Genet Evol* **11**, 1842–1861.

Razin, S., Yogev, D. & Naot, Y. (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev* **62**, 1094–1156.

Revollo, D. A. F. & Sáenz-Arroyo, A. (2012). The historical ecology of abalone in the Mexican Pacific. *Analisis* 89–112.

Rikihisa, Y. (2006). New findings on members of the family Anaplasmataceae of veterinary importance. *Ann N Y Acad Sci* **1078**, 438–445.

Roux, V. & Raoult, D. (1995). Phylogenetic Analysis of the Genus Rickettsia By 16S Rdna Sequencing. *Res Microbiol* **146**, 385–396.

Ruff-Roberts, A. L., Kuenen, J. G. & Ward, D. M. (1994). Distribution of cultivated and uncultivated cyanobacteria and Chloroflexus-like bacteria in hot spring microbial mats. *Appl Environ Microbiol* **60**, 697–704.

Rungrassamee, W., Klanchui, A., Maibunkaew, S., Chaiyapechara, S., Jiravanichpaisal, P. & Karoonuthaisiri, N. (2014). Characterization of intestinal bacteria in wild and domesticated adult black tiger shrimp (Penaeus monodon). *PLoS One* **9**, e91853.

SAGARPA. **(2009).** Sustentabilidad y pesca responsable en méxico. evaluación y manejo. Recuperado 09 de Octubre de: http://www.conapesca.sagarpa.gob.mx/wb/cona/pmregional_abulon

Sällström, B. & Andersson, S. G. E. (2005). Genome reduction in the α-proteobacteria. *Curr Opin Microbiol* **8**, 579–585.

Santos, A. S., Tilburg, J. J. H. C., Botelho, A., Barahona, M. J., N??ncio, M. S., Nabuurs-Franssen, M. H. & Klaassen, C. H. W. (2012). Genotypic diversity of clinical *Coxiella burnetii* isolates from Portugal based on MST and MLVA typing. *Int J Med Microbiol* **302**, 253–256. Elsevier GmbH.

Sawabe, T., Setoguchi, N., Inoue, S., Tanaka, R., Ootsubo, M., Yoshimizu, M. & Ezura, Y. (2003). Acetic acid production of *Vibrio halioticoli* from alginate: A possible role for establishment of abalone-V. halioticoli association. *Aquaculture* **219**, 671–679.

Sawabe, T., Inoue, S., Fukui, Y., Yoshie, K., Nishihara, Y. & Miura, H. (2007). Mass mortalities of Japanese abalone *Haliotis discus hannai* caused by *Vibrio harveyi* infection. *Microbes Environ* **22**, 300–308.

Schreiber, F., Pick, K., Erpenbeck, D., Wörheide, G. & Morgenstern, B. (2009). OrthoSelect: a protocol for selecting orthologous groups in phylogenomics. *BMC Bioinformatics* **10**, 219.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S. W., Huttenhower, C., Golub, T., Petricoin, E. & other authors. **(2011).** Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, 60.

Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**, 1114–1116.

Sjödin, A., Svensson, K., Ohrman, C., Ahlinder, J., Lindgren, P., Duodu, S., Johansson, A., Colquhoun, D. J., Larsson, P. & Forsman, M. (2012). Genome characterisation of the genus *Francisella* reveals insight into similar evolutionary paths in pathogens of mammals and fish. *BMC Genomics* **13**, 268.

Speel, E. J., Hopman, a H. & Komminoth, P. (1999). Amplification methods to increase the sensitivity of in situ hybridization: play card(s). *J Histochem Cytochem* **47**, 281–288.

Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. & Musser, J. M. (1997). Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* **94**, 9869–9874.

Stackebrandt, E., Stackebrandt, E., Frederiksen, W., Frederiksen, W., Garrity, G. M., Garrity, G. M., Grimont, P. a D., Grimont, P. a D., Ka, P. & other authors. **(2002).** Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* 1043–1047.

Steinbeck, J.R., Groff, J.M., Friedman, C.S., McDowell T. and Hedrick R.P. (1992). Investigations into a mortaliy among population of the California blach abalone *Haliotis cracherodii* on the Central Coast of California, U.S.A. In S.A. Shepherd of the World. Biology Fisheries and Cultive. Proceeding of 1st. International Symposium on Abalone. Fishing News Book. The University Press. 203-213. Tissot

Sweet, M. J. & Bateman, K. S. (2015). Diseases in marine invertebrates associated with mariculture and commercial fisheries. *J Sea Res* **104**, 16–32.

Swofford, D. L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. *Sinauer Assoc Sunderland, Massachusetts*.

Tanaka, R., Sugimura, I., Sawabe, T., Yoshimizu, M. & Ezura, Y. (2003). Gut microflora of abalone *Haliotis discus hannai* in. *Fish Sci* **69**, 951–958.

Tanaka, R., Ootsubo, M., Sawabe, T., Ezura, Y. & Tajima, K. (2004). Biodiversity and in situ abundance of gut microflora of abalone (*Haliotis discus hannai*) determined by culture-independent techniques. *Aquaculture* **241**, 453–463.

Team, R. C. (2015). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL From: http://www.R-project.org/. *http://wwwR-ProjectOrg/*

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Hoste, B., Munn, C. B., Swings, J. & Naser, S. (2005). Phylogeny and Molecular Identification of Vibrios on the Basis of Multilocus Sequence Analysis. *Appl Environ Microbiol* **71**, 5107–5115.

Urbanová, Z. & Bárta, J. (2014). Microbial community composition and in silico predicted metabolic potential reflect biogeochemical gradients between distinct peatland types. *FEMS Microbiol Ecol* **90**, 633–646.

Valles-Ríos, H. (2000). Análisis histopatológico del abulón negro *Haliotis cracherodii* afectado por el síndrome de deterioro. *Cienc Pesq* **14**, 5–19.

Vannini, C., Petroni, G., Verni, F. & Rosati, G. (2005). A bacterium belonging to the Rickettsiaceae family inhabits the cytoplasm of the marine ciliate *Diophrys appendiculata* (Ciliophora, Hypotrichia). *Microb Ecol* **49**, 434–442.

Vilchis, I. L., Tegner, M. J., Moore, J. D., Friedman, C. S., Riser, T. T. & Dayton, P. K. (2005). Ocean Warming Effects on Growth , Reproduction , and Survivorship of Southern California Abalone. *Ecol Appl* **15**, 469–480.

de Voogd, N. J., Cleary, D. F. R., Polónia, A. R. M. & Gomes, N. C. M. (2015). Bacterial community composition and predicted functional ecology of sponges, sediment and seawater from the thousand islands reef complex, West Java, Indonesia. *FEMS Microbiol Ecol* **91**, 1–12.

Wardwell, L. H., Huttenhower, C. & Garrett, W. S. (2011). Current Concepts of the Intestinal Microbiota and the Pathogenesis of Infection. *Curr Infect Dis Rep* **13**, 28–34.

Weinert, L., Werren, J. & Aebi, A. (2009). Evolution and diversity of *Rickettsia* bacteria. *BMC Biol* **7**, 6.

Westberg, J., Persson, A., Holmberg, A., Goesmann, A., Lundeberg, J. & Pettersson, B. (2004). The Genome Sequence of *Mycoplasma mycoides* subsp. *mycoides* SC Type Strain PG1 T , the Causative Agent of Contagious Bovine Pleuropneumonia (CBPP). *Genome Res* 221–227.

Wetchateng, T., Friedman, C. S., Wight, N. A., Lee, P. Y., Teng, P. H., Sriurairattana, S., Wongprasert, K. & Withyachumnarnkul, B. (2010). Withering syndrome in the abalone *Haliotis diversicolor supertexta*. *Dis Aquat Organ* **90**, 69–76.

Wolf, Y. I. & Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837.

Zhao, J., Shi, B., Jiang, Q. R. & Ke, C. H. (2012). Changes in gut-associated flora and bacterial digestive enzymes during the development stages of abalone (*Haliotis diversicolor*). *Aquaculture* **338–341**, 147–153.

Zhu, Y., Fournier, P.-E., Eremeeva, M. & Raoult, D. (2005). Proposal to create subspecies of *Rickettsia conorii* based on multi-locus sequence typing and an emended description of *Rickettsia conorii*. *BMC Microbiol* **5**, 11.

Annexes

Annex 1: GenBank accession numbers of DNA sequences used in phylogenetic reconstructions for the proposed genes (*: No data available; - Not applicable)

Taxon	ACCESSION	16S	23S	VirB11	VirD4	ftsZ
Anaplasma centrale strain Israel	CP001759	ACIS_01352	ACIS_01353	ACIS_00093	ACIS_00094	ACIS_00137
Anaplasma marginale strain Florida	CP001079	AMF_1059	AMF_1055	AMF_993	AMF_992	AMF_952
Anaplasma phagocytophilum strain HZ2	CP006616	YYU_04585	YYU_00420	YYU_06505	YYU_06500	YYU_05975
Candidatus Midichloria mitochondrii strain IricVA	CP002130	midi_00228	midi_00832	midi_00444	midi_00443	midi_00222
Candidatus Xenohaliotis californiensis	*	KU645900	KU645901	KU645904	KU645903	KU645902
Ehrlichia canis strain Jake	CP000107	Ecaj_R0013	Ecaj_R0039	Ecaj_0019	Ecaj_0018	Ecaj_0925
Ehrlichia chaffeensis strain Arkansas	CP000236	ECH_0919	ECH_0222	ECH_0041	ECH_0040	ECH_1153
Ehrlichia muris strain AS145	CP006917	EMUR_00930	EMUR_03980	EMUR_00045	EMUR_00040	EMUR_04680
Ehrlichia ruminantium strain Welgevonden	CR767821	*	*	Erum0270	Erum0260	Erum8800
Endosymbiont of *Acanthamoeba* sp., strain UWC8	CP004403	I862_r07731	I862_r07733	I862_04760	I862_04765	I862_02320
Neorickettsia helminthoeca strain Oregon	CP007481	NHE_0048	NHE_0763	NHE_0702	NHE_0701	NHE_0331
Neorickettsia risticii strain Illinois	CP001431	NRI_16SA	NRI_23SB	NRI_0714	NRI_0713	NRI_0334
Neorickettsia sennetsu strain Miyayama	CP000237	NSE_0973	NSE_0976	NSE_0740	NSE_0739	NSE_0342
Orientia tsutsugamushi strain Ikeda DNA	AP008981	OTT_RNA006	OTT_RNA022	OTT_1568	OTT_1567	OTT_0905
Rickettsia bellii strain OSU 85-389	CP000849	A1I_r08027	*	A1I_02290	A1I_02285	A1I_01985
Rickettsia conorii strain Malish 7	AE006914	RCRNA26	RCRNA07	RC0390	RC0391	RC1015
Rickettsia prowazekii strain Dachau	CP003394	MA3_r04405	MA3_r04407	MA3_01435	MA3_01440	MA3_03235
Rickettsia rickettsii strain Morgan	CP006010	RRM_04885	RRM_01540	RRM_02145	RRM_02150	RRM_05280
Rickettsia typhi strain Wilmington	AE017197	RT0602	RT0200	RT0283	RT0284	RT0658
Rickettsiales bacterium strain Ac37b	CP009217.2	NOVO_03225	NOVO_06390	NOVO_RS03175	NOVO_03380	NOVO_04175
Wolbachia endosymbiont of *Cimex lectularius*	AP013028	WCLE_003560	WCLE_006080	WCLE_00080	WCLE_00090	WCLE_00510
Wolbachia endosymbiont of *Drosophila melanogaster*	AE017196	WD_Wp16SA	WD_Wp23SB	WD_0007	WD0008	WD_0723
Wolbachia endosymbiont of *Drosophila simulans* strain wNo	CP003883	wNo_r001	wNo_r002	wNo_10750	wNo_10760	wNo_02790
Wolbachia endosymbiont strain TRS of *Brugia malayi*	AE017321	Wbm9003	Wbm9001	Wbm0282	Wbm0283	Wbm0602
Wolbachia sp. strain wRi	CP001391	WRi_r11990	WRi_r01850	WRi_000080	WRi_000090	WRi_007520
Agrobacterium radiobacter strain K84	CP000628	Arad_5100	Arad_5101	-	-	Arad_2986
Sinorhizobium fredii strain USDA 257	CP003563	USDA257_c50180	USDA257_c50150	-	-	USDA257_c44940
Xanthomonas sacchari strain R1	CP010409	-	-	SB85_04360	SB85_04380	-
Xanthomonas citri strain UI6	CP008992	-	-	J158_02787	J158_02792	-

Annex 2: Nucleotide and amino acid (AA) evolution models used in the phylogenetic reconstructions of any proposed locus and for concatenated tree.

Locus	**Nucleotide evolution model**	**AA evolution model**
16S rRNA	GTR + I + G	-
23S rRNA	GTR + G	-
ftsZ	GTR + G	LG + G
virB11	TIM3 + I + G	LG + G
virD4	GTR + I + G	LG + G
Concatenated	GTR + I + G	LG + G + I

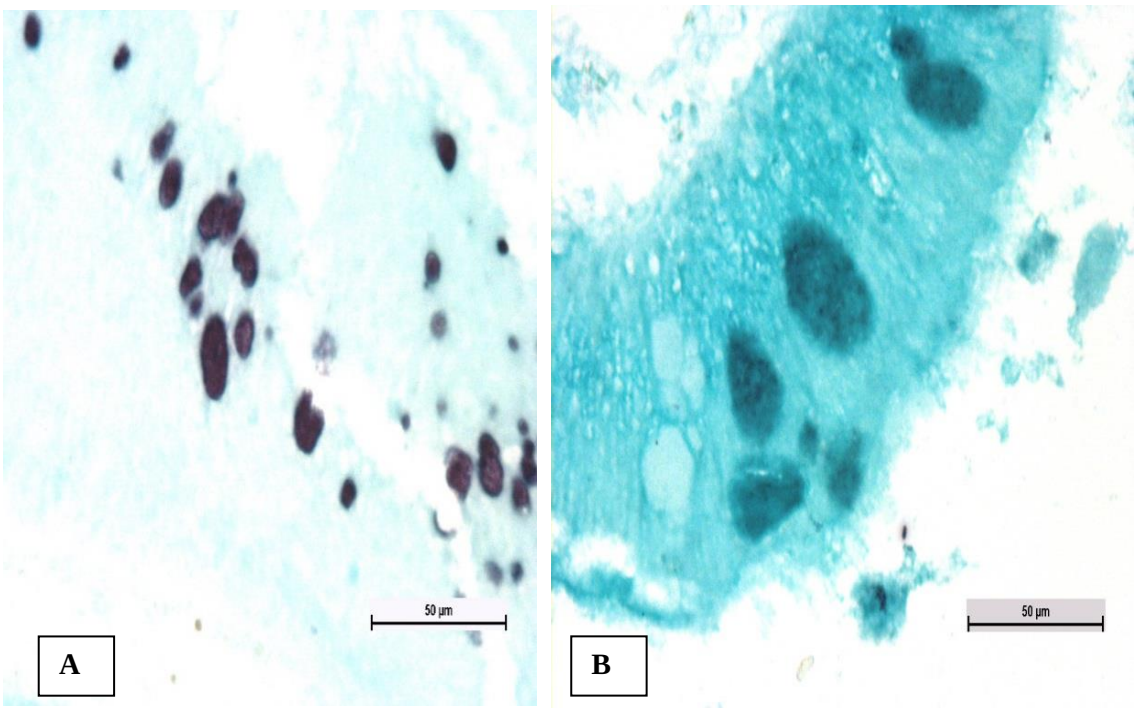
Annex 3: Loci and primers designed for the order *Rickettsiales*

Locus	Name	5' - Sequence - 3'	Length	Expected fragment size (bp)
23S	23S.F7	TGTTCGCCGATTAAAGTG	18	550
	5S.R2	GYTTMACDTYCRRGTTC	17	
virD4	VirD4.F1	GGKTCWGGDAARGGTGT	17	800
	VirD4.R1	ARYTGYTSNGTATCTT	16	
virB11	VirB11.F1	GARGTRTGGRTWGARAA	17	760
	VirB11.R1	YYKGCYTGYATWACCA	16	
ftsZ	ftsZ.F2	ATHACDGCNGGDATGG	16	550
	ftsZ.R1	GCNKYRTCMACYTCAAA	17	
tpiA	tpiA.nF1	GTWGCWAAYTGGAARATGAATGG	23	700
	tpiA.nR1	GCACTRCCAAYYAAWACHCC	20	
coxA	CoxA.nF2	GGWACRGGWTGGACWTTRTAYCC	23	700
	CoxA.nR1	CCCACATRGTKGCWATCCARCT	22	
gap	gap.nF1	TAGGAATWAAYGGYYTWGGTMG	22	1000
	gap.nR1	AAGCCCAYTCRTTRTCRTACC	21	
groEL	groEL.nF7	ATTAGYAAGCCYTATGGKKCVCC	23	650
	groEL.nR8	TTCACCYTCWACRTCTTCWGCDA	23	
atpD	atpD.nF1	GGWAARRTTGGYYTRTTTGGTGG	23	900
	atpD.nR1	CCMACCATRTAAAARGCMGCTTC	23	
tkt	tkt.nF2	GGARGGWATHAGYCAYGAAGCWGC	24	1000
	tkt.nR1	GCCATRCCRTTCATACAHGCKGC	23	
recA	RecA.nF2	CTTGCDYTKCAYGTWATTGCTGA	23	400
	RecA.nR7	CMGTVGTAGTYTCAGGATTWCCA	23	
rpoB	rpoB.nF1	GTKGCYTTTATGTCDTGGCARGG	23	800
	rpoB.nR6	TGYCKACCAGMCATYTTATCHCC	23	
rpoC	rpoC.F2	TGARCCKGGTACDCA	15	1000
	rpoC.R2	ATHACYTCKATGTGYTTRTY	20	
rpoD	rpoD.UCF1	GGMAATATYGGTCTKATGAARGC	23	550
	rpoD.UCR2	GYCTRATTCTYTCMCKTGTTAC	22	
omp	OMP.fF1	CCTTYATAGACMRTAAYYTATTGGTAC	28	820
	OMP.fR1	CCAATGCCRRAAAKRTCAAACCC	23	

Annex 4. Loci, number of observed haplotypes, and BLAST search hits with % similarity for the DNA sequences produced with primers designed for the order Rickettsiales

Locus	Haplotypes	BLAST	% Similarity
16S rRNA	2	*Wolbachia* spp. *16S rRNA*	85%
		Neorickettsia spp. *16S rRNA*	80%
ftsZ	2	*Ehrlichia* spp. *ftsZ*	72%
		Bacteria spp. *ftsZ*	-
virD4	2	*Wolbachia* spp. *virD4*	78%
		Rickettsia spp. *virD4*	79%
virB11	1	*Wolbachia* spp. *virB11*	76%
coxA	4	*Anaplasma* spp. *coxA*	72%
		Wolbachia spp. *coxA*	73%
		Rickettsia spp. *coxA*	76%
		Other organisms	-
tpiA	3	*Wolbachia* spp. *tpiA*	100% +
		Bacteria spp. *tpiA*	-
		Other organisms	-
gap	3	*Anaplasma* spp. *gap*	100% +
		Vibrio spp. *gap*	74%
		Other organisms	-
23S rRNA	4	*Wolbachia* spp. *23S rRNA*	87%
		Neorickettsia spp. *23S rRNA*	82%
		Microvibrio spp. *23S rRNA*	73%
		Bacteria spp. *23S rRNA*	-
tkt	3	*Wolbachia* spp. *tkt*	68%
		Neorickettsia spp. *tkt*	69%
		Bacteria spp. *tkt*	-
atpD	1	*Mollicutes* spp. *atpD*	80%
rpoD	3	*Pseudomonas* spp. *rpoD*	83%
		Alteromonas spp. *rpoD*	81%
		Bacteria spp. *rpoD*	-
groEL	1	Bacteria spp. *groEL*	-
omp	3	Other organisms	-

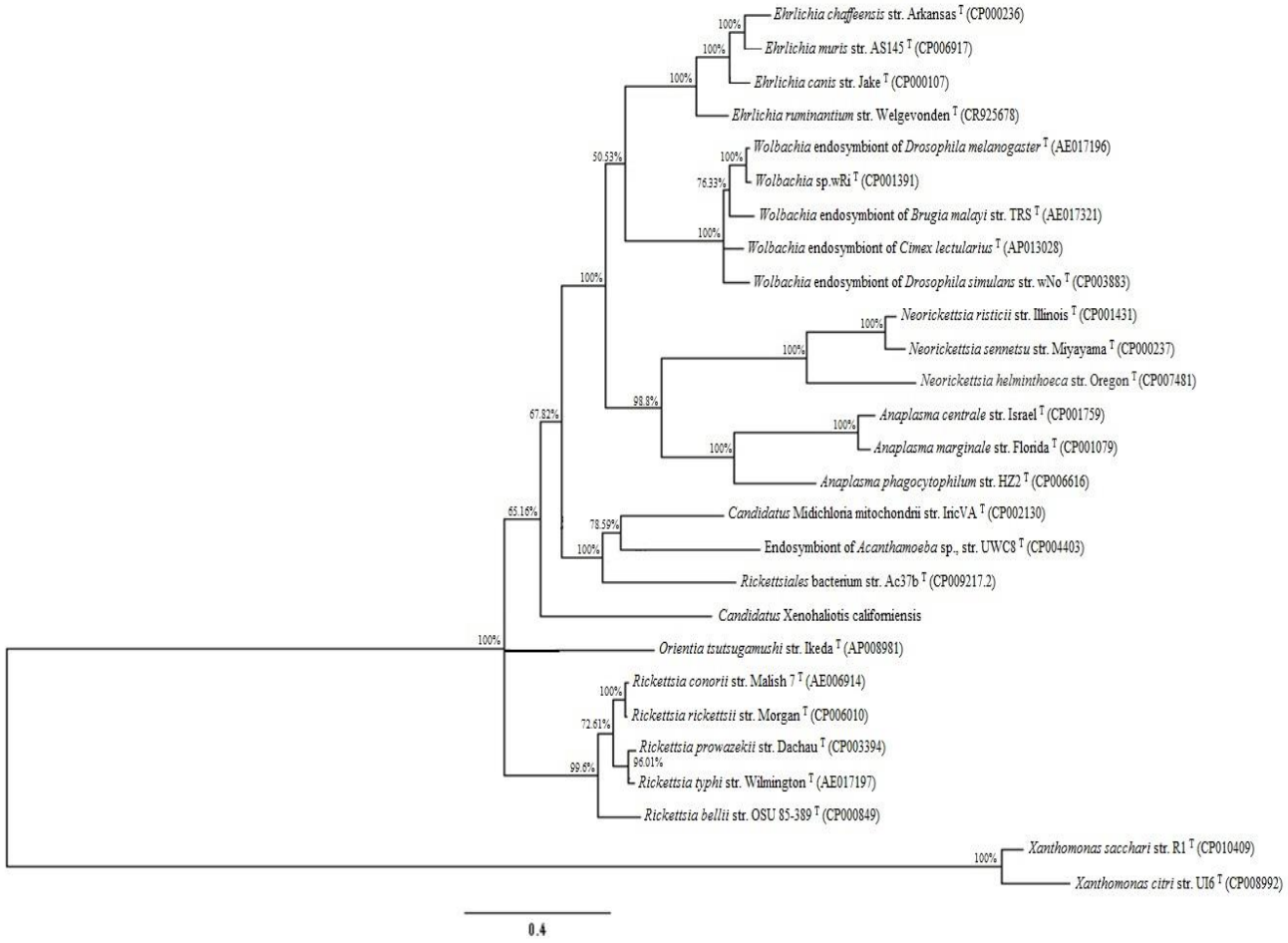
Annex 5. Micrographs of histological slides of post-esophageal tissue from red abalone (*Haliotis rufescens*) infected with *Candidatus* Xenohaliotis californiensis showing positive staining of bacterial inclusions with *In situ* hybridization probes for the ribosomal genes *16S rRNA* (A) and *23S rRNA* (B).



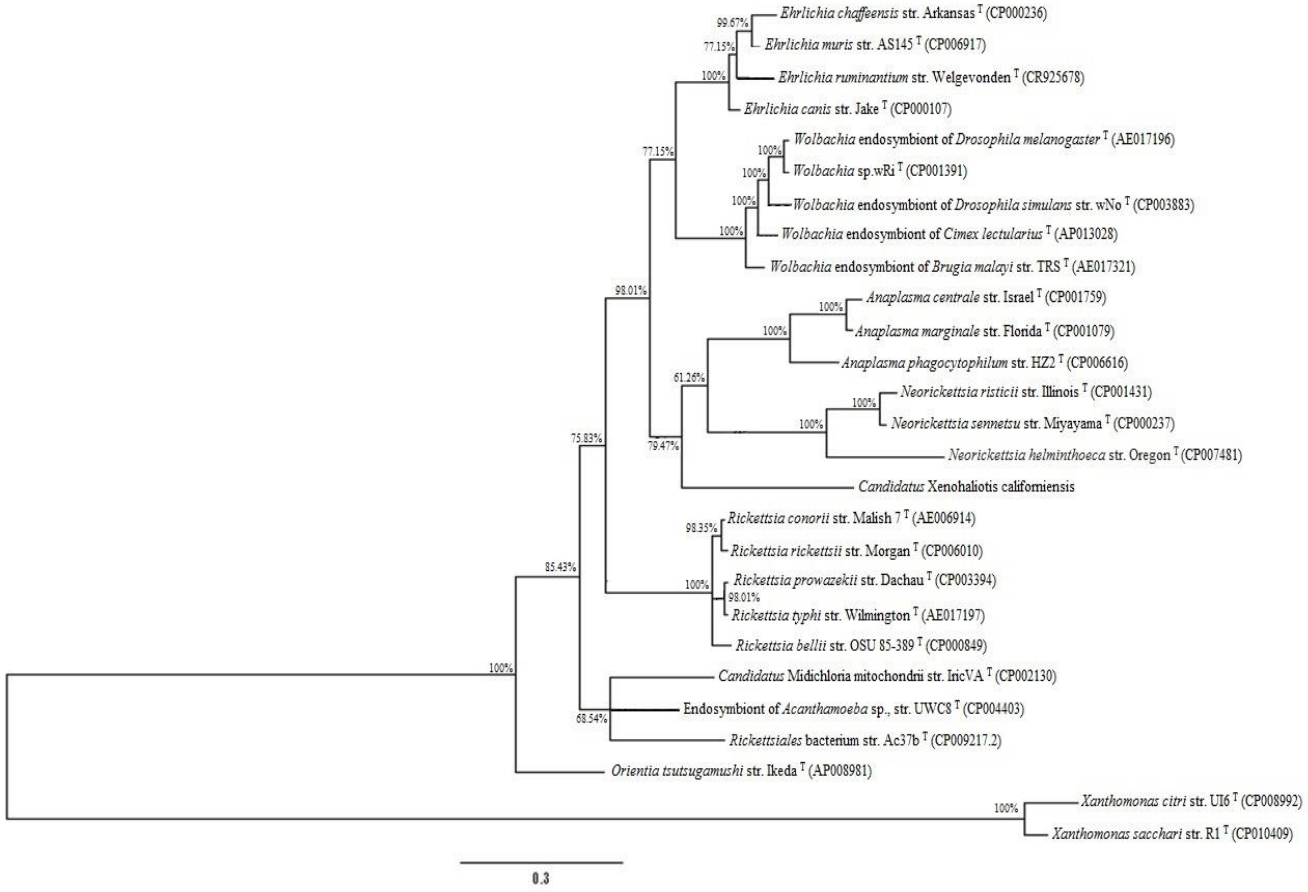
Annex 6: *Candidatus* Xenohaliotis californiensis detection and co-amplification pattern using 3 abalone host species *H. rufescens* (HR) farmed at the UC Davis Bodega Marine Laboratory (Bodega Bay, California, USA) and wild *H. corrugata* (HC) and *H. fulgens* (HF) sampled from the west coast of Baja California (Mexico). SSU amplification involved the ss16S primers set (see text for details). ID = specimen ID; GA = Geographic Area, 1 **=** positive amplification and 0 = no amplification.

ID	HR0001	HR0002	HR0003	HR0004	HR0005	HR0006	HR0007	HR0008	HC0001	HC0140	HC0007	HC0147	HC0154	HC0177	HC0023	HC0042	HC0054	HC0060	HC0091	HC0118	HC0111	HC0122	HC0159	HC0178	HF0004	HF0020	HF0037	HF0105	HF0048	HF0049	HF0073	HF0001e	HF0079e	HF0189	HF0194	HF0089	HF0091	HF0136	HF0144
GA	Bodega Bay	Bodega Bay	Bodega Bay	Bodega Bay	Bodega Bay	Bodega Bay	Bodega Bay	Bodega Bay	Bahia Tortuga	Bahia Tortuga	La Purissima	La Purissima	Emancipacion	Emancipacion	California	California	Leys De Reforma	Leys De Reforma	Punta Abreojos	Progresso	Progresso	Progresso	Buzos y Pescadores	Emancipacion	Bahia Tortuga	Bahia Tortuga	La Purissima	La Purissima	Emancipacion	Emancipacion	California	California	Leys De Reforma	Punta Abreojos	Punta Abreojos	Progresso	Progresso	Buzos y Pescadores	Buzos y Pescadores
16S	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
23S	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
ftsZ	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
virB 11	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
vir D4	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1

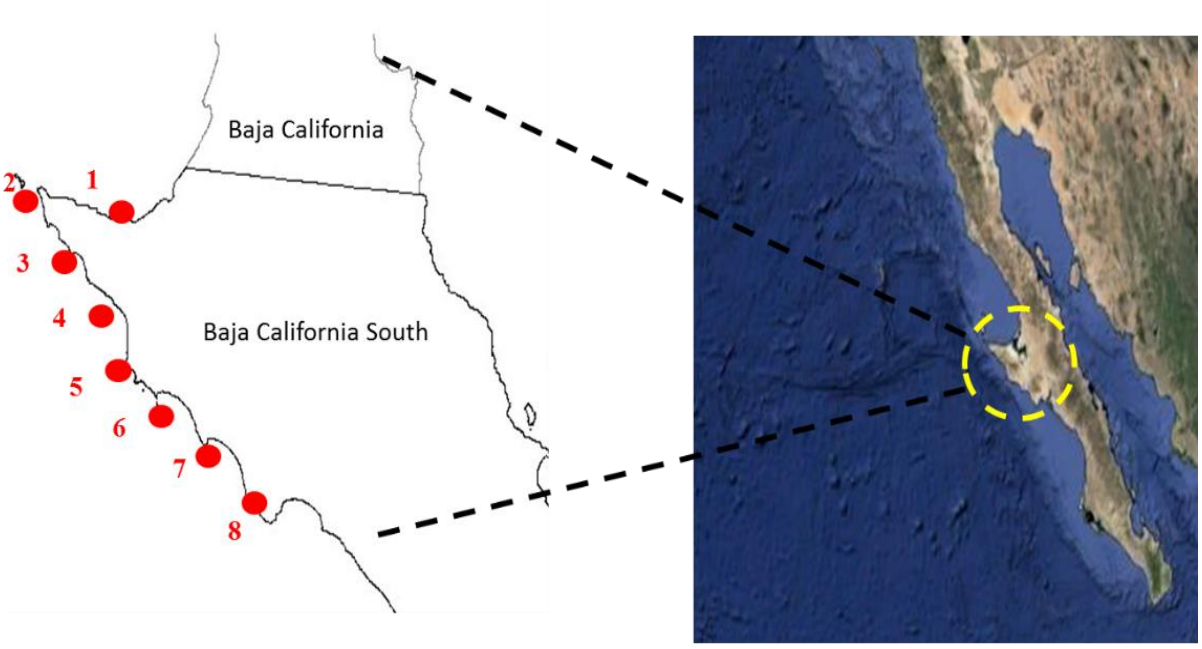
Annex 7. Bayesian phylogenetic tree (Mr Bayes version 3.2) obtained with the *virD4* nucleotide sequence (topology E). Nodes labels show posterior probabilities.



Annex 8. Bayesian phylogenetic tree (Mr Bayes version 3.2) obtained with the *virB11* amino acidic sequence (topology F). Nodes labels show posterior probabilities.



Annex 9. Sampling localities for *Haliotis fulgens* and *Haliotis corrugata*. 1 La Purísima, 2 Buzos y Pescadores, 3 Bahía Tortugas, 4 Emancipación, 5 California, 6 Leyes de Reforma, 7 Progreso, 8. Punta Abreojos. Each site represents an independent fishing cooperative.



Annex 10. List of *H. corrugata* (HC), *H. fulgens* (HF) and *H. rufescens* (HR) abalone used as source of *Candidatus* Xenohaliotis californiensis DNA for each tested locus. Abalobe ID = specimen ID; HS= Health state: WS = presence of morphological signs of WS; - = absence of morphological signs of WS; 1 **=** positive amplification; 1n = amplified by nested PCR; na = No PCR amplification; nt = not tested.

Abalone ID	Locality	HS	16S	23S	ITS.1	ftsZ	VirB11	VirD4	ITS.2
HC_0004	La Purissima	-	1	1	1	1	1	1	1
HC_0005	La Purissima	-	1	1	1	1	na	na	na
HC_0007	La Purissima	-	1	1	1	1	1	1	1
HC_0012	La Purissima	-	1	1	1	1	1n	1n	1n
HC_0023	California	-	nt	nt	nt	nt	1	1	1
HC_0024	California	-	1	1	1	1	na	na	na
HC_0026	California	-	1	1	1	1	na	na	na
HC_0027	California	WS	1	1	1	1	na	na	na
HC_0028	California	-	nt	nt	nt	nt	1	1	1
HC_0031	California	-	1	1	1	1	1n	1n	1n
HC_0051	Leys De Reforma	-	1	1	1	1	1n	1n	1n
HC_0051e	Buzos y Pescadores	-	1	1	1	1	1	1	1
HC_0054	Leys De Reforma	-	nt	nt	nt	1	1	1	1
HC_0059	Leys De Reforma	-	1	1	1	1	na	na	na
HC_0061	Leys De Reforma	-	1	1	1	1	1n	1n	1n
HC_0073	Leys De Reforma	-	nt	nt	nt	nt	1	1	1
HC_0082	Punta Abreojos	-	1	1	1	1	1	1	1
HC_0084	Punta Abreojos	-	1	1	1	1	1	1	1
HC_0092	Punta Abreojos	-	1	1	1	1	1	1	1
HC_0109	Progresso	-	1	1	1	1	na	na	na
HC_0110	Progresso	WS	1	1	1	1	1	1	1
HC_0112	Progresso	-	1	1	1	1	1	1	1
HC_0118	Progresso	WS	1	1	1	1	1	1	1
HC_0140	Bahia Tortuga	-	1	1	1	1	1n	1n	1n
HC_0152	Emancipacion	-	1	1	1	1	na	na	na
HC_0153	Emancipacion	-	1	1	1	1	na	na	na
HC_0154	Emancipacion	-	1	1	1	1	1	1	1
HC_0155	Emancipacion	-	1	1	1	1	1	1	1
HC_0156	Emancipacion	-	nt	nt	nt	nt	1	1	1
HC_0161	Buzos y Pescadores	-	1	1	1	1	1	1	1
HC_0166	Buzos y Pescadores	-	1	1	1	1	na	na	na
HC_0168	Buzos y Pescadores	-	1	1	1	1	1	1	1
HF_0001e	California	WS	1	1	1	1	na	na	na
HF_0004	Bahia Tortuga	-	1	1	1	1	1n	1n	1n
HF_0020	Bahia Tortuga	-	1	1	1	1	1n	1n	1n
HF_0022	Bahia Tortuga	-	1	1	1	1	1	1	1
HF_0024	Bahia Tortuga	-	1	1	1	1	na	na	na
HF_0031	La Purissima	-	1	1	1	1	1n	1n	1n
HF_0032	La Purissima	-	1	1	1	1	1	1	1

HF_0038	La Purissima	-	1	1	1	1	na	na	na
HF_0040	Emancipacion	WS	1	1	1	1	1	1	1
HF_0048	Emancipacion	WS	1	1	1	1	1	1	1
HF_0049	Emancipacion	WS	1	1	1	1	na	na	na
HF_0053	California	-	1	1	1	1	1	1	1
HF_0073	California	-	1	1	1	1	1	1	1
HF_0074	California	-	1	1	1	1	1	1	1
HF_0079e	Leys De Reforma	WS	1	1	1	1	1	1	1
HF_0081	Punta Abreojos	-	1	1	1	1	1n	1n	1n
HF_0083	Punta Abreojos	-	1	1	1	1	1	1	1
HF_0089	Progresso	WS	1	1	1	1	1	1	1
HF_0091	Progresso	-	1	1	1	1	1	1	1
HF_0092	Progresso	-	1	1	1	1	na	na	na
HF_0094	Progresso	-	1	1	1	1	1	1	1
HF_0101	Emancipacion	WS	1	1	1	1	1	1	1
HF_0105	La Purissima	WS	1	1	1	1	1	1	1
HF_0134	Buzos y Pescadores	-	1	1	1	1	1	1	1
HF_0135	Buzos y Pescadores	-	1	1	1	1	na	na	na
HF_0136	Buzos y Pescadores	WS	1	1	1	1	1	1	1
HF_0142	Buzos y Pescadores	-	1	1	1	1	1	1	1
HF_0144	Buzos y Pescadores	-	nt	nt	nt	1	na	na	na
HF_0194	Punta Abreojos	-	1	1	1	1	1	1	1
HR_0001	Bodega Bay	-	na	na	na	na	na	na	na
HR_0002	Bodega Bay	-	na	na	na	na	na	na	na
HR_0003	Bodega Bay	-	na	na	na	na	na	na	na
HR_0004	Bodega Bay	-	1	1	1	1	1	1	1
HR_0005	Bodega Bay	-	na	na	na	na	na	na	na
HR_0006	Bodega Bay	-	1	1	1	1	1	1	1
HR_0007	Bodega Bay	-	1	1	1	1	1	1	1
HR_0008	Bodega Bay	-	1	1	1	1	1	1	1

Annex 11. Number of DNA sequences obtained by Sanger sequencing for each locus in *H. corrugata* (HC), *H. fulgens* (HF) and *H. rufescens* (HR) per locality. - = Not sampled.

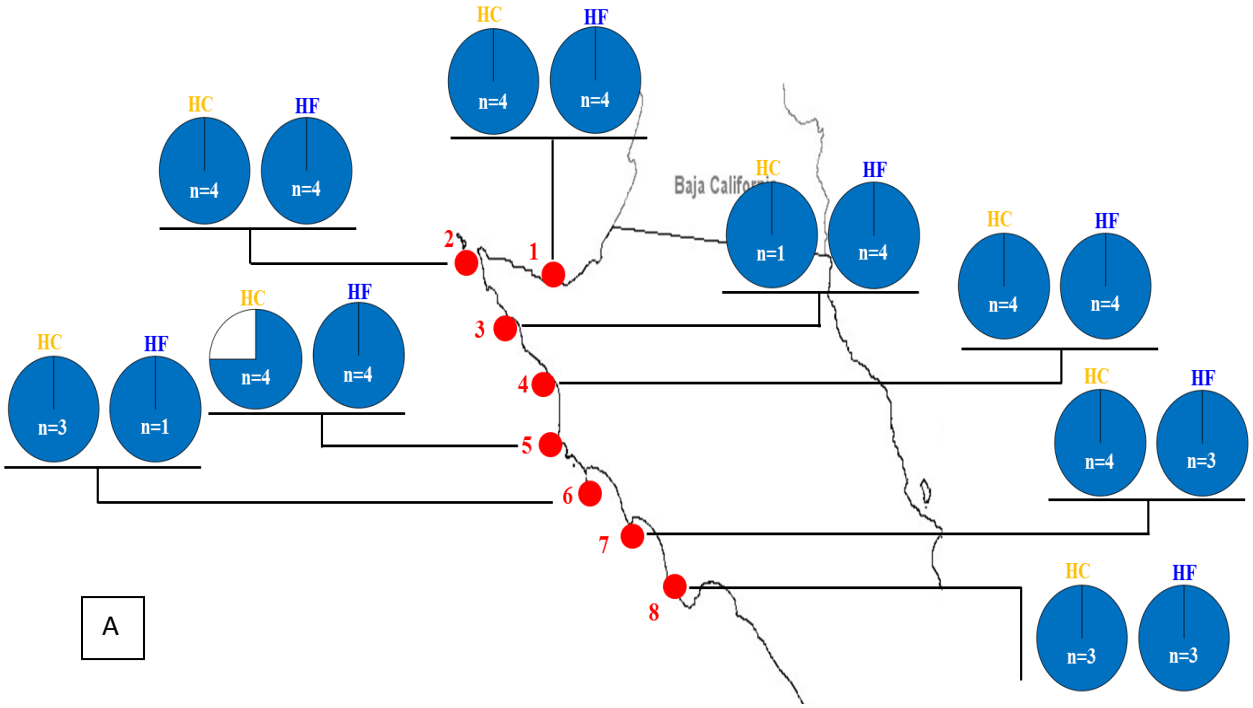
Locality	16S			23S			ITS.1			ftsZ			VirB11			VirD4			ITS.2		
	HC	HF	HR	HC	HF	HR	HC	HF	HR	HC	HF	HR	HC	HF	HR	HC	HF	HR	HC	HF	HR
Bahia Tortuga	1	4	-	1	4	-	1	4	-	1	4	-	1	3	-	1	3	-	1	3	-
Buzos y Pescadores	4	4	-	4	4	-	4	4	-	4	4	-	3	3	-	3	3	-	3	3	-
California	4	4	-	4	4	-	4	4	-	4	4	-	3	3	-	3	3	-	3	3	-
Emancipacion	4	4	-	4	4	-	4	4	-	4	4	-	3	3	-	3	3	-	3	3	-
La Purissima	4	4	-	4	4	-	4	4	-	5	4	-	3	3	-	3	3	-	3	3	-
Leys De Reforma	3	1	-	3	1	-	3	1	-	4	1	-	3	1	-	3	1	-	3	1	-
Progresso	4	4	-	4	4	-	4	4	-	4	4	-	3	3	-	3	3	-	3	3	-
Punta Abreojos	3	3	-	3	3	-	3	3	-	3	3	-	3	3	-	3	3	-	3	3	-
Bodega Bay	-	-	3	-	-	3	-	-	3	-	-	3	-	-	3	-	-	3	-	-	3

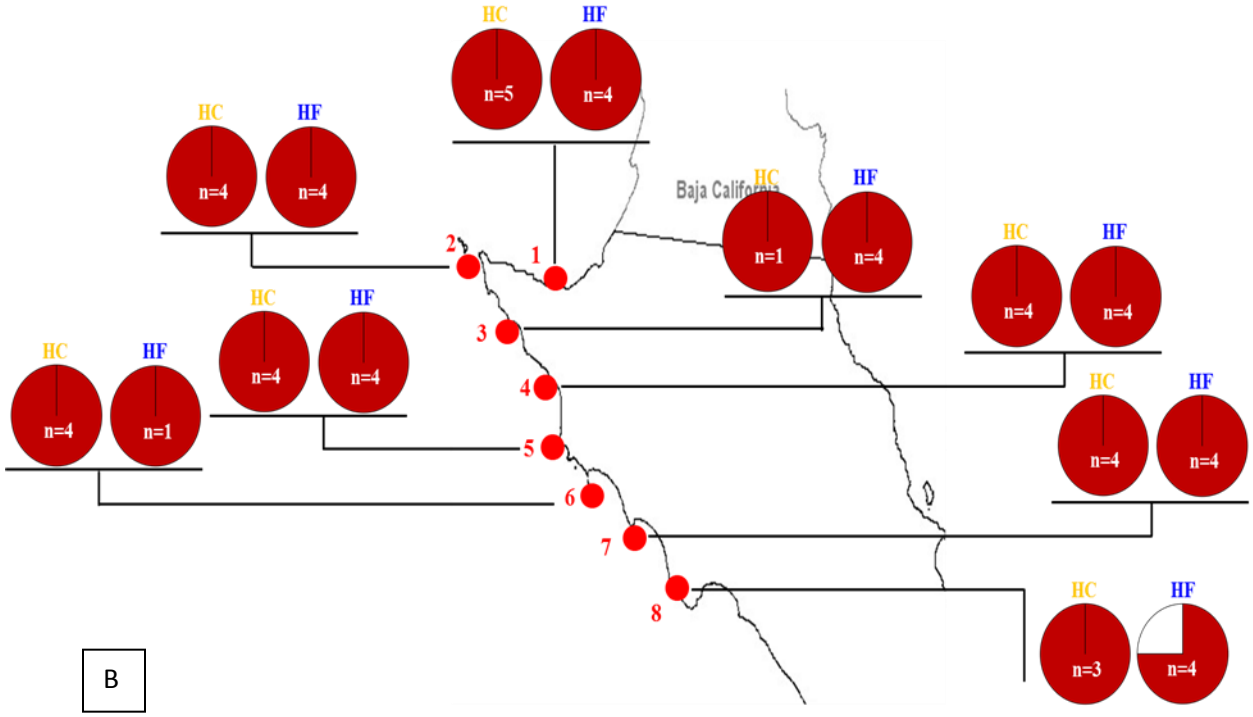
Annex 12. List of *H. corrugata* (HC) and *H. fulgens* (HF) abalone subject to 454 Pyrosequencing. ID = specimen ID, ID; HS= Health state: WS = presence of morphological and hystological signs of WS; - = absence of morphological signs; R1 and R2 number of reads number of ribotypes 1 (R1) and 2 (R2) in each abalone.

HC_ID	Locality	HS	R1	R2
HC_0140	Bahía Tortugas	-	3	0
HC_0157	Buzos y Pescadores	WS	0	0
HC_0159	Buzos y Pescadores	WS	0	0
HC_0168	Buzos y Pescadores	-	10	0
HC_0170	Buzos y Pescadores	-	0	0
HC_0179	Buzos y Pescadores	WS	0	0
HC_0181	Buzos y Pescadores	WS	1	0
HC_0186	Buzos y Pescadores	WS	0	0
HC_0023	California	-	2	0
HC_0029	California	-	0	0
HC_0040	California	-	48	0
HC_0042	California	-	2	0
HC_0152	Emancipación	-	15	0
HC_0153	Emancipación	-	212	8
HC_0154	Emancipación	-	957	4
HC_0155	Emancipación	-	2296	30
HC_0156	Emancipación	-	10	0
HC_0177	Emancipación	WS	0	0
HC_0178	Emancipación	-	151	9
HC_0006	La Purísima	-	0	0
HC_0147	La Purísima	-	68	4
HC_0054	Leyes de Reforma	-	0	0
HC_0056	Leyes de Reforma	-	0	0
HC_0060	Leyes de Reforma	-	0	0
HC_0063	Leyes de Reforma	-	0	0
HC_0064	Leyes de Reforma	-	1	0
HC_0072	Leyes de Reforma	-	0	0
HC_0074	Leyes de Reforma	-	2	0
HC_0075	Leyes de Reforma	-	1	0
HC_0109	Progreso	WS	2	0
HC_0111	Progreso	WS	0	0
HC_0116	Progreso	WS	0	0
HC_0122	Progreso	-	5	0
HC_0129	Progreso	-	0	0
HC_0089	Punta Abreojos	-	0	0
HC_0091	Punta Abreojos	-	19	1
HC_0094	Punta Abreojos	-	0	0

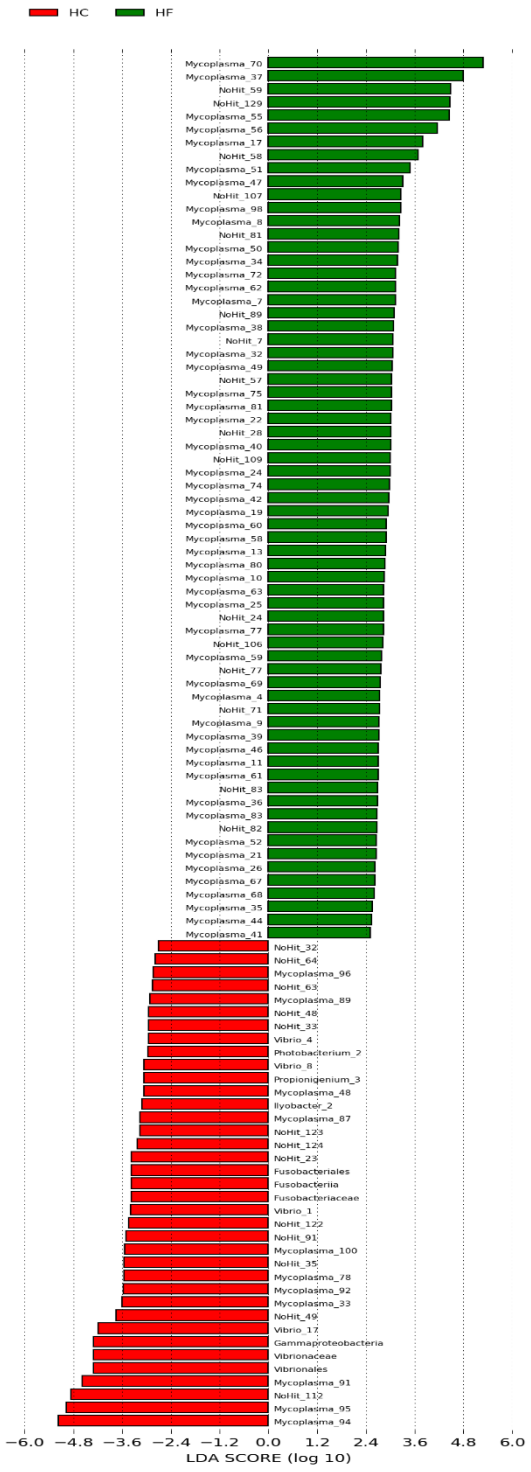
HF_ID	Locality	HS	R1	R2
HF_0004	Bahía Tortugas	-	1	0
HF_0009	Bahía Tortugas	-	4	0
HF_0020	Bahía Tortugas	-	6	1
HF_0021	Bahía Tortugas	-	0	0
HF_0022	Bahía Tortugas	-	13	0
HF_0023	Bahía Tortugas	-	9	0
HF_0024	Bahía Tortugas	-	0	0
HF_0028	Bahía Tortugas	-	7	0
HF_0029	Bahía Tortugas	-	5	0
HF_0030	Bahía Tortugas	-	3	0
HF_0009e	Buzos y Pescadores	WS	0	0
HF_0136	Buzos y Pescadores	WS	0	0
HF_0142	Buzos y Pescadores	-	0	0
HF_0144	Buzos y Pescadores	WS	3	0
HF_0001e	California	WS	87	4
HF_0073	California	-	21	1
HF_0196	California	-	0	0
HF_0197	California	-	0	0
HF_0039	Emancipación	WS	1	0
HF_0040	Emancipación	WS	4	0
HF_0042	Emancipación	WS	267	1
HF_0043	Emancipación	WS	123	2
HF_0046	Emancipación	WS	107	2
HF_0047	Emancipación	WS	71	3
HF_0048	Emancipación	WS	0	0
HF_0049	Emancipación	WS	13	0
HF_0099	Emancipación	WS	2	0
HF_0100	Emancipación	WS	766	1
HF_0101	Emancipación	WS	48	1
HF_0102	Emancipación	WS	343	2
HF_0116	Emancipación	WS	446	2
HF_0153	Emancipación	WS	1	0
HF_0158	Emancipación	WS	1	0
HF_0032	La Purísima	-	1	0
HF_0037	La Purísima	-	0	0
HF_0038	La Purísima	-	0	0
HF_0104	La Purísima	WS	26	0
HF_0105	La Purísima	-	9	0
HF_0079e	Leyes de Reforma	WS	763	33
HF_0089	Progreso	WS	38	0
HF_0097	Progreso	-	3	0
HF_0179	Progreso	-	5	1
HF_0080	Punta Abreojos	-	0	0
HF_0083	Punta Abreojos	-	5	0
HF_0086	Punta Abreojos	-	0	0
HF_0087	Punta Abreojos	-	0	0
HF_0088	Punta Abreojos	-	0	0
HF_0189	Punta Abreojos	-	0	0
HF_0194	Punta Abreojos	-	0	0

Annex 13: *Candidatus* Xenohaliotis californiensis *16S rDNA* [A] and *ftsZ* [B] haplotype distribution found in *H. corrugata* (HC) and *H. fulgens* (HF) abalone in the Pacific coast of Baja California Sur. Localities (names refer to fishing cooperatives): 1= La Purísima, 2= Buzos y Pescadores, 3= Bahía Tortugas, 4= Emancipación, 5= California, 6= Leyes de Reforma, 7= Progreso, 8= Punta Abreojos

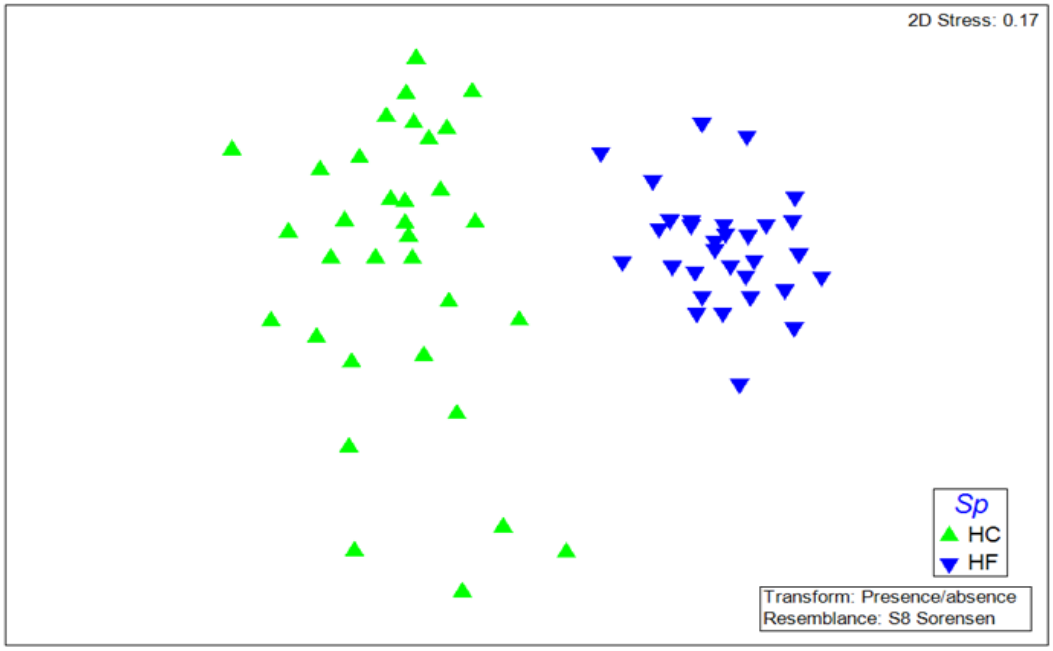




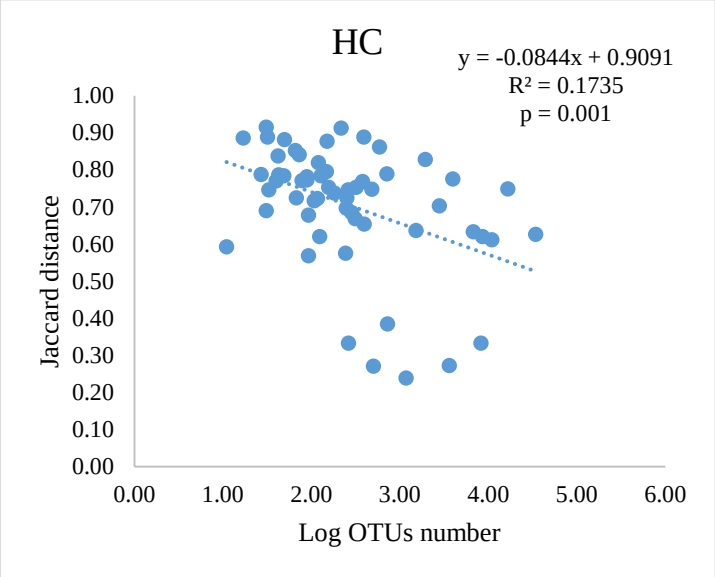
Annex 14. OTUs showing differential abundances between yellow (HC) and blue (HF) abalone with statistical and biological significance obtained from LEfSe analyses. OTUs are ranked according to its effect size (LDA score).

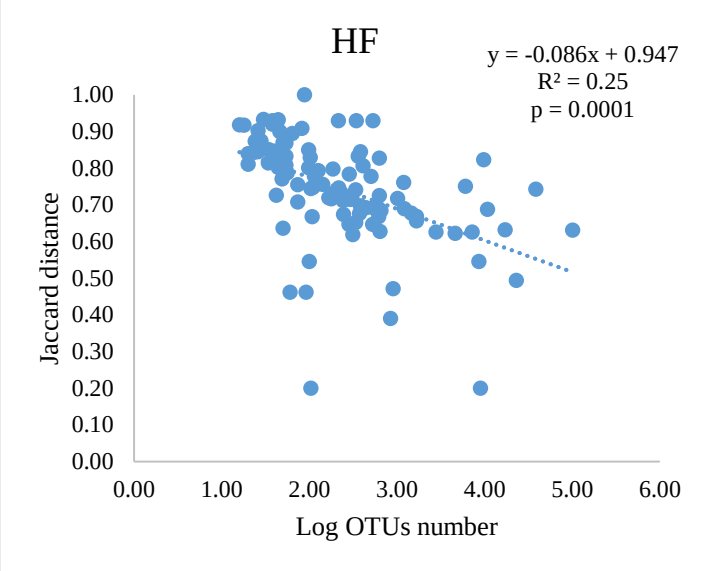


Annex 15. Non-parametric multidimensional scaling (MDS) of Sorensen similarity index based on presence and absence of OTUs assembled at 97% similarity cut-off of the gut microbiota of *Haliotis fulgens* (HF) and *H. corrugata* (HC).



Annex 16. Regression of the average Jaccard distance based on genetic groups of bacterial OTUs (10% similarity) as a function of their abundance –log scale – (HF: *Haliotis fulgens*, HC: *Haliotis corrugata*).





Annex 17. Ranks of the KO genes according to their KO count number (log scale) obtained with the script *categorize by function* in PICRUSt.