



Universidad Nacional del Comahue  
Centro Regional Universitario Bariloche

**Diversidad Molecular de Bacterias y  
Virus Marinos en Aguas  
Superficiales de Bahía Engaño  
(Chubut, Argentina)**

Trabajo de Tesis para optar al Título de Doctor en  
Biología

**Lic. Andrea Y. Calvo**

Director: Dr. Leandro R. Jones  
Co-Director: Dra. Bioq. Julieta M. Manrique

**2019**

---

# Agradecimientos

A mis directores, Dr. Leandro Jones y Dra. Julieta Manrique, por permitirme realizar este trabajo bajo su dirección y por acompañarme en este trayecto del doctorado. Gracias por brindarme las herramientas metodológicas y el material necesario para concretar este trabajo. Les agradezco por la predisposición, el tiempo y el esfuerzo invertido en mi formación, desde la elaboración del proyecto, y en cada una de los pasos que siguieron en todo este trayecto. Gracias por inculcarme a ser rigurosa y crítica en cada etapa de trabajo. Les agradezco a ambos, porque me han enseñado a respetar y querer lo que hago, me abrieron el camino para conocer y abrazar el mundo de la Ciencia. Con miedo a escribir una cursilería, debo agradecerles a ambos por hacerme crecer no sólo como profesional, sino como persona. Sin dudas no soy la misma que empezó este camino de doctorado.

A mis compañeras de Doctorado, por la compañía y el tiempo compartido.

Al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por el apoyo económico para la realización de este trabajo. Al Laboratorio de Virología y Genética Molecular de la Universidad Nacional de la Patagonia San Juan Bosco por proporcionarme el espacio de trabajo.

A mi familia, porque sin ella este trabajo no se podría haber concretado.

# Índice general

<b>Índice de Tablas</b>	<b>IV</b>
<b>Índice de Figuras</b>	<b>V</b>
<b>Resumen</b>	<b>1</b>
<b>Abreviaturas</b>	<b>7</b>
<b>1. Introducción</b>	<b>9</b>
1.1. Microorganismos Marinos . . . . .	10
1.1.1. Importancia de los microorganismos en los sistemas acuáticos	13
1.2. Métodos independientes de cultivo . . . . .	16
1.2.1. Concepto . . . . .	16
1.2.2. Características del gen ARNr 16S . . . . .	19
1.2.3. Aproximaciones metagenómicas . . . . .	22
1.3. Objetivos . . . . .	27
<b>2. Materiales y Métodos</b>	<b>28</b>
2.1. Zona de estudio y Muestreo . . . . .	28
2.2. Aislamiento de metagenomas . . . . .	31
2.3. Análisis genéticos . . . . .	33
2.3.1. Extracción de ADN . . . . .	33
2.3.2. Amplificación mediante PCR . . . . .	34
2.3.3. Metabarcoding . . . . .	38
2.3.4. Secuenciación directa . . . . .	43
2.3.5. Electroforesis en gradiente desnaturalizante (DGGE) . . . . .	44
2.4. Análisis de secuencias . . . . .	44
2.4.1. Curado de secuencias crudas . . . . .	44

---

2.4.2. Asignación taxonómica . . . . .	46
2.4.3. Análisis evolutivos . . . . .	47
2.5. Aislamiento en cultivo . . . . .	49
2.6. Preparación de bacterias competentes . . . . .	49
2.7. Análisis estadísticos . . . . .	50
<b>3. Resultados</b>	<b>51</b>
3.1. Análisis del bacterioplancton . . . . .	51
3.2. Secuencias locales no clasificables . . . . .	60
3.3. Análisis complementarios de picoeucariotas . . . . .	88
3.4. Análisis virológicos . . . . .	90
<b>4. Discusión y Conclusiones</b>	<b>97</b>
4.1. Discusión . . . . .	97
4.2. Conclusiones . . . . .	103
<b>A. Anexo: <i>Primers</i></b>	<b>105</b>
<b>Bibliografía</b>	<b>127</b>

# Índice de tablas

1.1. Clasificación de los organismos acuáticos según criterio taxonómicos.	11
1.2. Clasificación del plancton según su tamaño . . . . .	12
1.3. Técnicas utilizadas en el estudio del ADN ambiental . . . . .	24
2.1. Esquema de muestreo y análisis . . . . .	30
3.1. Diversidad del gen del ARNr 16S en la zona de estudio . . . . .	55
3.2. Regresión entre distancias genéticas y ambientales (dbLM) . . . . .	59
3.3. Correlación entre distancias genéticas y ambientales (test de Mantel)	60
3.4. OTUs no clasificables (uOTUs) de la zona de estudio . . . . .	71
3.5. Secuencias y OTUs no clasificables: Sanger Vs. Pirosecuenciación .	73
3.6. uOTUs: similitudes con GenBank . . . . .	74
3.7. uOTUs: similitudes con microorganismos cultivados . . . . .	76
3.8. uOTUs: similitudes con material tipo . . . . .	78
3.9. uOTUs cosmopolitas . . . . .	80
3.10. Análisis extendidos de las uOTUs cosmopolitas: resumen . . . . .	86
3.11. Variantes de <i>Ostreococcus virus</i> (OV) de la zona de estudio . . . . .	91
3.12. Dataset de <i>Prasinovirus</i> . . . . .	92
3.13. Variabilidad en <i>networks</i> de OV . . . . .	95
3.14. Anagénesis y sorporte filogenético en clados de OV . . . . .	95
A.1. Secuencias correspondientes a los <i>primers</i> utilizados en este trabajo.	105

# Índice de figuras

1.1. Loop microbiano . . . . .	14
1.2. Diversidad global del bacteriplancton . . . . .	16
1.3. Estado del conocimiento de la diversidad microbiana . . . . .	18
1.4. Estructura del ribosoma procariota . . . . .	19
1.5. Estructura secundaria del ARN ribosomal . . . . .	21
1.6. Métodos independientes de cultivo . . . . .	22
1.7. Estrategias independientes de cultivo utilizadas en este trabajo. . .	26
2.1. Zona de estudio . . . . .	29
2.2. Procesamiento de las muestras . . . . .	32
2.3. Diseño de PCR en gradiente . . . . .	35
2.4. PCRs en gradiente . . . . .	36
2.5. Mapa del vector de clonado . . . . .	40
2.6. <i>Screening</i> de clones recombinantes . . . . .	42
2.7. Curado de secuencias Sanger . . . . .	45
3.1. Filogenia de secuencias Sanger . . . . .	52
3.2. Filogenia de Pirosecuencias . . . . .	53
3.3. Detección molecular de cianobacterias . . . . .	54
3.4. Análisis mediante DGGE (gen del ARNr 16S) . . . . .	57
3.4. Análisis mediante DGGE (gen del ARNr 16S) . . . . .	58
3.5. Relación entre distancias genéticas (16S) y ambientales . . . . .	59
3.6. Análisis de OTUs no clasificables (uOTUs): Diagrama de flujo . . .	61
3.7. Análisis filogenético directo (AFD) de uOTUs . . . . .	64
3.8. Ubicación global de secuencias no clasificables . . . . .	66
3.9. Filogenia de uOTUs y 6194 secuencias de referencia . . . . .	69
3.10. Filogenia de piceucariotas fotosintéticos . . . . .	89
3.11. Detección molecular de virus de piceucariotas . . . . .	91

---

3.12. Filogenia de <i>Prasinovirus</i> de la zona de estudio y otras regiones . . .	96
---	----

# Resumen

El término microorganismos abarca las formas de vida cuyos tamaños no superan los 200  $\mu\text{m}$ . Los microorganismos marinos presentan una gran abundancia. Los virus son el componente numéricamente más abundante, mientras que las bacterias pueden llegar a representar alrededor del 90 % de la biomasa total. Gracias a esta abundancia, tienen una participación relevante en funciones vitales del ecosistema marino, como el bucle de carbono y el cortocircuito viral, las cuales pueden alcanzar magnitudes biogeoquímicas. Sin embargo, se cuenta con un conocimiento limitado de la identidad de los microorganismos marinos debido a que la gran mayoría de estos no pueden cultivarse. Actualmente se acepta que la descripción de los microbiomas constituye un paso fundamental para la futura comprensión de los ecosistemas marinos. Si bien la plataforma continental en el Atlántico sudoccidental es una de las mayores y más productivas del Mundo, los estudios microecológicos son escasos, en particular para la región patagónica.

En el presente trabajo de tesis se estudió por primera vez la diversidad del gen 16S en muestras de picoplancton de agua superficial de Bahía Engaño (Chubut, Argentina). Los datos obtenidos (2218 secuencias curadas) revelaron que el microbioma de la zona de estudio es dominado por unos pocos grupos mayoritarios (*SAR11*, *Rhodobacterales*, *Acidimicrobiales* y *Oceanospirillales*), tal como se ha observado en ecosistemas similares del hemisferio norte. Así mismo, la abundancia de los diferentes taxa detectados presentó variaciones temporales. Se observó también una cantidad importante ( $\sim 10\%$  de las *OTUs*) de secuencias no clasificables. No se observaron secuencias correspondientes a cianobacterias pero, interesantemente, se observaron muchas secuencias correspondientes a cloroplastos de picoeucariotas fotosintéticos (*PF*).

La presencia de células de *PF* en las muestras se corroboró, por un lado, mediante un análisis del gen 18S. Se amplificó dicho gen a partir del picoplancton de una de las muestras y los amplicones obtenidos fueron clonados en *E. coli* mediante un vector *pGem*. En concordancia con el análisis del gen 16S, de un total



---

de 35 clones moleculares secuenciados 32 presentaron secuencias pertenecientes a una especie de *Ostreococcus* (*Mamiellophyceae*). Estos resultados fueron además consistentes con experimentos de aislamiento microbiológico los cuales permitieron obtener en cultivo dos especies de *PF* de la zona de estudio, las cuales correspondieron a una especie de *Ostreococcus* y una de *Micromonas* (*Mamiellophyceae*) según se determinó mediante observación microscópica y análisis genéticos.

La presencia de *PF* también fue estudiada en forma indirecta mediante experimentos orientados a la detección y caracterización de sus virus. Por un lado se realizó un análisis por clonado de secuencias del gen *pol* de la familia *Phycodnaviridae* (la cual contiene el género *Prasinovirus* que infecta *PF*) basado en PCR con iniciadores universales. Este análisis, que fue realizado a partir de una muestra de picoplancton, reveló exclusivamente secuencias de *Ostreococcus Virus* (*OV*). Interesantemente un análisis detallado de dichas secuencias reveló que las secuencias de la zona de estudio corresponden a un clado de *OV* que no ha sido observado en otras regiones del planeta. El análisis, que incluyó secuencias de otras partes del mundo, también reveló que la especie presenta siete linajes que podrían corresponder a especies crípticas. La detección de *Prasinovirus* (PCR + secuenciación en extractos de picoplancton y concentrados de filtración tangencial) se extendió al resto de las muestras estudiadas (13 muestras tomadas en el transcurso de 2 años), lo cual reveló la presencia de *OV* o *Micromonas Virus* en  $\sim 50\%$  de estas.

La variación temporal de la composición taxonómica revelada por el análisis de las bibliotecas génicas del gen 16S fue reevaluada mediante electroforesis en geles con gradiente desnaturante (*DGGE*; del inglés *denaturing gradient gel electrophoresis*). Estos análisis fueron aplicados a 9 muestras colectadas con una frecuencia aproximadamente mensual. Interesantemente, los electroferogramas de algunos meses temporalmente contiguos presentaron variaciones sustanciales. Además, se corroboró la existencia de una variación estacional.

Las secuencias no clasificables de la zona de estudio fueron analizadas en forma detallada utilizando una aproximación filogenética, a través de metodologías desarrolladas en este trabajo. Éstas permitieron mejorar en gran medida las asignaciones taxonómicas y demostraron ser capaces de promover el descubrimiento de nuevos grupos. Interesantemente, se observó que seis de las *OTU* no clasificables (*uOTUs*) locales corresponden a grupos cosmopolitas desconocidos; esto es, fueron significativamente similares a secuencias no clasificables de otras regiones del mundo. Los análisis filogenéticos realizados en el marco de 6418 secuencias de referencia mostraron que estas *uOTUs* cosmopolitas posiblemente correspondan a cinco linajes de *Alphaproteobacteria* y a un clado desconocido de *Gracilibacteria*.

---

El presente trabajo constituye el primer estudio de la diversidad del picoplankton y femtoplankton de la zona de estudio.

---

## Summary

The term microorganisms comprise all the organisms with dimensions up to 200  $\mu\text{m}$ . Marine microorganisms, particularly bacteria and viruses, have a great abundance within these ecosystems: while viruses are the most abundant component, bacteria represent 90 % of the total biomass of the microorganisms that are present in marine ecosystems. Taking into account its great abundance, it is evident their relevant participation in vital functions of this ecosystems, reaching those of biogeochemical magnitude as the microbial loop and carbon pump, which result in the transfer and sinking of large amounts of organic carbon. While it is known that the ocean is the largest and oldest ecosystem of the entire planet, the abundance, diversity, structure and interaction of the communities that are part of them are still in the process of being unraveled.

Having knowledge about the different aspects of the microorganisms that govern ecosystems is of great relevance since it allows generating accurate assessments and the generation of policies that allow the rational and sustainable use of natural resources. In this sense, although the continental shelf in the southwestern Atlantic is one of the largest and most productive in the world, microecological studies are scarce, particularly for the Patagonian region. Additionally, the microecology of coastal regions in general has received less attention than that of the oceans. Previous work indicates there would be differences between both types of environment, however, it is necessary to extend these studies to other regions and latitudes.

In this thesis work we studied the microbial diversity in superficial sea waters near the mouth of the Chubut River in the Argentine Patagonia. The viral communities that are involved in the microecology of the nanoplankton marine communities were also studied as well as the fluctuations in the taxonomic composition.

Thirteen samples collected at different times were analyzed at a sampling point in the Bahía Engaño region. The samples were processed in order to isolate the picoplankton environmental genomic DNA. The studies involved the use of techniques to obtain bacterial 16S ribosomal gene sequences: generation of gene libraries by cloning along with Sanger sequencing and pyrosequencing, 18S ribosomal gene sequences and viral sequences, which were analyzed by various bioinformatics techniques. For the analysis of fluctuations in a monthly time scale a combination of community profiles techniques with statistical analysis and grouping and direct sequencing of amplicons were used.

Sequencing of ribosomal genes showed a substantial diversity dominated by

---

major groups: SAR11, Rhodobacterales, Acidimicrobiales and Oceanospirillales, together with low frequency taxa: Bacillales, Xanthomonadales, SAR86, OCS116, Rhizobiales, among others and the presence of undescribed groups (*uOTUs*). In addition, the results indicated that the taxonomic composition presents temporary fluctuations. Interestingly, the analyzes indicated significant amounts of 16S gene sequences of photosynthetic picoeucariote chloroplasts (PP, Mamiellophyceae Family). In contrast, cyanobacterial sequences were not detected. These results led to a more detailed study through the analysis of eukaryotic environmental sequences, isolation in culture, the molecular detection of cyanobacteria and their viruses by means of specific primers, and the identification of PP viruses by PCR and direct amplicon sequencing. Analysis of the environmental 18S eukaryotic sequences indicated that, during the summer season, 86 % of the library belonged to local PP from *Ostreococcus* species. Culture Isolation studies also confirmed the presence of this species of *Mamiellophyceae* during summer, while they revealed that during the winter season time the predominant species corresponds *Micromona* sp. Complementary virological analyzes resulted in a confirmation of the direct analyzes results described so far, both in the cellular fraction plus in the concentrates of the samples by tangential filtration of seven of the samples, the presence of *Ostreococcus Virus* and *Micromona Virus* was evidenced. When temporal variations in the composition were analyzed at a shorter time scale it could be demonstrated that the bacterial community can vary substantially between some months of the year, indicating a marked seasonal variation.

As mentioned above, the evaluation of the bacteriome from the 16S ribosomal gene showed the presence of undescribed local groups, unclassifiable *OTUs* or *uOTUs*. Despite contributing significantly to the diversity of the ecosystems, rare groups have received very little attention and have been poorly studied. In this thesis work, a detailed analysis of *uOTUs* was carried out using a phylogenetic approach, through methodologies developed in this work. These methodologies made possible to greatly improve taxonomic affiliations and proved to be capable of promoting the discovery of new groups. In particular, in this work they showed that several of the local *uOTUs* correspond to cosmopolitan groups, possibly belonging to five Alphaprotobacteria lineages and to an unknown Gracilibacteria clade.

Considering the abundance of PP and its viruses in the studied microbiome, the viral community was characterized by comparison with viral sequences from other parts of the world. The phylogenetic analyzes of the viral sequences showed a high similarity with each other, as well as consequences of *Ostreococcus* virus from other regions of the world. The results would indicate that the sequences of

---

*Ostreococcus* virus of the region under study correspond to a single lineage.

The present work constitutes the first study of the diversity of picoplankton and femtoplankton in the study area.

# Abreviaturas

A: adenina

ADN: ácido desoxirribonucleico

AFD: Análisis Filogenético Directo

ARN: ácido ribonucleico

ATP: adenosin trifosfato

BLAST: *Basic Local Alignment Search Tool*

C: citocina

CRISPR: *Clustered regularly interspaced short palindromic repeats*

c.s.p.: cantidad suficiente para

CTAB: bromuro de cetil trimetil amonio

ddATP:didesoxiadenina trifosfato

ddCTP:didesoxicitocina trifosfato

ddGTP:didesoxiguanina trifosfato

ddNTP: didesoxinucleótido trifosfato

ddTTP:didesoxitimina trifosfato

DGGE: *Denaturing gradient gel electrophoresis*

dNTP: desoxinucleótido trifosfato

DOM: dissolved organic matter

DTT: ditioneitol

EDTA: ácido etilen diamino tetracético

*e.g.: exempli gratia*

*et al.: et alia*

FFT: *fast fourier transform*

G: guanina

*i.e.: id est*

IPTG: isopropil tio-galactósido

kb: kilobases

kPa: kiloPascal

---

LB: Luria-Bertani  
M: molar  
ml: mililitros  
mM: milimolar  
MV: máxima verosimilitud  
N: normal  
ng: nanogramo/s  
NNI: *nearest neighbor interchange*  
O: Oeste  
OTU: *operational taxonomic unit*  
pb: pares de bases  
PCR: *Polymerase Chain Reaction*  
PE: Parsimonia Estadística POM: *particulate organic matter*  
r.p.m.: revoluciones por minuto  
rARN: ácido ribonucleico ribosomal  
S: Sur  
sp.: especie  
SPR: *subtree pruning regrafting*  
T: timidina  
TBE: Tris-borato-EDTA  
U: unidades  
uOTU: OTU no clasificable  
uSEQ: secuencia no clasificable  
V: voltios  
WNBC: clasificador naïve bayesiano de Wang (del inglés *Wang's Naïve Bayesian Classifier*)  
X-gal: bromo-cloro-indolyl-galactopiranosido  
 $\mu$ l: microlitro/s  
 $\mu$ m: micrómetro/s  
°C: grados centígrados

# 1. Introducción

La vida en la Tierra se originó hace, aproximadamente, 3700 millones de años en el medio marino (Nutman et al., 2016). Desde ese entonces y hasta la actualidad los microorganismos marinos han tenido una importancia esencial en el moldeado de la atmósfera, lo cual ha permitido la evolución de la vida como la conocemos actualmente. Además, los microorganismos son partícipes fundamentales en las redes tróficas de los ecosistemas marinos dada su gran abundancia, participando activamente en los ciclos de fijación de nutrientes e interactuando con otros organismos (Bolhuis y Cretoiu, 2016).

Como se detalla más adelante, los métodos basados en ADN ambiental han revelado que la mayoría los microorganismos del planeta corresponden a especies que habían pasado desapercibidas durante mucho tiempo. En consecuencia, hoy sabemos que se cuenta con conocimientos limitados de la filogenia, metabolismo y ecología de la mayoría de los microbios del planeta (Lloyd et al., 2018). Actualmente, la descripción de los microbiomas marinos se considera fundamental para el posterior desarrollo de modelos aplicables al aprovechamiento y monitoreo de los correspondientes servicios ecosistémicos (Liquete et al., 2013; Grant et al., 2013; Peralta et al., 2014).

La plataforma continental de Argentina es la más extensa de América del Sur y una de las mayores del Mundo, con alrededor de 2 millones de km<sup>2</sup>, siendo una de las zonas marítimas más productivas del planeta (Parker et al., 1997; Dogliotti et al., 2014). Si bien su importancia es notable y existen antecedentes de estudios en la región patagónica (Alvarez et al., 1997; Peressutti et al., 2008; Cristóbal et al., 2011; Lozada et al., 2014; Calderoli et al., 2018; Espínola et al., 2018), no existen estudios independientes de cultivo en la región de Bahía Engaño. Esta región corresponde a la zona costera de la provincia del Chubut donde desemboca el río homónimo. La región de Bahía Engaño es claramente importante por diversas razones. La primera es que, como ya se mencionó, en ella se encuentra la desembocadura del Río Chubut, el cual no sólo es importante por su extensión



---



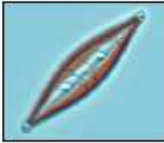


(más de 800 km), sino también porque en su desembocadura se asienta uno de los puertos pesqueros más importantes de la provincia, Puerto Rawson. Este puerto es de gran importancia socio-económica, ya que las especies capturadas en la zona (*e.g* merluza, langostino patagónico) son muy apreciadas en el mercado nacional e internacional. Asimismo, Bahía Engaño alberga el balneario Playa Unión, el cual experimenta una no despreciable actividad turística, sobre todo durante la época estival.

La comprensión de la dinámica del microbioma en las regiones costeras es fundamental para los modelos biogeoquímicos ya que esta fracción es la responsable de la alta productividad en dichas regiones (Lucas et al., 2011). Asimismo, en las costas que existen descargas de agua dulce, como en la desembocadura del Río Chubut, la productividad se ve reforzada por el aporte de materia orgánica y de nutrientes. Tanto los microorganismos como los recursos mencionados pueden ser trasladados hacia mar abierto, aumentando de este modo la influencia de las zonas productivas más allá de las costas (Fortunato et al., 2013; Satinsky et al., 2017).

## 1.1. Microorganismos Marinos

Los organismos que habitan en sistemas acuáticos, tanto marinos como continentales, se diferencian, entre otras cosas, por el modo de vida. De acuerdo a esto, es posible clasificar a los organismos en cuatro categorías. El *neuston* abarca los organismos que viven en la interfase agua-aire, el *plancton* incluye a los organismos que viven 'suspendidos' en la columna de agua, el *necton* a aquellos que nadan activamente y el *bentos* a los que viven en relación con el sustrato. Los organismos acuáticos también pueden clasificarse según su filiación en virioplancton (virus), bacterioplancton (bacterias), fitoplancton (algas) y zooplancton (animales, flagelados heterotróficos y ciliados) (Callieri y Stockner, 2002; Suttle, 2007), y según su tamaño (Sieburth et al., 1978) (Tablas 1.1 y 1.2).

Tabla 1.1: Clasificación de los organismos acuáticos según criterio taxonómicos.

	ORGANISMOS	
<b>Bacterioplancton</b>	 <p>a)</p>	
<b>Fitoplancton</b>	 <p>b)</p>	 <p>c)</p>
<b>Zooplancton</b>	 <p>d)</p>	 <p>e)</p>

Imágenes tomadas de: (a)[http://www.icm.csic.es/bio/outmol\\_s.htm](http://www.icm.csic.es/bio/outmol_s.htm);










(b)<http://hubpages.com/hub/Spirulina-The-Worlds-Healthiest-Food>;

(c)<http://www.la-farga.org/biosfera/Album/organismos01/original/diatomea.html>;

(d)<http://puertovallarta.olx.com.mx/daphnia-pulex-pulga-de-agua-alimento-vivo-peces-alevines-iid-134820375>;

(e)<http://supremaorden.blogspot.com/2010/10/zooplancton-el-nuevo-y-el-conocido.html>.

Tabla 1.2: Clasificación del plancton según su tamaño

	TAMAÑO / EJEMPLO	
<b>Femtoplanton</b>	0,02 a 0,2 $\mu\text{m}$	 Fago Synechococcus
<b>Picoplanton</b>	0,2 a 2 $\mu\text{m}$	   <i>Dunaliella</i> sp.    Picocianobacteria <i>Ostreococcus</i> sp.
<b>Nanoplanton</b>	2 a 20 $\mu\text{m}$	 <i>Gephyrocapsa oceanica</i>
<b>Microplanton</b>	20 a 200 $\mu\text{m}$	  <i>Eupodiscus radiatus</i> <i>Actinopterychus heliopena</i>
<b>Macroplanton</b>	200 a 2000 $\mu\text{m}$	  <i>Cyclops</i> sp.    Larva náyade

Imágenes tomadas de: (a) <https://bit.ly/2OImB2U>; (b) <https://bit.ly/2K5jIuv>; (c) <https://bit.ly/2Vh9nNk>; (d) <https://bit.ly/2Vb6ldn>; (e) <https://bit.ly/2YNhRO7>

---

El término **microorganismo** hace referencia a aquellos organismos que poseen un tamaño menor a 2000  $\mu\text{m}$ . Los microorganismos forman parte de todas las categorías mencionadas arriba, con excepción del megaplancton y el necton. Comprenden los virus y formas de vida pertenecientes a los tres Dominios de organismos celulares: *Bacteria*, *Eukarya* y *Archaea*. Cabe aclarar que algunos autores prefieren no referirse a los virus como *organismos*. En este trabajo serán incluidos dentro de dicho término dado que este uso no es raro en la literatura especializada y además facilitará la descripción de los resultados.

### 1.1.1. Importancia de los microorganismos en los sistemas acuáticos

Los océanos y mares cubren alrededor del 70% de la superficie de la Tierra, y los microorganismos existen en todos los ecosistemas marinos. Estos organismos pueden encontrarse desde formas de vida libre (Flombaum et al., 2013) hasta viviendo en el interior de otros organismos, e interactuando entre ellos y con el medio abiótico que los rodea. En relación a la interacción con el medio abiótico, y tal como se mencionó anteriormente, los microorganismos participan activamente en los procesos biogeoquímicos de muchos elementos, lo cual los convierte en actores críticos en el funcionamiento de estos dinámicos y complejos ecosistemas.

Para entender la participación de los microorganismos marinos en los ciclos de elementos, es importante destacar el concepto de *loop* microbiano que introdujeron Azam *et al.* en el año 1983 (Azam et al., 1983). Este concepto se aplica para describir cómo fluye la materia orgánica (tanto disuelta como particulada) proveniente de la producción primaria dentro de las redes tróficas de los microorganismos, y el papel de éstos en el ciclo de nutrientes. La materia orgánica es aprovechada por la fracción microbiana, la cual luego es reintroducida a las redes tróficas donde participan organismos macroscópicos (Fenchel, 2008) (Fig. 1.1).

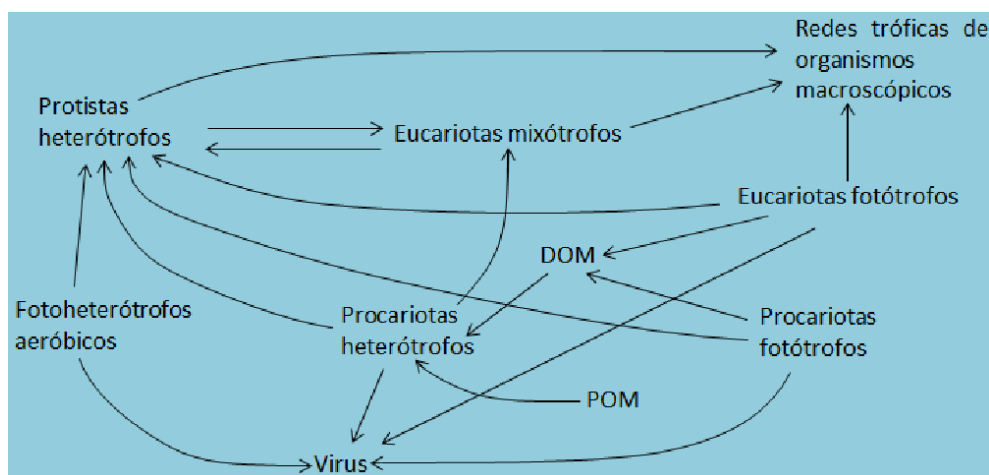


Figura 1.1: *Loop* microbiano en ecosistemas acuáticos, donde DOM es la materia orgánica disuelta (por sus siglas en inglés, *dissolved organic matter*) y POM es la materia orgánica particulada (*particulate organic matter*). Modificado a partir de Fenchel, 2008.

El componente procariota (fotótrofos, heterótrofos y fotoheterótrofos) y los virus dentro de estas redes tróficas conforman la principal vía de flujo de carbono. Esto radica en la gran abundancia que presentan. El componente viral es el más abundante en los ecosistemas acuáticos, alcanzando una densidad de entre  $1 \times 10^6$  y  $1 \times 10^{11}$  partículas por cada litro de agua de mar, y su biomasa total corresponde a aproximadamente el 5% de la biomasa de microorganismos oceánicos (Brussaard et al., 2016). Los virus marinos presentan un muy amplio rango de hospedadores, logrando infectar bacterias (Leigh et al., 2017; Kalatzis et al., 2017; Fang et al., 2019), archaeas (Krupovic et al., 2018), protistas (Weynberg et al., 2017; Coy et al., 2018), invertebrados (Laffy et al., 2018; Kawato et al., 2018; Barthélémy et al., 2019), peces (Breyta et al., 2016; Van Doorslaer et al., 2018; Crane y Hyatt, 2011), reptiles (Mejía-Radillo et al., 2019; Fichi et al., 2016) y mamíferos (Kluge et al., 2016; Rodrigues et al., 2019)).

Los virus, particularmente los que infectan microorganismos, afectan directamente la estructura de las comunidades microbianas mediante lisis celular. Dicha estructura se vería afectada en una disminución de la abundancia de linajes específicos de microorganismos. En el caso de las bacterias, se estima que los virus producen una mortalidad selectiva tan importante como los organismos de pastoreo (*i.e.* zooplancton, (Talmy et al., 2019)), mientras que en otros casos serían los responsables predominantes de la mortalidad microbiana (Lara et al., 2017). La

---

disminución de los linajes afectados de bacterias produciría fluctuaciones dinámicas en las poblaciones microbianas, tanto virales como bacterianas y eucarióticas. De este modo, mediante la lisis celular, los virus redireccionan el carbono disuelto y otros nutrientes al medio circundante; efecto conocido como *viral shunt* o cortocircuito viral (Weynberg, 2018).

Las bacterias son la población más numerosa de microorganismos marinos, alcanzando un total de  $1 \times 10^{29}$  células en el mar abierto (Overmann y Lepleux, 2016), representando un 90 % de la biomasa total de microorganismos oceánicos (Bar-On et al., 2018). Sin duda, su abundancia y su diversidad metabólica, fisiológica y bioquímica, les permite cumplir con un rol preponderante en ciclos biogeoquímicos (participando en la fijación del nitrógeno, en la oxidación del amonio y en la oxidación del metano [por ejemplo (Delmont et al., 2018; Klawonn et al., 2019; Kalyuzhnaya et al., 2019)] y también en las redes tróficas de los ecosistemas marinos.

Las características de tamaño y abundancia, sumadas a la gran capacidad de dispersión que presentan, ha llevado a pensar que los microorganismos marinos no reflejan un ensamble estructurado en comunidades locales, sino más bien, una distribución cosmopolita (Fenchel y Finlay, 2004). Sin embargo, estudios más exhaustivos han sugerido que existe un alto grado de endemismo en las comunidades de bacterias marinas (Ghiglione et al., 2012), aunque los grupos taxonómicos principales están presentes en todas las comunidades [Figura 1.2; (Pommier et al., 2007)]. Entre los grupos dominantes de las comunidades de bacterioplancton se encuentran: *Alfaproteobacteria*, *Gamaproteobacteria* y *Bacteroidetes* Farnelid et al. (2016); Ghosh (2018).

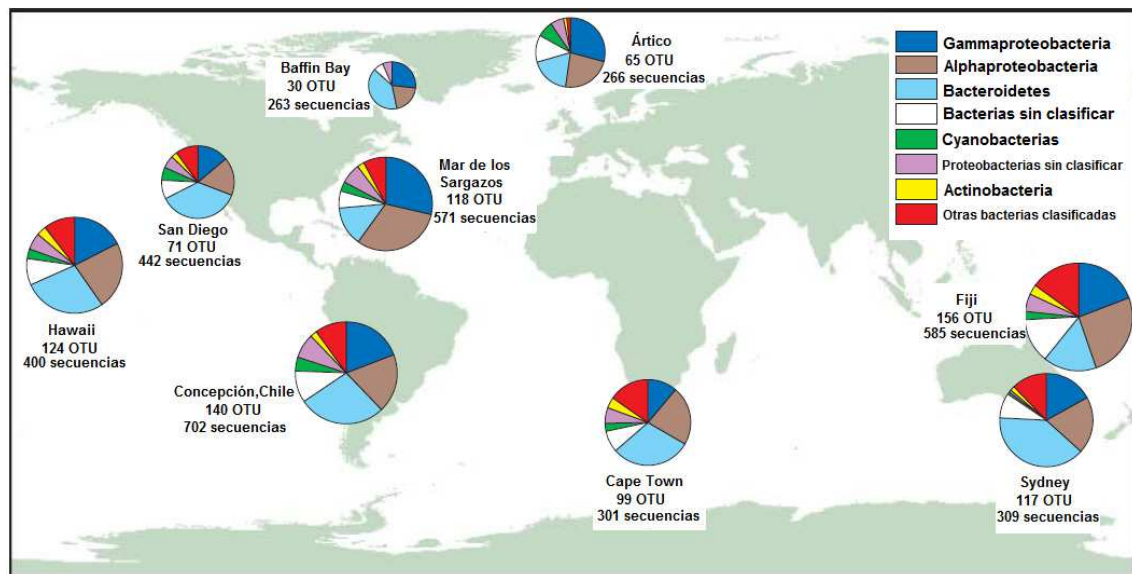


Figura 1.2: Muestreo a nivel mundial de comunidades planetónicas procariotas marinas. Para cada sitio muestreado se indican los principales grupos taxónomicos en los gráficos de torta, el número de OTUs y el número de clones secuenciados. Tomado y modificado de Pommier *et al.*, 2007.

## 1.2. Métodos independientes de cultivo

### 1.2.1. Concepto

El estudio de los microorganismos ha experimentado grandes transformaciones desde sus comienzos. Desde el invento del microscopio realizado por Antonie van Leeuwenhoek, la observación y el cultivo de microorganismos constituía la única herramienta disponible para obtener información acerca de la composición microbiana de cualquier ambiente. Sin embargo, hoy sabemos que sólo entre el 0,1 y 1% de los microorganismos presentes en los sistemas naturales han sido cultivados en condiciones de laboratorio (Solden *et al.*, 2016; Lloyd *et al.*, 2018).

En un principio, todo lo conocido sobre la vida de los microorganismos estaba basado en información proveniente exclusivamente de aquellos organismos cultivados. Sin embargo, los caracteres bioquímicos y los caracteres morfológicos no siempre resultan útiles para la identificación taxonómica de microorganismos. Asimismo, diversas observaciones indicaron que el cultivo no reflejaba la diversidad de microorganismos presentes en un determinado ecosistema (Amann *et al.*,

---

1995). Es por esto que la comunidad científica comenzó a notar que las técnicas clásicas no eran suficientes para describir una comunidad microbiana. En particular, la mayor parte de los microorganismos acuáticos no pueden aislarse en cultivo, por lo cual estos comenzaron estudiándose mediante técnicas indirectas, como por ejemplo, la epifluorescencia directa (Hobbie et al., 1977). Los resultados obtenidos por medio de dichas técnicas, lograron que comenzara a centrarse la atención en los microorganismos como componentes fundamentales de todos los sistemas naturales.

Hacia fines de la década del 70 y comienzo de los 80, el grupo de C. Woese demostró que es posible inferir la historia evolutiva de una diversa gama de microorganismos a niveles filogenéticos muy profundos (Fox et al., 1980). Estos estudios, además de revelar la división ancestral entre *Bacteria* y *Archaea*, sugirieron que es posible determinar la filiación taxonómica de un microorganismo sin la necesidad del cultivo. En el año 1985, Lane y colaboradores mostraron que, además, era posible generar secuencias del gen 16S sin la necesidad de aislar o clonar el gen, para lo cual utilizaron síntesis reversa a partir de ARN ribosomal obtenido directamente de cultivos celulares (Lane et al., 1985). Estos trabajos mostraron que podían generarse secuencias de diversos taxa con un conjunto limitado de iniciadores (*o primers*), sentando las bases para el desarrollo de los métodos independientes de cultivo (MIC).

Con el advenimiento de los métodos de amplificación de ácidos nucleicos fue posible identificar microorganismos prescindiendo de su cultivo en escala masiva. En microbiología marina, pueden mencionarse los trabajos seminales del grupo de S. Giovannoni, como por ejemplo (Giovannoni et al., 1990). Una de las ventajas que presentan los MIC es que, en comparación con los métodos tradicionales, permiten estudiar una mayor cantidad de taxa (Troussellier et al., 2017). Los métodos moleculares han dado lugar a mejoras en la comprensión del componente procariótico de la diversidad presente en los ecosistemas (Kirchman et al., 2010; Shan et al., 2015; Liu et al., 2018; Reza et al., 2018).

Los MIC han comenzado a llenar el vacío en el conocimiento de los microbiomas, con lo cual se abrieron nuevas perspectivas en la microbiología de sistemas ambientales, y particularmente de sistemas acuáticos. El desarrollo de las tecnologías de secuenciación, ha permitido dar cuenta de la gran abundancia de taxones desconocidos en los ecosistemas marinos (Lloyd et al., 2018). Además, la aplicación de nuevos enfoques que prescinden de cultivo han permitido encontrar nuevos sistemas CRISPR-Cas (Burstein et al., 2017), relacionar genes desconocidos a grupos de genes de referencia (*i.e.* genes conservados en organismos cultivados, pero



que presentan homólogos divergentes en sistemas naturales) (Lopez et al., 2015), como así también nuevas familias de genes ligados a funciones metabólicas específicas, como el caso de genes de bacterias que viven asociadas a esponjas marinas y que son responsables de la producción de la mayoría de los compuestos bioactivos vinculados a ellas (Lok, 2015).

La información obtenida hasta el momento a partir de secuenciación ambiental sugiere que más del 50 % de los genes presentes en muestras ambientales no sólo pertenecen a organismos desconocidos sino que además presentan funciones desconocidas (Figura 1.3; (Bernard et al., 2018)). Esto ha originado el concepto de *materia oscura microbiana*, ya que representa una cantidad abrumadora de microorganismos de los cuales no se tiene mayor conocimiento que de secuencias genómicas.<sup>1</sup>

		FUNCIÓN	
		Conocida	Desconocida
TAXÓN	Conocido	11,26%	13,91%
	Desconocido	22,19%	52,67% Materia Oscura Microbiana

Figura 1.3: Clasificación de las secuencias ambientales en base a su taxonomía y su función. Modificado a partir de Bernard *et. al* 2018.

<sup>1</sup>En Astronomía se conoce como *materia oscura* a un hipotético tipo de materia que no interactúa con la luz, pero interactúa gravitacionalmente, y que representaría el 90 % de toda la materia presente en el Universo (Bauer y Plehn, 2019).

## 1.2.2. Características del gen ARNr 16S

Los organismos celulares poseen complejos moleculares encargadas de la síntesis proteica denominados ribosomas. Estos están formados por dos subunidades, la subunidad mayor y la subunidad menor. Las dos subunidades tienen una estructura general que es similar dentro de los tres dominios biológicos (*Bacteria*, *Archaea*, *Eukarya*), presentando, además, diferencias características asociadas a cada uno de ellos. En los organismos procariontes la subunidad mayor (50S) está conformada por el ARNr ribosomal 5S y el ARNr ribosomal 23S y 34 proteínas ribosomales, mientras que la subunidad menor (30S) está conformada por el ARNr 16S y 21 proteínas ribosomales (Figura 1.4). Las secuencias de ARN ribosomal desempeñan un papel central tanto en el estudio de la evolución como en la ecología microbiana.

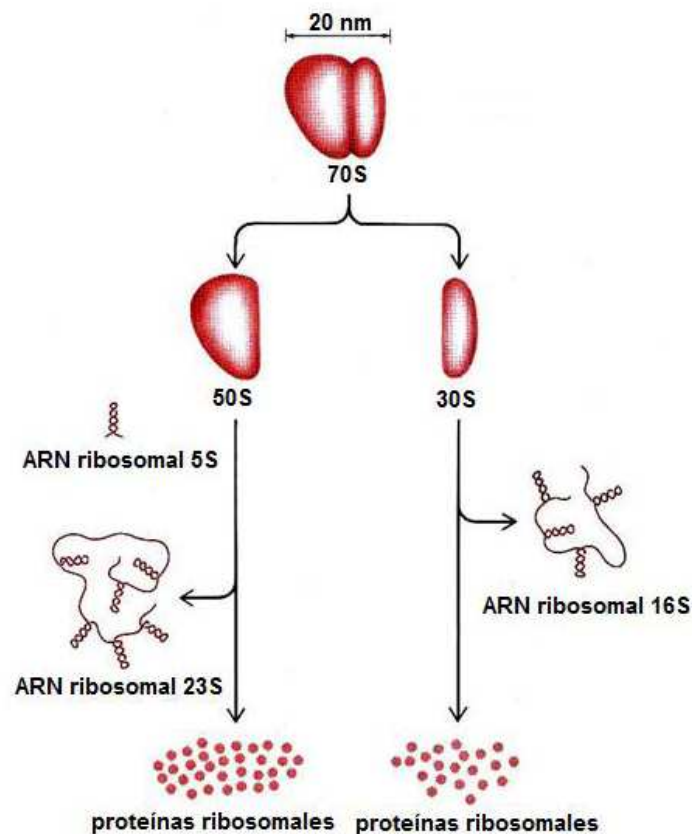


Figura 1.4: Esquema de la conformación de un ribosoma procarionte (Extraído y modificado de <http://www.uic.edu/classes/phys/phys461/phys450/ANJUM04/>).

---

El ARNr ha experimentado un cambio evolutivo constante en algunas de las regiones de la estructura primaria. Sin embargo, tanto otras regiones de la estructura primaria como las estructuras secundaria y terciaria, se mantienen relativamente constantes, con lo cual las funciones moleculares no se ven alteradas. La estructura secundaria de los ARNr consiste en el plegamiento de la secuencia, lo que forma estructuras denominadas tallos y bucles (Figura 1.5). Las secuencias de algunos bucles están conservadas en todas las especies cercanas de bacterias, debido a que cumplen funciones esenciales, mientras que las partes estructurales son principalmente variantes y específicas de una o más clases (Wang y Qian, 2009). Es por esto que estos plegamientos desempeñan un papel fundamental dentro de la estructura secundaria, como también en la estructura terciaria.

Los genes ARNr 16S combinan varias propiedades que los hacen especialmente adecuados para aplicaciones tan diversas. Primero, el hecho de que estén universalmente distribuidos permite realizar estudios filogenéticos y taxonómicos entre grupos muy divergentes (Wang & Qian, 2009). En segundo lugar, por ser parte de un núcleo de genes informativos, están débilmente afectados por transferencia horizontal de genes, por lo que sus relaciones proporcionan un marco sólido para evaluar cambios evolutivos en los linajes (Espejo y Plaza, 2018). Gracias a las características mencionadas, es posible realizar análisis comparativos de estas moléculas de manera muy efectiva. Así, los genes ARNr 16S son útiles para determinar relaciones filogenéticas, evaluar la diversidad presente en un ambiente determinado, y para detectar y cuantificar poblaciones específicas (Welsh et al., 2016; Bukin et al., 2019).

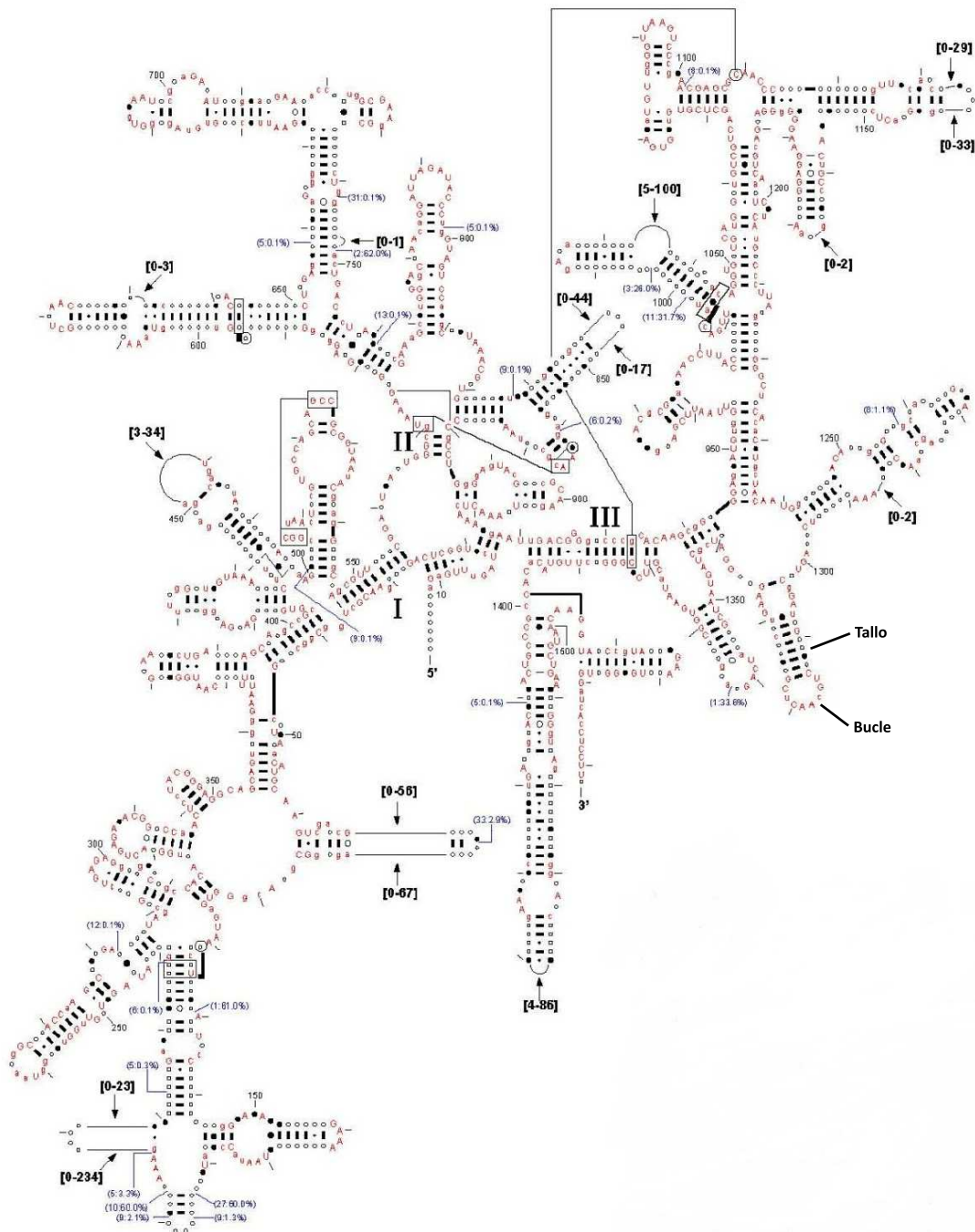


Figura 1.5: Estructura del ARNr de *Escherichia coli*. Se detallan las interacciones secundarias (líneas cortas) e interacciones terciarias (líneas largas). Las bases en mayúscula corresponden a posiciones conservadas en más del 98 % de las especies; las que están en minúscula corresponden a posiciones conservadas en el 90-98 % de las especies. Las posiciones indicadas con un punto negro corresponden a aquellas posiciones que presentan un 80-90 % de conservación, mientras que las indicadas con puntos grises corresponden a posiciones conservadas en el 90 % de las especies. Se indica a modo de ejemplo, un tallo y un bucle en la estructura del gen (Extraído y modificado de la página de la Universidad de Texas, <http://www.rna.icbm.utexas.edu>).

---

### 1.2.3. Aproximaciones metagenómicas

El metagenoma es el conjunto de todos los genomas que están presentes en una comunidad en un momento determinado. El estudio de la composición taxonómica de una comunidad microbiana puede abordarse desde distintas perspectivas, entre ellas el *metabarcoding* y la *metagenómica* (Figura 1.6).

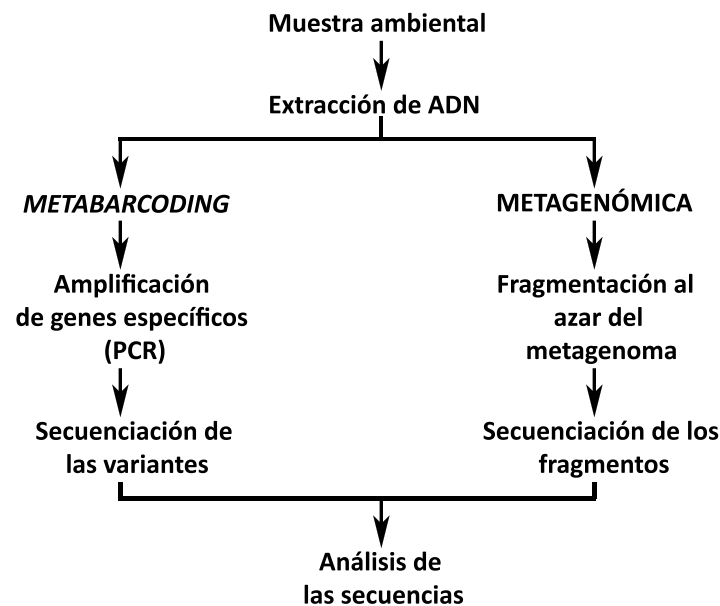


Figura 1.6: Esquema de los pasos principales de dos estrategias utilizadas en el estudio de metagenomas.

El *metabarcoding* consiste en aislar marcadores genéticos universales de una mezcla compleja de ácidos nucleicos, determinar sus secuencias e inferir la composición taxonómica de la muestra a partir de estas últimas 1.3. Actualmente, el aislamiento se realiza mediante PCR (*Polymerase Chain Reaction*) a partir del ADN aislado de la comunidad en estudio (ADN metagenómico). Con los productos de PCR es posible realizar una caracterización utilizando DGGE (del inglés *Denaturing Gradient Gel Electrophoresis*), técnica que permite separar fragmentos de ADN de la misma longitud, pero con diferencias en sus secuencias. Además se puede construir una biblioteca génica del marcador y posterior secuenciación mediante técnicas tradicionales [*i.e.* secuenciación de Sanger; (Giovannoni et al., 1990)], o se puede trabajar directamente sobre este marcador mediante técnicas de secuenciación masiva (Figura 1.7, pág. 26). Las bibliotecas génicas se construyen mediante clonado molecular. Esto implica ligar los fragmentos amplificados

---

en vectores de clonación, luego introducir estos vectores recombinantes en bacterias que, al multiplicarse, generen clones del fragmento que contienen. Luego, se analizan cada uno de los clones generados para identificar si tienen los fragmentos deseados (los cuales representan distintas variantes presentes en el metagenoma) y se secuencian mediante técnicas tradicionales (*i.e.* secuenciación de Sanger).

Mediante la aproximación conocida como *metagenómica* el ADN de los genomas que fueron aislados de la comunidad en estudio se fragmentan, luego dichos fragmentos son secuenciados y, finalmente, las secuencias obtenidas se ensamblan para realizar los análisis deseados.

La ventaja de estudiar una comunidad con una aproximación al metagenoma de dicha comunidad, es que los métodos utilizados (tanto en *Metabarcoding* como en *Metagenómica*) permiten un muestreo aleatorio de secuencias a partir del ADN de la muestra en estudio. Por otro lado, contar con el conocimiento previo de secuencias homólogas, disponibles en bases de datos, permite estudiar los grupos presentes en una comunidad desde el punto de vista taxonómico, realizar estudios comparativos en espacio y tiempo, y utilizar métodos históricos basados en el análisis filogenético de las secuencias obtenidas (Yau y Seth-Pasricha, 2019; Zhang y Gi, 2018; Manrique y Jones, 2017; Park et al., 2018). La metagenómica es más adecuada para el estudio funcional de los metagenomas que para su estudio taxonómico. Por ello en el presente estudio se utilizó metabarcoding.

**Secuenciación de Sanger.** En el año 1977 Sanger y colaboradores desarrollaron una nueva técnica de secuenciación (Sanger et al., 1977a,b), la cual se convertiría en el método de secuenciación exclusivo de los siguientes 30 años. Esta nueva técnica mejoró notablemente los métodos utilizados hasta el momento (Maxam y Gilbert, 1977; Sanger y Coulson, 1975), gracias a la reducción del tiempo de trabajo necesario. En la actualidad el método se realiza de forma automatizada, pero en sus comienzos era realizada manualmente.

Este método automático se basa en copiar la molécula de ADN a determinar utilizando entre los desoxirribonucleótidos trifosfato (dNTPs), didesoxirribonucleótidos trifosfato (ddNTPs) marcados con un fluoróforo distinto: un color para ddATP, otro para ddTTP, otro para ddCTP y otro para ddGTP. Cuando la cadena en crecimiento incorpora uno de estos ddNTP's finaliza la incorporación. De este modo se obtienen fragmentos que se diferencian en el largo por 1 nucleótido. Mediante electroforesis en poliacrilamida o capilar se separan los fragmentos, los cuales son excitados con un láser cuando llegan a un punto del gel o del capilar. De este modo, se detecta la fluorescencia que emite el último ddNTP del fragmen-

Tabla 1.3: Descripción de las técnicas utilizadas en aproximaciones metagenómicas.

TÉCNICA		DESCRIPCIÓN
Extracción y purificación de ácidos nucleicos		Implica la obtención de ácidos nucleicos de manera aislada del resto de los componentes celulares y el material particulado presente en la muestra.
PCR ( <i>Poymerase Chain Reaction</i> )		Facilita la obtención de un gran número de copias de un fragmento de ADN o ARN específico.
DGGE ( <i>Denaturing Gradient Gel Electrophoresis</i> )		Permite separar fragmentos de ADN de la misma longitud, pero con diferencias en sus secuencias mediante un gradiente de urea y formamida en un gel de poliacrilamida. Los amplicones migran por el gel hasta que llegan a una determinada concentración desnaturalizante a la cual las cadenas se separan parcialmente.
Clonado	Ligación	El fragmento amplificado mediante PCR se inserta en un vector.
	Transformación	Se inserta el material ligado dentro de una bacteria.
Secuenciación		Permite obtener la secuencia de nucleótidos de un fragmento de ADN o ARN.
Método de Sanger		La secuenciación se lleva a cabo utilizando desoxinucleótidos trifosfato (dNTP's) y didesoxinucleótidos trifosfato (ddNTP's). Al final de las reacciones, se obtienen cuatro familias distintas de cadenas de polinucleótidos sintetizadas, una para cada ddNTP ( <i>i.e.</i> ddATP, ddTTP, ddCTP y ddGTP).
Secuenciación Masiva		Posibilita la secuenciación de millones de bases por experimento, debido a que se utiliza un sistema de secuenciación en paralelo.

---

to. Las lecturas generadas de los nucleótidos en el orden en el que se encuentran en la molécula original se pueden visualizar mediante un cromatograma, que es el gráfico resultante del procesamiento de las imágenes colectadas de los geles de poliacrilamida o de la serie de señales emitidas en la secuenciación en capilares. La secuencia de nucleótidos que se asigna a partir del cromatograma se conoce como *base calling* y se realiza automáticamente por parte del *software* de los equipos de secuenciación. Sin embargo, este proceso puede presentar errores debidos a diversos factores como por ejemplo la existencia de una relación señal/ruido reducida (Fig. 2.7, pág. 45).

**Métodos de Secuenciación Masiva.** Los métodos de secuenciación masiva o de alta performance son plataformas que tienen la capacidad de generar una elevadísima cantidad de datos comparados con las técnicas tradicionales de secuenciación. En la actualidad existen diversas plataformas de secuenciación masiva (*e.g.* Illumina MiSeq y HiSeq, SOLiD) y todas se caracterizan por realizar análisis en donde hay interacciones muy complejas entre bioquímica, óptica de alta resolución e ingeniería informática (D'Amore et al., 2016). Es necesario tener en cuenta, no obstante, que para algunos autores la tecnología Sanger continúa siendo el *gold standard* de la biología molecular (Mu et al., 2016). Recientemente se han desarrollado plataformas que han aumentado en gran medida las prestaciones de la química Sanger<sup>2</sup>.

La pirosecuenciación es un método de secuenciación de alta performance que se basa en el principio de *secuenciación por síntesis* (Margulies et al., 2005). Cuando la ADN polimerasa incorpora un nucleótido durante la extensión de la cadena de ADN, se libera pirofosfato, lo cual desencadena una serie de reacciones que generan una señal lumínica detectable. Esta señal, es proporcional al número de nucleótidos incorporados. El método comienza con la preparación de la biblioteca del ADN metagenómico (fragmentos del metagenoma o amplicones obtenidos a partir del mismo), luego cada fragmento individual de la biblioteca se inmoviliza en un soporte sólido y es amplificado. Finalmente se realiza la pirosecuenciación y se ensamblan computacionalmente las secuencias obtenidas. La plataforma Roche 454, que fue la utilizada en este trabajo, se caracteriza por realizar la amplificación en una emulsión. El primer paso consiste en la ligación de los fragmentos (amplicones) a adaptadores. Luego se mezcla un aceite con una solución acuosa, la cual contiene los fragmentos de ADN, microesferas recubiertas de oligonucleótidos afines a los adaptadores antes mencionados, y los reactivos de PCR. Por agitación

---

<sup>2</sup>Ver por ejemplo: <https://www.youtube.com/watch?v=htoWfeQvysc>



se forma una emulsión, en cuyas microgotas (conocidas como microreactores) sólo hay una perla y un fragmento, donde ocurre la PCR. Posteriormente, se colocan las microesferas en microplacas que incorporan sólo una microesfera por pocillo (*o well*) y se lleva a cabo la secuenciación en paralelo que permite generar millones de bases en una corrida.

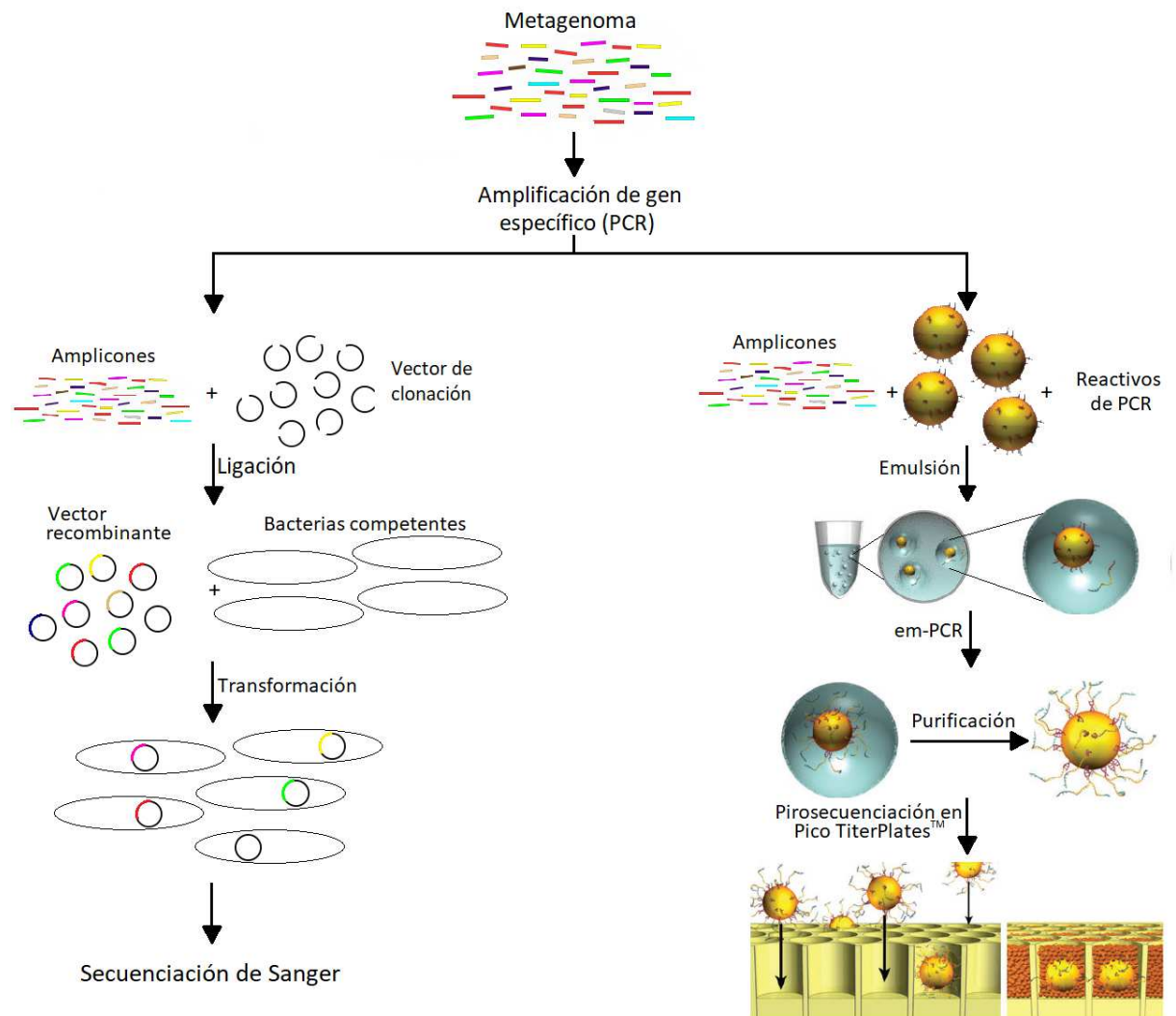


Figura 1.7: Estrategias independientes de cultivo utilizadas en este trabajo.

---

## 1.3. Objetivos

- **Generales**

El presente trabajo de Tesis Doctoral se enmarca en el objetivo general de describir la composición, estructura y dinámica de microorganismos marinos de aguas superficiales en la zona de Bahía Engaño.

- **Específicos**

1. Estudiar la diversidad molecular del bacterioplancton marino de las aguas superficiales del estuario del Río Chubut a lo largo de las estaciones del año y comparar estos datos con datos correspondientes a otras regiones del Mundo. Dependiendo de los resultados, se prevee estudiar otros componentes del picoplancton.
2. Estudiar la diversidad de bacteriófagos marinos de las aguas superficiales del estuario del Río Chubut a lo largo de las estaciones del año y comparar estos datos con datos correspondientes a otras regiones del Mundo. Dependiendo de los resultados, se prevee estudiar otros componentes del femtoplancton.

## 2. Materiales y Métodos

### 2.1. Zona de estudio y Muestreo

La Figura 2.1 y la Tabla 2.1 muestran la localización de la zona de estudio en Bahía Engaño (Chubut, Argentina) y la geolocalización exacta de cada una de las muestras analizadas en el presente trabajo. La toma de las muestras se realizó siempre durante la marea alta a aproximadamente 2 Km hacia el mar con respecto a la boca del río Chubut. La colección, de  $\sim 5$  L de agua superficial, se realizó desde una embarcación utilizando recipientes opacos, previamente lavados con una solución de HCl 0,1N seguida de tres enjuagues con agua destilada. Al momento del muestreo se registró la temperatura del agua. Las muestras fueron trasladadas al laboratorio y procesadas en forma inmediata.

Los datos de caudal del río fueron obtenidos de la Subsecretaría de Recursos Hídricos del Ministerio del Interior, Obras Públicas y Vivienda de la Nación.

No fue posible tomar muestras en los meses de Mayo, Junio y Octubre debido a que el estado del mar y/o el clima dificultaron la navegación.

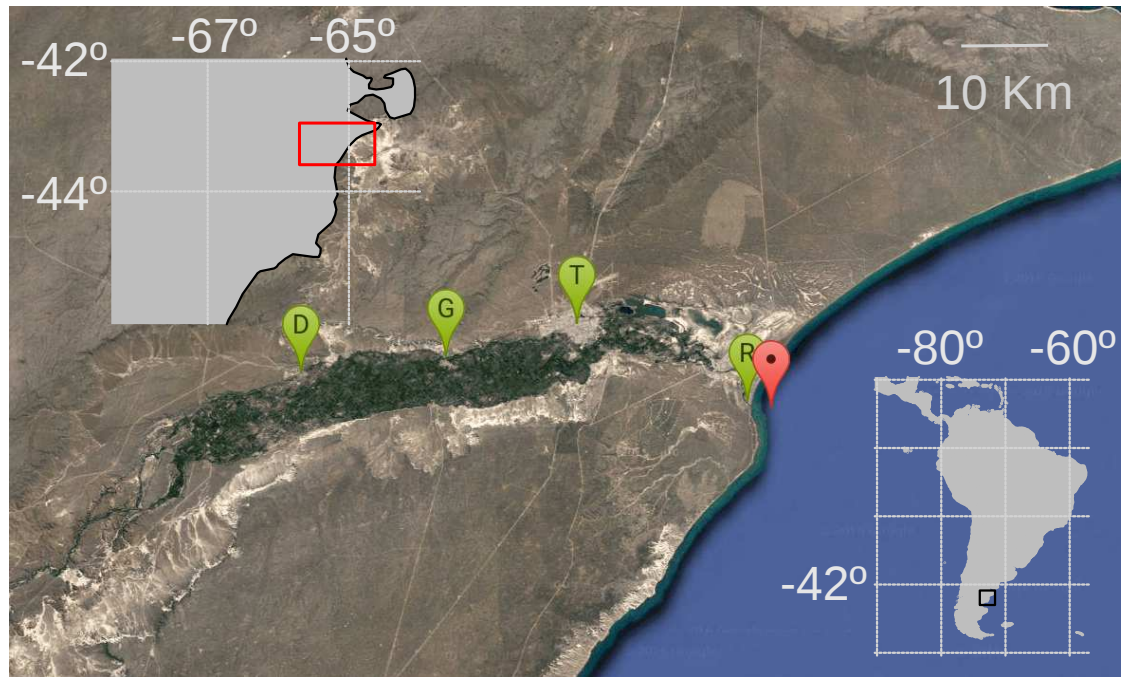


Figura 2.1: Localización de la zona de estudio. La marca roja con un punto indica la ubicación aproximada en la cual se tomaron las 13 muestras estudiadas en este trabajo. El resto de las marcas muestran la localización de centros urbanos en el curso inferior del Río Chubut (D Dolavon, G Gaiman, T Trelew, R Rawson). La imagen satelital fue descargada de la base de datos Google Earth mediante el programa RGoogleMaps (Loecher y Ropkins, 2015).

Tabla 2.1: Muestras estudiadas en esta tesis y análisis realizados a partir de cada una de las mismas.

MUESTRA	VER11	OTO11	INV11	PRI11	E12 <sup>4</sup>	F12 <sup>4</sup>	M12 <sup>4</sup>	A12 <sup>4</sup>	J12 <sup>4</sup>	Ag12 <sup>4</sup>	S12 <sup>4</sup>	N12 <sup>4</sup>	D12 <sup>4</sup>
Fecha	2/11	6/11	8/11	11/11	1/12	2/12	3/12	4/12	7/12	8/12	9/12	11/12	12/12
Estación	Ver	Oto	Inv	Prim	Ver	Ver	Ver	Oto	Inv	Inv	Inv	Prim	Prim
Coordenadas	43.35S- 65.01O	43.36S- 65.01O	43.35S- 65.00O	43.35S- 65.03O	43.35S- 65.01O	43.35S- 64.95O	43.33S- 65.01O	43.40S- 65.00O	43.35S- 65.00O	43.36S- 65.00O	43.36S- 64.98O	43.35S- 65.01O	43.35S- 65.01O
Caudal <sup>1</sup>	0.32	0.41	0.75	1.51	0.31	0.28	0.43	0.19	1.68	1.47	2.13	1.25	0.80
C. histórico <sup>2</sup>	0.25	1.12	1.69	1.54	0.44	0.25	0.21	0.26	1.5	1.69	1.71	1.54	0.85
Temp. (°C)	14	8	8	11	16	14	15	14	5	8	10	13	14
Análisis <sup>3</sup>	Sg, C, Pol, F, Pr	Sg, C, F, Pr	Sg, C, F, Pr	P, C, F, Pr	D, C, F, Pr, A	D, C, F, Pr, A	D, C, F, Pr, A	D, C, F, Pr, A	D, C, F, Pr, A	D, C, F, Pr, A	D, C, F, Pr, A	D, C, F, Pr, A	D, C, F, Pr, A

<sup>1</sup> Caudal índice: razón entre el caudal medio mensual y anual del correspondiente año de muestreo.

<sup>2</sup> Caudal histórico: razón, para cada mes, entre caudales índices medios mensuales y anuales a lo largo de un período de 74 años (1943 - 2015).

<sup>3</sup> *Sg* Secuenciado mediante clonado molecular seguido de secuenciación Sanger, *P* secuenciado mediante pirosecuenciación, *D* Analizado mediante DGGE, *C* Detección molecular (PCR específica) de cianobacterias, *Pol* Análisis virológico mediante clonado molecular y secuenciación Sanger de ADN polimerasa de virus de picoeucariotas, *F* detección de virus (Fagos) de cianobacterias, *Pr* Detección molecular de prasinovirus, *A* aislamiento en cultivo de picoeucariotas.

<sup>4</sup> Nombre que indica el mes y el año en el que se tomó cada muestra.

---

## 2.2. Aislamiento de metagenomas

Para enriquecer las muestras en ADN correspondiente al bacterioplancton, estas fueron tamizadas forzando su pasaje a través de filtros de diferente tamaño de poro (filtración por impacto; Fig. 2.2). El procedimiento permitió separar el picoplancton más una fracción del nanoplancton del resto de los elementos celulares del plancton. El remanente de las muestras luego de todo el proceso de filtración por impacto fue sometido a concentración mediante filtración tangencial con el objeto de aumentar la cantidad de genomas virales disponible por unidad de volumen, lo cual constituye un procedimiento estándar en el análisis virológico molecular independiente de cultivo. Los detalles técnicos de estos procedimientos se describen a continuación.

**Filtración por impacto** La filtración de cada una de las muestras se realizó aplicando vacío mediante una bomba de membrana. Dos litros de agua de cada muestra se filtraron de forma secuencial a través de filtros de policarbonato de distinto tamaño de poro aplicando una presión suave ( $\sim 20$  mmHg). Se comenzó con un filtro de  $20 \mu\text{m}$  de tamaño de poro y una vez filtrada la muestra se retiró el filtro y se lavó el equipo con agua destilada. El agua filtrada de la muestra se reservó en un recipiente limpio. Se colocó un filtro de  $10 \mu\text{m}$  de tamaño de poro en el equipo y se filtró nuevamente el volumen reservado del paso anterior. Este procedimiento se repitió con filtros de  $5$  y  $0,22 \mu\text{m}$  de tamaño de poro. De este modo, en cada filtro quedaron retenidas distintas fracciones planctónicas, siendo el filtro de  $0,22 \mu\text{m}$  donde quedó retenida la fracción de interés (Fig. 2.2). Los filtros fueron procesados inmediatamente o almacenados a  $-20^\circ\text{C}$  hasta su procesamiento.

**Filtración tangencial** El filtrado obtenido luego del pasaje por los filtros de  $0,22 \mu\text{m}$  (conteniendo el femtoplancton) fue concentrado mediante filtración tangencial utilizando un elemento de un punto de corte nominal de  $30$  KDa. Esto se realizó mediante un equipo Vivaflow (Sartorius) siguiendo las recomendaciones del fabricante, y una bomba peristáltica. Luego de esta etapa los concentrados fueron dialfiltrados con el objeto de eliminar la sal contenida en la muestra, y así evitar interferencias en los métodos moleculares utilizados posteriormente. La concentración fue de  $2$  l a  $2$  ml.

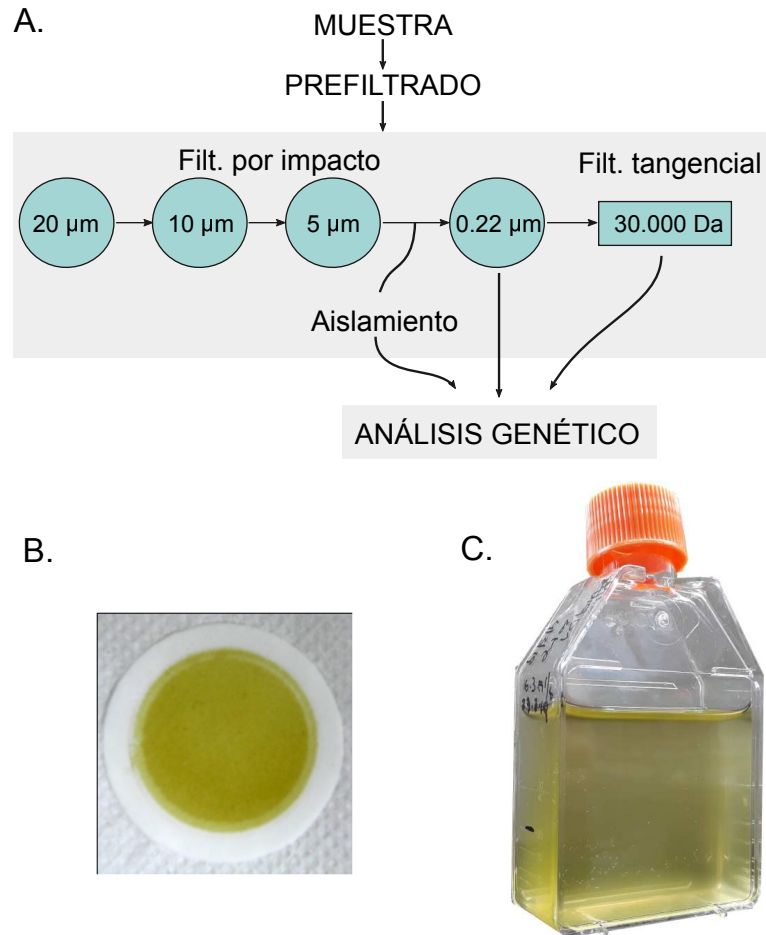


Figura 2.2: Esquema del procesamiento de las muestras (A), incluyendo la filtración por impacto y el concentrado de virus mediante filtración tangencial (sección 2.2), los análisis genéticos (sección 2.3) y el aislamiento microbiológico (sección 2.5). Los paneles (B) y (C) muestran el aspecto de un filtro de  $0,22 \mu\text{m}$  conteniendo picroalgas y un cultivo de *Micromonas* sp., respectivamente.

---

## 2.3. Análisis genéticos

### 2.3.1. Extracción de ADN

**Filtros.** El ADN de los microorganismos retenidos en los filtros de  $0,22 \mu\text{m}$  se extrajo mediante un protocolo basado en CTAB de la siguiente manera. Cada filtro se incubó a  $60^\circ\text{C}$  durante 40 minutos con 720  $\mu\text{l}$  de buffer de lisis CTAB (CTAB 2 %, NaCl 1,4M, beta-mercaptoetanol 0,2 %, EDTA 20mM, Tris-Cl pH 8 100mM) precalentado. El lisado obtenido se trató con un volumen de cloroformo/alcohol isoamílico en una relación 24:1, y se centrifugó a 30000 G durante 15 minutos. El sobrenadante se incubó con RNasa A a una concentración final de 10  $\mu\text{g}/\text{ml}$  durante 1 hora a  $37^\circ\text{C}$ . Luego de la incubación se agregó un volumen de isopropanol frío a la solución y se incubó a  $4^\circ\text{C}$  durante 1 hora, y posteriormente se centrifugó a  $4^\circ\text{C}$  a 30000xg por 30 minutos en una microcentrífuga refrigerada. El pellet se lavó con etanol 70 %, se secó y se resuspendió en agua de grado biología molecular. El ADN obtenido fue procesado inmediatamente o almacenado a  $-20^\circ\text{C}$  hasta su procesamiento. La cantidad de ADN obtenida fue calculada a partir de imágenes digitalizadas de geles de agarosa 0,8 %, mediante comparaciones densitométricas con un patrón de masa, utilizando el programa ImageJ2 (Rueden et al., 2017).

**Concentrados de filtración tangencial.** Las extracciones de los ácidos nucleicos presentes en los concentrados de femtoplankton fueron realizadas mediante protocolos estándar (Sambrook et al., 1989). Brevemente, se mezcló un volumen de concentrado (aproximadamente 750  $\mu\text{l}$ ) con un volumen de buffer de extracción (Tris 10 mM, EDTA disódico 1  $\mu\text{M}$ , SDS 0,5 %, proteinasa K 50  $\mu\text{g}/\text{ml}$  pH 7,8) y la mezcla se incubó durante 1 hora a  $56^\circ\text{C}$ . Luego se adicionó medio volumen de fenol pH 7,9 y medio de cloroformo:alcohol isoamílico (24:1). La mezcla fue homogenizada mediante vortex y se centrifugó a máxima velocidad en microcentrífuga durante dos minutos para separar la fase acuosa de la orgánica. Se recuperó la fase acuosa y se agregó un volumen de cloroformo-alcohol isoamílico (24:1), se homogeneizó y se separaron las fases nuevamente mediante centrifugación. La fase acuosa fue mezclada con 0,1 volumen de NaAc 3 M (pH 5,2) y 2,5 volúmenes de etanol 100 %. El material genético fue precipitado mediante centrifugación a máxima velocidad en microcentrífuga durante 30 minutos y el pellet obtenido fue lavado con 1 ml de etanol 70 %. El sobrenadante fue retirado, y el precipitado fue secado a temperatura ambiente y resuspendido en 10  $\mu\text{l}$  de agua de grado biología molecular. Las suspensiones fueron utilizadas inmediatamente o almacenadas a



---

-20°C hasta su utilización.

**Cultivos.** El ADN proveniente de cultivos (sección 2.5) fue obtenido mediante un procedimiento estándar similar al utilizado con los concentrados de filtración tangencial, salvo que se partió de un pellet celular obtenido mediante la aplicación de un pulso de 10 s a máxima velocidad en microcentrifuga a un volumen de 500  $\mu\text{L}$  de cultivo. El pellet fue lavado con agua de grado biología molecular mediante resuspensión, nuevamente pelleteado mediante un pulso de 10 s en microcentrifuga, y resuspendido en agua de grado biología molecular. Luego se procedió a la extracción de ADN.

### 2.3.2. Amplificación mediante PCR

Las reacciones de amplificación mediante PCR de los genes ribosomales 16S y 18S y el gen de la polimerasa viral fueron realizadas utilizando en todos los casos volúmenes finales de 50  $\mu\text{l}$  de reacción. Los desoxinucleótidos (dNTPs) fueron incorporados en las reacciones en una concentración final de 1 mM. Los iniciadores siempre fueron utilizados en una concentración final de 1  $\mu\text{M}$  cada uno. En todos los casos se utilizaron los *buffer* de reacción provistos por los fabricantes siguiendo las instrucciones de los mismos. Para los análisis de *metabarcoding* se utilizó la ADN polimerasa *Taq AccuPrime High Fidelity* (Invitrogen), mientras que el resto de los ensayos se realizaron mediante la enzima *Go Taq DNA Polymerase* (Promega).

En las corridas de electroforesis en geles de agarosa se utilizó el marcador de peso molecular 1kb Plus (Invitrogen), cuyas bandas tienen un rango de 15 kb a 100 pb.

Las secuencias de los *primers* utilizados en este trabajo se listan en la tabla A.1 (pág. 105).

**Optimización mediante PCR en gradiente** Uno de los parámetros más importantes que influyen en la especificidad, sensibilidad y rendimiento de una reacción de PCR es la temperatura de anillado o hibridación (o *annealing*) (Roux, 2009). Esto resulta particularmente importante en los casos en donde la amplificación se lleva a cabo utilizando un molde multi-copia, como los que se utilizan en el presente trabajo de tesis (Ishii y Fukui, 2001). Con el objeto de determinar la temperatura óptima de anillado para cada par de iniciadores utilizados en las diferentes reacciones de PCR que se describen, se realizaron análisis de amplificación en gradiente de temperatura, utilizando una cantidad fija de ADN molde. Como

se observa en la Figura 2.3, el diseño en gradiente consiste en establecer distintas temperaturas de anillado en un mismo ciclado, de acuerdo con la posición en la placa del termociclador. De este modo se selecciona un rango de temperatura de forma que, en cada pocillo, el paso de anillado de la reacción se lleva a una temperatura específica. Se realizan reacciones en paralelo manteniendo las mismas condiciones a excepción de las diferentes de temperatura de anillado, y luego son evaluadas por electroforesis en geles de agarosa.

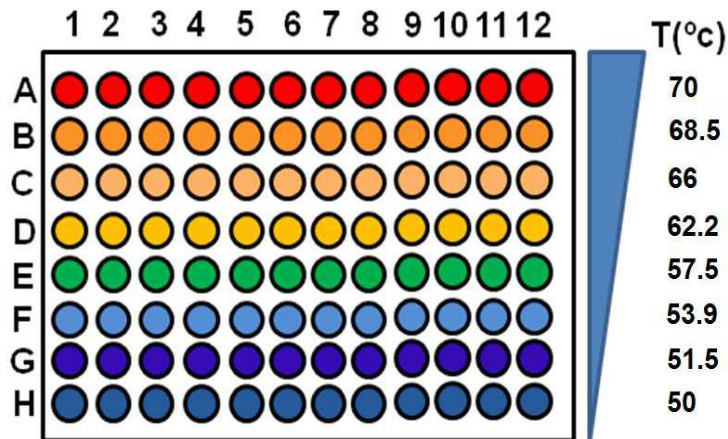


Figura 2.3: Esquema representativo de una placa de 96 pocillos indicando las condiciones de temperatura de anillado en los experimentos de gradiente térmico. Cada fila de pocillos adopta una temperatura de hibridación diferente, como se detalla en la parte derecha de la figura.

Para cada par de iniciadores se utilizaron las condiciones detalladas en la descripción específica de cada uno, utilizando una rampa de temperatura de 50 a 70°C y repitiendo las reacciones para confirmar los resultados. En la Figura 2.4 se detallan los resultados obtenidos de las reacciones de optimización para la amplificación de los diferentes genes.

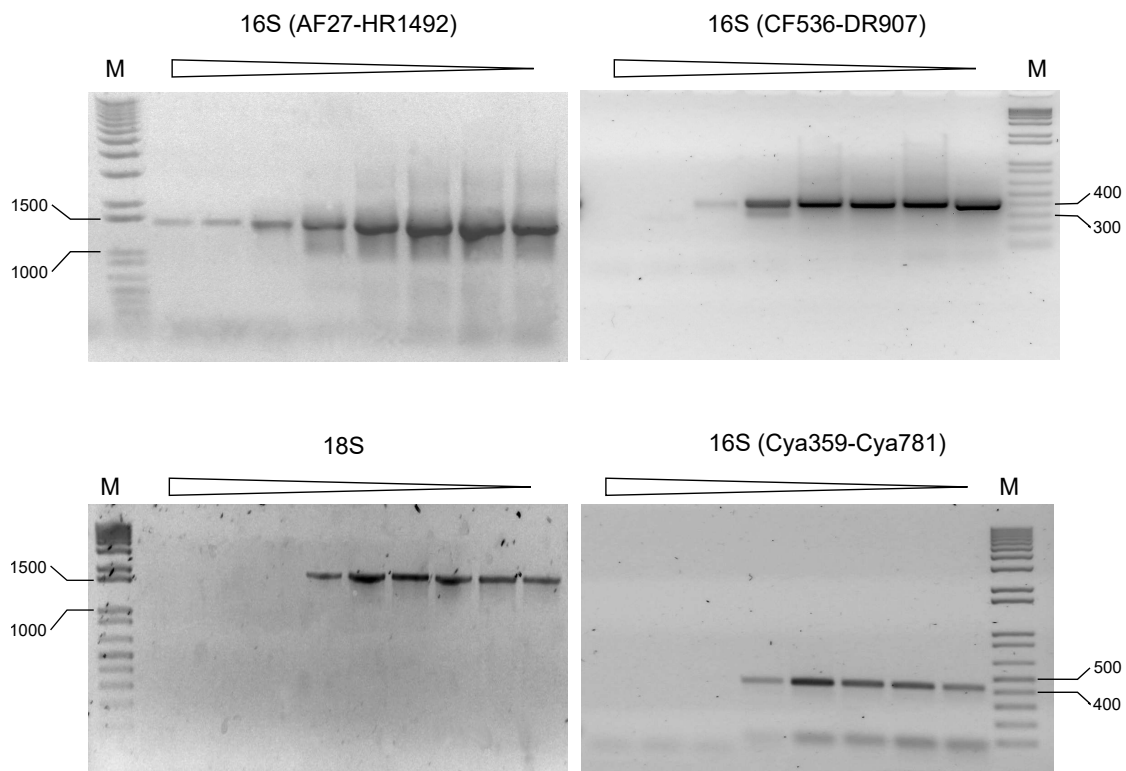


Figura 2.4: Desempeño de los diferentes ensayos de PCR, evidenciado a partir de las reacciones de en gradiente. Los paneles corresponden a la evaluación por electroforesis en geles de agarosa 1 y 2%, revelados con un colorante para ADN y observación bajo UV. En cada panel se detalla el nombre del gen amplificado (16S y 18S). Para el gen ARNr 16S se utilizaron tres iniciadores diferentes: AF27-HR1492 para obtener secuencias del gen casi completo utilizando las reacciones de secuenciación Sanger; el par CF536-DR907 para la obtención de fragmentos compatibles con los análisis de DGGE e iniciadores específicos para cianobacterias (Cya359-Cya781). Los moldes utilizados y condiciones de reacción utilizados en cada uno de los experimentos se encuentran detallados más abajo. En todos los casos se utilizaron condiciones de gradiente descritos en la Figura 2.3, los cuales se indican como triángulos en la parte superior de cada gel, indicando la dirección del gradiente de alta (izquierda) a baja (derecha) temperatura. Las temperaturas que resultaron más adecuadas en cada caso son las que se describen en las respectivas descripciones que se encuentran más abajo, como se mencionó anteriormente. *M* marcador de peso molecular *1Kb Plus* (bandas 100, 200, 300, 400, 500, 650, 850, 1000, 1500, 2000-15000 bases), se indican bandas de referencia.

---

Como puede observarse, los diferentes iniciadores presentan diferente desempeño, inclusive en las reacciones a diferente temperatura de anillado, en algunos casos pueden observarse dímeros de los iniciadores mientras que en otros la presencia de bandas inespecíficas. La temperatura óptima fue seleccionada según los siguientes criterios: ausencia de bandas inespecíficas junto con una buena cantidad de producto específico.

**Optimización de temperaturas de anillado** En los casos en que se indica en los siguientes puntos, la temperatura de anillado (o *annealing*) fue optimizada mediante PCR en gradiente.

**Gen ARNr 16S: metabarcoding** A partir del ADN obtenido de los filtros de 0,22  $\mu\text{m}$  de cada muestra se amplificó la longitud casi completa del gen mediante PCR utilizando el set de primers universales AF-27 y HR-1492 (Weisburg et al., 1991), . La temperatura de anillado se optimizó mediante PCR en gradiente. Se utilizó el siguiente esquema de ciclado: un ciclo inicial de 60 segundos de desnaturalización a 94°C, seguido de 35 ciclos de desnaturalización (30 segundos a 94°C), hibridación (*annealing*) optimizada, 10 segundos a 53,9°C) y extensión (100 segundos a 68°C). Luego se mantuvo a 4°C hasta su utilización o almacenamiento a -20 °C.

**Gen ARNr 16S: Electroforesis en gradiente desnaturalizante (DGGE)** Para los análisis por electroforesis en gradiente desnaturalizante (DGGE), los amplicones se obtuvieron con los iniciadores CF536 y DR907 (Weinbauer et al., 2002). La temperatura de anillado de los iniciadores CF536 y DR907 fue optimizada mediante PCR en gradiente (95°C por 1' de desnaturalización; 35 ciclos de 95°C 15", gradiente de 50 a 70 °C por 15" y 72°C durante 50", y una extensión final de 50" a 72°C). Las condiciones finales de la PCR consistieron en una etapa de desnaturalización de 1' a 95°C, 35 ciclos de 95°C por 15", 15" a temperatura de anillado optimizada (58,1°C), y una extensión final a 72°C durante 50".

**Gen 18S** Las secuencias correspondientes al gen ribosomal 18S fueron amplificadas mediante los iniciadores universales UNI\_17F y UNI\_1534R (Moon-van der Staay et al., 2001). La correspondiente temperatura de anillado fue optimizada mediante gradiente, obteniendo un óptimo de 62,2 °C. El ciclado consistió en una desnaturalización a 95°C durante 1', seguida de 35 ciclos de 95°C por 15", 15" a

---

temperatura de anillado optimizada (58,1°C). Luego de dichos ciclos se realizó una polimerización final a 72°C durante 50”.

**Amplificación de secuencias virales** La presencia de bacteriófagos de los géneros *Podovirus* y *Cyanopodovirus* fue evaluada mediante los iniciadores Podo-F y Podo-R2 (Labonté et al., 2009) y CP-DNAP-349F y 533Ra (Chen et al., 2009), respectivamente, utilizando las condiciones descriptas en las correspondientes publicaciones. Las secuencias del gen de la polimerasa de *Prasinovirus* fueron amplificadas mediante los iniciadores AVS-1 y AVS-2 (Chen y Suttle, 1995). Los análisis iniciales realizados con estos iniciadores fueron realizados utilizando las condiciones descriptas en el citado trabajo. Una vez que se obtuvo una muestra positiva (concentrado de filtración tangencial) se realizó una optimización utilizando la misma como molde y mediante gradiente utilizando la estrategia ya descripta. La temperatura óptima de anillado fue de 51,3 °C. Para el ciclado se realizó una desnaturalización a 94 °C durante 60” seguida de 35 ciclos de desnaturalización durante 60”, annealing optimizado durante 30” y 1’ de extensión a 68 °C.

**Detección de cianobacterias** La detección molecular de cianobacterias se realizó mediante los iniciadores específicos Cya359 y Cya781 (Nubel et al., 1997; Agarwal et al., 2017). Las condiciones de esta PCR consistieron en una etapa de desnaturalización de 1’ a 95°C, 35 ciclos de 95°C durante 15”, 15” a una temperatura de anillado optimizada por gradiente de 57.5°C y 72°C por 50”, con una extensión final de 5’ a 72°C. El ADN molde utilizado para la puesta a punto de la temperatura de anillado de esta PCR correspondió a una cepa de la colección institucional (*Nostoc* sp. Vaucher ex Bornet et Flahault; código LMPA5, colección de la Facultad de Ciencias Naturales y Ciencias de la Salud sede Trelew, Universidad Nacional de la Patagonia San Juan Bosco).

### 2.3.3. Metabarcoding

El metabarcoding (Taberlet et al., 2012) se realizó utilizando clonado molecular seguido de secuenciación Sanger y mediante Pirosecuenciación.

#### 2.3.3.1. Clonado molecular y secuenciación Sanger

**Purificación a partir de geles preparativos** Las purificaciones a partir de geles de agarosa se realizaron mediante un procedimiento basado en el método de Tautz & Renz (Tautz y Renz, 1989) (también llamado *freeze-squeeze*), según

---

se describe a continuación. Primero se separaron los fragmentos de ADN del gen del ARNr 16S mediante electroforesis en gel de agarosa, luego se cortó del gel la porción de agarosa conteniendo el fragmento de interés y se incubó en una solución de acetato de sodio 0,3 M pH 7,0 durante 30 minutos a temperatura ambiente, en oscuridad. Posteriormente, el taco de agarosa se colocó en una columna comercial con el fondo de fibra de vidrio de baja unión al ADN. Luego se incubó la columna a -20°C hasta que la agarosa se congeló completamente (~ 15'). Seguidamente, la columna se introdujo en un tubo tipo eppendorf de 1,5  $\mu$ l y se centrifugó a 14000 rpm durante 5 minutos en microcentrífuga. La agarosa quedó retenida en la fibra de vidrio mientras que la solución de ADN se recuperó en el segundo tubo. Se agregó 0,1 volumen de acetato de sodio 3 M (pH 5,2), se agitó y se agregaron 2,5 volúmenes de etanol absoluto y se incubó a -20°C durante toda la noche. El ADN se sedimentó a 14000 rpm durante 20 minutos a máxima velocidad en microcentrífuga y luego se resuspendió en agua de grado biología molecular.

**Ligación** Los fragmentos purificados según lo descrito en el punto anterior se ligaron en el vector de clonación pGEM-T Easy (Figura 2.5).

Cada reacción del procedimiento de ligación se llevó a cabo mediante tres relaciones molares inserto:vector con el fin de determinar las condiciones de reacción que permitiesen obtener la mayor cantidad de clones recombinantes posible. Estas relaciones fueron 1:1, 2:1 y 3:1. Para cada reacción se calculó la cantidad necesaria de inserto mediante la siguiente fórmula:

$$ng\ de\ inserto = \left( \frac{ng\ de\ vector * largo\ del\ inserto(kb)}{largo\ del\ vector(kb)} \right) * i : v \quad (2.1)$$

donde  $i : v$  corresponde a la relación molar inserto:vector.

En cada reacción se utilizaron 3 U de la enzima T4 DNA ligasa (Promega), 5 $\mu$ l del buffer recomendado por el fabricante [300 mM Tris-HCl (pH 7,8), 100 mM MgCl<sub>2</sub>, 100 mM DTT y 10 mM ATP], 0,4  $\mu$ l de ATP 10 mM y agua de grado biología molecular (c.s.p. 10  $\mu$ l). En cada reacción se utilizaron 50 ng de vector y la cantidad de inserto calculada según la ecuación 2.1 para cada relación molar. Las reacciones se llevaron a cabo a 4°C durante toda la noche.

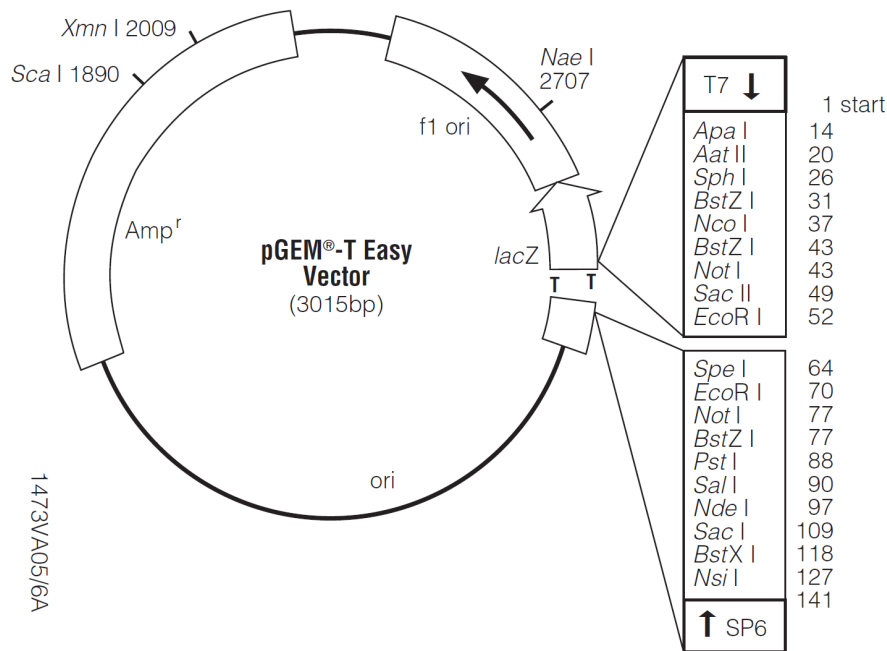


Figura 2.5: Mapa del vector pGEM-T *Easy*. Se detalla el sitio de inserción, o "polilinker", el cual se encuentra flanqueado por timidinas 3' terminales (T) que facilitan el clonado de los fragmentos de PCR. La región de clonado múltiple interrumpe el gen *lacZ*, que codifica para la enzima B-galactosidasa. También se detalla el gen *Amp<sup>r</sup>* que codifica un gen que confiere resistencia a la ampicilina, y la región correspondiente a los promotores T7 y Sp6 que flanquean el sitio de clonado múltiple.

**Transformación** El producto de las ligaciones fue utilizado para transformar bacterias competentes (sección 2.6). Para ello se descongelaron a temperatura ambiente 50  $\mu$ l de una suspensión de bacterias competentes y se mantuvieron en hielo. Luego se agregaron las bacterias a 5  $\mu$ l de mezcla de cada ligación y se incubó en hielo durante 30 minutos. Posteriormente se aplicó un shock térmico, el cual consistió en incubar la mezcla a 42°C durante exactamente 30 segundos y luego 2 minutos en hielo. Para su recuperación, las bacterias se incubaron en 200  $\mu$ l de LB durante 1 hora a 37°C. Las mezclas de transformación se plaquearon en medio LB-agar 1,5% conteniendo 50  $\mu$ g/ml de ampicilina, IPTG y X-Galactósido (X-Gal) y se incubaron durante toda la noche a 37°C.

---

**Identificación de plásmidos recombinantes** La identificación de plásmidos recombinantes se realizó mediante el método *blue/white screening*, el cual consiste en que las bacterias que fueron transformadas con plásmidos conteniendo el inserto de interés presentan el gen *lacZ* interrumpido (Fig. 2.5) y por lo tanto no degradan el X-Gal, permaneciendo las correspondientes colonias de color blanco (Fig. 2.6) . El clonado molecular resultó en más de 1700 colonias blancas y muy pocas colonias azules para todas las muestras, indicando que se logró un ajuste adecuado de la relación inserto:vector y que el grado de pureza del ADN inserto fue el suficiente como para afectar mínimamente la función de la ligasa y la incorporación del producto ligado en las bacterias competentes.

A partir de las colonias blancas de cada placa de clonado, se realizaron 120 estrías sobre placas con medio LB-agar 1,5% conteniendo 50  $\mu\text{g}/\text{ml}$  de ampicilina. Las placas se incubaron durante toda la noche a 37°C.





Figura 2.6: Biblioteca génica del gen ribosomal 16S. El método de construcción de las bibliotecas permite diferenciar los clones recombinantes (colonias blancas) de los clones que no presentan ningún inserto (colonias azules).

La purificación del ADN plasmídico se llevó a cabo utilizando el método "*turboprep*" (Woodford y K., 1991). Brevemente, se tomó en forma estéril una porción de una estría y se homogenizó mediante vórtex en 200  $\mu$ l de agua de grado biología molecular. Luego se trató la suspensión en un baño de agua hirviendo durante 90 segundos para desorganizar la estructura celular de las bacterias y así liberar los ácidos nucleicos. La preparación se centrifugó en microcentrífuga a máxima velocidad durante 15 segundos con el objeto de sedimentar los restos celulares.

El sobrenadante obtenido a partir de cada colonia se utilizó como molde en las PCR. Cada mezcla de reacción consistió en 1,25 U de la enzima Go Taq DNA Polymerase, 10  $\mu$ l del buffer Colorless Buffer Go Taq 5X (1,5mM MgCl<sub>2</sub> pH 8,5),

---

1  $\mu$ l de dNTP's (10  $\mu$ M cada uno; Invitrogen), los primers Sp6 y T7 (0,5  $\mu$ M c/u; Fig. 2.5), 1  $\mu$ l de la preparación plasmídica y agua de grado biología molecular (c.s.p. 50  $\mu$ l). Las condiciones de ciclado en la PCR fueron las siguientes: 1 ciclo de 60 segundos de desnaturalización a 95°C, 35 ciclos de desnaturalización (30 segundos a 95°C), *annealing* optimizado y extensión (100 segundos a 72°C), y luego se mantuvo a 4°C.

Se fueron haciendo tandas de 15 reacciones hasta que, para cada muestra, se pudieron identificar un mínimo de 100 clones portando el tamaño esperado.

**Secuenciación** La secuenciación se llevó a cabo por el servicio Macrogen Inc. (World Meridian Venture Center 10F, #60-24, Gasan-dong, Geumcheon-gu, Korea) utilizando los iniciadores utilizados en las PCRs. En general se secuenciaron ambas cadenas de los insertos, salvo en el caso de la muestra INV11, la cual fue secuenciada con el primer reverso ya que se utilizó un formato de secuenciación en placa.

Ver también sección 2.4.1.

### 2.3.3.2. Secuenciación de alta performance

A partir del ADN obtenido según se describe en la sección 2.3.1, se amplificaron secuencias correspondientes al gen del ARNr 16S mediante primers universales y las condiciones descritas en la sección 2.3.2. Previo a la pirosecuenciación, los amplicones se sometieron a 30 ciclos de PCR con el coctel HotStarTaq Plus de Qiagen, los primers universales 27F y 534R y condiciones de ciclado consistentes en 94 °C por 3 minutos, seguidos de 28 ciclos de 94 °C por 30 segundos, 53 °C por 40 segundos y 72 °C por 60 segundos y una etapa de extensión final de 5 minutos a 72 °C. Los productos obtenidos fueron purificados y secuenciados con el primer 27F unido a un *barcode* y la plataforma y química Roche 454 FLX Titanium.

Ver también sección 2.4.1.

### 2.3.4. Secuenciación directa

La secuenciación directa de amplicones obtenidos a partir de muestras ambientales se utilizó únicamente en los casos en que previamente se corroboró, mediante metabarcoding, la preponderancia de una secuencia mayoritaria (explicado en el texto). Los productos de PCR obtenidos directamente de ADN metagenómico fueron purificados y secuenciados mediante secuenciación tradicional utilizando los iniciadores de la PCR.

---

Ver también sección 2.4.1.

### 2.3.5. Electroforesis en gradiente desnaturante (DGGE)

Los amplicones, obtenidos mediante los iniciadores CF536 y DR907 (pág. 37), fueron purificados mediante un kit comercial siguiendo las instrucciones del fabricante. Los productos purificados ( $1 \mu\text{L}$ ) fueron mezclados con  $2 \mu\text{L}$  de una solución de azul de bromofenol 0,01 % y glicerol 70 % en buffer TBE y sembrados en un gel de acrilamida/poliacrilamida (37,5:1) 7,5 % con un gradiente lineal desnaturante 50-80 % (donde 100 % corresponde a 7M urea y 40 % formamida). La electroforesis se llevó a cabo a 120 V durante 16 hs. a una temperatura de  $60^\circ\text{C}$ , con buffer TBE. Los geles fueron posteriormente incubados durante 12 hs. en una solución de fijado (55 % etanol y 4 % metanol en agua de grado biología molecular), revelados mediante GelRed (Biotum) y observados en transiluminador de luz ultravioleta.

**Obtención de distancias genéticas.** Las distancias genéticas se obtuvieron mediante el índice de Bray-Curtis implementado en el paquete de R Ecodist, a partir de los datos de presencia/densidad de las bandas observadas en los electroferogramas. Los perfiles electroforéticos se cuantificaron mediante análisis densitométrico. Esto permitió tener una medida cuantitativa de la intensidad de cada banda para ser ingresada en el índice de Bray-Curtis. La digitalización de las bandas y su cuantificación fueron realizadas con la asistencia del programa ImageJ2 (Rueden et al., 2017).

## 2.4. Análisis de secuencias

### 2.4.1. Curado de secuencias crudas

**Secuenciación Sanger** El control de calidad de las secuencias se realizó mediante el programa BioEdit v7.0.5 (Hall, 1999), el cual posee un visor de cromatogramas que permite evaluar la calidad de las lecturas de acuerdo a la altura y equidistancia entre los picos, así como la presencia de *background*. En cada secuencia parcial se eliminaron las regiones donde el cromatograma mostraba baja calidad (Fig. 2.7). Una vez removidas las regiones de baja calidad, las secuencias directa y reversa fueron ensambladas mediante el programa MAFFT con la estrategia *L-INS-i* (Kato et al., 2002). Para identificar potenciales secuencias quiméricas se utilizó el

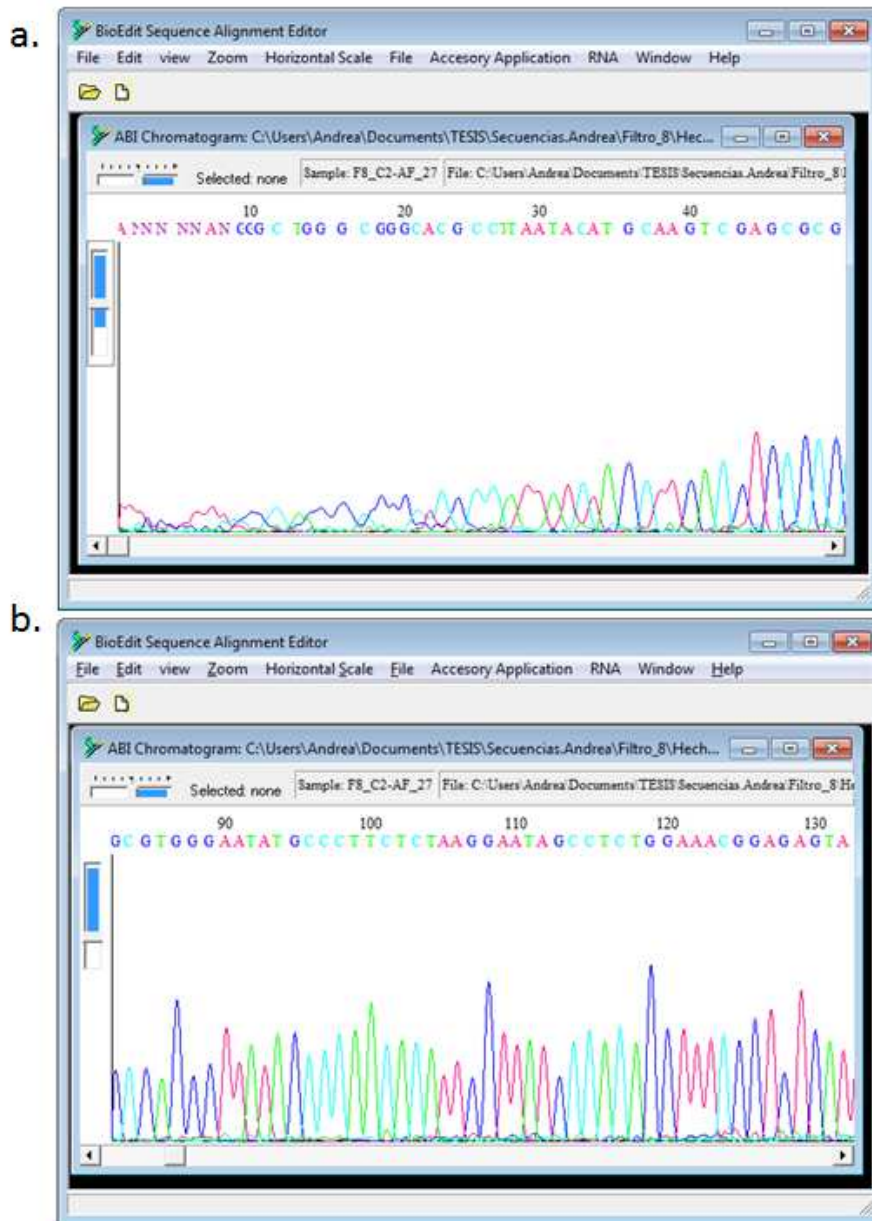


Figura 2.7: Análisis de cromatogramas mediante el programa BioEdit. Se muestran dos paneles en los que se observan diferentes calidades de secuencia. **a** Sector 5' de la secuencia, donde los picos aparecen en forma irregular y se presenta una proporción importante de ruido (o *background*) en relación a la señal. **b** Porción media de una lectura, de buena calidad, donde existe una alta relación señal/ruido y los picos aparecen en forma regular.

---

modo *de novo* del programa UCHIME (Edgar et al., 2011). Las secuencias que resultaron positivas fueron eliminadas de los análisis.

**Pirosecuenciación** Los datos crudos generados por el secuenciador fueron sometidos a un estricto control de calidad según la metodología propuesta en la referencia (Schloss et al., 2011) utilizando el programa Mothur (Shloss et al., 2009). Brevemente, los flujogramas que presentaron diferencias de 3 y 2 bases en las secuencias del iniciador o el *barcode*, respectivamente, trectos homopoliméricos de más de 9 bases y menos de 360 o más de 720 flujos fueron eliminados, luego de lo cual los datos restantes fueron depurados mediante el algoritmo *Pyronoise* (Quince et al., 2011). Luego de eliminar las regiones correspondientes a los iniciadores y los *barcodes*, las secuencias fueron alineadas con la base de datos SILVA (<http://www.arb-silva.de>) mediante Mothur, con el objeto de detectar y eliminar las secuencias que no correspondieron a la región blanco (región V1-V3 del gen). Luego de esto, se fusionaron las secuencias que se hallaron a menos de 2 sustituciones de secuencias más abundantes con el fin de reducir aún más los posibles errores, y se identificaron y eliminaron las potenciales quimeras mediante el algoritmo UCHIME implementado en Mothur.

#### 2.4.2. Asignación taxonómica

**Gen 16S** Las secuencias del gen del ARNr 16S obtenidas mediante secuenciación tradicional fueron analizadas en forma separada de aquellas obtenidas mediante pirosecuenciación para evitar potenciales problemas relacionados con diferencias en la variabilidad e información taxonómica de las regiones del gen determinadas en cada caso (Wang et al., 2007; Liu et al., 2008). En primer lugar, se obtuvieron distancias no corregidas ( $p$ ) mediante el programa Mothur. Los gaps terminales fueron ignorados mientras que los internos fueron ponderados como una única mutación independientemente de su largo con el objeto de evitar potenciales problemas debidos a errores en la determinación de trectos homopoliméricos. Las matrices obtenidas se utilizaron para generar agrupamientos (*clusters*) de 97% de similitud mediante el método de unión al vecino promedio (*average neighbor linkage*), y los datos obtenidos fueron procesados con el comando `make.shared` de Mothur utilizando los parámetros default. Las secuencias representativas de cada OTU fueron generadas con el comando `get.oturep` de Mothur el cual, para cada OTU, identifica la secuencia que presenta la mínima distancia al resto de las secuencias incluidas en el mismo. Se obtuvieron 250 OTUs correspondiendo a 1997

---

pirosecuencias de alta calidad (177 OTUs) y 241 secuencias Sanger (73 OTUs) también de alta calidad. Las asignaciones taxonómicas propiamente dichas fueron realizadas mediante la versión del clasificador naïve bayesiano de Wang (WNBC, del inglés *Wang's Naïve Bayesian Classifier*) (Wang et al., 2007) implementada en el programa Mothur con un valor de *cutoff* de 80 % de probabilidad posterior. Como dataset de referencia para el WNBC se utilizó la base de datos SILVA, la cual fue descargada localmente.

**Gen 18S** Las secuencias del gen 18S fueron clasificadas en primera instancia mediante comparaciones con GenBank con el programa BLAST. A cada secuencia se le asignó la taxonomía de la secuencia de GenBank que presentó el mayor score en los análisis mediante BLAST. Además, se realizaron análisis filogenéticos y mediante Parsimonia Estadística (sección 2.4.3).

**ADN polimerasa viral** Las secuencias de la polimerasa viral de virus de la familia Phycodnaviridae fueron identificadas mediante BLAST y análisis filogenético como se describe en la sección 3.

### 2.4.3. Análisis evolutivos

**Alineamiento** Los alineamientos de secuencias fueron obtenidos mediante el programa MAFFT (Kato et al., 2002). En general se utilizó la opción *L-INS-i* del programa, salvo en los casos en que se especifica lo contrario. Los alineamientos obtenidos se analizaron mediante los programas gráficos ClustalX (Larkin et al., 2007) y Jalview (Waterhouse et al., 2009) y el paquete de R *ips* (Heibl, 2014).

**Inferencia de árboles** La inferencia de árboles mediante Máxima Verosimilitud (*Maximum Likelihood*) se realizó con los programas *FastTree* (Price et al., 2009) o *PhyML* (Guindon et al., 2010). El primero se utilizó para analizar las matrices grandes (más de  $\sim 150$  secuencias). *FastTree* ajusta la intensidad de las búsquedas de manera automática realizando  $4 \times \log_2(N)$  ciclos de búsqueda mediante intercambio de vecinos más cercanos (NNI) y 2 rondas de poda e injerto de subárboles (SPR) utilizando el criterio de evolución mínima, y luego  $2 \times \log(N)$  ciclos de optimización del valor de verosimilitud (Likelihood) mediante SPR, donde  $N$  es el número de terminales del árbol. Los parámetros utilizados para los análisis mediante *PhyML* se detallan en cada caso particular. El soporte de ramas se obtuvo mediante bootstrap (*PhyML*) o soporte local basado en el test de Shimodaira-Hasegawa

---

(FastTree). Los modelos evolutivos fueron inferidos mediante MrAIC (Nylander, 2004) o se utilizaron modelos soportados por bibliografía, según se indica en cada paso particular.

**Análisis de árboles** Los árboles filogenéticos obtenidos fueron visualizados mediante los programas Dendroscope (Huson y Scornavacca, 2012) y *FigTree*<sup>1</sup> y analizados mediante el paquete *ape* (Paradis, 2012) del entorno estadístico R (R Core Team, 2016) y la suite de programas *Newick Utilities* (Junier y Zdobnov, 2010).

**Diferenciación genética** Los análisis de similitud entre secuencias y grupos de secuencias se realizaron como se describe en los trabajos de Rodríguez et al. (2009) y Shin et al. (2010). Brevemente, para un par de secuencias  $a$  y  $b$ , el número de sustituciones de nucleótidos o aminoácidos ( $D$ ) se obtiene de la siguiente manera:

$$D_{ab} = \sum_{i=1}^P f(a_i, b_i) \quad (2.2)$$

donde  $P$  es el número de posiciones alineadas, y  $f(a_i, b_i)$  se obtiene según la ecuación:

$$F(a_i, b_i) = \begin{cases} 0 & \text{si } a_i = b_i \\ 1 & \text{si } a_i \neq b_i \end{cases}$$

Las comparaciones dentro de los clusters se promediaron por el número de comparaciones de secuencias realizadas en cada caso.

El grado de diferenciación genética entre clados fue ponderado siguiendo lineamientos propuestos por Salgado-Salazar et al. (2010) para el reconocimiento de especies filogenéticas (*Phylogenetic Species Recognition*, PSR). Brevemente, sea  $r_i$  el largo de la rama que separa al clado de interés del resto del árbol de máxima verosimilitud y  $R$  la mediana de todos los largos de rama del árbol. Luego la relación

$$\frac{r_i}{R}$$

es un estimador de qué tan diferenciado está el linaje de interés,  $i$ , en el marco de la filogenia del grupo de estudio.

---

<sup>1</sup><http://tree.bio.ed.ac.uk/programa/figtree/>

---

**Parsimonia estadística (PE)** La obtención de *networks* mediante PE (Templeton et al., 1992) fue realizada mediante el programa TCS (Clement et al., 2000) bajo un límite de conexión de 95 %, el cual constituye un estándar en la implementación del método que es cuantitativamente confiable en una variedad de grupos y diversos marcadores moleculares (Hart y Sunday, 2007).

## 2.5. Aislamiento en cultivo

Las muestras de agua fueron sometidas a filtración por impacto según se describe al comienzo de la sección en el punto 2.2 (pág. 31), salvo que luego del pasaje por las membranas de 5  $\mu\text{m}$  el filtrado se sometió a un tamizado mediante un filtro de 1,2  $\mu\text{m}$  de tamaño de poro. Luego de ese punto el proceso se detuvo. Se tomaron alicuotas de 30ml y las mismas fueron mezcladas con un volumen de medio *Keller* (Keller et al., 1987) suplementado con estreptomicina (10  $\mu\text{g}/\text{ml}$ ), ampicilina (10  $\mu\text{g}/\text{ml}$ ) y kanamicina (20  $\mu\text{g}/\text{ml}$ ). La preparación fue incubada en una cámara de cultivo y se fueron realizando observaciones microscópicas y conteos de las células de picoeucariotas mediante un hemocitómetro cada cuatro días. Cada vez que las células alcanzaron una densidad de 20-30 x 10<sup>6</sup> células mL<sup>-1</sup>, los cultivos fueron subcultivados en forma aseptica mediante dilución 1:10 en medio fresco suplementado.

## 2.6. Preparación de bacterias competentes

Se utilizaron bacterias de la cepa *Escherichia coli* One Shot TOP 10 [F<sup>-</sup> *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\varphi$ 80*lacZ* $\Delta$ M15  $\Delta$ *lacX74* *recA1* *araD139*  $\Delta$ (*ara-leu*) 7697 *galU galK rpsL* (Str<sup>R</sup>) *endA1 nupG*  $\lambda$ -]. A partir de una placa conteniendo colonias aisladas, se repicó una colonia en 10 ml de medio Luria-Bertani (LB; triptona 1 %, extracto de levadura 0,5 %, NaCl 1 %) y se incubó con agitación (190-200 rpm) a 37°C durante 10 horas en un baño termostático. Luego se realizó una dilución 1/100 del cultivo en medio fresco y se incubó con agitación (190-200 rpm) a 37°C hasta alcanzar una DO<sub>550</sub> de 0,35-0,6. Este cultivo fue incubado en hielo durante 15 minutos y se centrifugó a 1000 G durante 10 minutos en rotor oscilante (Gelec G 1615). Posteriormente las bacterias fueron resuspendidas en 10 ml de buffer RF1 (100mM ClK, 50 mM Cl<sub>2</sub>Mn, 30 mM AcK pH 7,5, 10 mM Cl<sub>2</sub>Ca, 15 % glicerol), incubadas en hielo durante 15 minutos y centrifugadas a 1000 g durante 10 minutos. El pellet de bacterias se resuspendió en una solución de buffer RF2



---

(10 mM MOPS pH 6,8, 10 mM ClK, 10 mM Cl<sub>2</sub>Ca, 15 % glicerol), se incubaron en hielo durante 15 minutos e inmediatamente se almacenaron a -20°C.

## 2.7. Análisis estadísticos

Los análisis de componentes principales (ACP) y de agrupamiento (AA) fueron realizados mediante el programa PAST<sup>2</sup>. Las distancias ecológicas fueron calculadas como diferencias absolutas entre los valores registrados para las dos variables ambientales consideradas (temperatura y caudal). Con el objeto de eliminar posibles efectos de escala, las variables fueron estandarizadas sustrayendo la media y dividiendo por el desvío estándar correspondientes. Para estimar las disimilitudes entre las muestras se utilizaron distancias euclidianas. En el análisis de agrupamiento se utilizó el método aglomerativo de Ward, el cual agrupa las muestras de forma tal de minimizar la varianza dentro de cada grupo (Murtagh y Legendre, 2014). Los análisis multivariados de regresión y correlación se realizaron mediante estimación de modelos lineales en matrices de distancia (dbLM) (Lichstein, 2007) y el test de Mantel de correlación en matrices de distancia (Goslee, 2010) implementados en el paquete *Ecodist* (Goslee y Urban, 2007) del entorno estadístico R (R Core Team, 2017). Tanto para el dbLM como en el test de Mantel se utilizaron 10.000 permutaciones para evaluar la significancia estadística.

---

<sup>2</sup><https://folk.uio.no/ohammer/past/>

## 3. Resultados

### 3.1. Análisis del bacterioplancton

Se generaron 241 secuencias de alta calidad de alrededor de 900 bases, de las cuales 102 correspondieron a la muestra VER11, 71 a OTO11 y 68 a INV11.

Asimismo, se detectaron 33 secuencias del cloroplasto de células de *Ostreococcus sp.* retenidas en el filtro de 0,22  $\mu\text{m}$ . Las secuencias pudieron agruparse en 73 OTUs. La pirosecuenciación de la muestra PRI11 se realizó con una cobertura nominal de 3000. A partir de las correspondientes lecturas se realizó un control de calidad que permitió generar 1997 secuencias de alta calidad. Las mismas pudieron agruparse en 177 OTUs.

El análisis de las secuencias reveló un microbioma dominado por pocos grupos taxonómicos, los cuales estuvieron acompañados por una plétora de grupos numéricamente menos importantes. Estos resultados se resumen en las Figuras 3.1 y 3.2, las cuales además indican la frecuencia de cada OTU. Se observó una cantidad importante ( $\sim 13\%$ ) de OTUs no clasificables <sup>1</sup>.

---

<sup>1</sup>En la subsección 3.2 se presenta un análisis detallado de estas OTUs

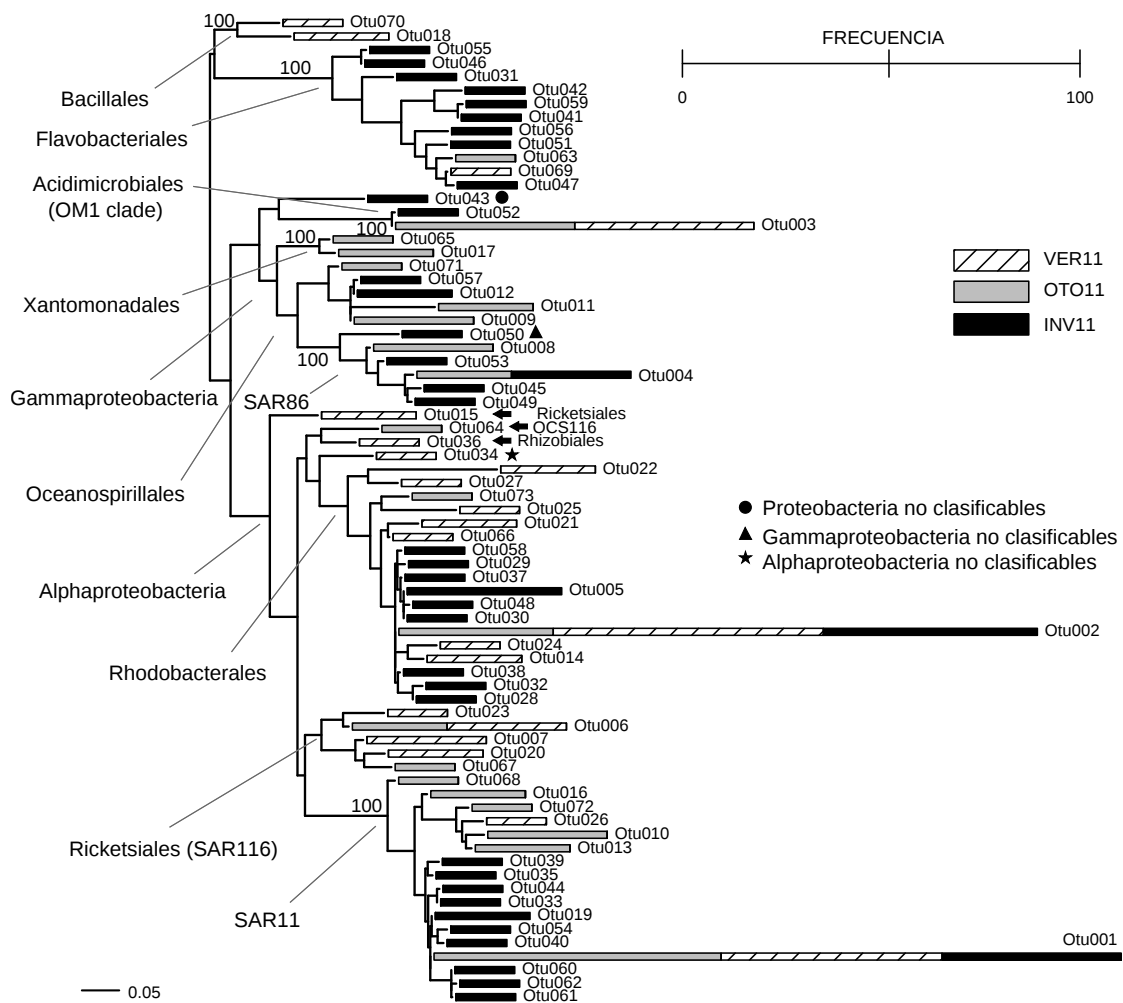


Figura 3.1: Árbol filogenético de OTUs correspondientes a los datos obtenidos mediante secuenciación Sanger. La taxonomía corresponde a la clasificación obtenida mediante WNBC (*cutoff* = 80 %). Las barras en los terminales del árbol indican la abundancia de cada OTU (número de secuencias), según se indica en la escala ubicada en el ángulo superior derecho de la figura. Los números ubicados cerca de las ramas indican valores de bootstrap (n=100). La escala en el ángulo inferior izquierdo de la figura está en sustituciones por posición alineada.

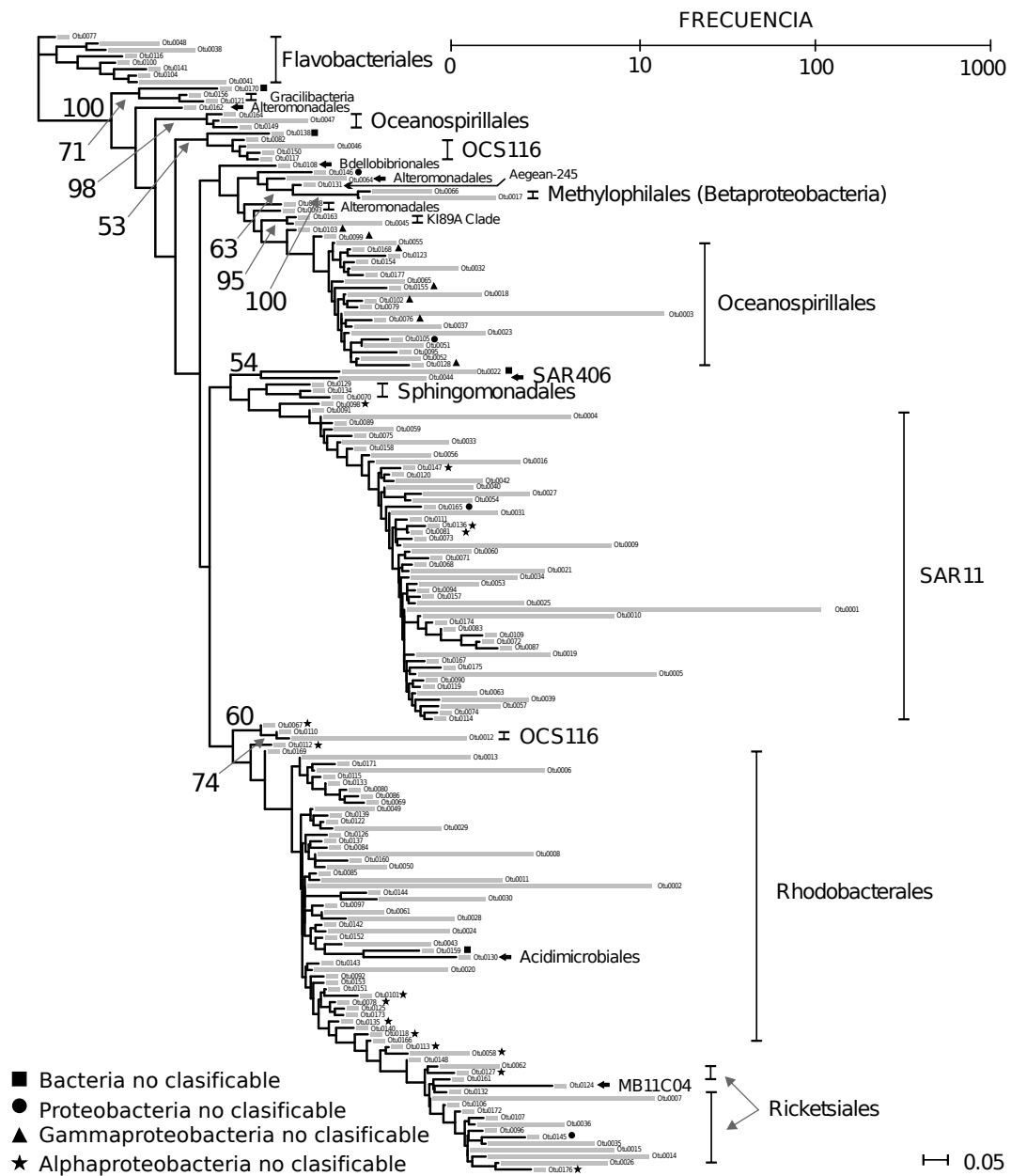


Figura 3.2: Árbol filogenético de OTUs correspondientes a secuencias del gen del ARNr 16S obtenidas mediante pirosecuenciación. Las barras en los terminales del árbol indican la abundancia de cada OTU (número de secuencias), según se indica en la escala ubicada en la parte superior de la figura. Los números ubicados cerca de las ramas indican valores de bootstrap (n=100). La escala en el ángulo inferior derecho de la figura está en sustituciones por posición alineada.

Ninguna de las muestras estudiadas presentó secuencias de cianobacterias. Estos resultados fueron corroborados mediante la detección molecular de estos microorganismos mediante PCRs específicas. Las mismas resultaron positivas en algunas muestras, aunque en varias de las mismas sólo se observaron bandas relativamente tenues (Fig. 3.3), posiblemente debido a una baja abundancia.

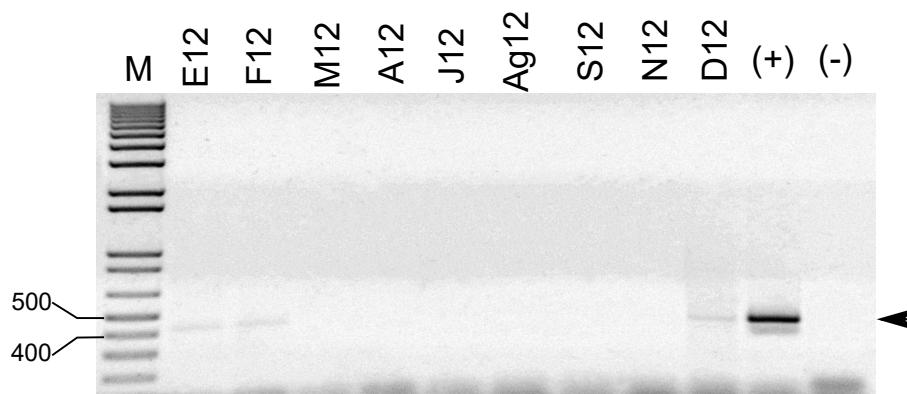


Figura 3.3: Detección de cianobacterias mediante PCR en muestras tomadas con una frecuencia aproximadamente mensual (E12, F12, M12, A12, J12, Ag12, S12, N12 y D12). *M* marcador de peso molecular 1 Kb Plus donde se indican como referencia las bandas de 400 y 500 pb (bandas 100, 200, 300, 400, 500, 650, 850, 1000, 1500, 2000-15000 pb), + control positivo, - control negativo. La flecha indica la posición de la banda correspondiente al ADN blanco.

Si bien los análisis de las bibliotecas génicas del gen del ARNr 16S revelaron que 3 taxa (*Pelagibacterales*, *Rhodobacterales* y *Oceanospirillales*) estuvieron presentes en forma abundante en todas las estaciones del año, las proporciones de los mismos fueron sustancialmente variables. Las *Pelagibacterales* (Clase *Alphaproteobacteria*, Phylum *Proteobacteria*) representaron frecuencias de entre 13,1% y 63%. Por su parte las *Rhodobacterales* (Clase *Alphaproteobacteria*, Phylum *Proteobacteria*) constituyeron entre el 8,6 y el 38,1% y las *Oceanospirillales* (*Gammaproteobacteria*, Phylum *Proteobacteria*) alcanzaron un máximo de 15,7% y estuvieron ausentes en una de las muestras. Estos resultados se encuentran resumidos en la Tabla 3.1. Además, como puede apreciarse en la Figura 3.1, los grupos minoritarios también presentaron fluctuaciones.

Tabla 3.1: Resumen del análisis de secuencias ambientales del gen ARNr 16S y detección molecular de cianobacterias.

MUESTRA	VER11	OTO11	INV11	PRI11	E12	F12	M12	A12	J12	Ag12	S12	N12	D12
Picoalgas <sup>1</sup>	30.3 %	2.8 %	0 %	0 %	NA <sup>2</sup>	NA	NA	NA	NA	NA	NA	NA	NA
Pelagibacterales	13,1 %	51,4 %	28,5 %	63 %	NA	NA	NA	NA	NA	NA	NA	NA	NA
Rhodobacterales	32,3 %	8,6 %	38,1 %	20,3 %	NA	NA	NA	NA	NA	NA	NA	NA	NA
Acidimicrobiales	7,1 %	1 %	1,6 %	0,1 %	NA	NA	NA	NA	NA	NA	NA	NA	NA
Oceanospirillales	0 %	15,7 %	14,9 %	9 %	NA	NA	NA	NA	NA	NA	NA	NA	NA
Otros <sup>3</sup>	17,2 %	20,5 %	16,9 %	7,6 %	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cyanobacteria <sup>4</sup>	+	-	-	-	+	+	-	-	-	-	-	-	+

<sup>1</sup> Proporción de secuencias del gen del ARNr 16S correspondientes a la Clase Mamiellophyceae (secuencias de cloroplastos).

<sup>2</sup> No Aplicable.

<sup>3</sup> Frecuencias que presentaron el conjunto de secuencias correspondientes a Bacillales, Flavobacteriales, Xantomonadales, Oceanospirillales, SAR86, OCS116, Rhizobiales, SAR116 y otros Rickettsiales, MB11C04, Flavobacteriales, Bdellobibrionales, Alteromonadales, Aegean-245, Methylophilales, KI89A, SAR406, Sphingomonadales y Gracilibacteria (también llamados BD1-5 y GN02) y grupos no clasificables.

<sup>4</sup> Detección por PCR a partir de células inmovilizadas mediante filtración por impacto; + positivo, - negativo.

---

**Análisis mediante DGGE.** Con el objeto de estudiar las fluctuaciones de la diversidad en una escala temporal menor se realizaron análisis mediante DGGE de muestras tomadas con una frecuencia aproximadamente mensual (muestras E12, F12, M12, A12, J12, Ag12, S12, N12 y D12; Tabla 2.1, sección 2.1).

Los resultados obtenidos indicaron que la composición de la comunidad bacteriana de la zona de estudio puede variar sustancialmente entre algunos meses (Fig. 3.4 A). Los electroferogramas de enero, febrero, marzo, abril y diciembre presentaron entre dos y tres bandas principales (bandas A, B, C; Fig. 3.4 A) más 2 bandas de baja intensidad en el caso de la muestra D12 (bandas D e I). Las muestras de invierno y la de noviembre también presentaron la banda A, aunque en menor intensidad, mientras que la banda B sólo se observó para la muestra J12. Los electroferogramas correspondientes a los meses de julio, agosto, septiembre y noviembre presentaron otras bandas (E, F, G, H y J), las cuales estuvieron ausentes en el resto de las muestras.

Con el objeto de evaluar cuantitativamente estas similitudes y diferencias entre los electroferogramas de las muestras, se obtuvieron ordenamientos de las mismas mediante análisis de los componentes principales (ACP) y de agrupamiento (AA), ambos basados en distancias genéticas calculadas a partir de datos de presencia/intensidad de las bandas en los perfiles de DGGE. Dichos análisis fueron coincidentes con el análisis visual de los electroferogramas; esto es, indicaron que las bandas observadas en estos últimos, particularmente las bandas A, B, F, J, G y H, permiten agrupar las muestras en dos grupos principales: por un lado las muestras E12, F12, M12, A12 y D12 y por otro las J12, Ag12, S12 y N12 (Fig. 3.4, paneles D y E). El C1 explicó un 75,6 % de la varianza, mientras que la proporción de la misma explicada por el C2 fue de 12,2 %.

El análisis combinado de los perfiles de DGGE con datos de temperatura y caudal del río (los cuales presentan una importante variación estacional), sugirió una asociación entre los cambios genéticos entre las muestras y estas variables ambientales (u otros parámetros estacionalmente variables). La presencia de las bandas A y B fue aproximadamente coincidente con caudales bajos del río y temperaturas altas, mientras que el aumento del caudal y descenso de la temperatura fue coincidente con la desaparición de dichas bandas y la emergencia de las bandas E, F, G, H y J (Fig. 3.4 paneles A, B y C). La banda C presentó un comportamiento más errático, habiéndose detectado tanto en meses fríos (agosto y julio) como a fines de la primavera (Diciembre) y con caudales superiores e inferiores a la media anual.

De manera interesante, los análisis mediante ACP y AA (Fig. 3.4; paneles D y E,

respectivamente) mostraron que las muestras de meses templados se agruparon por un lado y las de meses frío por otro, lo cual también coincide aproximadamente con momentos de bajo y alto caudal, respectivamente. El componente 1 (C1) del ACP discriminó preferencialmente las muestras de los meses de mayor caudal (todos ubicados en valores negativos del C1) de las correspondientes a los meses de menor caudal (valores positivos de C1). El componente 2 mostró una correspondencia más marcada con la temperatura, aunque la muestra N12 quedó muy separada del resto

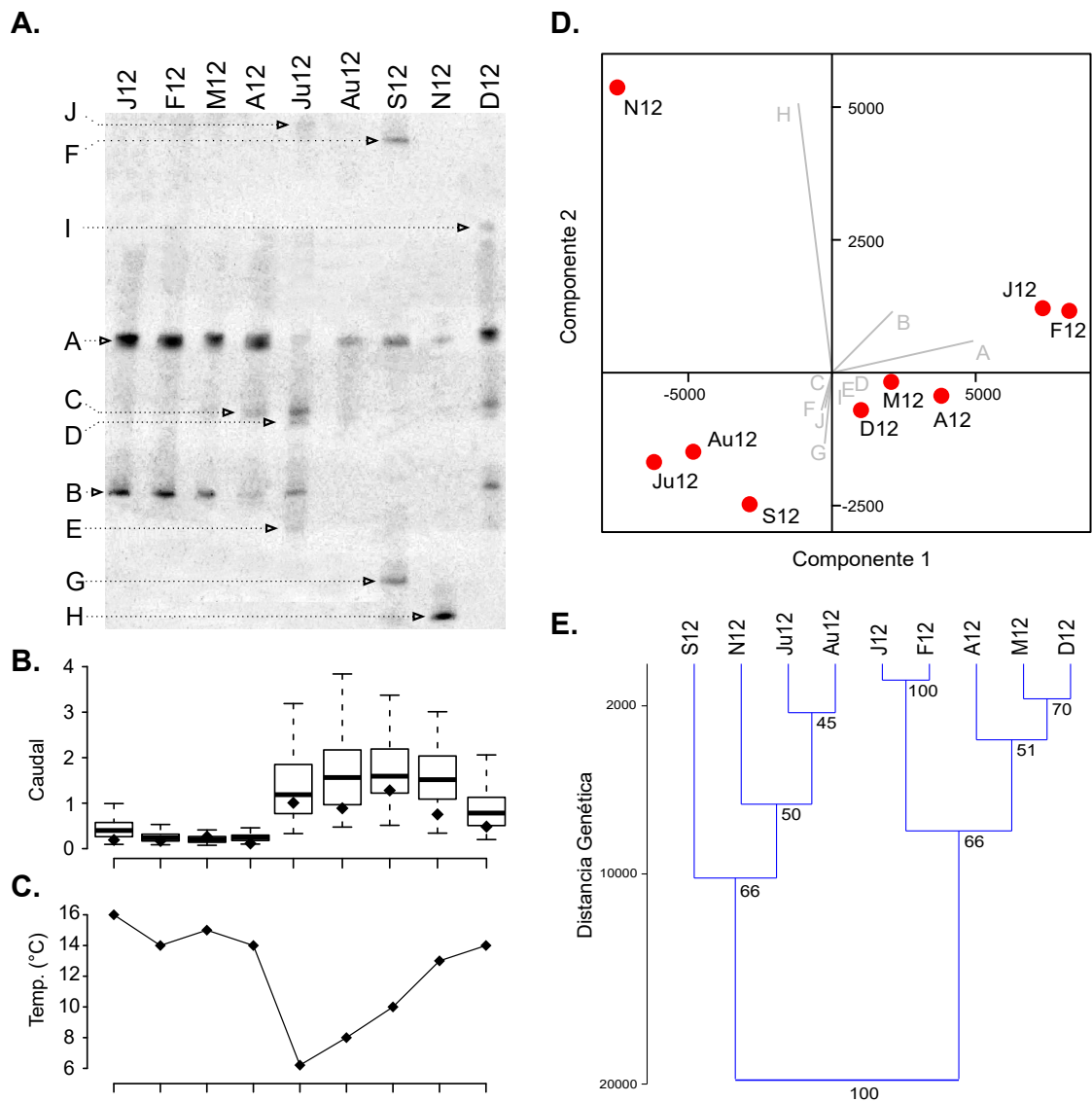


Figura 3.4: Leyenda en la página siguiente.



---

Figura 3.4: Análisis de la fluctuación la diversidad del gen del ARNr 16S mediante DGGE. El panel **A** muestra los electroferogramas de las 9 muestras colectadas para este análisis con una frecuencia aproximadamente mensual (E12, F12, M12, A12, J12, Ag12, S12, N12 y D12). A la izquierda se encuentran indicadas las bandas. El panel **B** indica el estado hidrológico del río en los meses en los que se tomó cada una de las muestras (puntos) y boxplots correspondientes a medidas históricas. El caudal está expresado en medidas no dimensionales (Materiales y Métodos). La leyenda del eje x de los paneles B y C corresponde al indicado en el panel A. El panel **C** muestra las temperaturas registradas en el agua superficial al momento de la toma de las muestras. Los paneles **D** y **E** corresponden a los resultados del análisis de los componentes principales (D) y de agrupamiento (E) de las muestras en base a distancias genéticas calculadas a partir de los correspondientes electroferogramas. La contribución de cada banda al ordenamiento obtenido mediante APC se representa mediante vectores (gris). Los vectores correspondientes a las bandas C, D, E e I no se visualizan (son muy pequeños) ya que los mismos discriminan comparativamente poco con respecto al resto de las bandas. Los números en el árbol corresponden a valores de bootstrap de cada una de las divisiones y subdivisiones propuestas por el algoritmo; por ejemplo, el valor de 100 en la base del árbol indica la confianza estadística de la división en los grupos  $\{E12, F12, M12, A12, D12\}$  y  $\{J12, Ag12, S12, N12\}$ .

de las muestras de los meses templados.

La relación entre los electroferogramas de las muestras y el caudal y la temperatura se estudió también mediante análisis de regresión. En primer lugar, se graficaron las distancias genéticas en función de las distancias ecológicas obtenidas a partir de los datos de temperatura y caudal por separado. Tal como sugirieron la inspección directa de los datos y los análisis multivariados, los gráficos obtenidos mostraron que ambas covariables se asocian con las distancias genéticas (Fig. 3.5). Esta asociación también es soportada por los resultados del modelado mediante regresión parcial en matrices de distancia (dbLM) (Lichstein, 2007), los cuales se muestran en la Tabla 3.2. El test parcial de Mantel (Goslee, 2010) también soportó una correlación significativa con el caudal (Tabla 3.3).

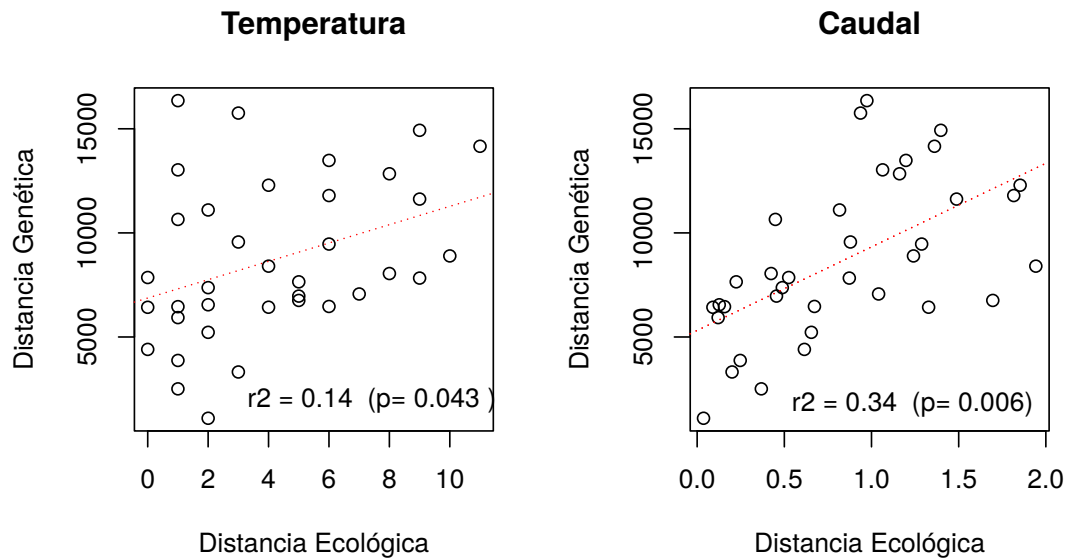


Figura 3.5: Análisis gráfico de las relaciones entre las distancias genéticas entre las muestras y las correspondientes distancias ecológicas, calculadas a partir de datos de temperatura del agua de mar y caudal del río. Las líneas rojas y los coeficientes corresponden a modelos de regresión simple en matrices de distancia (dbLM) (Lichstein, 2007). Ver también análisis de regresión parcial en Tabla 3.2.

Tabla 3.2: Modelo de regresión parcial en matrices de distancia (dbLM) (Lichstein, 2007) para la relación entre las distancias genéticas entre las muestras y las correspondientes distancias ecológicas. Ver también Figura 3.5.

	DGGE	valor- $p$
$\hat{\alpha}$	5169.15	0.995
Temperatura ( $\hat{\beta}_1$ )	93.90	0.700
Caudal ( $\hat{\beta}_2$ )	3727.19	0.012

Tabla 3.3: Análisis de regresión entre las distancias genéticas entre muestras mensuales y la temperatura del agua (T) y el caudal del río (C) mediante el test parcial de Mantel (Goslee, 2010).

	R	$H_0^{(1)}$			IC 95 % <sup>(2)</sup>	
		$R \leq 0$	$R \geq 0$	$R = 0$	L. inf.	L. sup.
$C T^{(3)}$	0.495	0.007	0.994	0.007	0.376	0.750
$T C^{(4)}$	0.082	0.346	0.655	0.712	-0.177	0.454

<sup>1</sup> Hipótesis nula.

<sup>2</sup> Intervalo de confianza del 95 %. *L. inf.* límite inferior; *L. sup.* límite superior.

<sup>3</sup> Relación entre el perfil genético y el caudal, dada la temperatura.

<sup>4</sup> Relación entre el perfil genético y la temperatura, dado el caudal.

## 3.2. Secuencias locales no clasificables

En esta subsección se describe el análisis detallado de las OTUs no clasificables (uOTUs) detectadas en la zona de estudio. Se consideraron no clasificables a las OTU que no pudieron asignarse a una categoría equivalente o inferior al rango Orden mediante WNBC, con un valor de confianza igual o superior a 80 %. El interés se centró en identificar uOTUs que fueran similares a secuencias observadas en otras regiones del Mundo, ya que los taxa cosmopolitas tienen una mayor probabilidad de representar grupos *condicionalmente raros* (Shade et al., 2014) y ser importantes en ambientes fragmentarios (Sunagawa et al., 2010), además de tener un mayor impacto potencial en la diversidad microbiana futura (Choi et al., 2017; Shade et al., 2014).

Como primer aproximación al problema, se implementó una aproximación filogenética (a partir de aquí Análisis Filogenético Directo, *AFD*) consistente en mapear la clasificación obtenida mediante un clasificador automático, el clasificador naïve de Wang, sobre una filogenia de las OTUs clasificables y las uOTUs.

Por otro lado, las uOTUs fueron comparadas con secuencias disponibles en GenBank utilizando el algoritmo BLAST de una forma personalizada de forma tal que dicho algoritmo discrimine las secuencias ambientales de secuencias de material cultivado y de material tipo.

Finalmente, la información derivada de ambos análisis (*i. e.* AFD y BLAST personalizado) fue utilizada para constuir un dataset conteniendo 6194 secuencias de referencia, principalmente de material tipo, el cual fue combinado con las uOTUs con el objeto de intentar determinar a qué grupos podrían corresponder estas

---

últimas.

La Figura 3.6 muestra un esquema del procedimiento completo.

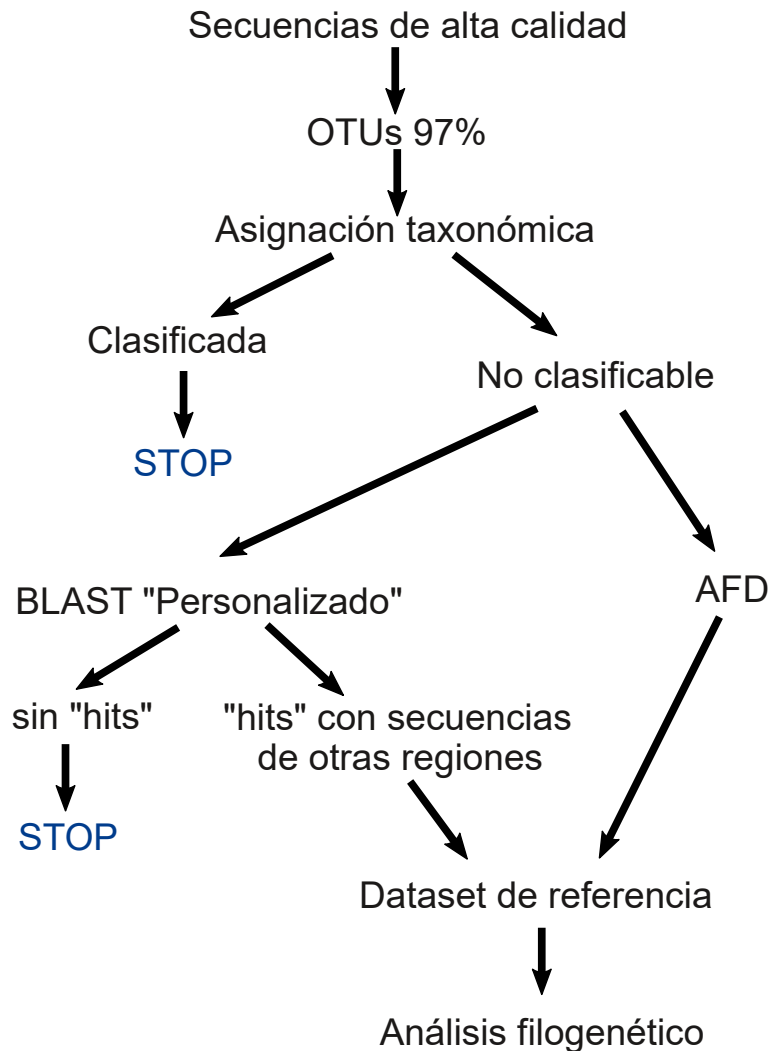


Figura 3.6: Diagrama de flujo del análisis realizado para estudiar las secuencias no clasificables. *OTU* Unidad Taxonómica Operativa; *BLAST* Herramienta Básica de Búsqueda de Alineamientos Locales (del inglés *Basic Local Alignment Search Tool*); *AFD* Análisis Filogenético Directo (pág. 62)

**El 13% de las OTU resultaron ser no clasificables.** Como se mencionó brevemente en la sección 3.1, de un total de 250 OTUs, 33 no pudieron clasificarse (Tabla 3.4; pág. 71). Cuatro de ellas no pudieron clasificarse en ningún Phylum co-

---

nocido y seis fueron asignadas a *Proteobacteria*. Siete uOTUs no pudieron asignarse a ningún taxon dentro de la clase *Gammaproteobacteria*, mientras que 16 de ellas correspondieron a grupos desconocidos dentro de *Alphaproteobacteria*. No se observaron diferencias significativas entre las proporciones de secuencias no clasificables (uSEQs) observadas para los datos obtenidos mediante secuenciación tradicional y secuenciación de alta performance (Tabla 3.5; pág. 73). Sin embargo, la proporción de OTUs Sanger no clasificables fue levemente menor que las correspondientes a los datos de alta performance (SAP), posiblemente debido a la mayor resolución de la secuenciación Sanger (las secuencias obtenidas mediante pirosecuenciación tienen entre 300 y 350 bases, mientras que las obtenidas mediante secuenciación Sanger tienen alrededor de 800 bases).

**El AFD indicó posibles filiaciones de las uOTUs.** Como se mencionó brevemente al comienzo de la sección, la filiación de las uOTUs fue reevaluada mediante análisis filogenético de las secuencias ambientales seguido del mapeo de la clasificación obtenida mediante WNBC (AFD, pág. 60).

El alineamiento de las secuencias previo al análisis filogenético fue realizado en etapas y utilizando la información provista por la taxonomía de las secuencias (clasificables) como guía. Dado que esto no se trata de un procedimiento estándar sino que fue desarrollado en este estudio, el mismo es explicado en esta sección.

El gen del ARNr 16S es particularmente difícil de alinear dada la profusión que existe en el mismo de variaciones en la longitud de su secuencia. Además, en el presente estudio se buscó poder utilizar toda la evidencia disponible; es decir, alinear las secuencias Sanger y las secuencias de material tipo con las lecturas relativamente cortas obtenidas mediante SAP. Incluso mediante los métodos más intensivos disponibles en el programa MAFFT [estrategias *linsi* y *ginsi*; (Katoh y Standley, 2014)], no fue posible obtener un alineamiento directamente a partir de las secuencias no alineadas. Por lo tanto, se utilizó una serie de herramientas disponibles en dicho programa las cuales permiten agregar las secuencias desalineadas a un alineamiento de referencia previamente generado.

Como primer paso, se alinearon las secuencias Sanger mediante MAFFT utilizando el refinamiento iterativo basado en la suma pesada de los scores pareados y scores de consistencia obtenidos a partir de alineamientos locales (Katoh y Standley, 2014). Las penalidades de apertura y extensión de *gap* fueron fijadas en 0.1 para reflejar la naturaleza variable del largo de las secuencias. El alineamiento obtenido fue utilizado para obtener una filogenia preliminar mediante Máxima Verosimilitud con el programa PhyML. Se realizaron 10 árboles iniciales mediante secuencias de

---

adición al azar (RAS) seguidas de reordenamientos por medio de intercambio de vecinos más cercanos (NNI) y poda y reinjerto de ramas (SPR) (Guindon et al., 2010). La filogenia obtenida que se observa en la Figura 3.1 (pág. 52) mostró que la mayoría de los grupos identificados mediante el clasificador automático pudieron ser recuperados. Sin embargo, las regiones hipervariables del gen resultaron estar pobremente alineadas en comparación con las regiones conservadas. Por lo tanto, con el objeto de mejorar la asignación de homología primaria [*sensu* M. de Pinna (1991)], se realizó un alineamiento en etapas en el cual cada taxón fue alineado por separado para luego combinar los alineamientos parciales obtenidos mediante MAFFT (opción *-addprofile*). Esto se basó, por un lado, en que el análisis filogenético preliminar resultó ser altamente congruente con la asignación taxonómica, indicando que la segunda puede ser útil para reemplazar (o complementar) el/los árbol guía estándar del proceso de alineamiento dinámico. En segundo lugar, el alineamiento de cada taxón por separado es menos complejo que el alineamiento de todos ellos en un único paso ya que la divergencia en el largo de las secuencias dentro de cada grupo bacteriano es mucho menor que las diferencias de largo observadas entre todos los grupos de bacterias. Una vez generado el alineamiento de referencia, las secuencias representativas de las uOTUs fueron agregadas al mismo utilizando la opción *-seed* del programa MAFFT y se procedió a realizar el AFD.

De forma interesante, el análisis indicó que algunas de las uOTUs podrían pertenecer a, o estar relacionadas con, grupos conocidos. Las uOTUs 0081, 0098, 0136, 0147 and 0165 se agruparon dentro del clado SAR11, el cual además presentó un soporte relativamente elevado (Fig. 3.7 a). Dos uOTUs, (0022 y 0138) resultaron estar distantemente relacionadas con el clado SAR406 del orden *Deferribacterales* (*Marinimicrobia*) (Fig. 3.7 b). De manera similar, la uOTU 0170 se relacionó con secuencias clasificadas como *Gracilibacteria* (Fig. 3.7 b). Tres uOTUs formaron un grupo monofilético con secuencias clasificadas como *Flavobacteriales*, OCS116 (*Rhodobacterales*) y SAR86 (*Oceanospirillales*) (uOTUs 0159, 0067 y 050, respectivamente; Fig. 3.7 a). Varias uOTUs se agruparon con secuencias correspondientes a *Oceanospirillales* (Fig. 3.7 a). De manera similar, muchas uOTUs se localizaron en el árbol cerca de las secuencias correspondientes a *Rickettsiales* y *Rhodobacterales* (Fig. 3.7 a). Dos uOTUs (0103 y 0146) no formaron agrupamientos con ningún grupo descrito y resultaron ser muy divergentes respecto al resto de las secuencias (Fig. 3.7 a, c). Estos resultados se resumen en la Tabla 3.4 (columna *AFD*).

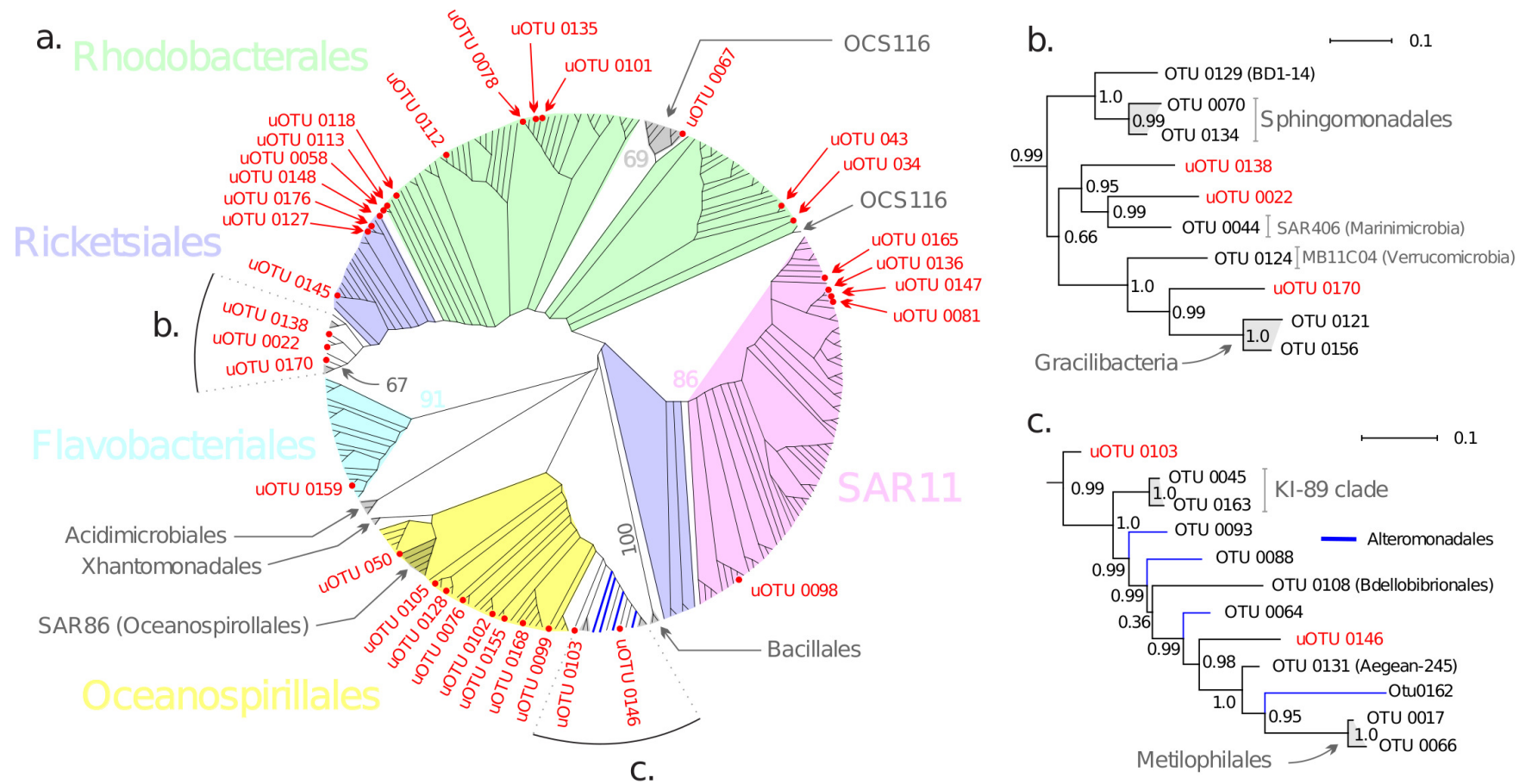


Figura 3.7: Análisis de las OTU no clasificables mediante AFD. El panel **a** muestra un árbol filogenético de las secuencias ambientales con la taxonomía obtenida mediante WNBC mapeada sobre el mismo. Las OTU no clasificables (uOTU) se indican en rojo. Los paneles **b** y **c** muestran vistas detalladas de los clados marcados con las letras **b** y **c** en el panel **a**. Los números ubicados cerca de las ramas corresponden a valores de bootstrap ( $n=100$ , panel **a**) o a soportes locales (*aBayes*; paneles **b** y **c**). Las unidades de las escalas provistas en los paneles **b** y **c** son sustituciones por posición alineada.

---

**Las uOTUs locales fueron similares a uSEQs de locaciones remotas.** Como se mencionó brevemente al comienzo de la sección, las uOTUs fueron comparadas con secuencias existentes en GenBank con el objeto de detectar similitudes con secuencias no clasificables (uSEQs) de otras regiones del Mundo. Esto, además, permitió evaluar la potencial existencia de similitudes con taxa no representados entre las secuencias utilizadas en el AFD. Se realizaron comparaciones contra la base de datos completa y con los subconjuntos de la misma correspondientes a organismos cultivados y a ejemplares tipo (Federhen, 2015).

Las posibles relaciones taxonómicas sugeridas por los análisis mediante BLAST se resumen en la Tabla 3.4 y se presentan en detalle en las Tablas 3.6, 3.7 y 3.8. Estos análisis, además, revelaron que 8 de las uOTUs locales son muy similares a secuencias de sitios remotos (Fig. 3.8, Tabla 3.9; pág 80). La uOTU 0022 presentó un 100 % de similitud con a uSEQ de California y del 99 % con una secuencia de un aislamiento obtenido en el estuario del río Jiulong (China). Esta uOTU también presentó similitudes superiores a 97 % con una gran cantidad de secuencias de otras regiones del Mundo (ver Tabla 3.9). Las uOTUs 0078 y 0081 resultaron ser muy similares a secuencias amplificadas en fiordos de Noruega y la isla de Vancouver (Canadá), respectivamente. Además, la uOTU 0081 presentó una alta similitud con una secuencia obtenida en la region subártica del océano Pacífico. Las uOTUs 0136 y 0138 presentaron una similitud de 97 % con secuencias del norte del Pacífico y Australia, respectivamente. Las uOTUs 0127 y 0148 fueron muy similares a secuencias de aislamientos del golfo de Maine clasificadas como *Alphaproteobacteria* y con 65 secuencias ambientales de diversos orígenes. La uOTU 0170 mostró una similitud del 97 % con secuencias ambientales de Italia.



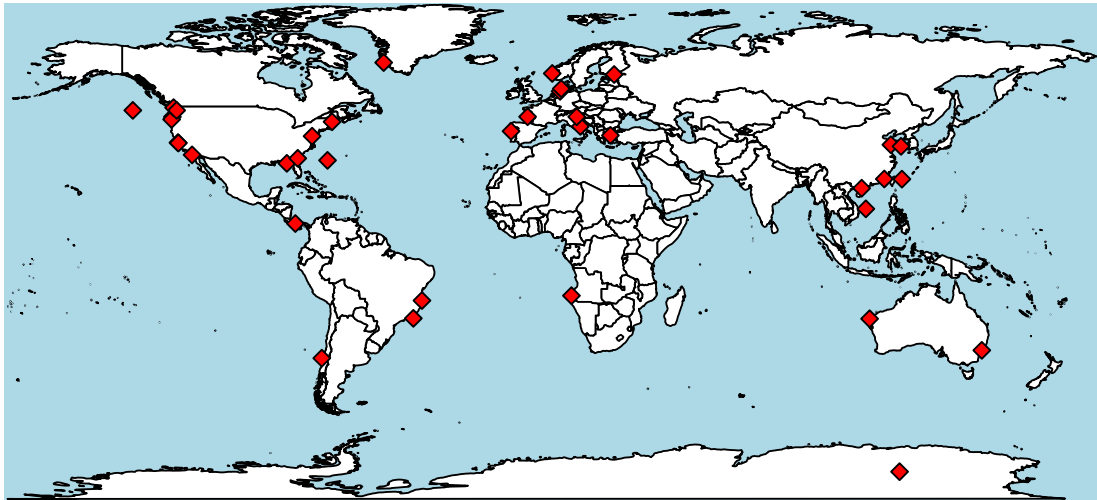


Figura 3.8: Mapa que muestra las locaciones de secuencias de otras regiones del Mundo que presentaron similitudes  $> 97\%$  con las uOTUs de la región de Bahía Engaño. Ver también Tabla 3.9.

**Las similitudes con material tipo fueron muy concordantes con el AFD.** Como se esperaba, la similitud de las secuencias de las uOTU locales con respecto a las secuencias provenientes de material tipo fue bastante baja. Sin embargo, se observó una muy buena concordancia con los resultados obtenidos mediante AFD. Por ejemplo, tanto el AFD como los BLAST contra secuencias cultivadas y material tipo indicaron una potencial relación de la uOTU 0058 con el Orden *Rhodobacterales* (Tabla 3.10).

**Las uSEQs cosmopolitas podrían pertenecer a linajes desconocidos de *Alphaproteobacteria* y *Gracilibacteria*.** Como se detalla más arriba, el AFD sugirió potenciales relaciones entre las uOTU cosmopolitas (Tabla 3.9) y varios grupos de la Clase *Alphaproteobacteria* y los Phyla *Gracilibacteria*, *Verrucomicrobia* y *Marinimicrobia*. En base a dichos resultados, las secuencias de las uOTU y las uSEQs de otras partes del mundo, fueron analizadas filogenéticamente en el marco de 6194 secuencias de referencia pertenecientes a dichos taxa. La finalidad de este procedimiento fue, por un lado, averiguar si las secuencias no clasificables presentan relaciones estrechas con algún grupo conocido, no detectadas por la clasificación automática. Por otro lado, se buscó determinar si cada grupo de secuencias cosmopolitas similares (*e.g.* la uOTU 0022 y las 159 secuencias de otras

---

regiones del mundo; Tabla 3.9, pág. 80) consisten en grupos monofiléticos, lo cual indicaría que corresponden a potenciales taxa no descriptos.

A continuación se detalla la forma en que se compiló el dataset de referencia. Para simplificar la presentación de los resultados, se categorizará a las uOTU cosmopolitas en cuatro categorías: *simil-SAR11*, *simil-Rhodobacterales*, *simil-Rickettsiales* y *Clado-b*. Las uOTU *simil-SAR11* comprenden aquellas uOTU que se agruparon con secuencias del clado SAR11 en el análisis mediante AFD y resultaron ser similares a secuencias provenientes de material tipo de dicho grupo (0081 y 0136; Fig. 3.7 a; Tablas 3.4 y 3.8, pp. 71 y 78, respectivamente). El grupo *simil-Rhodobacterales* incluye la uOTU 0078, la cual se agrupó con secuencias correspondientes al orden *Rhodobacterales* (Fig. 3.7 a). Las *simil-Rickettsiales* incluyen a las uOTUs que agruparon junto a secuencias clasificadas como *Rickettsiales* (uOTUs 0148 y 0127 Fig. 3.7 a). Finalmente, el *Clado-b* incluye las uOTUs 0022, 0138 y 0170, las cuales se agruparon en una rama del árbol conteniendo OTUs de la clase *Alphaproteobacteria* (BD1-14 clade y *Sphingomonadales*) y los phyla *Deferribacteres*, *Verrucomicrobia* (MB11C04) y *Gracilibacteria* (Fig. 3.7 b).

Estudios previos han indicado que los órdenes *Rhodobacterales* y *Rickettsiales* y las SAR11 se agrupan consistentemente en un clado dentro de *Alphaproteobacteria* que también alberga los órdenes *Holosporales*, *Rhodospirillales*, *Sphingomonadales*, *Rhizobiales* y *Caulobacterales* (Ferla et al., 2013; Lee, 2005). Ferla y colaboradores (Ferla et al., 2013) han propuesto dividir dicho clado en dos subclases llamadas *Rickettsidae* (*Rickettsiales* + SAR11 o *Pelagibacterales*) y *Caulobacteridae* (*Holosporales* + *Rhodospirillales* + *Sphingomonadales* + *Rhizobiales* + *Rhodobacterales* + *Caulobacterales*). Por lo tanto se decidió incluir miembros representativos de ambas subclases en el dataset de referencia, así como representantes de una tercer subclase, *Magnetoocccidae*, también propuesta por Ferla y colaboradores (Ferla et al., 2013).

Como ya se ha explicado, el *Clado-b* incluye OTUs asignadas a los phyla *Marinimicrobia*, *Deferribacteres*, *Verrucomicrobia* y *Gracilibacteria* (Fig. 3.7), por lo cual también se incluyeron secuencias de referencia de dichos grupos. También se incluyeron secuencias de referencia de los phyla *Crenarchaeota*, *Euryarchaeota* y *Thaumarchaeota*, las cuales fueron utilizadas para enraizar el árbol. Las secuencias de referencia (n=6194) fueron obtenidas de la base de datos GenBank. Sólo se utilizaron secuencias provenientes de microorganismos cultivados y, en todos los casos en que estuvieron disponibles, sólo aquellas correspondientes a material tipo.

Las secuencias de referencia junto con las uOTU locales y las uSEQs de otras regiones del Mundo (Tabla 3.9, pág. 80) fueron alineadas mediante la estrategia

---

taxonómicamente guiada descrita en la página 62. Se comenzó alineando las secuencias de cada familia en forma individual. Los alineamientos de cada familia fueron combinados para obtener alineamientos correspondientes a cada orden incluido en el dataset de referencia. Luego se obtuvieron alineamientos para cada clase y así sucesivamente para obtener un alineamiento de Bacteria con el grupo externo. Las secuencias correspondientes a las uOTU y las uSEQ de otras regiones del Mundo fueron agregadas a los alineamientos de referencia utilizando las subrutinas *-addfragments* o *-seed*, dependiendo de la calidad de los alineamientos obtenidos en cada caso, los que fueron analizados mediante Jalview. El alineamiento obtenido (6439 secuencias) presentó 3482 posiciones de las cuales 2370 fueron potencialmente informativas, según se determinó con el paquete de R *ips* (Heibl, 2014).

La inferencia filogenética se realizó mediante el programa *FastTree* bajo el modelo GTR + CAT (Ferla et al., 2013). La Figura 3.9 muestra una versión simplificada del árbol obtenido. El árbol que se muestra en la figura fue obtenido mediante el condensado de varios de los grupos monofiléticos presentes en el árbol obtenido mediante *FastTree*. Como puede apreciarse, se recuperó la mayor parte de los taxa incluidos en el dataset de referencia. Las únicas excepciones fueron los órdenes *Rhodospirillales*, *Rhizobiales* y *Rhodobacterales*. Los *Rhodospirillales* se agruparon en un único clado conteniendo a la especie *Terasakiella pusilla* y varias especies de *Reseospirillum*, que actualmente se clasifican en las familias *Methylocystaceae* y *Rhodobiaceae*, respectivamente, del orden *Rhizobiales*. El clado conteniendo el resto de los *Rhizobiales* también incluyó secuencias del género *Ahrensia*, el cual se clasifica en la familia *Rhodobacteraceae* (*Rhodobacterales*). Cabe destacar que el árbol obtenido es consistente con la división de *Alphaproteobacteria* en las 3 subclases propuestas por Ferla y colaboradores (Ferla et al., 2013).

El análisis mostró que la uOTU 0078 y dos secuencias provenientes de Noruega (Tabla 3.9, pág. 80), posiblemente pueda incluirse en el orden *Rhodobacterales*, consistentemente con lo sugerido por el AFD (Fig. 3.7). Así mismo, indicó que las uOTUs 0127 y 0148 se encuentran estrechamente emparentadas con el orden *Rickettsiales*, junto a 66 secuencias de varios puntos remotos del planeta como la Bahía de San Diego, el Mar Adriático, la Bahía Jiaozhou, la Isla de Vancouver y la Ría de Aveiro, entre muchos otros (ver detalles en Tabla 3.9). Las *simil-Rhodobacterales* junto a dos uSEQs obtenidas en un fiordo de Noruega (Newbold et al., 2012) se agruparon en el clado principal del orden *Rhodobacterales* (1030 secuencias). Las *simil-SAR11* y dos uSEQs obtenidas en el Océano Pacífico (Allers et al., 2013) y en la isla de Vancouver (Walsh et al., 2009) se ubicaron como grupo

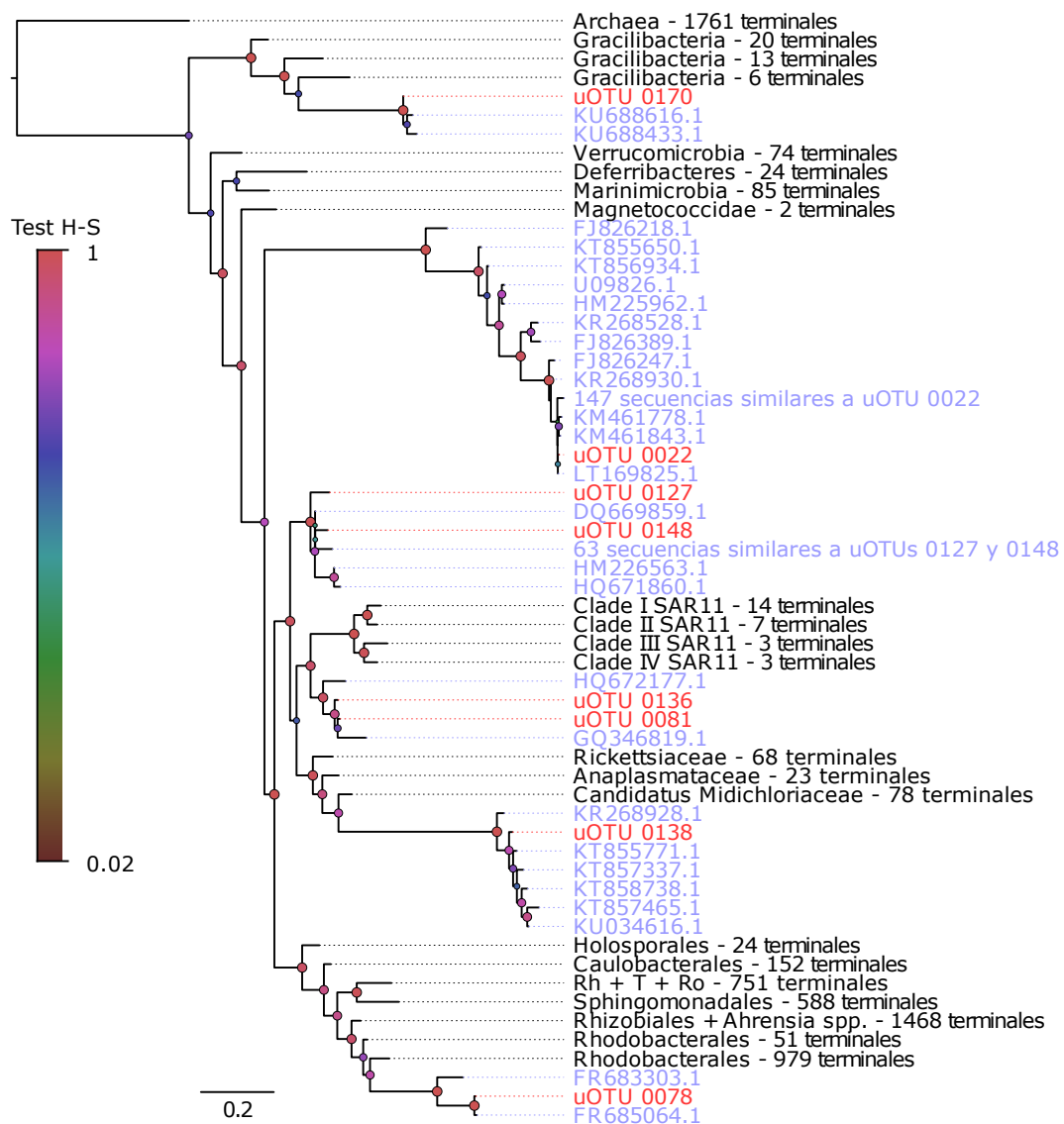


Figura 3.9: Árbol filogenético de 6194 secuencias de referencia y 244 secuencias locales y de otras regiones correspondientes a linajes desconocidos. Las secuencias correspondientes a las uOTU locales se destacan en color rojo y las secuencias de linajes desconocidos de otras partes del Mundo en azul. Los colores asignados a los nodos del árbol corresponden a soportes de H-S según la escala vertical localizada a la izquierda de la figura. La escala ubicada debajo del árbol corresponde a un largo de rama equivalente a 0.2 sustituciones por posición alineada. Muchos grupos monofiléticos fueron condensados para permitir la demostración gráfica del árbol. En dichos casos, el número de terminales colapsados se indica en el nombre de correspondiente terminal. Rh + T + Ro: *Rhodospirillales* + *Terasekiella pusilla* + *Roseospirillum* sp.

---

hermano del clado SAR11. Una de las uOTU del *Clado b* (uOTU 0170) junto a dos uSEQs similares provenientes de Italia (Mancuso et al., 2016), se ubicaron en el clado correspondiente a *Gracilibacteria*. El resto del mismo se agrupó con las *Alphaproteobacteria*. La uOTU 0138 y 4 uSEQs del golfo de Finlandia (Laas et al., 2016) y dos de Austria (Scherr et al., 2016) y Sydney se ubicaron como grupo hermano de la familia *Candidatus Midichloriaceae* del orden *Rickettsiales*. La uOTU 0022 y 159 secuencias de diversos puntos del planeta constituyeron un clado separado de las tres subclases de *Alphaproteobacteria* propuestas por Ferla y colaboradores (Ferla et al., 2013).

En concordancia con el AFD, el análisis filogenético en el marco de secuencias de referencia indicó que la uOTU 0078 junto a las uSEQs similares de otras regiones (Tabla 3.9) podría corresponder a un linaje divergente de *Rhodobacterales* y que las uOTUs 0127 y 0148, junto a 66 secuencias de varios lugares del Mundo (Tabla 3.9) podrían pertenecer a una familia divergente dentro del orden *Rickettsiales*. Además, también pudo corroborarse que las uOTUs 0136 y 0081 junto a dos uSEQs de la isla de Vancouver y el noreste del Océano Pacífico se relacionan estrechamente con el clado SAR11. La uOTU 0138 y seis uSEQs del resto del Mundo se agruparon en una rama larga relativamente cerca de la familia *Candidatus Midichloriaceae* (*Rickettsiales*). Interesantemente, la uOTU 0022 y 159 uSEQs provenientes de diversos puntos geográficos remotos alrededor del Mundo, se ubicaron en un clado divergente dentro de la clase *Alphaproteobacteria*, claramente separado de las 3 subclases propuestas por Ferla y colaboradores (Ferla et al., 2013). La uOTU 0170 y dos uSEQs obtenidas de bacterias marinas epífitas provenientes de Italia (Mancuso et al., 2016) se agruparon en un clado divergente dentro de *Gracilibacteria*.

Tabla 3.4: Unidades taxonómicas operativas no clasificables (uOTUs) y resumen de los análisis de las mismas mediante AFD y BLAST.

uOTU	WNBC <sup>1</sup>	AFD <sup>2</sup>	BLAST <sup>3</sup>	Taxon <sup>4</sup>	% Identidad
034	Alphaproteobacteria	Rhodobacterales	JN591806.1	Clado PS1	93 %
043	Proteobacteria	Rhodobacterales	EU802859.1	Bacteria	90 %
050	Gammaproteobacteria	SAR86	EU802840.1	Bacteria	91 %
0022	Bacteria	Deferribacterales	JQ199557.1	Bacteria	100 %
0058	Alphaproteobacteria	Rhodobacterales	EU799247.1	Bacteria	94 %
0067	Alphaproteobacteria	OCS116	JN591913.1	Clado PS1	94 %
0076	Gammaproteobacteria	Oceanospirillales	FR686176.1	Bacteria	96 %
0078	Alphaproteobacteria	Rhodobacterales	FR685064.1	Bacteria	99 %
0081	Alphaproteobacteria	SAR11	GQ346819.1	Alphaproteobacteria	97 %
0098	Alphaproteobacteria	SAR11	EU802479.1	Bacteria	95 %
0099	Gammaproteobacteria	Oceanospirillales	GU234924.1	Bacteria	96 %
0101	Alphaproteobacteria	Rhodobacterales	FR683303.1	Bacteria	93 %
0102	Gammaproteobacteria	Oceanospirillales	KX935466.1	Bacteria	95 %
0103	Gammaproteobacteria	-	HQ163109.1	Gammaproteobacteria	95 %
0105	Proteobacteria	Oceanospirillales	FR684004.1	Bacteria	93 %
0112	Alphaproteobacteria	Rhodobacterales	DQ234144.2	Rhodobacteraceae	92 %
0113	Alphaproteobacteria	Rhodobacterales	EU799247.1	Bacteria	95 %
0118	Alphaproteobacteria	Rhodobacterales	FJ825915.1	Bacteria	93 %
0127	Alphaproteobacteria	Rickettsiales	JF488534.1	Alphaproteobacteria	98 %
0128	Gammaproteobacteria	Oceanospirillales	JQ269288.1	Bacteria	92 %
0135	Alphaproteobacteria	Rhodobacterales	FR685064.1	Bacteria	96 %
0136	Alphaproteobacteria	SAR11	HQ672177.1	Bacteria	97 %
0138	Bacteria	Deferribacterales	KU034616.1	Bacteria	97 %

Tabla 3.4: (Continuación)

uOTU	WNBC <sup>1</sup>	AFD <sup>2</sup>	BLAST <sup>3</sup>	Taxon <sup>4</sup>	% Identidad
0145	Proteobacteria	Rickettsiales	EU802860.1	Bacteria	90 %
0146	Proteobacteria	-	JN986470.1	Bacteria	96 %
0147	Alphaproteobacteria	SAR11	NR_074224.1	SAR11	95 %
0148	Alphaproteobacteria	Rickettsiales	JF488580.1	Alphaproteobacteria	98 %
0155	Proteobacteria	Oceanospirillales	HQ163230.1	Actinobacteria	94 %
0159	Bacteria	Flavobacteriales	KC336640.1	Bacteria	89 %
0165	Proteobacteria	SAR11	NR_074224.1	SAR11	95 %
0168	Gammaproteobacteria	Oceanospirillales	JQ269279.1	Bacteria	94 %
0170	Bacteria	Gracilibacteria	KU688616.1	Bacteria	97 %
0176	Alphaproteobacteria	Rickettsiales	JQ199248.1	Bacteria	93 %

<sup>1</sup> Asignación mediante clasificador naïve bayesiano de Wang (*cutoff* de 80 %).

<sup>2</sup> Taxon más cercano según el análisis filogenético.

<sup>3</sup> Número de acceso de la secuencia con el score de BLAST más alto en GenBank.

<sup>4</sup> Taxonomía anotada en la secuencia con el score de BLAST más alto en GenBank.

Tabla 3.5: Número de secuencias y OTUs clasificables y no clasificables en datos generados mediante secuenciación Sanger y de alta performance (*SAP*).

	OTUs		Secuencias	
	Sanger	SAP	Sanger	SAP
Clasificable	70	147	238	1960
No Clasificable	3 (4.1 %)*	30 (16.9 %)*	3 (0.8 %)	37 (1.8 %)
Total	73	177	241	1997

\*  $p = 0.01169$ ; Test  $\chi^2$  de Pearson.



Tabla 3.6: Comparación de las uOTU patagónicas con secuencias presentes en GenBank.

uOTU	Blast hit	Taxón <sup>1</sup>	% Identidad <sup>2</sup>	Fuente <sup>3</sup>	Origen <sup>4</sup>
034	JN591806.1	Clado PS1	93 %	Agua de Mar	Puget Sound
043	EU802859.1	Bacteria	90 %	Agua de Mar	Panamá
050	EU802840.1	Bacteria	91 %	Agua de Mar	Panamá
0022	JQ199557.1	Bacteria	100 %	Agua de Mar	California
0058	EU799247.1	Bacteria	94 %	Zona de Mezcla (estuario)	Newport
0067	JN591913.1	Clado PS1	94 %	Agua de Mar	Puget Sound
0076	FR686176.1	Bacteria	96 %	Fiordo	Noruega
0078	FR685064.1	Bacteria	99 %	Fiordo	Noruega
0081	GQ346819.1	Alphaproteobacteria	97 %	Fiordo	Isla de Vancouver
0098	EU802479.1	Bacteria	95 %	Agua de Mar	Panamá
0099	GU234924.1	Bacteria	96 %	Agua de Mar	Antártida
0101	FR683303.1	Bacteria	93 %	Fiordo	Noruega
0102	KT906757.1	Bacteria	95 %	Agua de Mar	Bahía Monterrey
0103	HQ163109.1	Gammaproteobacteria	95 %	Agua de Mar	Isla de Vancouver
0105	FR684004.1	Bacteria	93 %	Fiordo	Noruega
0112	DQ234144.2	Rhodobacteraceae	92 %	Estuario	China
0113	EU799247.1	Bacteria	95 %	Zona de Mezcla (estuario)	Newport
0118	FJ825915.1	Bacteria	93 %	Agua de Mar	Mar Amarillo
0127	JN986397.1	Bacteria	95 %	Agua de Mar	Giro del Pacífico Sur
0128	JQ269288.1	Bacteria	92 %	Estuario	China
0135	FR685064.1	Bacteria	96 %	Fiordo	Noruega
0136	HQ672177.1	Bacteria	97 %	Agua de Mar	Pacífico Norte
0138	KU034616.1	Bacteria	97 %	Acuífero subterráneo	Austria
0145	EU802860.1	Bacteria	90 %	Agua de Mar	Panamá

Tabla 3.6: (Continuación)

uOTU	Blast hit	Taxón <sup>1</sup>	% Identidad <sup>2</sup>	Fuente <sup>3</sup>	Origen <sup>4</sup>
0146	JN986470.1	Bacteria	96 %	Agua de Mar	Pacífico Sur
0147	FN435574.1	Bacteria	95 %	Agua de Mar	Mallorca
0148	KX933169.1	Bacteria	97 %	Agua de Mar	California
0155	HQ163230.1	Actinobacteria	94 %	Agua de Mar	Isla de Vancouver
0159	KC336640.1	Bacteria	89 %	Agua de Mar	ND
0165	FJ529365.1	Alphaproteobacteria	93 %	Reef sponge	Caribbean Sea
0168	HQ241956.1	Bacteria	96 %	Agua de Mar	Pacífico Nor Este
0170	KU688616.1	Bacteria	97 %	Algas	Italia
0176	JQ199248.1	Bacteria	93 %	Agua de Mar	California

<sup>1</sup> Clasificación anotada en la base de datos.

<sup>2</sup> Porcentaje de similitud con la secuencia de mayor escore de GenBank.

<sup>3</sup> Material muestreado.

<sup>4</sup> Origen geográfico.

Tabla 3.7: Comparación de las uOTU patagónicas con secuencias presentes en GenBank provenientes de microorganismos cultivados.

uOTU	BLAST hit	Taxón <sup>1</sup>	% Identidad <sup>2</sup>	Fuente <sup>3</sup>	Origen <sup>4</sup>
034	KM273259.1	Rhizobiales	92 %	Intestino de calamar	ND
043	KR003320.1	Bacteria	85 %	Agua de Mar	Francia
050	JQ269288.1	Bacteria	88 %	Estuario	China
0022	JQ269283.1	Bacteria	99 %	Estuario	China
0058	JX624253.1	Rhodobacterales	92 %	Tierra ganada	Corea
0067	HQ675302.1	Alphaproteobacteria	91 %	Agua de Mar	Hawai
0076	HQ675262.1	Oceanospirillales	94 %	Agua de Mar	Hawai
0078	EU346504.1	Bacteria	96 %	Esponja marina	California
0081	JX438790.1	Alphaproteobacteria	95 %	Agua de Mar	Pacífico occidental
0098	LN850157.1	SAR11	90 %	Agua de Mar	Mar Rojo
0099	HQ675262.1	Oceanospirillales	94 %	Agua de Mar	Hawai
0101	KU725768.1	Rhodobacterales	92 %	Algas	ND
0102	HQ675262.1	Oceanospirillales	94 %	Agua de Mar	Hawai
0103	KJ491976.1	Bacteria	91 %	Agua de Mar	Corea
0105	JQ269279.1	Bacteria	89 %	Estuario	China
0112	CP018572.1	Rhodobacterales	92 %	Agua de Mar	Taiwan
0113	FJ196061.1	Rhodobacterales	91 %	Agua de Mar	Antártida
0118	KJ492127.1	Bacteria	91 %	Agua de Mar	Corea
0127	JF488534.1	Alphaproteobacteria	98 %	Agua de Mar	Golfo de Maine
0128	JQ269288.1	Bacteria	92 %	Estuario	China
0135	KJ492127.1	Bacteria	94 %	Agua de Mar	Corea
0136	KJ492216.1	Bacteria	95 %	Agua de Mar	Corea
0138	JF488542.1	Alphaproteobacteria	95 %	Agua de Mar	Golfo de Maine

Tabla 3.7: (Continuación)

uOTU	BLAST hit	Taxón <sup>1</sup>	% Identidad <sup>2</sup>	Fuente <sup>3</sup>	Origen <sup>4</sup>
0145	HQ675326.1	Alphaproteobacteria	88 %	Agua de Mar	Hawai
0146	KJ493944.1	Bacteria	88 %	Agua de Mar	Mar Rojo
0147	KJ492216.1	Bacteria	94 %	Agua de Mar	Corea
0148	JF488580.1	Alphaproteobacteria	98 %	Agua de Mar	Golfo de Maine
0155	CP010552.1	Candidatus Thioglobus	89 %	Agua de Mar	Isla de Vancouver
0159	NR_113854.1	Cytophagales	85 %	ND	ND
0165	JX439616.1	Alphaproteobacteria	91 %	Agua de Mar	Pacífico occidental
0168	JQ269279.1	Bacteria	94 %	Estuario	China
0170	CP011268.1	SR1	80 %	Barro activado	Dinamarca
0176	JX439497.1	Alphaproteobacteria	93 %	Agua de Mar	Pacífico occidental

<sup>1</sup> Clasificación anotada en la base de datos.

<sup>2</sup> Porcentaje de similitud con la secuencia de mayor score de GenBank.

<sup>3</sup> Fuente de la muestra.

<sup>4</sup> Origen geográfico.

Tabla 3.8: Comparación de las uOTU patagónicas con secuencias presentes en GenBank provenientes de material tipo.

uOTU	BLAST hit	Taxón <sup>1</sup>	% Identidad <sup>2</sup>	Fuente <sup>3</sup>	Origen <sup>4</sup>
034	KM273259.1	Rhizobiales	92 %	Intestino de calamar	Corea
043	KT966463.1	Cellvibrionales	84 %	Agua de Mar	Taiwan
050	KM017973.1	Gammaproteobacteria	85 %	Sedimento de playa de marea	Corea
0022	FJ858267.1	Micromonas chloroplast	89 %	Agua de Mar	Pacífico Sur
0058	NR_043271.1	Rhodobacterales	92 %	Solar saltern	Corea
0067	NR_109323	Rhodobacterales	89 %	Sea squirt	Corea
0076	AP012273.1	Gammaproteobacteria	86 %	Hydrothermal vent	Okinawa Trough
0078	NR_133962.1	Rhodobacterales	94 %	Agua de Mar	Japón
0081	NR_074224.1	SAR11	95 %	Agua de Mar	Mar de los Sargazos
0098	NR_074269.1	Rhizobiales	88 %	Soil	ND
0099	NR_041001.1	Cellvibrionales	89 %	Sedimento de profundidad	Japón
0101	NR_133962.1	Rhodobacterales	90 %	Agua de Mar	Japón
0102	KM017973.1	Gammaproteobacteria	87 %	Sedimento de playa de marea	Corea
0103	NR_121684.1	Cellvibrionales	90 %	Agua de Mar	Germany
0105	NR_133962.1	Rhodobacterales	87 %	Agua de Mar	Japón
0112	KR185328.1	Rhodobacterales	92 %	Agua de Mar	Corea
0113	NR_134070.1	Rhodobacterales	91 %	Agua de Mar	Corea
0118	NR_133962.1	Rhodobacterales	91 %	Agua de Mar	Japón
0127	KJ782430.2	Parvularculales	91 %	Hypersaline lake	China
0128	NR_041567.1	Oceanospirillales	88 %	Whale carcasses	Japón
0135	NR_133962.1	Rhodobacterales	92 %	Agua de Mar	Japón
0136	NR_074224.1	SAR11	95 %	Agua de Mar	Mar de los Sargazos
0138	CP010969.1	Rickettsiales	85 %	Tick	Russia

Tabla 3.8: (Continuación)

uOTU	BLAST hit	Taxón <sup>1</sup>	% Identidad <sup>2</sup>	Fuente <sup>3</sup>	Origen <sup>4</sup>
0145	NR_117412.1	Cellvibrionales	84 %	Sand from seashore	Corea
0146	NR_044243.1	Cellvibrionales	86 %	Marine solar saltern	Corea
0147	NR_074224.1	SAR11	95 %	Agua de Mar	Mar de los Sargasos
0148	CP012403.1	Rhodospirillales	89 %	Sulfide spring	Russia
0155	CP010552.1	Gammaproteobacteria	89 %	Agua de Mar	Isla de Vancouver
0159	NR_113854.1	Cytophagales	85 %	ND	ND
0165	NR_074224.1	SAR11	95 %	Agua de Mar	Mar de los Sargasos
0168	NR_044243.1	Cellvibrionales	86 %	Marine solar saltern	Corea
0170	NR_041633.1	Acidimicrobiales	78 %	Estuario sediment	Japón
0176	NR_104756.1	Rhizobiales	88 %	ND	ND

<sup>1</sup> Clasificación anotada en la base de datos.

<sup>2</sup> Porcentaje de similitud con la secuencia de mayor escore de GenBank.

<sup>3</sup> Fuente de la muestra.

<sup>4</sup> Origen geográfico.

Tabla 3.9: Secuencias de otras regiones del Mundo que presentaron similitudes significativas ( $> 97\%$ ) con las uOTUs de la zona de estudio.

uOTU	GI	Clasificación	Origen	Referencia
	JQ195135.1, JQ195183.1, JQ195408.1, JQ195795.1, JQ195800.1, JQ195880.1, JQ195899.1, JQ196115.1, JQ196134.1, JQ196139.1, JQ196143.1, JQ196401.1, JQ196458.1, JQ196467.1, JQ196553.1, JQ196580.1, JQ196597.1, JQ196599.1, JQ196623.1, JQ196683.1, JQ196706.1, JQ196734.1, JQ196826.1, JQ196916.1, JQ197059.1, JQ197067.1, JQ197696.1, JQ197826.1, JQ197851.1, JQ197961.1, JQ197991.1, JQ198077.1, JQ198214.1, JQ198301.1, JQ198399.1, JQ198413.1, JQ198452.1, JQ198526.1, JQ199275.1, JQ199296.1, JQ199360.1, JQ199370.1, JQ199397.1, JQ199545.1, JQ199557.1, JQ199863.1, JQ200035.1, JQ200190.1	Bacteria	Bahía de San Diego	(Bik et al., 2016)
	KX935074.1, KX938295.1-KX938300.1	Eukaryota (plastido)	Bahía Monterrey	(Choi et al., 2017)
	HQ241941.1, HQ242127.1, HQ242190.1, HQ242218.1, HQ242640.1	Cyanobacteria/Bacteria	Pacífico Nor Este	ND
	HQ427471.1, HQ427470.1	Bacteria	San Diego	ND
0022	U09826.1	Bacteria	ND	ND
	HQ836711.1	Bacteria	Brazil	ND
	GU170776.1, GU170747.1	Bacteria	Golfo de México	(Olapade, 2010)

Tabla 3.9: (Continuación)

uOTU	GI	Clasificación	Origen	Referencia
	GU200174.1, GU200446.1, GU200430.1, GU200623.1, GU200577.1, GU200607.1, GU200603.1 GU119870.1	Bacteria	Brasil	(de Castro et al., 2010)
		Chloroplast	Panamá	(Sunagawa et al., 2010)
	GU061480.1, GU061828.1, GU061467.1, GU061462.1, GU061835.1, GU061481.1, GU061470.1, GU061460.1	Bacteria	Mar de China	(Zhang et al., 2011)
	GU184525.1	Bacteria	Australia	(Ceh et al., 2011)
	KT856934.1, KT855650.1	Bacteria	Golfo de Finlandia	ND
	KT318649.1	Bacteria	Golfo de México	ND
	KU578831.1	Bacteria	Golfo de México	ND
	JX405718.1, JX406109.1, JX406123.1, JX405857.1, JX405793.1, JX405907.1, JX406129.1	Bacteria	Bahía Zhanjiang	ND
	KC197669.1	Bacteria	Mar de los Sargazos	(Sjostedt et al., 2014)
	AB294987.1	Bacteria	Japón	(Hirayama et al., 2007)



Tabla 3.9: (Continuación)

uOTU	GI	Clasificación	Origen	Referencia
	JF968597.1	Eukaryota	Mar de los Sargasos	ND
	AF406501.1, AF406507.1	Prasinophyte	Norte de Mar Egeo	(Moeseneder et al., 2005)
	KF928825.1	Bacteria	Antártida	ND
	KJ870962.1	Cyanobacteria	Antártida	(Michaud et al., 2014)
	AB974149.1, AB974153.1	Bacteria	China Sea	ND
	AF001660.1	Eukaryota	Costa de Oregon	(Rappe et al., 1998)
	AY628654.1, AY628655.1	Prasinophyceae	Bahía Tillamook	(Bernhard et al., 2005)
	AY663945.1	Eukaryota	China Sea	ND
	EU236308.1	Bacteria	California	(Sipkema et al., 2009)
	EU258745.1, EU258747.1, EU258747.1, EU258748.1, EU259787.1, EU259794.1, EU259802.1	Bacteria	Mar Amarillo	ND
	EU394554.1, EU394571.1	Prasinophyceae	Bahía de Biscay	ND
	FJ826218.1, FJ826315.1	Bacteria	Mar Amarillo	
	FJ826247.1, FJ826264.1, FJ826280.1, FJ826389.1, FJ826491.1	Eukaryota	Mar Amarillo	(Liu et al., 2013)
	FJ875452.1, FJ875497.1	Bacteria	California	ND
	GQ246247.1, GQ246281.1	Bacteria	Mar Amarillo	ND
	HM134422.1,	Eukaryote	Mar de China	ND
	HM134442.1, HM134692.1	Bacteria	Mar de China	ND
	HM225962.1, HM226488.1	Bacteria	Brasil	(Cury et al., 2011)

Tabla 3.9: (Continuación)

uOTU	GI	Clasificación	Origen	Referencia
	HM437689.1	Bacteria	China	(Liu et al., 2011)
	JQ269283.1, JQ269287.1	Bacteria	China	ND
	KM461778.1, KM461843.1	Cyanobacteria	Chile	ND
	KR268528.1, KR268930.1	Bacteria	Sydney Harbour	ND
	LT169087.1, LT169825.1	Bacteria	Mar del Norte	(Lokmer y Mat- hias Wegner, 2015)
0078	FR685064.1, FR683303.1	Bacteria	Noruega	(Newbold et al., 2012)
0081	GQ346819.1	Alphaproteobacteria	Isla de Vancouver	(Walsh et al., 2009)
	HQ672177.1	Bacteria	Subártico del Pacífico	(Allers et al., 2013)
0127	JF488534.1, JF488513.1, JF488440.1, JF488580.1, JF488510.1, JF488499.1, JF488497.1, JF488494.1	Alphaproteobacteria	Golfo de Maine	(Martinez-Garcia et al., 2012)
0136	HQ672177.1	Bacteria	Subártico del Pacífico	(Allers et al., 2013)
	GQ346819.1	Alphaproteobacteria	Isla de Vancouver	(Walsh et al., 2009)
0138	KT855771.1, KT857337.1, KT858738.1, KT857465.1	Bacteria	Golfo de Finlandia	(Laas et al., 2016)
	KR268928.1	Rickettsiales	Sydney Harbour	ND
	KU034616.1	Bacteria	Austria	(Scherr et al., 2016)

Tabla 3.9: (Continuación)

uOTU	GI	Clasificación	Origen	Referencia
	KX933013.1, KX933407.1, KX934581.1, KX935622.1, KX935625.1	KX933169.1, KX933408.1, KX934584.1,	Bacteria	Bahía Monterrey (Choi et al., 2017)
	KJ094115.1		Bacteria	ND
	JQ195189.1, JQ196077.1, JQ196209.1, JQ197130.1, JQ197232.1, JQ197249.1, JQ199838.1		Bacteria	Bahía de San Diego (Bik et al., 2016)
	HE863612.1		Bacteria	Groenlandia ND
	JX016138.1, JX016921.1, JX017087.1, JX017131.1, JX017132.1		Bacteria	Helgoland (Teeling et al., 2012)
	JX864168.1		Bacteria	Mar Adriático (Tinta et al., 2015)
	JN591813.1-JN591816.1, JN591917.1,		Rhodospirillales	Puget Sound (Iverson et al., 2012)
0148	JN591918.1			
	HQ242266.1		Alphaproteobacteria	Pacífico Nor Este ND
	HQ671860.1		Bacteria	Subártico del Pacífico (Allers et al., 2013)
	HM226563.1		Bacteria	Brazil (Cury et al., 2011)
	HM802383.1		Alphaproteobacteria	Isla Sapelo (Mou et al., 2011)
	HM437421.1		Bacteria	China (Liu et al., 2011)
	FR684591.1		Bacteria	Noruega (Newbold et al., 2012)
	GU460770.1		Bacteria	Atlántico Sur (Morris et al., 2010)
	GQ350083.1, GQ349665.1, GQ348758.1	GQ350058.1,	Alphaproteobacteria	Isla de Vancouver (Walsh et al., 2009)

Tabla 3.9: (Continuación)

uOTU	GI	Clasificación	Origen	Referencia
	FJ826263.1, FJ826475.1, FJ826160.1, FJ826078.1	Bacteria	Mar Amarillo	(Liu et al., 2013)
	FJ745251.1, FJ745111.1	Alphaproteobacteria	Isla Sapelo	(Poretsky et al., 2010)
	EU394570.1	Alphaproteobacteria	Bahía de Biscay	ND
	EU801601.1, EU801601.1, EU799325.1	Bacteria	Bahía de Chesapeake	(Shaw et al., 2008)
	EU258753.1	Bacteria	ND	ND
	EU131348.1	Alphaproteobacteria	Noruega	ND
	AY627381.1	SAR11	Bahía Monterrey	(Suzuki et al., 2004)
	AY033298.1	SAR116	Bahía Monterrey	(Suzuki et al., 2001)
	AY499449.1	Alphaproteobacteria	Portugal	(Henriques et al., 2004)
	AY476783.1	Alphaproteobacteria	ND	ND
	DQ669859.1	Alphaproteobacteria	ND	(Pommier et al., 2007)
	LT549205.1	Bacteria	Boothbay Harbor	ND
0170	KU688616.1, KU688433.1	Bacteria	Italia	(Mancuso et al., 2016)

Tabla 3.10: Resumen de los resultados obtenidos mediante WNBC, AFD, y BLAST personalizados.

uOTU	WNBC <sup>1</sup>	AFD <sup>2</sup>	BLAST-A <sup>3</sup>	BLAST-C <sup>4</sup>	BLAST-T <sup>5</sup>
034	Alphaproteobacteria	Rhodobacterales	PS1	Rhizobiales	Rhizobiales
043	Proteobacteria	Rhodobacterales	Bacteria	Bacteria	Cellvibrionales
050	Gammaproteobacteria	SAR86	Bacteria	Bacteria	Gammaproteobacteria
0022	Bacteria	Deferribacterales	Bacteria*	Bacteria*	Micromonas
0058	Alphaproteobacteria	Rhodobacterales	Bacteria	Rhodobacterales	Rhodobacterales
0067	Alphaproteobacteria	OCS116	PS1	Alphaproteobacteria	Rhodobacterales
0076	Gammaproteobacteria	Oceanospirillales	Bacteria	Oceanospirillales	Gammaproteobacteria
0078	Alphaproteobacteria	Rhodobacterales	Bacteria*	Bacteria	Rhodobacterales
0081	Alphaproteobacteria	SAR11	Alphaproteobacteria*	Alphaproteobacteria	SAR11
0098	Alphaproteobacteria	SAR11	Bacteria	SAR11	Rhizobiales
0099	Gammaproteobacteria	Oceanospirillales	Bacteria	Oceanospirillales	Cellvibrionales
0101	Alphaproteobacteria	Rhodobacterales	Bacteria	Rhodobacterales	Rhodobacterales
0102	Gammaproteobacteria	Oceanospirillales	Bacteria	Oceanospirillales	Gammaproteobacteria
0103	Gammaproteobacteria	-	Gammaproteobacteria	Bacteria	Cellvibrionales
0105	Proteobacteria	Oceanospirillales	Bacteria	Bacteria	Rhodobacterales
0112	Bacteria	Deferribacterales	Bacteria	Alphaproteobacteria	Rickettsiales
0113	Alphaproteobacteria	Rhodobacterales	Bacteria	Rhodobacterales	Rhodobacterales
0118	Alphaproteobacteria	Rhodobacterales	Bacteria	Bacteria	Rhodobacterales
0127	Alphaproteobacteria	Rickettsiales	Bacteria	Alphaproteobacteria*	Parvularculales
0128	Gammaproteobacteria	Oceanospirillales	Bacteria	Bacteria	Oceanospirillales
0135	Alphaproteobacteria	Rhodobacterales	Bacteria	Bacteria	Rhodobacterales
0136	Alphaproteobacteria	SAR11	Bacteria*	Bacteria	SAR11
0138	Bacteria	Deferribacterales	Bacteria*	Alphaproteobacteria	Rickettsiales
0145	Proteobacteria	Rickettsiales	Bacteria	Alphaproteobacteria	Cellvibrionales

Tabla 3.10: (Continuación)

uOTU	WNBC <sup>1</sup>	AFD <sup>2</sup>	BLAST-A <sup>3</sup>	BLAST-C <sup>4</sup>	BLAST-T <sup>5</sup>
0146	Bacteria	-	Bacteria	Bacteria	Cellvibrionales
0147	Alphaproteobacteria	SAR11	Bacteria	Bacteria	SAR11
0148	Alphaproteobacteria	Rickettsiales	Bacteria <sup>*</sup>	Alphaproteobacteria <sup>*</sup>	Rhodospirillales
0155	Proteobacteria	Oceanospirillales	Actinobacteria	Gammaproteobacteria	Gammaproteobacteria
0159	Bacteria	Flavobacteriales	Bacteria	Cytophagales	Cytophagales
0165	Proteobacteria	SAR11	Alphaproteobacteria	Alphaproteobacteria	SAR11
0168	Gammaproteobacteria	Oceanospirillales	Bacteria	Bacteria	Cellvibrionales
0170	Bacteria	Gracilibacteria	Bacteria <sup>*</sup>	SR1	Acidimicrobiales
0176	Alphaproteobacteria	Rickettsiales	Bacteria	Alphaproteobacteria	Rhizobiales

<sup>1</sup> Asignación mediante WNBC.

<sup>2</sup> Pariente más cercano según el AFD.

<sup>3</sup> Taxonomía del hit de mayor escore resultante de los BLAST contra la base de datos GenBank completa.

<sup>4</sup> Taxonomía del hit de mayor escore resultante de los BLAST contra las secuencias presentes en GenBank correspondientes a organismos cultivados.

<sup>5</sup> Taxonomía del hit de mayor escore resultante de los BLAST contra las secuencias presentes en GenBank correspondientes a material tipo.

<sup>\*</sup> Significativamente similar (>97 %).

---

### 3.3. Análisis complementarios de picoeucariotas

Además de secuencias bacterianas, las bibliotecas génicas del gen de ARNr 16S presentaron secuencias de cloroplastos de picoeucariotas. Esta observación fue confirmada mediante análisis complementarios basados en *metabarcoding* del gen ARNr 18S y aislamiento microbiológico de dichos microorganismos.

Para el análisis mediante *metabarcoding*, se amplificaron secuencias correspondientes al gen de ARNr 18S a partir del ADN metagenómico de pico-nanoplancton (filtro de 0,22  $\mu\text{m}$ ) correspondiente a la muestra VER11. El producto obtenido fue clonado y se secuenciaron 35 clones elegidos al azar. Treinta y dos clones presentaron secuencias que resultaron ser muy similares a secuencias de especies de *Ostreococcus* de otras regiones. Estas 32 secuencias correspondieron a 10 haplotipos (secuencias únicas). Las secuencias locales fueron combinadas con secuencias de referencia de otras *Mamiellophyceae*, incluyendo los clados de *Ostreococcus* descritos con anterioridad (Guillou et al., 2004; Rodríguez et al., 2005), y sometidas a análisis filogenético y mediante Parsimonia Estadística. El dataset presentó 1623 posiciones alineadas, de las cuales 194 fueron variables y 154 filogenéticamente informativas. El análisis filogenético confirmó que las secuencias locales corresponden a una especie de *Ostreococcus* (Fig. 3.10).

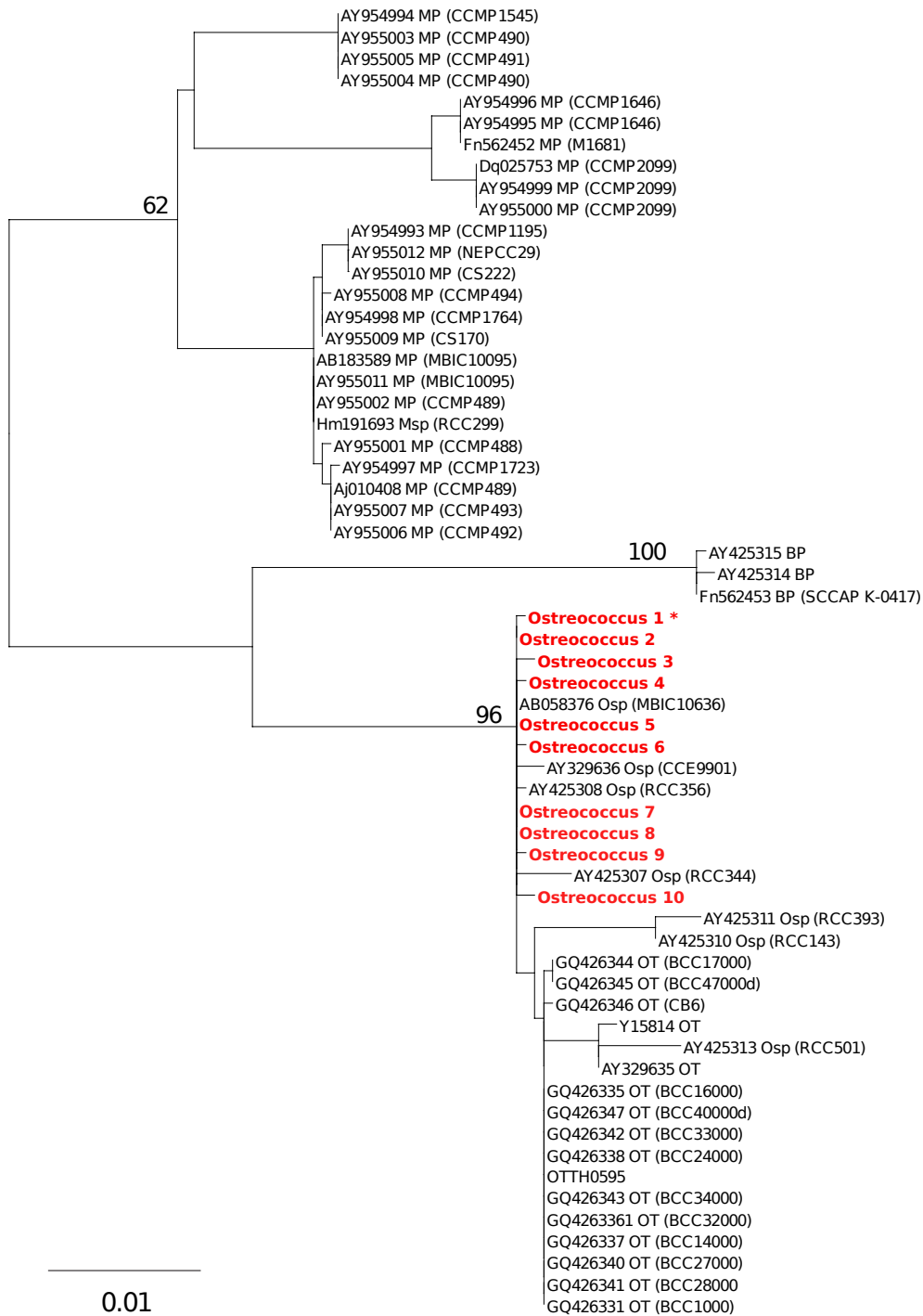


Figura 3.10: Árbol filogenético de secuencias del gen ARNr 18S de picoeucariotas de la zona de estudio (rojo) y secuencias homólogas de especies de la familia *Mamiellophyceae*, obtenido mediante el criterio de máxima verosimilitud. El asterisco indica el haplotipo que presentó el cultivo obtenido en este trabajo. El alineamiento utilizado para inferir el árbol presentó 981 posiciones. Los números sobre las ramas corresponden a soportes de *bootstrap* (n=100). El número de acceso a GenBank de las secuencias de referencia se indica en los terminales del árbol, seguido de la especie y cepa correspondiente en los casos en que el mismo figura en la base de datos. MP *Micromonas pusilla*, Msp *Micromonas* sp., BP *Bathycoccus prasinus*, Osp *Ostreococcus* sp., OT *Ostreococcus tauri*. La escala en el ángulo inferior izquierdo está en sustituciones por posición alineada. Cada una de las secuencias en rojo corresponde a uno de los haplotipos (secuencias únicas) locales.



---

El aislamiento microbiológico, seguido de análisis genético mediante secuenciación del gen de ARNr 18S, corroboró la presencia de *Ostreococcus* sp. y *Micromonas* sp., respectivamente, en las muestras E12 y J12 (Tabla 3.1, pág. 55).

### 3.4. Análisis virológicos

Con el objeto de evaluar la presencia de virus de cianobacterias se realizaron PCR específicas contra los mismos a partir de todas las muestras disponibles. En todos los casos estas reacciones brindaron resultados negativos, lo cual es compatible con la escasez de secuencias de sus huéspedes evidenciadas por los análisis descriptos en la sección 3.1 (resumido en la Tabla 3.1).

Por el contrario, la presencia de virus de picoalgas pudo evidenciarse mediante PCR en 7 muestras: VER11, OTO11, INV11, E12, F12, J12 y D12. Los resultados obtenidos a partir de muestras tomadas durante el 2012 pueden apreciarse en la Figura 3.11. En algunas muestras, dichas secuencias correspondieron a *Ostreococcus Virus* (VER11, OTO11, E12, F12, D12) y en otras a *Micromonas Virus* (INV11, J12). Estos resultados son compatibles con lo obtenidos a partir de secuenciación de los genes ARNr 16S y 18S y el aislamiento microbiológico, los cuales indicaron que la presencia de picoeucariotas de la familia *Mamiellophyceae* es común en la zona de estudio.

Se procedió a caracterizar en detalle los *Ostreococcus virus* (OV) presentes en las muestras correspondientes a meses de verano, las cuales mostraron la mayor cantidad de secuencias de picoeucariotas.

La estrategia utilizada consistió en amplificar los virus retenidos en las membranas de 0,22  $\mu\text{m}$  (virus intracelulares) mediante iniciadores dirigidos a la ADN polimerasa viral. Posteriormente, los amplicones obtenidos fueron clonados y se secuenciaron 84 clones mediante secuenciación Sanger. Las secuencias obtenidas correspondieron a 7 haplotipos (h1-h7; Tabla 3.11).

Las secuencias patagónicas fueron combinadas con secuencias de otras regiones del mundo, incluyendo otros *O. virus* y secuencias de virus de especies de *Bathycoccus* y *Micromonas*, las cuales se utilizaron como grupo externo (Tabla 3.12). Una vez alineadas las secuencias, se obtuvo un dataset de 478 posiciones de las cuales 238 presentaron polimorfismos y 217 fueron filogenéticamente informativas.

El análisis mediante PE reveló la existencia de 10 *clusters* y 8 *singletones* (secuencias que no agrupan con ninguna otra). Las secuencias correspondientes al grupo externo (virus de *Bathycoccus* y *Micromonas* spp.) se agruparon en forma separada del grupo interno. Las dos secuencias de *Bathycoccus prasinus Virus* (BV)

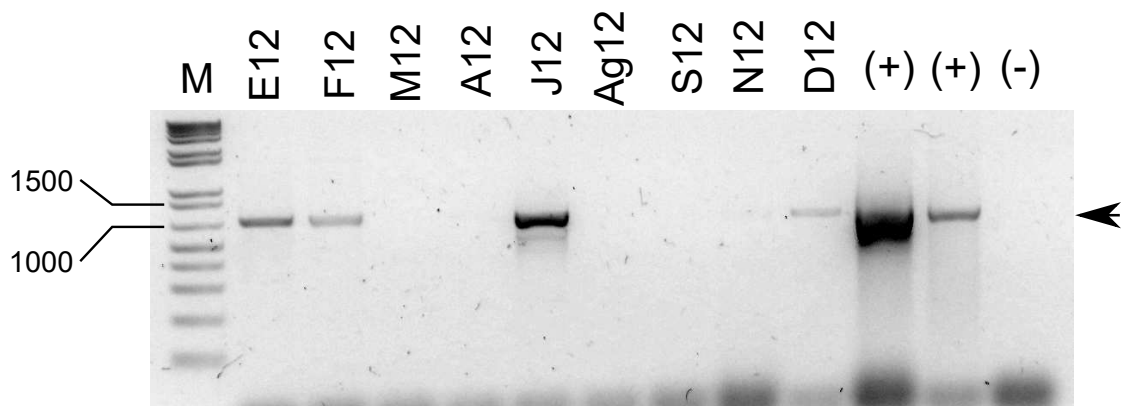


Figura 3.11: Detección de *Prasinovirus* mediante PCR en muestras tomadas con una frecuencia aproximadamente mensual (E12, F12, M12, A12, J12, Ag12, S12, N12 y D12). *M* marcador de peso molecular 1 Kb Plus donde se indican como referencia las bandas de 1000 y 1500 pb (bandas 100, 200, 300, 400, 500, 650, 850, 1000, 1500, 2000-15000 pb); + control positivo (izquierda: ADN proveniente de cultivo de *Ostreococcus* sp. infectado con *O. virus*; derecha: concentrado de filtración tangencial de la muestra VER11); - control negativo. La flecha indica la posición de la banda correspondiente al ADN blanco. Ver también Tabla 3.1.

Tabla 3.11: Haplotipos de *Ostreococcus virus* presentes en la muestra VER11.

Haplotipo	Clones
h1	27
h2	2
h3	9
h4	1
h5	1
h6	8
h7	36

y una secuencia de *Micromonas pusilla Virus* (MV) constituyeron 3 *singletons* separados, mientras que el resto de las secuencias de MV se agruparon en un único *network*. Cinco secuencias de OV no pudieron conectarse con ningún otro haplotipo. El resto de las secuencias de OV conformaron 9 *networks*, uno de los cuales, interesantemente, incluyó exclusivamente a los virus de la zona de estudio (*network* 3; Tabla 3.12). Los *networks* 4, 5 y 6 fueron relativamente similares entre sí, lo que sugirió que las correspondientes secuencias podrían corresponder a un linaje viral único (Tabla 3.13). De hecho, la similitud media entre todas las secuencias incluidas en estos tres *networks* (11,83) fue similar a las similitudes observadas dentro de cada uno del resto de los grupos. Los *networks* 1 y 2 también fueron relativamente similares entre sí.

Tabla 3.12: Secuencias virales analizadas junto a las secuencias patagónicas de *Ostreococcus Virus* y resumen de los resultados. Se indican los números de acceso (ID), el resultado del análisis por Parsimonia Estadística (*Network*), el agrupamiento según el análisis filogenético (*Linaje*, el análisis geográfico de las cepas (*MM* Mar Mediterraneo, *OAS* Atlántico Sur, *OAN* Atlántico Norte, *OPN* Pacífico Norte, *CM* Canal de la Mancha, *OPS* Pacífico Sur) y el huésped de los *O. virus* (clado de *Ostreococcus. sp.*).

ID	Cepa	Network	Linaje	Origen	Huésped
NC_014765	BpV1	-	-	MM	-
HM004430	BpV2	-	-	MM	-
NC_014767	MpV1	-	-	MM	-
MPU32975	SP1	-	-	OAN	-
MPU32976	SP2	-	-	OAN	-
MPU32982	PL1	-	-	OPN	-
MPU32981	SG1	-	-	OPN	-
MPU32980	PB8	-	-	OAN	-
MPU32979	PB7	-	-	OAN	-
MPU32978	PB6	-	-	OAN	-
MPU32977	GM1	-	-	OPN	-
NC_014766	OIV1	1	2	MM	A
GQ412099	OIV158	1	2	MM	A
GQ412100	OIV164	1	2	MM	A
GQ412091	OIV462	2	2	OAN	A
GQ412095	OIV467	2	2	OAN	A

Tabla 3.12: (Continuación)

ID	Cepa	Network	Linaje	Origen	Huésped
GQ412092	OIV464	2	2	OAN	A
GQ412093	OIV465	2	2	OAN	A
GQ412094	OIV466	2	2	OAN	A
GQ412098	OIV458	2	2	OAN	A
Esta Tesis	h1	3	1	OAS	A
Esta Tesis	h2	3	1	OAS	A
Esta Tesis	h3	3	1	OAS	A
Esta Tesis	h4	3	1	OAS	A
Esta Tesis	h5	3	1	OAS	A
Esta Tesis	h6	3	1	OAS	A
Esta Tesis	h7	3	1	OAS	A
GQ412082	OIV349	4	5	CM	A
GQ412083	OIV350	4	5	CM	A
GQ412090	OIV470	4	5	CM	A
GQ412089	OIV468	4	5	CM	A
GQ412088	OIV402	4	5	CM	A
GQ412084	OIV359	4	5	OPS	A
NC_014789	Otv-2	4	5	CM	B
JCVI_READ_1092963530480	-	4	5	OAN	-
GQ412085	OIV360	5	5	OPS	A
GQ412096	OIV536	5	5	MM	A
GQ412097	OIV537	5	5	MM	A
GQ412101	OIV155	5	5	MM	A
GQ412086	OIV364	6	5	OPS	A
GQ412087	OIV368	6	5	OPS	A
FJ267509	OTV102	7	7	MM	C
FJ267510	OTV113	7	7	MM	C
FJ267504	OTV3	7	7	MM	C
FJ267498	OTV23	7	7	MM	C
FJ267512	OTV126	7	7	MM	C
FJ267496	OTV21	7	7	MM	C
FJ267505	OTV29	7	7	MM	C
FJ267508	OTV78	7	7	MM	C
FJ267506	OTV52	7	7	MM	C
FJ267507	OTV72	7	7	MM	C
NC_010191	Otv5	7	7	MM	C

Tabla 3.12: (Continuación)

ID	Cepa	Network	Linaje	Origen	Huésped
NC_013288	OTV-1	7	7	CM	C
FJ267511	OTV121	7	7	MM	C
FJ267497	OTV22	7	7	MM	C
FJ267502	OTV64	8	6	MM	D
FJ267499	OTV66	8	6	MM	D
FJ267500	OTV67	8	6	MM	D
FJ267501	OTV63	9	4	MM	D
FJ267503	OTV65	9	4	MM	D
AF405581	BSA99_5	-	3	OPN	-
EU889370	KBvp_17	-	3	OPN	-
JCVI_READ_1092955067637	-	-	6	OAN	-
JCVI_READ_1091140189807	-	-	-	OAN	-
JCVI_READ_1092246500722	-	-	3	OAN	-

El análisis evolutivo indicó que todos los *networks*, salvo el 6, corresponden a grupos monofiléticos. No obstante, la topología obtenida y los análisis de diferenciación genética ( $r_i/R$ ; Materiales y métodos) mostraron que algunos de los mismos no se encuentran lo suficientemente diferenciados o soportados como para considerarlos linajes independientes (Fig. 3.4; Tabla 3.14). Los *networks* 1 y 2 mostraron una baja diferenciación y se agruparon en una rama bien soportada (Linaje 2). De manera similar, los *networks* 4, 5 y 6 fueron muy similares entre sí y conformaron un clado diferenciado y bien soportado del árbol (Linaje 5). Los análisis mostraron además que 3 singletons corresponden a un único linaje (Linaje 3) y un cuarto (JCVI\_READ\_1092955067637) podría corresponder a una variante divergente del Linaje 6 o un octavo linaje.

En base a estas observaciones, las secuencias estudiadas de virus de *Ostreococcus* spp. pueden agruparse en siete linajes o genotipos (resumido en la Tabla 3.12). Además de presentar un fuerte soporte en los análisis filogenéticos, cada uno de estos linajes fue sustancialmente divergente con respecto a cualquiera de los otros.

Tabla 3.13: Número medio de sustituciones nucleotídicas (diagonal inferior) y amino acídicas (diagonal superior) entre *networks* de *Ostreococcus Virus*. La diagonal indica las distancias nucleotídicas ( $p$ ) intra-*network*.

Network	1	2	3	4	5	6	7	8	9
1	0.00	0.94	8.86	5.76	5.53	5.14	10.04	6.90	6.85
2	21.63	1.00	9.06	6.85	6.72	6.46	11.14	7.89	7.38
3	69.54	72.46	1.71	10.10	9.93	9.60	10.32	9.50	5.86
4	78.12	89.67	94.14	1.60	0.00	0.00	8.21	8.32	8.47
5	74.07	86.88	86.72	21.15	4.66	0.00	8.14	8.00	8.00
6	71.14	85.07	84.53	12.11	11.11	0.00	8.00	7.42	7.20
7	85.67	90.25	96.72	72.65	70.54	69.17	10.40	6.69	7.93
8	79.20	86.21	88.81	78.96	79.92	74.00	60.34	15	6.57
9	79.71	84.92	81.86	85.05	86.22	77.60	80.20	65.42	0.00

Tabla 3.14: Índices de diferenciación genética y soportes filogenéticos de los *networks* y linajes de virus de *Ostreococcus* spp. identificados en este estudio.

	$r_i/R$	SB <sup>1</sup>	
Network	1	4.7	94
	2	18	95
	3	80	100
	4	11.4	94
	5	3.3	38
	6	0.009	100
	7	20.9	90
	8	13.3	66
	9	82.1	100
Linaje	1	80.2	100
	2	32.3	92
	3	50.3	98
	4	82.1	100
	5	48.4	100
	6	35.1	94
	7	20.9	98

<sup>1</sup> Soporte (bootstrap).

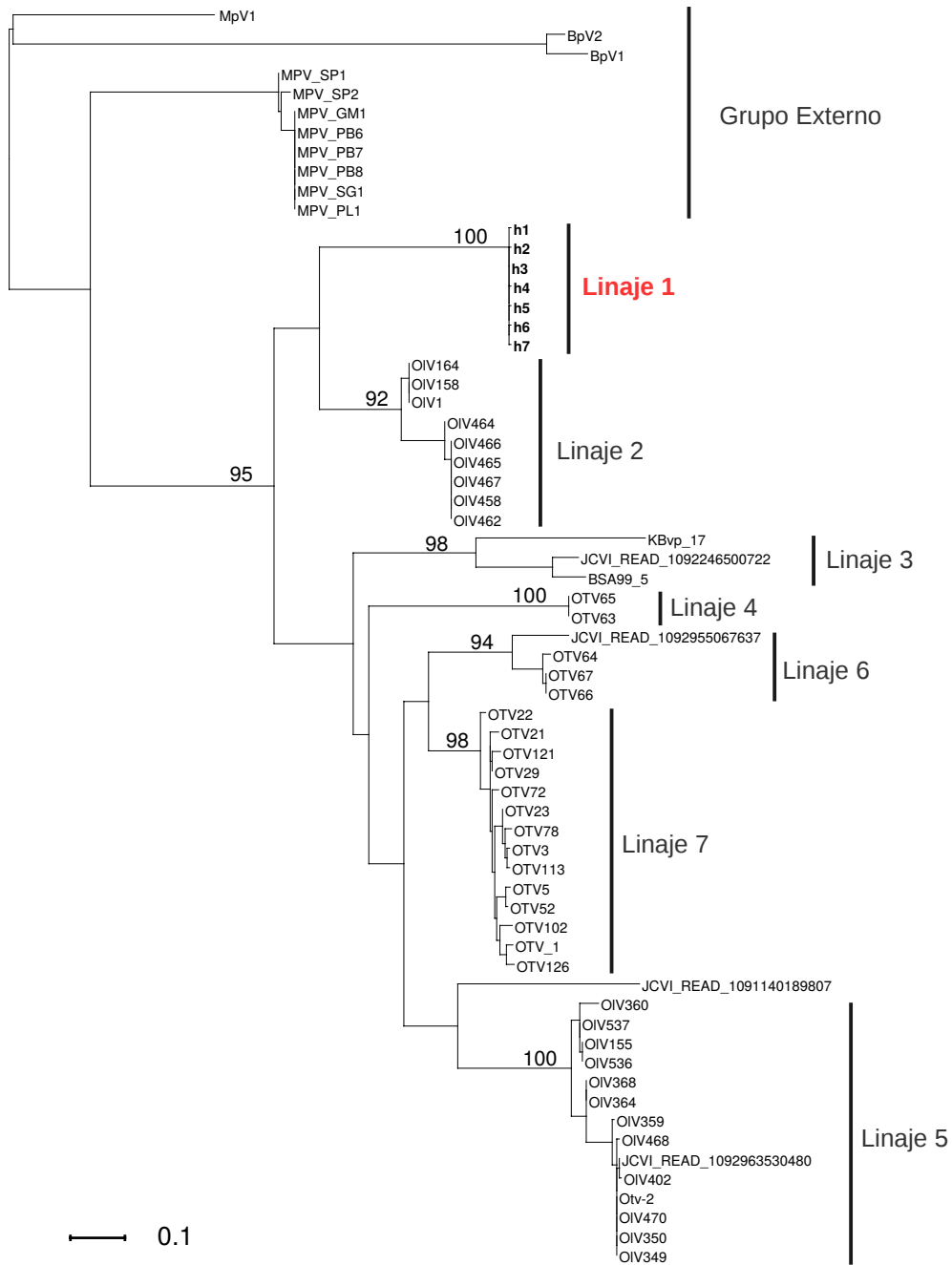


Figura 3.12: Árbol filogenético de secuencias de prasinovirus de la zona de estudio (h1-7; Linaje 1) y de otras regiones del mundo, obtenido mediante el criterio de máxima verosimilitud. Los números en las ramas corresponden a soportes de bootstrap (n=100). La escala en la parte inferior izquierda de la figura corresponde a 0.1 sustituciones por posición alineada.

## 4. Discusión y Conclusiones

### 4.1. Discusión

El bacterioma de la zona de estudio presentó una composición similar a la observada en otros ambientes marinos, destacándose grupos mayoritarios como *Pelagibacterales*, *Rhodobacterales*, *Acidimicrobiales* y *Oceanospirillales* (Campbell et al., 2009; Feng et al., 2009; Fortunato et al., 2013; Liu et al., 2015). Por ejemplo, en el estuario Pearl (Hong Kong) han estudiado la distribución de las comunidades bacterianas, y se encontró que los Órdenes *Rhodobacterales* y *Acidimicrobiales* fueron dominantes en las aguas marinas del estuario (Liu et al., 2015). *Rhodobacterales* es un grupo de gran importancia en ambientes costeros, ya que tendría un rol destacado en la colonización primaria de la superficie en dichos ambientes (Dang et al., 2008). Por otro lado, el grupo de *Pelagibacterales* está presente en todos los océanos del mundo y representan el 25% de todas las células planctónicas (Morris et al., 2002; Zhao et al., 2010). Además son organismos muy activos, no sólo por su importancia en ciclos biogeoquímicos, como el carbono, sino que estarían implicados en la asimilación de aminoácidos (Sun et al., 2011; Malmstrom et al., 2004). El Orden *Oceanospirillales* se distribuye en los océanos de todo el mundo y se ha encontrado que viven en relación simbiótica con un amplio rango de invertebrados marinos (Verna et al., 2010; Morrow et al., 2012; Fiore et al., 2015). Se ha observado que tienen capacidad para metabolizar compuestos orgánicos complejos (Hazen et al., 2010). Asimismo, se observó la presencia de taxa de baja frecuencia como los *Bacillales*, *Xantomonadales*, SAR86, OCS116, *Rhizobiales*, *Rickettsiales*, MB11C04, *Bdellobibrionales*, *Alteromonadales*, Aegean-245, *Methylophilales*, KI89A, SAR406, *Sphingomonadales* y *Gracilibacteria* (también conocido como BD1-5 y GN02). La presencia de numerosos grupos relativamente escasos también es coincidente con lo observado en estudios microecológicos previos de otras regiones. Los taxa minoritarios han sido poco estudiados en compara-



---

ción con los taxa abundantes (Schloss et al., 2016). No obstante, en publicaciones recientes se ha destacado que algunos grupos minoritarios pueden comportarse como condicionalmente raros, es decir grupos que bajo ciertas condiciones pueden tornarse muy numerosos (Shade et al., 2014), y ser supernumerarios en algunos ambientes circunscriptos (Sunagawa et al., 2010). Recientemente se ha presentado evidencia de que los grupos minoritarios podrían representar reservorios de biodiversidad ante futuros cambios climáticos globales, resaltando la importancia de su descripción y estudio (Choi et al., 2017).

La prevalencia de cianobacterias de los géneros *Prochlorococcus* y *Synechococcus* en ambientes marinos ha sido bien documentada (Scanlan et al., 2009). Sin embargo, los resultados presentados en este trabajo apoyan la idea de que las picroalgas eucariotas son componentes importantes del picoplancton fotosintético de la zona de estudio, mientras que para evidenciar la presencia de cianobacterias fue necesario utilizar PCR mediante iniciadores específicos. Esto sugiere que las mismas se encontrarían en menores cantidades respecto a otros representantes del picoplancton, por lo cual quedan poco o no representadas entre los amplicones generados con iniciadores universales. Esta hipótesis coincide con observaciones realizadas en algunos ecosistemas similares del hemisferio Norte, en los cuales se demostró que los picoeucariotas fotosintéticos (PF) son prevalentes, muchas veces en detrimento de otros productores primarios picoplanctónicos (Biegala et al., 2003; Lovejoy et al., 2007; Santic et al., 2014; Wu et al., 2014). Estudios recientes sugieren que las cianobacterias marinas preferirían ambientes más oligotróficos que las picroalgas, lo cual posiblemente sea la principal causa de por qué se seleccionan unas u otras (Bouman et al., 2006; Wang et al., 2019). En conjunto, estos resultados soportan el concepto de que los niveles más bajos de las redes tróficas oceánicas y costeras pueden ser sostenidas por diferentes taxa.

Los resultados de esta Tesis indican que la presencia y taxonomía de las picroalgas de la zona de estudio varían a lo largo del año, ya que se encontró evidencia directa (molecular y aislamiento en cultivo) o indirecta (sus virus) de la presencia de estas en 7 de las 13 muestras estudiadas y se aislaron dos especies diferentes de PF. Es posible que la no detección de PF en algunas muestras responda a fluctuaciones de la abundancia de estos microorganismos, tal como recientemente se ha documentado en una región costera de la bahía de Blanes (España) (Giner et al., 2018). No obstante, también debe tenerse en cuenta que a partir de la muestra INV11 se amplificaron secuencias de *Micromonas Virus* pero no se observaron secuencias del gen 16S de *Micromonas* sp. Por el contrario, la detección de *Ostreococcus Virus* siempre se asoció con la presencia de secuencias del gen 16S de

---

*Ostreococcus* spp. en el ADN ambiental. Esto posiblemente se deba a que los iniciadores utilizados para amplificar las secuencias del gen 16S no amplifican con la misma eficiencia las secuencias de estos dos taxa, y por ende que podría ocurrir lo mismo con otros picoeucariotas. Tampoco puede descartarse la presencia de grupos de picoalgas que forman parte de la comunidad microbiana y no son detectados con los primers para 16S (como el caso de *Micromonas* sp.). Al momento de clasificar y describir la comunidad, estos grupos podrían estar pasando inadvertidos.

En concordancia con los datos obtenidos en el presente estudio, la existencia de fluctuaciones estacionales en la composición de los microbiomas marinos ha sido previamente documentada en otras regiones del Mundo (Campbell et al., 2009; Salter et al., 2015; Giner et al., 2018). Las causales precisas de estas variaciones aún están en una etapa de descubrimiento. No obstante, se acepta que estas corresponden a variaciones en la disponibilidad de determinados nutrientes y/o a fluctuaciones estacionales de parámetros físicos como la temperatura. La zona de estudio de esta Tesis parece no ser una excepción, ya que se observó una relación clara y estadísticamente significativa entre parámetros estacionalmente variables y el perfil genético del gen 16S en el pico-nanoplancton. En estudios previos realizados en la cuenca inferior del Río Chubut se observó que las disponibilidades de nutrientes como amonio, silicato, nitrato y fosfato son proporcionales al caudal del río, lo cual se debe a que en las épocas de mayores precipitaciones el río arrastra una mayor cantidad de solutos debido al lavado por escorrentía (Sastre et al., 1998). Los estudios de la cuenca inferior indicaron también que las variaciones estacionales del caudal se correlacionan con el desarrollo y la diversidad planctónica, presumiblemente debido a las ya mencionadas variaciones de la disponibilidad de nutrientes. No es posible descartar, por lo tanto, que la correlación observada en el presente estudio entre el caudal del río y el perfil genético del nano-picoplancton se relacione con el estado hidrológico del primero. Sin embargo, las diferentes especies pueden tener diferentes óptimos de crecimiento en relación a la temperatura, ocasionando que exista una sucesión de especies a lo largo del año. Es necesario realizar nuevos estudios que contemplen el análisis químico del aporte de nutrientes del mismo, con el objeto de determinar si éste influye sobre el microbioma. Esto permitirá determinar si la correlación descubierta en el presente estudio refleja una relación causal entre la carga de nutrientes del río y el microbioma, los efectos de otras variables ambientales que fluctúan estacionalmente, o una combinación de ambas cosas. Si esto fuera correcto, podría ser posible explicar las condiciones del microbioma a partir de datos hidrológicos y/o químicos del río, constituyendo un interesante aporte al futuro modelado del ecosistema del estuario y zona de

---

influencia.

Como ya se ha explicado, los microorganismos marinos constituyen un enorme y complejo ensamble de grupos taxonómicos (Mucko et al., 2018). Esta enorme biodiversidad se halla en una etapa de descubrimiento por lo cual las OTU no clasificables (uOTUs) constituyen un problema común en el estudio de secuencias ambientales. El problema afecta en forma particular a los taxa de baja frecuencia, los cuales a pesar de contribuir la mayor diversidad de los microbiomas marinos han sido muy poco estudiados en comparación con los grupos abundantes (Schloss et al., 2016). Estudios recientes han demostrado que los taxa de baja frecuencia pueden ser influyentes en la dinámica microecológica (Shade et al., 2014) y actuar como reservorios de biodiversidad resultando por lo tanto relevantes ante un escenario de cambio global (Choi et al., 2017). Además, se ha observado que algunos grupos que son numéricamente escasos en el agua de mar pueden ser mayoritarios en algunos habitats marinos particulares (Sunagawa et al., 2010).

Alrededor del 13% de las OTU detectadas en la zona de estudio resultaron ser no clasificables. El hecho de que no se detectaran diferencias significativas entre las proporciones de uSEQs observadas entre los datos generados mediante secuenciación Sanger y SAP, y que la proporción de uOTUs fué sólo levemente mayor para el caso de los datos de SAP corrobora el valor de la región V1-V3 del gen del ARNr 16S en el análisis taxonómico de secuencias ambientales (Liu et al., 2008, 2007). La identificación de uSEQs de puntos remotos que fueron similares a muchas de las uSEQs locales representa una evidencia fuerte de que las mismas corresponden a linajes reales del gen ARNr 16S (*i.e.* no a artefactos del proceso de secuenciado).

Los análisis presentados en la sección 3.2 mostraron que varias de las uOTU locales corresponden a grupos cosmopolitas no descriptos, posiblemente pertenecientes a cinco linaje de *Alphaproteobacteria* y a un clado desconocido de *Gracilibacteria*. Estos hallazgos soportan el concepto de que los taxa que son relativamente escasos a nivel local pueden ser relevantes a una escala global (Choi et al., 2017; Shade et al., 2014; Sunagawa et al., 2010).

Las dificultades observadas en la asignación taxonómica de secuencias ambientales generalmente es atribuida al conocimiento limitado que se tiene de la sistemática de los microorganismos marinos (Schloss et al., 2016). Sin embargo, las limitaciones metodológicas también podrían representar un problema. Entre ellas pueden mencionarse la persistencia de secuencias quiméricas no detectables por los métodos disponibles, taxa no monofiléticos (esto es clasificaciones artificiales)

---

y fenómenos no tenidos en cuenta por los algoritmos de clasificación automática como la transformación genética y la coalescencia profunda.

Uno de los resultados de este trabajo que resultan interesantes desde el punto de vista metodológico es la fuerte correspondencia observada entre los análisis filogenéticos y aquellos realizados mediante BLAST. En especial, que la correspondencia fue mayor cuando los análisis de similitud se realizaron utilizando ejemplares tipo como referencia, lo cual puede atribuirse a una anotación más verosímil de la taxonomía de los mismos en las bases de datos. Esto indica que el material tipo tiene una utilidad en la clasificación automática de secuencias ambientales, la cual ha sido poco aprovechada hasta el momento.

El uso de métodos filogenéticos para clasificar secuencias ambientales también ha sido limitado hasta el momento con respecto a uso de técnicas basadas en similitud (*e.g.* BLAST) y composición (*e.g.* WNBC implementado en la base de datos SILVA) de las mismas (Hiraoka et al., 2016). El presente estudio mostró que el AFD y las comparaciones filogenéticas *de novo* con secuencias de referencia pueden mejorar en gran medida la asignación taxonómica de secuencias que no son clasificables mediante algoritmos automáticos. Además, también se evidenció que dicho tipo de análisis puede impulsar el descubrimiento de nuevos grupos. Estos hallazgos sugieren un potencial desaprovechado del análisis filogenético en el estudio sistemático de secuencias ambientales. En base a los resultados obtenidos en el presente estudio pueden diagramarse algunos de los desafíos que pueden plantear la implementación de estos métodos. Uno de los más conspicuos es que algunos taxa pueden no ser recuperados en el análisis filogenético, tal como se observó en el presente trabajo para los casos de los órdenes *Rickettsiales*, *Rhodobacterales* y *Rhizobiales*. Esto puede deberse a que algunos taxa no son monofiléticos o bien a problemas de los métodos analíticos o los marcadores utilizados. La implementación de estadísticos que permitan medir la confianza en los resultados provistos por el análisis filogenético también es complicada. El soporte de ramas puede parecer una solución fácil y simple. No obstante, dependiendo de qué marcador y método se utilicen algunos grupos pueden no ser recuperados como clados bien soportados. Además, debe tenerse en cuenta que la probabilidad de recuperar un clado determinado por azar (incluso sin soporte o con uno muy bajo) es extremadamente baja y depende del número de secuencias del grupo de interés y del total de secuencias en el dataset. El problema se complica aún más si se toma en consideración la incertidumbre filogenética, esto es la existencia de múltiples filogenias plausibles/óptimas. Se requerirá una cantidad sustancial de investigaciones y análisis de datos empíricos para poder desarrollar el tema de una manera acabada.

---

Hasta hace no mucho tiempo los estudios de los virus ambientales (VA) se limitaban a observaciones realizadas mediante microscopía electrónica o epifluorescencia. Si bien dichos métodos son de reconocida utilidad, tienen un valor limitado si se desea considerar la diversidad viral, ya que los caracteres virales que pueden evidenciarse mediante observación directa (abundancia, tamaño y morfología de la partícula viral) son poco resolutivos a la hora de realizar determinaciones taxonómicas. El estudio de la diversidad de los VA ha sido fuertemente impactado por los métodos genéticos independientes de cultivo (Angly et al., 2011; Clasen y Suttle, 2008). Actualmente se sabe que los virus ambientales son mucho más diversos de lo que se creía hace algunos años y que son responsables de funciones ecológicas relevantes (Allen y Wilson, 2008; Gimenes et al., 2012; Short et al., 2011; Williamson et al., 2008; Yang et al., 2019).

La familia *Phycodnaviridae* comprende seis géneros (*Chlorovirus*, *Prasinovirus*, *Prymnesiovirus*, *Phaeovirus*, *Coccolithovirus* y *Raphidovirus*) de virus icosaédricos y genoma a ADN cadena doble que infectan una variedad de algas marinas y de agua dulce (Van Etten et al., 2002; Wilson et al., 2009; Maruyama y Ueki, 2016). Los miembros del género *Prasinovirus* infectan algas de la clase *Mamiellophyceae*, las cual incluye los picoeucariotas fotosintéticos (PF) más pequeños conocidos (Weynberg et al., 2017; Monier et al., 2016). El picoplancton fotosintético es responsable de una importante proporción de la productividad primaria en los océanos (Chavez et al., 2011). En regiones oligotróficas, dicho grupo es dominado por las cianobacterias de los géneros *Synechococcus* y *Prochlorococcus*. Sin embargo, los PF de los géneros *Bathycoccus*, *Micromonas* y *Ostreococcus* pueden ser importantes en aguas estuarinas y costeras (Biegala et al., 2003; Giner et al., 2018). Los *Prasinovirus* han recibido mucha atención ya que pueden producir la lisis celular, interfiriendo de esta forma con el rol ecológico de estos PF (Suttle, 2007).

Los resultados presentados en esta Tesis mostraron que los *Ostreococcus Virus* (OV) de la zona de estudio corresponden a un clado divergente y que el resto de los OVs incluidos en los análisis corresponden a otros seis clados bien soportados y divergentes. Una de las secuencias incluidas entre los datos analizados (JCVI READ 1091140189807) no se agrupó con ninguna otra secuencia, sugiriendo la existencia de un octavo linaje. El marcador utilizado en este estudio, la ADN polimerasa viral de la familia B, se encuentra altamente conservada a nivel de su secuencia aminoacídica pero presenta una variabilidad sustancial a nivel nucleotídico. Esto la convierte en un marcador ideal para la realización de estudios evolutivos en varios rangos taxonómicos. De manera interesante, analizando los datos genómicos

---

disponibles para algunas de las cepas incluidas en el presente estudio, se observó que las variaciones en el gen de la polimerasa tienen un correlato en el resto del genoma. Como se detalla a continuación, parecen existir regiones exclusivas de algunos linajes, lo cual sugiere la posesión de funciones metabólicas ausentes en otros grupos. El genoma de la cepa de OV Otv-2, la cual pertenece al Linaje 5 identificado en el presente estudio, codifica un citocromo b5, un factor sigma para ARN polimerasa, un transportador de fosfato de alta fidelidad, una desoxicitidylato desamina putativa, una 6-fosfogluconato deshidrogenasa putativa, una oxigenasa tipo 2OG-Fe(II) putativa y una proteína de ensamble de fibra de cola viral putativa, ninguna de las cuales esta presente en las cepas Otv-1 y Otv-5 (Weynberg et al., 2011), la cual según el presente estudio corresponde al Linaje 7. Los genomas de las cepas Otv-1 y Otv-5 también codifican proteínas que están ausentes del genoma de la cepa Otv-2 .

## 4.2. Conclusiones

- Los grupos mayoritarios en la zona de estudio fueron SAR11, *Rhodobacterales*, *Acidimicrobiales* y *Oceanospirillales* (Clases *Alphaproteobacteria*, *Actinobacteria* y *Gammaproteobacteria*). Estos estuvieron acompañados por grupos de menor frecuencia, que incluyeron *Bacillales*, *Flavobacteriales*, representantes de las Clases *Gammaproteobacteria* (*Xantomonadales*, *Oceanospirillales*, SAR86, *Alteromonadales*, KI89A), *Alphaproteobacteria* (OCS116, *Rhizobiales*, SAR116 y *Rickettsiales*, Aegean-245, *Sphingomonadales*), *Betaproteobacteria* (*Methylophilales*), y los grupos MB11C04, *Bdellovibrionales*, SAR406, y *Gracilibacteria* (también llamados BD1-5 y GN02).
- La diversidad del gen 16S puede variar sustancialmente a lo largo del tiempo y presentó un patrón estacional muy marcado en la zona de estudio.
- Las secuencias de cianobacterias son difíciles de detectar en la zona de estudio, lo que sugiere que las mismas son raras. No se detectaron secuencias de cianófagos.
- Se identificaron secuencias de 5 clados cosmopolitas de bacterias no clasificables, los cuales podrían corresponder a un linaje no descrito de *Gracilibacteria* y a una familia, dos Órdenes y una Subclase desconocidos de *Alphaproteobacteria*.

- 
- La zona de estudio presenta un linaje aparentemente exclusivo de *Ostreococcus Virus*.
  - *Ostreococcus Virus* es un complejo de al menos siete linajes monofiléticos que podrían corresponder a especies crípticas.

## A. Anexo: *Primers*

Tabla A.1: Secuencias correspondientes a los *primers* utilizados en este trabajo.

NOMBRE	SENTIDO	SECUENCIA
AF27	<i>Forward</i>	5'-AGAGTTTGATCMTGGCTCAG-3'
HR1492	<i>Reverse</i>	5'-TACGGYTACCTTCTTACGACTT-3'
SP6	<i>Forward</i>	5'-ATTTAGGTGACACTATAG-3'
T7	<i>Reverse</i>	5'-TAATACGACTCACTATAGGG-3'
Podo-F	<i>Forward</i>	5'-GACACHCTYRTVHTGTCWMGWYTG-3'
Podo-R2	<i>Reverse</i>	5'-MCKACCRTCYARDCCYTTMAK-3'
CP-DNAP-349F	<i>Forward</i>	5'-CCAAAYCTYGCMCARGT-3'
CP-DNAP-533Ra	<i>Reverse</i>	5'-CTCGTCRTGSACRAASGC-3'
CP-DNAP-533Rb	<i>Reverse</i>	5'-CTCGTCRTGDATRAASGC-3'
AVS-1	<i>Forward</i>	5'-GARGGNGCNACNGTNYTNGAYGC-3'
AVS-2	<i>Reverse</i>	5'-GCNGCRTANCKYTTYTTNSWRTA-3'
POL	<i>Reverse</i>	5'-SWRTCNGCTRTCNCCRTA-3
UNI_17F	<i>Forward</i>	5'-ACCTGGTTGATCCTGCCAG-3'
UNI_1534R	<i>Reverse</i>	5'-TGATCCTTCYGCAGGTTCCAC-3'
CF536	<i>Forward</i>	5'-CAGCMGCCGCGGTAATWC-3'
DR907	<i>Reverse</i>	5'-CCCCGTCAATTCCTTTGAGTTT-3'
Cya359	<i>Forward</i>	5'-GGGGAATYTTCCGCAATGGG-3'
Cya781ab	<i>Reverse</i>	5'-GACTACTGGGGTATCTAATCCCATT-3'
Cya781bb	<i>Reverse</i>	5'-GACTACAGGGGTATCTAATCCCTTT-3'



# Bibliografía

- V. Agarwal, J.M. Blanton, S. Podell, A. Taton, M.A. Schorn, J. Busch, Z. Lin, E.W. Schmidt, P.R. Jensen, V.J. Paul, J.S. Biggs, J.W. Golden, E.E. Allen, y B.S. Moore. Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nature Chemical Biology*, 13:537–543, 2017.
- M.J. Allen y W.H. Wilson. Aquatic virus diversity accessed through omic techniques: a route map to function. *Current Opinion in Microbiology*, 11:226–232, 2008.
- E. Allers, J.J. Wright, K.M. Konwar, C.G. Howes, E. Beneze, S.J. Hallam, y M.B. Sullivan. Diversity and population structure of marine group a bacteria in the northeast subarctic pacific ocean. *ISME Journal*, 7:256–268, 2013.
- H. M. Alvarez, O. H. Pucci, y A. Steinbüchel. Lipid storage compounds in marine bacteria. *Applied Microbiology and Biotechnology*, 47:132–139, 1997.
- R. I. Amann, W. Ludwig, y K. H. Schleifer. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiology and Molecular Biology Reviews*, 59:143–169, 1995.
- F.E. Angly, B. Felts, M. Breitbart, P. Salamon, R.A. Edwards, C. Carlson, A.M. Chan, M. Haynes, S. Kelley, H. Liu, J.M. Mahaffy, J.E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C.A. Suttle, y F. Rohwer. The marine viromes of four oceanic regions. *PLoS Biology*, 4:e368, 2011.
- F. Azam, T. Fenchel, J.G. Field, J.S. Gray, L.A. Meyer-Reil, y F. Thingstad. The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, 10:257–263, 1983. doi: 10.3354/meps010257.
- Y. M. Bar-On, R. Phillips, y R. Milo. The biomass distribution on earth. *PNAS*, 115:6506–6511, 2018.

- 
- R.-M. Barthélémy, E. Faure, y T. Goto. Serendipitous discovery in a marine invertebrate (phylum chaetognata) of the longest giant viruses reported till date. *Virology: Current Research*, 3:110–122, 2019.
- M. Bauer y T. Plehn. *Yet another introduction to dark matter*. Springer International Publishing, 2019.
- G. Bernard, J. S. Pathmanathan, R. Lannes, P. Lopez, y E. Bapteste. Microbial dark matter investigations: how microbial studies transform biological knowledge and empirically sketch a logic of scientific discovery. *Genome Biology and Evolution*, 10:707–715, 2018.
- A.E. Bernhard, D. Colbert, J. McManus, y K.G. Field. Microbial community dynamics based on 16s rRNA gene profiles in a Pacific Northwest estuary and its tributaries. *FEMS Microbiology Ecology*, 52:115–128, 2005.
- I.C. Biegala, F. Not, D. Vaultot, y Simon N. Quantitative assessment of pico-eukaryotes in the natural environment by using taxon-specific oligonucleotide probes in association with tyramide signal amplification-fluorescence in situ hybridization and flow cytometry. *Applied and Environmental Microbiology*, 69:5519–5529, 2003.
- E.M. Bik, E.K. Costello, A.D. Switzer, B.J. Callahan, S.P. Holmes, R.S. Wells, K.P. Carlin, E.D. Jensen, S. Venn-Watson, y D.A. Relman. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. *Nature Communications*, 7:10516, 2016.
- H. Bolhuis y M. S. Cretoiu. What is so special about marine microorganisms? introduction to the marine microbiome—from diversity to biotechnological potential. En *The Marine Microbiome*, pages 3–20. Springer International Publishing, 2016.
- H. A. Bouman, A. Ulloa, D. J. Scanlan, K. Zwirgmaier, W. K. Li, T. Platt, V. Stuart, R. Barlow, O. Leth, L. Clementson, V. Lutz, M. Fukasawa, S. Watanabe, y S. Sathyendranath. Oceanographic basis of the global surface distribution of prochlorococcus ecotypes. *Science*, 312:918–921, 2006.
- R. Breyta, A. Black, J. Kaufman, y G. Kurath. Spatial and temporal heterogeneity of infectious hematopoietic necrosis virus in Pacific Northwest salmonids. *Infection, Genetics and Evolution*, 45:347–358, 2016.

- 
- C.P.D. Brussaard, A-C. Baudoux, y F. Rodríguez-Valera. Marine viruses. En *The Marine Microbiome*, pages 155–183. Springer International Publishing, 2016. doi: 10.1007/978-3-319-33000-6<sub>5</sub>.
- Y. S. Bukin, Y. P. Galachyants, I. V. Mozorov, A. V. Bukin, A. S. Zakharenko, y T. I. Zemskaya. The effect of 16s rna region choice on bacterial community metabarcoding results. *Scientific Data*, 6:190007, 2019. doi: 10.1038/sdata.2019.7.
- D. Burstein, L. B. Harrington, S. C. Strutt, A. J. Probst, K. Anantharaman, B. C. Thomas, J. A. Doudna, y J. F. Banfield. New crispr-cas systems from uncultivated microbes. *Nature*, 542:237–241, 2017.
- P. A. Calderoli, F. J. Espínola, H. M. Dionisi, M. N. Gil, J. K. Jansson, y M. Lozada. Predominance and high diversity of genes associated to denitrification in metagenomes of subantarctic coastal sediments exposed to urban pollution. *PLoS One*, 75:123–139, 2018.
- C. Callieri y J. G. Stockner. Freshwater autotrophic picoplankton: a review. *Journal of Limnology*, 61(1):1, feb 2002. doi: 10.4081/jlimnol.2002.1.
- B.J. Campbell, L. Yu, T.R.A. Straza, y D.L. Kirchman. Temporal changes in bacterial rna and rna genes in delaware (usa) coastal waters. *Aquatic Microbial Ecology*, 57:123–135, 2009.
- J. Ceh, M. Van Keulen, y D.G. Bourne. Coral-associated bacterial communities on ningaloo reef, western australia. *FEMS Microbiology Ecology*, 75:134–144, 2011.
- F.P. Chavez, M. Messié, y J.T. Pennington. Marine primary production in relation to climate variability and change. *Annual Review of Marine Sciences*, 3:227–260, 2011.
- F. Chen y C. A. Suttle. Amplification of dna polimerase gene fragments from viruses infecting microalgae. *Applied and Environmental Microbiology*, 61:1274–1278, 1995.
- F. Chen, K. Wang, S. Huang, H. Cai, M. Zhao, N. Jiao, y K. E. Wommack. Diverse and dynamic populations of cyanobacterial podoviruses in the chesapeake bay unveiled through dna polymerase gene sequences. *Environmental Microbiology*, 11:2884–2892, 2009.

- 
- C.J. Choi, C. Bachy, G.S. Jaeger, C. Poirier, L. Sudek, V.V. Sarma, A. Mahadevan, S.J. Giovannoni, y A.Z. Worden. Newly discovered deep-branching marine plastid lineages are numerically rare but globally distributed. *Current Biology*, 27:R15–R16, 2017.
- J.L. Clasen y C.A. Suttle. Identification of freshwater phycodnaviridae and their potential phytoplankton hosts, using dna pol sequence fragments and a genetic-distance analysis. *Applied and Environmental Microbiology*, 75:991–997, 2008.
- M. Clement, D. Posada, y K. A. Crandall. Tcs: a computer program to estimate gene genealogies. *Molecular Ecology*, 9:1657–1659, 2000.
- S. Coy, E. Gann, H. Pound, S. Short, y S Wilhelm. Viruses of eukaryotic algae: Diversity, methods for detection, and future directions. *Viruses*, 10(9):487, sep 2018. doi: 10.3390/v10090487.
- M. Crane y A. Hyatt. Viruses of fish: An overview of significant pathogens. *Viruses*, 3(11):2025–2046, oct 2011. doi: 10.3390/v3112025.
- H. A. Cristóbal, M. A. López, E. Kothe, y C. M. Abate. Diversity of protease-producing marine bacteria from sub-antarctic environments. *Journal of Basic Microbiology*, 51:590–600, 2011.
- J.C. Cury, F.V. Araujo, S.A. Coelho-Souza, R.S. Peixoto, J.A. Oliveira, H.F. Santos, A.M. Davila, y A.S Rosado. Microbial diversity of a brazilian coastal region influenced by an upwelling system and anthropogenic activity. *PLoS One*, 6:e16553, 2011.
- R D’Amore, U. Z. Ijaz, M. Schirmer, J. G. Kenny, R. Gregory, A. C. Darby, M. Shakya, M. Podar, C. Quince, y N. Hall. A comprehensive benchmarking study of protocols and sequencing platforms for 16s rRNA community profiling. *BMC Genomics*, 17(1), jan 2016. doi: 10.1186/s12864-015-2194-9.
- H. Dang, T. Li, M. Chen, y G. Huang. Cross-ocean distribution of rhodobacterales bacteria as primary surface colonizers in temperate coastal marine waters. *Applied and Environmental Microbiology*, 74:52–60, 2008.
- A.P. de Castro, S.D. Araujo, A.M. Reis, R.L. Moura, R.B. Francini-Filho, G. Pappas, T.B. Rodrigues, F.L. Thompson, y R.H. Kruger. Bacterial community associated with healthy and diseased reef coral *mussismilia hispida* from eastern brazil. *Microbial Ecology*, 59:658–667, 2010.

- 
- M.G.G. de Pinna. Concepts and tests of homology in the cladistic paradigm. *Cladistics*, 7:367–384, 1991.
- T.O. Delmont, C. Quince, A. Shaiber, Ö. C. Esen, S. T. M. Lee, M S. Rappé, S. L. McLellan, S. Lücker, y A. M. Eren. Nitrogen-fixing populations of planctomycetes and proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, 3:804–813, 2018.
- A.I. Dogliotti, V.A. Lutz, y V. Segura. Estimation of primary production in the southern argentine continental shelf and shelf-break regions using field and remote sensing data. *Remote Sensing of Environment*, 140:497–508, 2014.
- R. C. Edgar, B. J. Haas, J. C. Clemente, C. Quince, y R. Knight. Uchime improves sensitivity and speed of chimera detection. *Bioinformatics*, 27:2194–2200, 2011.
- R. Espejo y N. Plaza. Multiple ribosomal rna operons in *Bacteria*; their concerted evolution and potential consequences on the rate of evolution of their 16s rna. *Frontiers in Microbiology*, 9:1232–1237, 2018.
- F. J. Espínola, H. M. Dionisi, S. Borglin, C. J. Brislawn, J. K. Jansson, W. P. Mac Cormack, J. Carroll, S. Sjöling, y M. Lozada. Analysis of subtidal sediments from polar and subpolar coastal environments highlights the relevance of anaerobic hydrocarbon degradation processes. *Microbial Ecology*, 75:123–139, 2018.
- X. Fang, Y. Liu, Y. Zhao, Y. Chen, R. Liu, Q-L. Qin, G. Li, Y-Z. Zhang, W. Chan, W. R. Hess, y Q. Zeng. Transcriptomic responses of the marine cyanobacterium prochlorococcus to viral lysis products. *Environmental Microbiology*, jan 2019. doi: 10.1111/1462-2920.14513.
- H. M. Farnelid, K. A. Turk-Kubo, y J. P. Zehr. Identification of associations between bacterioplankton and photosynthetic picoeukaryotes in coastal waters. *Frontiers in Microbiology*, 7:–, 2016.
- S. Federhen. Type material in the ncbi taxonomy database. *Nucleic Acids Research*, 43:D1086–1098, 2015.
- T. Fenchel. The microbial loop – 25 years later. *Journal of Experimental Marine Biology and Ecology*, 366(1-2):99–103, nov 2008. doi: 10.1016/j.jembe.2008.07.013.
- T. Fenchel y B. J. Finlay. The ubiquity of small species: patterns of local and global diversity. *Bioscience*, 54:777–784, 2004.

- 
- B.W. Feng, X.R. Li, J.H. Wang, Z.Y. Hu, H. Meng, L.Y. Xiang, y Z.X. Quan. Bacterial diversity of water and sediment in the changjiang estuary and coastal area of the east china sea. *FEMS Microbiology Ecology*, 70:80–92, 2009.
- M.P. Ferla, J.C. Thrash, S.J. Giovannoni, y W.M. Patrick. New rrna gene-based phylogenies of the alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS One*, 8:e83383, 2013.
- G. Fichi, G. Cardeti, A. Cersini, C. Mancusi, M. Guarducci, G. Di Guardo, y G. Terracciano. Bacterial and viral pathogens detected in sea turtles stranded along the coast of tuscan, italy. *Veterinary Microbiology*, 185:56–61, 2016.
- C. L. Fiore, M. Labrie, J. K. Jarett, y M. P. Lesser. Transcriptional activity of the giant barrel sponge, *xestospongia muta* holobiont: molecular evidence for metabolic interchange. *Frontiers in Microbiology*, 6:364, 2015.
- P. Flombaum, J. L. Gallegos, R. A: Gordillo, J. Rincón, L. L. Zabala, N. Jiao, D. M. Karl, W. K. W. Li, M. W. Lomas, D. Veneziano, C. S. Vera, Vrugt J. A., y A. C. Martiny. Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Science*, 110:9824–9829, 2013.
- C.S. Fortunato, A. Eiler, L. Herfort, J.A. Needoba, T.D. Peterson, y B.C. Crump. Determining indicator taxa across spatial and seasonal gradients in the columbia river coastal margin. *ISME Journal*, 7:1899–1911, 2013.
- G. Fox, E Stackebrandt, R. Hespell, J Gibson, J Maniloff, T. Dyer, R. Wolfe, W. Balch, R. Tanner, L. Magrum, L. Zablen, R Blakemore, R Gupta, L Bonen, B. Lewis, D. Stahl, K. Luehrsen, K. Chen, y C. Woese. The phylogeny of prokaryotes. *Science*, 209(4455):457–463, jul 1980. doi: 10.1126/science.6771870.
- J.-F. Ghiglione, P. E. Galand, T. Pommier, C. Pedrós-Alió, E. W. Maas, K. Bakker, S. Bertilson, D. L. Kirchman, C. Lovejoy, P. L. Yager, y A. E. Murray. Pole-to-pole biogeography of surface and deep marine bacterial communities. *PNAS*, 109:17633–17638, 2012.
- P. Ghosh, A. Bhadury. Exploring biogeographic patterns of bacterioplankton communities across global estuaries. *Microbiology Open*, 00:e741:–, 2018.

- 
- M. V. Gimenes, P. M. A. Zanotto, C. A. Suttle, H. B. da Cunha, y D. U. Mehnert. Phylogenetics and movement of phycodnaviruses among aquatic environments. *The ISME Journal*, 6:237–247, 2012.
- C.R. Giner, V. Balagué, A.K. Krabberod, I. Ferrera, A. Reñón, E. Garcés, J.M. Gasol, Logares R., y R. Massana. Quantifying long-term recurrence in planktonic microbial eukaryotes. *Molecular Ecology*, 2019:doi: 10.1111/mec.14929, 2018.
- S. J. Giovannoni, T. B. Britschgi, C. L. Moyer, y K. G. Field. Genetic diversity in sargasso sea bacterioplankton. *Nature*, 345(6270):60–63, may 1990. doi: 10.1038/345060a0.
- S.C. Goslee. Correlation analysis of dissimilarity matrices. *Plant Ecology*, 206:279–286, 2010.
- S.C. Goslee y D.L. Urban. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22:1–19, 2007.
- S. M. Grant, S. L. Hill, P. N. Trathan, y E. J. Murphy. Ecosystem services of the southern ocean: trade-offs in decision-making. *Antarctic Science*, 25(05):603–617, jun 2013. doi: 10.1017/s0954102013000308.
- L. Guillou, W. Eikrem, M.-J. Chrétiennot-Dinet, F. Le Gall, R. Massana, K. Romari, C. Pedrós-Alió, y D. Vaultot. Diversity of picoplanktonic prasinophytes assessed by direct nuclear ssu rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist*, 155:193–204, 2004.
- S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, y O. Gascuel. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyml 3.0. *Systematic Biology*, 59:307–321, 2010.
- T. Hall. Bio-edit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/nt. *Nucleic Acids Symposium Series*, 41:95–98, 1999.
- M. W. Hart y J. Sunday. Things fall apart: biological species from unconnected parsimony networks. *Biology Letters*, 3:509–512, 2007.
- T. C. Hazen, E. A. Dubinsky, T. Z. DeSantis, G. L. Andersen, Y. M. Piceno, N. Singh, J. k. Jansson, A. Probst, S. E. Borglin, J. L. Fortney, W. T. Stringfellow, M. Bill,

- 
- M. E. Conrad, L. M. Tom, K. L. Chavarria, T. R. Alusi, r. Lamendella, D. C. Joyner, C. Spier, J. Baelum, M. Auer, M. L. Zemla, R. Chakraborty, E. L. Sonnenthal, P. D'haeseleer, H-Y. Holman, S. Osman, Z. Lu, J. D. Van Nostrand, Y. Deng, J. Zhou, y O. U. Mason. Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science*, 330:204–208, 2010.
- C. Heibl. ips: Interfaces to phylogenetic software in r, 2014. URL <https://cran.r-project.org/web/packages/ips/index.html>.
- I.S. Henriques, A. Almeida, A. Cunha, y A. Correia. Molecular sequence analysis of prokaryotic diversity in the middle and outer sections of the portuguese estuary ria de aveiro. *FEMS Microbiology Ecology*, 49:269–279, 2004.
- S. Hiraoka, C.C. Yang, y W. Iwasaki. Metagenomics and bioinformatics in microbial ecology: Current status and beyond. *Microbes and Environments*, 31:204–212, 2016.
- H. Hirayama, M. Sunamura, K. Takai, T. Nunoura, T. Noguchi, H. Oida, Y. Furushima, H. Yamamoto, T. Oomori, y K. Horikoshi. Culture-dependent and -independent characterization of microbial communities associated with a shallow submarine hydrothermal system occurring within a coral reef off taketomi island, japan. *Applied and Environmental Microbiology*, 73:7642–7656, 2007.
- J. E. Hobbie, R. J. Daley, y S. Jasper. Use of nucleopore filters for counting bacteria by fluorescence microscopy. *Applied and Environmental Microbiology*, 33:1225–1228, 1977.
- D. H. Huson y C. Scornavacca. Dendroscope3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, 61:1061–1067, 2012.
- K. Ishii y M. Fukui. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Applied and Environmental Microbiology*, 67(8):3753–3755, aug 2001. doi: 10.1128/aem.67.8.3753-3755.2001.
- V. Iverson, R.M. Morris, C.D. Frazar, C.T. Berthiaume, R.L. Morales, y E.V. Armbrust. Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. *Science*, 335:587–590, 2012.
- T. Junier y E.M. Zdobnov. The newick utilities: high-throughput phylogenetic tree processing in the unix shell. *Bioinformatics*, 26:1669–1670, 2010.



- 
- P. Kalatzis, N. Rørbo, D. Castillo, J. Mauritzen, J. Jørgensen, C. Kokkari, F. Zhang, P. Katharios, y M. Middelboe. Stumbling across the same phage: Comparative genomics of widespread temperate phages infecting the fish pathogen *Vibrio anguillarum*. *Viruses*, 9(5):122, may 2017. doi: 10.3390/v9050122.
- M. G. Kalyuzhnaya, O. A. Gomez, y J. C. Murrell. *The methane-oxidizing bacteria (methanotrophs)*. Springer, Cham, 2019.
- K. Katoh y D.M. Standley. Mafft: iterative refinement and additional methods. *Methods in Molecular Biology*, 1079:131–146, 2014.
- K. Katoh, K. Misawa, K. Kuma, y T. Miyata. Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Research*, 30:3059–3066, 2002.
- S. Kawato, S. Shitara, Y. Wang, R. Nozaki, H. Kondo, y I. Hirono. Crustacean genome exploration reveals the evolutionary origin of white spot syndrome virus. *Journal of Virology*, 93(3), nov 2018. doi: 10.1128/jvi.01144-18.
- M.D. Keller, R. C. Selvin, W. Claus, y R.R.L. Guillard. Media for the culture of oceanic ultraphytoplankton. *Journal of Phycology*, 23:633–638, 1987.
- D. L. Kirchman, M. T. Cottrell, y Lovejoy C. The structure of bacterial communities in the western arctic ocean as revealed by pyrosequencing of 16s rna genes. *Environmental Microbiology*, 12:1132–1143, 2010.
- I. Klawonn, S. Bonaglia, M. J. Whitehouse, S. Littman, D. Tienken, M. M. M. Kuypers, V. Brüchert, y H. Ploug. Untangling hidden nutrient dynamics: rapid ammonium cycling and single-cell ammonium assimilation in marine plankton communities. *The ISME Journal*, 2019. doi: 10.1038/s41396-019-0386-z.
- M. Kluge, F. S. Campos, M. Tavares, D. B. de Amorim, F. P. Valdez, Giongo A., P. M. Roehle, y A. C. Franco. Metagenomic survey of viral diversity obtained from feces of subantarctic and south american fur seals. *PLoS ONE*, 11:e0151921, 2016.
- M. Krupovic, V. Cvirkaite-Krupovic, I. Iranzo, D. Prangishvili, y E. V. Koonin. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research*, 244:181–193, jan 2018. doi: 10.1016/j.virusres.2017.11.025.
- P. Laas, E. Satova, I. Lips, U. Lips, J. Simm, V. Kisand, y M. Metsis. Near-bottom hypoxia impacts dynamics of bacterioplankton assemblage throughout water column of the gulf of finland (baltic sea). *PLoS One*, 11:e0156147, 2016.

- 
- J. M. Labonté, K. E. Reid, y C. A. Suttle. Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus dna polimerase gene sequences. *Applied Environmental Microbiology*, 75:3634–3640, 2009.
- P. W. Laffy, E. M. Wood-Charlson, D. Turaev, S. Jutz, C. Pascelli, E. S. Botté, S. C. Bell, T. E. Pierce, K. D. Weynberg, M. J. H. van Oppen, T. Rattei, y N. S. Webster. Reef invertebrate viromics: diversity, host specificity and functional capacity. *Environmental Microbiology*, 20(6):2125–2141, apr 2018. doi: 10.1111/1462-2920.14110.
- D. J. Lane, B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, y N. R. Pace. Rapid determination of 16s ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences*, 82(20):6955–6959, oct 1985. doi: 10.1073/pnas.82.20.6955.
- E. Lara, D. Vaqué, E. L. Sarà, J. A. Boras, A. Gomes, E. Borrul, C. Díez-Vives, E. Teira, M. C. Pernice, F. C. Garcia, I. Forn, Y. M. Castillo, A. Peiró, G. Salazar, X. A. G. Morán, R. Massana, T. S. Catalá, G. M. Luna, S. Agustí, M. Estrada, P. M. Gasol, y C. M. Duarte. Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Science Advances*, 3:e1602565, 2017.
- M. A. Larkin, G. Blackshields, Brown. N. P., R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, y D. G. Higgins. Clustal w and clustal x version 2.0. *Bioinformatics*, 23: 2947–2948, 2007.
- K.-B. Lee. The hierarchical system of the alphaproteobacteria: description of hyphomonadaceae fam. nov., xanthobacteraceae fam. nov. and erythrobacteraceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, 55(5): 1907–1919, sep 2005. doi: 10.1099/ijls.0.63663-0.
- B. Leigh, C. Karrer, J. Cannon, M. Breitbart, y L. Dishaw. Isolation and characterization of a shewanella phage–host system from the gut of the tunicate, *Ciona intestinalis*. *Viruses*, 9(3):60, mar 2017. doi: 10.3390/v9030060.
- J. Lichstein. Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecology*, 188:117–131, 2007.
- C. Liqueste, C. Piroddi, E. G. Drakou, L. Gurney, S. Katsanevakis, A. Charef, y B. Egoh. Current status and future prospects for the assessment of marine and coastal ecosystem services: A systematic review. *PLoS ONE*, 8(7):e67737, jul 2013. doi: 10.1371/journal.pone.0067737.

- 
- J. Liu, B. Fu, H. Yang, M. Zhao, B. He, y X.H. Zhang. Phylogenetic shifts of bacterioplankton community composition along the pearl estuary: the potential impact of hypoxia and nutrients. *Frontiers in Microbiology*, 6:64, 2015.
- M. Liu, Y. Dong, Y. Zhao, G. Zhang, W. Zhang, y T. Xiao. Structures of bacterial communities on the surface of ulva prolifera and in seawaters in an ulva blooming region in jiaozhou bay, china. *World Journal of Microbiology and Biotechnology*, 27:1703–1712, 2011.
- M. Liu, T. Xiao, J. Sun, H. Wei, Y. Wu, Y. Zhao, y W. Zhang. Bacterial community structures associated with a natural spring phytoplankton bloom in the yellow sea, china. *Deep Sea Research. Part II Topical Studies in Oceanography*, 97:85–92, 2013.
- S. Liu, C. Wang, P. Wang, J. Chen, X. Wang, y Q. Yuan. Variation of bacterioplankton community along an urban river impacted by touristic city: with a focus on pathogen. *Ecotoxicology and Environmental Safety*, 165:573–581, 2018.
- Z. Liu, C. Lozupone, M. Hamady, F.D. Bushman, y R. Knight. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research*, 35:e120, 2007.
- Z. Liu, T.Z. DeSantis, G.L. Andersen, y R. Knight. Accurate taxonomy assignments from 16s rrna sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, 36:e120, 2008.
- K. G. Lloyd, A. D. Steen, J. Ladau, J. Yin, y L. Crosby. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems*, 3(5), sep 2018. doi: 10.1128/msystems.00055-18.
- M. Loecher y K. Ropkins. Rgooglemaps and loa: Unleashing r graphics power on map tiles. *Journal of Statistical Software*, 63:1–18, 2015.
- C. Lok. Mining the microbial dark matter. *Nature*, 522:270–273, 2015.
- A. Lokmer y K. Mathias Wegner. Hemolymph microbiome of pacific oysters in response to temperature, temperature stress and infection. *ISME Journal*, 9:670–682, 2015.
- P. Lopez, S. Halary, y E. Baptiste. Highly divergent ancient gene families in metagenomic samples are compatible with additional divisions of life. *Biology Direct*, 10:64–78, 2015.

- 
- C. Lovejoy, W.F. Vincent, S. Bonilla, S. Roy, M-J. Martineau, R. Terrado, M. Potvin, R. Massana, y C. Pedrós-Alió. Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas. *Journal of Phycology*, 43:78–89, 2007.
- M. Lozada, M. S. Marcos, M. G. Commendatore, M. N. Gil, y H. M. Dionisi. The bacterial community structure of hydrocarbon-polluted marine environments as the basis for the definition of an ecological index of hydrocarbon exposure. *Microbes and Environments*, 29:269, 2014.
- A. J. Lucas, C. L. Dupont, V. Tai, J. L. Largier, B. Palenik, y P. J. S. Franks. The green ribbon: Multiscale physical control of phytoplankton productivity and community structure over a narrow continental shelf. *Limnology and Oceanography*, 56:611–626, 2011.
- R. R. Malmstrom, R. P. Kiene, M. T. Cottrell, y D. L. Kirchman. Contribution of sar11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the north atlantic ocean. *Applied and Environmental Microbiology*, 70:4129–4135, 2004.
- F.P. Mancuso, S. D’Hondt, A. Willems, L. Airoidi, y O. De Clerck. Diversity and temporal dynamics of the epiphytic bacterial communities associated with the canopy-forming seaweed *Cystoseira compressa* (esper) gerloff and *nizamuddin*. *Frontiers in Microbiology*, 7:476, 2016.
- J. M. Manrique y L. R. Jones. Are ocean currents too slow to counteract sar11 evolution? a next-generation sequencing, phylogeographic analysis. *Molecular Phylogenetics and Evolution*, 107:324–337, 2017.
- M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, y J. M. Rothberg. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376–380, jul 2005. doi: 10.1038/nature03959.

- 
- M. Martinez-Garcia, D.M. Brazel, B.K. Swan, C. Arnosti, P.S. Chain, K.G. Reiten-  
ga, G. Xie, N.J. Poulton, M. Lluesma Gomez, D.E. Masland, B. Thompson, W.K.  
Bellows, K. Ziervogel, C.C. Lo, S. Ahmed, C.D. Gleasner, C.J. Detter, y R. Ste-  
panauskas. Capturing single cell genomes of active polysaccharide degraders: an  
unexpected contribution of verrucomicrobia. *PLoS One*, 7:e35314, 2012.
- F. Maruyama y S. Ueki. Evolution and phylogeny of large dna viruses, mimiviridae  
and phycodnaviridae including newly characterized heterosigma akashiwo virus.  
*Frontiers in Microbiology*, 30:1942, 2016.
- A. M. Maxam y W. Gilbert. A new method for sequencing dna. *Proceedings of the  
National Academy of Sciences USA*, 74:560–564, 1977.
- R. Mejía-Radillo, A. A. Zavala-Norzagaray, J. A. Chávez-Medina, A. A. Aguirre, y  
C. M. Escobedo-Bonilla. Presence of chelonid herpesvirus 5 (chhv5) in sea turtles  
in northern sinaloa, mexico. *Diseases of Aquatic Organisms*, 132:99–108, 2019.
- L. Michaud, A. Lo Giudice, M. Mysara, P. Monsieurs, C. Raffa, N. Leys, S. Amal-  
fitano, y R. Van Houdt. Snow surface microbiome on the high antarctic plateau  
(dome c). *PLoS One*, 9:e104505, 2014.
- M.M. Moeseneder, J.M. Arrieta, y G.J. Herndl. A comparison of dna- and rna-  
based clone libraries from the same marine bacterioplankton community. *FEMS  
Microbiology Ecology*, 51:341–352, 2005.
- A. Monier, A.Z. Worden, y T.A. Richards. Phylogenetic diversity and biogeography  
of the mamiellophyceae lineage of eukaryotic phytoplankton across the oceans.  
*Environmental Microbiology Reports*, 8:461–469, 2016.
- S.Y. Moon-van der Staay, R. De Wachter, y D. Vaultot. Oceanic 18s rdna sequences  
from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, 1:607–610,  
2001.
- R. M. Morris, M. S. Rappé, S. A. Connon, K. L. Vergin, W. A. Siebold, C. A.  
Carlson, y S. J. Giovannoni. Sar11 clade dominates ocean surface bacterioplankton  
communities. *Nature*, 420:806–810, 2002.
- R.M. Morris, B.L. Nunn, C. Frazar, D.R. Goodlett, Y.S. Ting, y G. Rocap. Compa-  
rative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization  
and energy transduction. *ISME Journal*, 4:673–685, 2010.

- 
- K. M. Morrow, A. G. Moss, N. E. Chadwick, y M. R. Liles. Bacterial associates of two caribbean coral species reveal species- specific distribution and geographic variability. *Applied and Environmental Microbiology*, 78:6438–6449, 2012.
- X. Mou, M. Vila-Costa, S. Sun, W. Zhao, S. Sharma, y M.A. Moran. Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. *Environmental Microbiology Reports*, 3:798–806, 2011.
- W. Mu, H-M. Lu, J. Chen, S. Li, y A. M. Elliott. Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing. *The Journal of Molecular Diagnostics*, 18(6):923–932, nov 2016. doi: 10.1016/j.jmoldx.2016.07.006.
- M. Mucko, S. Bosak, R. Casotti, C. Balestra, y Z. Ljubesic. Winter picoplankton diversity in an oligotrophic marginal sea. *Marine Genomics*, 42:14–24, 2018.
- F. Murtagh y P. Legendre. Ward’s hierarchical agglomerative clustering method: Which algorithms implement ward’s criterion? *Journal of Classification*, 31:274–295, 2014.
- L.K. Newbold, A.E. Oliver, T. Booth, B. Tiwari, T. Desantis, M. Maguire, G. Andersen, C.J. van der Gast, y A.S. Whiteley. The response of marine picoplankton to ocean acidification. *Environmental Microbiology*, 14:2293–2307, 2012.
- U. Nubel, F. Garcia-Pichel, y G. Muyzer. Pcr primers to amplify 16s rrna genes from cyanobacteria. *Applied and Environmental Microbiology*, 63:3327–3332, 1997.
- A. P. Nutman, V. C. Bennett, C. R. L. Friend, M. J. Van Kranendonk, y A. R. Chivas. Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature*, 537(7621):535–538, aug 2016. doi: 10.1038/nature19355.
- J. A. Nylander. Mr.aic.pl. program distributed by the author., 2004. URL <https://github.com/nylander/pMrAIC>.
- O.A. Olapade. Molecular analyses of the diversity in marine bacterioplankton assemblages along the coastline of the northeastern gulf of mexico. *Canadian Journal of Microbiology*, 56:853–863, 2010.
- J. Overmann y C. Lepleux. *Marine bacteria and archaea: diversity, adaptations, and culturability*. Springer, 2016.

- 
- E. Paradis. *Analysis of phylogenetics and evolution with R*. Springer-Verlag, 2 edition, 2012.
- D. S: Park, S. Worthington, y Z. Xi. Taxon sampling effects on the quantification and comparison of community phylogenetic diversity. *Molecular Ecology*, 27:1296–1308, 2018.
- G. Parker, M. C. Paterlini, y R. A. Violante. *El fondo marino*. INIDEP, 1 edition, 1997.
- A. L. Peralta, D. Stuart, A. D. Kent, y J. T. Lennon. A social–ecological framework for “micromanaging” microbial services. *Frontiers in Ecology and the Environment*, 12(9):524–531, nov 2014. doi: 10.1890/130308.
- S. R. Peressutti, N. L. Olivera, P. A. Babay, M. Costagliola, y H. M. Alvarez. Degradation of linear alkylbenzene sulfonate by a bacterial consortium isolated from the aquatic environment of argentina. *Journal of Applied Microbiology*, 105:476–484, 2008.
- T. Pommier, B. Canback, L. Riemann, K.H. Bostrom, K. Simu, P. Lundberg, A. Tunlid, y A. Hagstrom. Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, 16:867–880, 2007.
- R.S. Poretsky, S. Sun, X. Mou, y M.A. Moran. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environmental Microbiology*, 12:616–627, 2010.
- M. N. Price, P. S. Dehal, y A. P. Arkin. Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology Evolution*, 26: 1641–1650, 2009.
- C. Quince, A. Lanzen, R. J. Davenport, y P. J. Turnbaugh. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, 12:38, 2011.
- R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL <http://www.R-project.org/>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>.

- 
- M.S. Rappe, M.T. Suzuki, K.L. Vergin, y S.J. Giovannoni. Phylogenetic diversity of ultraplankton plastid small-subunit rRNA genes recovered in environmental nucleic acid samples from the Pacific and Atlantic coasts of the United States. *Applied and Environmental Microbiology*, 64:294–303, 1998.
- M. S: Reza, A. Kobiyama, Y. Yamada, Y. Ikeda, D. Ikeda, N. Mizusawa, K. Ikeo, S. Sato, T. Ogata, M. Jimbo, T. Kudo, S. Kaga, S. Watanabe, K. Naiki, Y. Kaga, K. Mineta, V. Bajic, T. Gojobori, y S. Watabe. Basin-scale seasonal change in marine free-living bacterioplankton community in the Ofunato Bay. *Gene*, 665: 185–191, 2018.
- T. C. S. Rodrigues, K. Subramaniam, S. D. McCulloch, J. D. Goldstein, A. M. Schaefer, P. A. Fair, J. S. Reif, G.D. Bossart, y T. B. Waltzek. Genomic characterization of a novel pegivirus species from free-ranging bottlenose dolphins (*Tursiops truncatus*) in the Indian River Lagoon, Florida. *Virus Research*, 263:98–101, 2019.
- F. Rodríguez, F. Derelle, L. Gollou, F. Le Gall, Vault D., y Moreau H. Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environmental Microbiology*, 7:853–859, 2005.
- S. M. Rodríguez, M. D. Golemba, R. H. Campos, K. Trono, y L. R. Jones. Bovine leukemia virus can be classified into seven genotypes: evidence for the existence of two novel clades. *The Journal of General Virology*, 90:2788–2797, 2009.
- K. H. Roux. Optimization and troubleshooting in PCR. *Cold Spring Harbor protocols*, 2009:pdb.ip66, April 2009. ISSN 1559-6095. doi: 10.1101/pdb.ip66.
- C.T. Rueden, J. Schindelin, M.C. Hiner, B.E. DeZonia, A.E. Walter, E.T. Arena, y Eliceiri K.W. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18:529, 2017.
- C. Salgado-Salazar, L. R. Jones, Á. Restrepo, y J. G. McEwen. The human fungal pathogen *Paracoccidioides brasiliensis* (Onygenales: Ajellomycetaceae) is a complex of two species: phylogenetic evidence from five mitochondrial markers. *Cladistics*, 26:613–624, 2010.
- I. Salter, P.E. Galand, S.K. Fagervold, P. Lebaron, I. Obernosterer, M.J. Oliver, M.T. Suzuki, y C. Tricoire. Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean Sea. *ISME Journal*, 9:347–360, 2015.



- 
- J. Sambrook, E. F. Fritsch, T. Maniatis, y H. Ludwig. Commonly used techniques in molecular cloning. En N Ford, C Nolan, y M (eds.) Ferguson, editors, *Molecular cloning. A laboratory manual. Second Edition*. Cold Spring Harbor Laboratory Press and New York. EEUU, 1989.
- F. Sanger y A. R. Coulson. A rapid method for determining sequences in dna by primed synthesis with dna polymerase. *Journal of Molecular Biology*, 94:441–448, 1975.
- F. Sanger, G. M. Air, B. G: Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison III, P. M. Slocombe, y M. Smith. Nucleotide sequence of bacteriophage phix174 dna. *Nature*, 265:687–695, 1977a.
- F. Sanger, S. Nicklen, y A. R. Coulson. Dna sequencing with chain-terminatin inhibitors. *Proceedings of the National Academy of Sciences USA*, 74:5463–5467, 1977b.
- D. Santic, S. Sestanovic, M. Solic, N. Krstulovic, G. Kuspilic, M. Ordulj, y Z. Nincevic-Gladan. Dynamics of the picoplankton community from coastal waters to the open sea in the central adriatic. *Mediterranean Marine Science*, 15:179–188, 2014.
- A.V. Sastre, N.H. Santinelli, S.H. Otano, y M.E. Isanissevich. Water quality in the lower section of the chubut river, patagonia, argentina. *Verhandlungen des Internationalen Verein Limnologie*, 26:951–955, 1998.
- B. M. Satinsky, C. B. Smith, S. Sharma, M. Landa, P. M. Medeiros, V. J. Coles, P. L. Yager, B. C. Crump, y M. A. Moran. Expression patterns of elemental cycling genes in the amazon river plume. *The ISME Journal*, 11:1852–1864, 2017.
- D.J. Scanlan, M. Ostrowski, S. Mazard, A. Dufresne, L. Garczarek, W.R. Hess, A.F. Post, M. Hagemann, I. Paulsen, y F. Partensky. Ecological genomics of marine picocyanobacteria. *Microbiol Molecular Biology Reviews*, 73:249–299, 2009.
- K.E. Scherr, D. Backes, A.G. Scarlett, W. Lantschbauer, y M. Nahold. Biogeochemical gradients above a coal tar dnapl. *Science of the Total Environment*, 563-564: 741–754, 2016.
- P. D. Schloss, D. Gevers, y S. L. Westcott. Reducing the effects of pcr amplification and sequencing artifacts on 16s rRNA-based studies. *PLoS One*, 6:e27310, 2011.

- 
- P.D. Schloss, R.A. Girard, T. Martin, J. Edwards, y J.C. Thrash. Status of the archaeal and bacterial census: an update. *MBio*, 7, 2016.
- A. Shade, S.E. Jones, J.G. Caporaso, J. Handelsman, R. Knight, N. Fierer, y J.A. Gilbert. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio*, 5:e01371–01314, 2014.
- D. Shan, G. Wei, M. Li, W. Wang, X. Li, Z. Gao, y Z. Shao. Distribution and diversity of bacterioplankton communities in subtropical seawater around xiamen island, china. *Microbiological Research*, 175:16–23, 2015.
- A.K. Shaw, A.L. Halpern, K. Beeson, B. Tran, J.C. Venter, y J.B. Martiny. It's all relative: ranking the diversity of aquatic bacterial communities. *Environmental Microbiology*, 10:2200–2210, 2008.
- Y. C. Shin, L. R. Jones, J. M. Manrique, W. Lauer, A. Carville, K. G. Mansfield, y R. C. Desrosiers. Glycoprotein gene sequence variation in rhesus monkey rhadinovirus. *Virology*, 400:175–186, 2010.
- P. D Shloss, S. L. Westcott, T. Ryabin, Hall J. R., M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, y C. F. Weber. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied Environmental Microbiology*, 75:7537–7541, 2009.
- S. M. Short, O. Rusanova, y M. A. Staniewski. Novel phycodnavirus genes amplified from canadian freshwater environments. *Aquatic Microbial Ecology*, 63:61–67, 2011.
- J. M. N. Sieburth, V. Smetacek, y J. Lenz. Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1. *Limnology and Oceanography*, 23(6):1256–1263, nov 1978. doi: 10.4319/lo.1978.23.6.1256.
- D. Sipkema, B. Holmes, S.A. Nichols, y H.W. Blanch. Biological characterisation of haliclona (?gellius) sp.: sponge and associated microorganisms. *Microbial Ecology*, 58:903–920, 2009.
- J. Sjostedt, J.B. Martiny, P. Munk, y L. Riemann. Abundance of broad bacterial taxa in the sargasso sea explained by environmental conditions but not water mass. *Applied and Environmental Microbiology*, 80:2786–2795, 2014.

- 
- L. Solden, K. Lloyd, y K. Wrighton. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Current Opinion in Microbiology*, 31:217–226, 2016.
- J. Sun, L. Steindler, J. C. Thrash, K. H. Halsey, D. P. Smith, A. E. Carter, Z. C. Landry, y S. J. Giovannoni. One carbon metabolism in sar11 pelagic marine bacteria. *Proceedings of the National Academy of Sciences ONE*, 6:e23973, 2011.
- S. Sunagawa, C.M. Woodley, y M. Medina. Threatened corals provide underexplored microbial habitats. *PLoS One*, 5:e9554, 2010.
- C.A. Suttle. Marine viruses - major players in the global ecosystem. *Nature Reviews*, 5:801–812, 2007.
- M.T. Suzuki, O. Beja, L.T. Taylor, y E.F. Delong. Phylogenetic analysis of ribosomal rna operons from uncultivated coastal marine bacterioplankton. *Environmental Microbiology*, 3:323–331, 2001.
- M.T. Suzuki, C.M. Preston, O. Beja, J.R. de la Torre, G.F. Steward, y E.F. DeLong. Phylogenetic screening of ribosomal rna gene-containing clones in bacterial artificial chromosome (bac) libraries from different depths in monterey bay. *Microbial Ecology*, 48:473–488, 2004.
- P. Taberlet, E. Coissac, M. Hajibabaei, y L. Rieseberg. Environmental dna. *Molecular Ecology*, 21:1789–1793, 2012.
- D. Talmy, S. J. Beckett, D. A. A. Taniguchi, Brussaard C. P. D., J. S. Weitz, y M. J. Follows. An empirical model of carbon flow through marine viruses and microzooplankton grazers. *Environmental Microbiology*, 21:2171–2181, 2019.
- D. Tautz y M. Renz. An optimized freeze-squeeze method for the recovery of dna fragments from agarose gels. *Analys of Biochemistry*, 132:14–19, 1989.
- H. Teeling, B.M. Fuchs, D. Becher, C. Klockow, A. Gardebrecht, C.M. Bennke, M. Kassabgy, S. Huang, A.J. Mann, J. Waldmann, M. Weber, A. Klindworth, A. Otto, J. Lange, J. Bernhardt, C. Reinsch, M. Hecker, J. Peplies, F.D. Bockelmann, U. Callies, G. Gerdts, A. Wichels, K.H. Wiltshire, F.O. Glockner, T. Schweder, y R. Amann. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science*, 336:608–611, 2012.

- 
- A. R. Templeton, K. A. Crandall, y C. F. Sing. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and dna sequence data. iii. cladogram estimation. *Genetics*, 132:619–633, 1992.
- T. Tinta, J. Vojvoda, P. Mozetic, I. Talaber, M. Vodopivec, F. Malfatti, y V. Turk. Bacterial community shift is induced by dynamic environmental parameters in a changing coastal ecosystem (northern adriatic, northeastern mediterranean sea)—a 2-year time-series study. *Environmental Microbiology*, 17:3581–3596, 2015.
- M. Troussellier, A. Escalas, T. Bouvier, y D. Mouillot. Sustaining rare marine microorganisms: macroorganisms as repositories and dispersal agents of microbial diversity. *Frontiers in Microbiology*, 8, 2017.
- K. Van Doorslaer, S. Kraberger, C. Austin, C. Farkas, M. Bergeman, E. Paunil, W. Davison, y A. Varsani. Fish polyomaviruses belong to two distinct evolutionary lineages. *Journal of General Virology*, 99(4):567–573, apr 2018. doi: 10.1099/jgv.0.001041.
- J.L. Van Etten, M.V. Graves, D.G. Müller, Boland W., y N. Delaroque. Phycodnaviridae—large dna algal viruses. *Archives of Virology*, 147:1479–1516, 2002.
- C. Verna, A. Ramette, H. Wiklund, T. G. Dahlgren, A. G. Glover, F. Gaill, y N. Dubilier. High symbiont diversity in the bone-eating worm *Osedax mucofloris* from shallow whale-falls in the north atlantic. *Environmental Microbiology*, 12:2355–2370, 2010.
- D.A. Walsh, E. Zaikova, C.G. Howes, Y.C. Song, J.J. Wright, S.G. Tringe, P.D. Tortell, y S.J Hallam. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science*, 326:578–582, 2009.
- Q. Wang, G. M. Garrity, J. M. Tiedje, y J. R. Cole. Naïve bayesian classifier for rapid assignment of rrnasequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73:5261–5267, 2007.
- Y. Wang y P.-Y. Qian. Conservative fragments in bacterial 16s rrna genes and primer design for 16s ribosomal dna amplicons in metagenomic studies. *PLoS One*, 4:e7401, 2009.

- 
- Z. Wang, D. L. Juarez, J-F. Pan, S. K. Blinebry, J. Gronniger, J. C. Clark, Z. I. Johnson, y D. E. Hunt. Microbial communities across nearshore to offshore coastal transects are primarily shaped by distance and temperature. *Environmental Microbiology*, 21:3862–3872, 2019.
- A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, y G.J. Barton. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25:1189–1191, 2009.
- M.G. Weinbauer, I. Fritz, D.F. Wenderoth, y Hoffe M.G. Simultaneous extraction from bacterioplankton of total rna and dna suitable for quantitative structure and function analyses. *Applied and Environmental Microbiology*, 68:1082–1087, 2002.
- W. G. Weisburg, S. M. Barns, D. A. Pelletier, y D. J. Lane. 16s ribosomal dna amplification for phylogenetic study. *Journal of Bacteriology*, 173:697–703, 1991.
- R. M. Welsh, J. R. Zaneveld, S. M. Rosales, J. P. Payet, D: E. Burkepile, y R. V. Thurber. Bacterial predation in a marine host-associated microbiome. *The ISME Journal*, 10:1540–1544, 2016.
- K. D. Weynberg. Viruses in marine ecosystems: from open ocean waters to coral reefs. *Advances in Virus Research*, 101, 2018.
- K. D. Weynberg, M. J. Allen, I. C. Gilg, D. J. Scanlan, y W. H. Wilson. Genome sequence of *Ostreococcus tauri* virus otv-2 throws light on the role of picoeukaryote niche separation in the ocean. *Journal of Virology*, 85:4520–4529, 2011.
- K.D. Weynberg, M.J. Allen, y W.H. Wilson. Marine prasinoviruses and their tiny plankton hosts: A review. *Viruses*, 9:E43, 2017.
- S.J. Williamson, D.B. Rusch, S. Yooseph, A.L. Halpern, K.B. Heidelberg, J.I. Glass, C. Andrews-Pfannkoch, D. Fadrosh, C.S. Miller, G. Sutton, M. Frazier, y J.C. Venter. The sorcerer ii global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One*, 3:e1456, 2008.
- W.H. Wilson, J.L. Van Etten, y M.J. Allen. The phycodnaviridae: the story of how tiny giants rule the world. *Current Topics in Microbiology and Immunology*, 328: 1–42, 2009.
- K. Woodford y Usdin K. 'turboprep' dna: ultraquick preparation of plasmid dna from single colonies for dna sequencing. *Nucleic Acids Research*, 19:6652, 1991.

- 
- W. Wu, B. Huang, y C. Zhong. Photosynthetic picoeukaryote assemblages in the south china sea from the pearl river estuary to the seats station. *Aquatic Microbial Ecology*, 71:271–284, 2014.
- Y. Yang, X. Gu, S.H. Te, S.G. Goh, K. Mani, Y. He, y K.Y. Gin. Occurrence and distribution of viruses and picoplankton in tropical freshwater bodies determined by flow cytometry. *Water Research*, 149:342–350, 2019.
- S. Yau y M. Seth-Pasricha. Viruses of polar aquatic environments. *Viruses*, 11: 189–203, 2019.
- Q.-Y. Zhang y J.-F. Gi. Diversity, evolutionary contribution and ecological roles of aquatic viruses. *Science China Life Sciences*, 61, 2018. doi: 10.1007/s11427-018-9414-7.
- Y. Zhang, N. Jiao, Z. Sun, A. Hu, y Q. Zheng. Phylogenetic diversity of bacterial communities in south china sea mesoscale cyclonic eddy perturbations. *Research in Microbiology*, 162:320–329, 2011.
- X. Zhao, C. L. Schwartz, J. Pierson, S. J. Giovannoni, J. R. McIntosh, y D. Nicastro. Three-dimensional structure of the ultraoligotrophic marine bacterium "*Candidatus pelagibacter ubique*". *Applied and Environmental Microbiology*, 83:e02807, 2010.

# Rare unclassified 16S rRNA operational taxonomic units from the uncharted Engaño Bay (Argentinean Patagonia)

Andrea Y. Calvo, Julieta M. Manrique, and Leandro R. Jones

**Abstract:** Rare microbes make up most of the diversity of marine microbiomes, and recent works have highlighted their importance for microbial community dynamics and in fragmented habitats. Rare taxa have been infrequently studied in comparison with abundant groups, and rare unclassified sequences are common in culture-independent studies. Here, we describe a detailed analysis of nonclassifiable sequences from the Chubut river estuary at the Argentinean Patagonia. Standard taxonomic assignments of environmental 16S rRNA sequences resulted in about 13% unclassified operational taxonomic units (OTUs). The potential affiliations of these OTUs could be narrowed by mapping the classification software assignments on a phylogeny obtained directly from our environmental sequence data. Customized BLAST analyses were remarkably consistent with these phylogenetic assignments, especially when the unclassified OTUs were blasted against sequences from cultured and type microorganisms. In addition, our BLAST analyses revealed significant similarities between several unclassified OTUs and a plethora of unclassified sequences from around the world. Further phylogenetic comparisons with 6194 carefully selected reference sequences showed that these unclassified sequences may correspond to 5 unnamed groups, possibly encompassing ranks from subclass to family inside the *Alphaproteobacteria*, and to an unknown *Gracilibacteria* lineage. Overall, these results demonstrate the value of straight phylogenetic analysis, customized BLAST searches, and comparisons with sequences from type material, for the systematic study of rare unclassified sequences.

*Key words:* 16S, environmental DNA, taxonomy, picoplankton, Patagonia.

**Résumé :** Les microbes rares composent la plus grande partie de la diversité du microbiome marin et des travaux récents ont souligné leur importance dans la dynamique de la communauté microbienne et dans les habitats fragmentés. Les taxons rares ont été peu étudiés comparativement aux groupes abondants et les séquences rares non classées sont fréquentes dans les études ne reposant pas sur la culture des souches. Les auteurs décrivent ici une analyse détaillée de séquences qui ne peuvent être classées provenant de l'estuaire du Rio Chubut en Patagonie Argentine. Les attributions taxonomiques standard de séquences d'ARNr 16S de l'environnement produisaient environ 13 % d'UTO qui ne pouvaient être classées. Les affiliations potentielles de ces UTO pouvaient être resserrées en cartographiant les attributions obtenues avec un logiciel de classification dans une phylogénie obtenue directement des données de séquences environnementales. Les analyses BLAST sur mesure étaient remarquablement cohérentes avec ces attributions phylogénétiques, particulièrement lorsque les UTO non classées étaient alignées avec des séquences de microorganismes cultivés et des microorganismes types. De plus, les analyses BLAST révélaient des similarités significatives entre plusieurs UTO non classées et un grand nombre de séquences non classées à travers le monde. D'autres comparaisons phylogénétiques de 6194 séquences de référence soigneusement choisies ont montré que ces séquences non classées pourraient correspondre à cinq groupes non identifiés, englobant possiblement des rangs allant des sous-classes aux familles à l'intérieur des *Alphaproteobacteria* et un lignage inconnu des *Gracilibacteria*. Globalement, ces résultats démontrent la valeur d'une analyse phylogénétique directe, de recherches BLAST sur mesure et de comparaisons de séquences de matériel type pour l'analyse systématique de séquences rares qui ne peuvent être classées. [Traduit par la Rédaction]

*Mots-clés :* 16S, ADN environnemental, taxonomie, picoplancton, Patagonie.

Received 1 June 2017. Revision received 23 October 2017. Accepted 26 October 2017.

**A.Y. Calvo, J.M. Manrique, and L.R. Jones.** Laboratorio de Virología y Genética Molecular, Facultad de Ciencias Naturales y Ciencias de la Salud (FCNyCS), Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB), 9 de Julio y Belgrano s/n (9100), Trelew, Chubut, Argentina; Consejo Nacional de Investigaciones Científicas y Técnicas, Avenida Rivadavia 1917 (C1083ACA) Buenos Aires, Argentina.

**Corresponding author:** Leandro R. Jones (emails: [lrj000@gmail.com](mailto:lrj000@gmail.com); [ljones@conicet.gov.ar](mailto:ljones@conicet.gov.ar)).

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from [RightsLink](https://www.nrcresearchpress.com/cjm).

Marine bacteria constitute a large and complex assemblage of taxa. This huge diversity is in a discovery stage, and unclassified operational taxonomic units (uOTUs) are common in environmental DNA studies. This problem especially affects rare taxa, which despite contributing a large share of marine microbial diversity, remain untamed compared with abundant groups (Schloss et al. 2016). Recent works have shown that rare taxa can be important to community dynamics (Shade et al. 2014) and act as a biodiversity reservoir; hence, they are potentially relevant for future environmental change (Choi et al. 2017). Furthermore, it has been observed that some groups that are rare in seawater can be important in particular marine habitats (Sunagawa et al. 2010).

In this work, we set out to shed light on the systematics of rare uOTUs identified among 16S ribosomal gene (rRNA) environmental sequences (eSEQs) from the Argentinean Patagonia. We were particularly interested in identifying unclassified sequences (uSEQs) presenting similarities to sequences from elsewhere because cosmopolitan lineages are more likely to represent conditionally rare groups (Shade et al. 2014), to be important in fragmentary habitats (Sunagawa et al. 2010), and to potentially impact the future microbial biodiversity (Shade et al. 2014; Choi et al. 2017). We combined phylogenies obtained directly from our 16S rDNA sequence data with assignments obtained by a standard classification software and performed customized BLAST analyses and further phylogenetic comparisons with carefully selected reference sequences. The workflow implemented in this study is outlined in supplemental Fig. S1<sup>1</sup>. Several uOTUs were closely related to previously described uSEQs from around the world. We provide evidence that these cosmopolitan uSEQs may belong to 5 undescribed *Alphaproteobacteria* lineages and an unknown *Gracilibacteria* clade.

A seawater sample (N, supplemental Table S1<sup>1</sup>) collected in the uncharted Engaño Bay at about 2 km seawards from the mouth of the Chubut River (43.34°S, 65.03°W) was used to generate high-quality 16S rRNA sequences by high-throughput sequencing (HTS; supplemental Materials and Methods<sup>1</sup>). These sequences were grouped in 97% similarity OTUs, and the obtained OTUs were classified up to the Order rank with Wang's Naïve Bayesian Classifier (WNBC) implemented in Mothur (Schloss et al. 2011) using an 80% cutoff (supplemental Materials and Methods<sup>1</sup>). For comparison, 3 independent samples (F, J, A; supplemental Table S1<sup>1</sup>) were sequenced by molecular cloning and Sanger sequencing and analyzed in an equivalent manner (supplemental Materials and Methods<sup>1</sup>).

#### About 13% of the Patagonian OTUs could not be classified

The WNBC analyses reflected a microbial diversity similar to that of environments akin from elsewhere. A few bacterial groups dominated the bacterioplankton and were accompanied by several taxa present at relatively low frequencies (supplemental Table S1<sup>1</sup>). We also observed significant amounts of 16S sequences from picoalgae plastids. Out of 250 OTUs, 33 could not be classified (Table 1; supplemental Table S2<sup>1</sup>). Four OTUs could not be classified in any bacterial phylum and 6 were assigned to *Proteobacteria*. Additionally, 7 OTUs could not be assigned beyond *Gammaproteobacteria* and 16 OTUs seemingly corresponded to unknown *Alphaproteobacteria*. There were no significant differences between the number of unclassified Sanger and HTS sequences (supplemental Table S3<sup>1</sup>). However, the proportion of unclassified Sanger OTUs was slightly lower than the proportion of HTS ones (supplemental Table S3<sup>1</sup>), probably due to the higher resolution of Sanger sequences (longer sequences).

#### Straight phylogenetic analysis (SPA) shed light on the affiliations of the uOTUs

We reassessed the affiliation of our uOTUs by establishing their phylogenetic relations with OTUs that could be classified by the WNBC analyses. To this aim, we inferred a phylogeny, and the WNBC taxonomic assignments were mapped on the obtained tree (Fig. 1). To align the sequences, we used a multistep, taxonomy-guided approach, in which we fed the sequence alignment algorithm with taxonomic information to facilitate homology detection, after which, the obtained alignment was submitted to maximum-likelihood phylogenetic analysis (supplemental Materials and Methods<sup>1</sup>). The obtained results indicated that some uOTUs may belong or be related to known groups. The uOTUs 0081, 0098, 0136, 0147, and 0165 were clustered in the SAR11 clade (*Pelagibacterales*), which was supported by a relatively high bootstrap support (Fig. 1a). Two uOTUs (0022 and 0138) were distantly related to the SAR406 clade of *Deferribacterales* (*Marinimicrobia*) (Fig. 1b), and likewise, the uOTU 0170 was distantly related to *Gracilibacteria* sequences (Fig. 1b). Three extra uOTUs formed monophyletic groups with the *Flavobacteriales* (uOTU 0159), OCS116 (*Rhodobacterales*; uOTU 0067) and SAR86 (*Oceanospirillales*) (uOTU 050) (Fig. 1a). Several uOTUs were close to sequences classified as *Oceanospirillales* (Fig. 1a). Likewise, many uOTUs were located close to sequences classified as *Rickettsiales* and *Rhodobacterales* (Fig. 1a). Two uOTUs (0103 and 0146) did not cluster monophyletically with other sequences and were very divergent with respect to the rest of sequences (Figs. 1a and 1c). These results are summarized in Table 1 and supplemental Table S2<sup>1</sup>.

<sup>1</sup>Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/cjm-2017-0342>.



**Table 1.** Operational taxonomic units that could not be classified by WNBC analyses (uOTUs), and summary of SPA and BLAST results.

uOTU	WNBC <sup>a</sup>	SPA <sup>b</sup>	BLAST hit <sup>c</sup>	Taxon <sup>d</sup>	Homology <sup>e</sup>
034	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	JN591806.1	PS1 clade	93%
043	<i>Proteobacteria</i>	<i>Rhodobacterales</i>	EU802859.1	<i>Bacteria</i>	90%
050	<i>Gammaproteobacteria</i>	SAR86	EU802840.1	<i>Bacteria</i>	91%
0022	<i>Bacteria</i>	<i>Deferribacterales</i>	JQ199557.1	<i>Bacteria</i>	<b>100%</b>
0058	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	EU799247.1	<i>Bacteria</i>	94%
0067	<i>Alphaproteobacteria</i>	OCS116	JN591913.1	PS1 clade	94%
0076	<i>Gammaproteobacteria</i>	<i>Oceanospirillales</i>	FR686176.1	<i>Bacteria</i>	96%
0078	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	FR685064.1	<i>Bacteria</i>	<b>99%</b>
0081	<i>Alphaproteobacteria</i>	SAR11	GQ346819.1	<i>Alphaproteobacteria</i>	<b>97%</b>
0098	<i>Alphaproteobacteria</i>	SAR11	EU802479.1	<i>Bacteria</i>	95%
0099	<i>Gammaproteobacteria</i>	<i>Oceanospirillales</i>	GU234924.1	<i>Bacteria</i>	96%
0101	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	FR683303.1	<i>Bacteria</i>	93%
0102	<i>Gammaproteobacteria</i>	<i>Oceanospirillales</i>	KX935466.1	<i>Bacteria</i>	95%
0103	<i>Gammaproteobacteria</i>	—	HQ163109.1	<i>Gammaproteobacteria</i>	95%
0105	<i>Proteobacteria</i>	<i>Oceanospirillales</i>	FR684004.1	<i>Bacteria</i>	93%
0112	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	DQ234144.2	<i>Rhodobacteraceae</i>	92%
0113	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	EU799247.1	<i>Bacteria</i>	95%
0118	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	FJ825915.1	<i>Bacteria</i>	93%
0127	<i>Alphaproteobacteria</i>	<i>Rickettsiales</i>	JF488534.1	<i>Alphaproteobacteria</i>	<b>98%</b>
0128	<i>Gammaproteobacteria</i>	<i>Oceanospirillales</i>	JQ269288.1	<i>Bacteria</i>	92%
0135	<i>Alphaproteobacteria</i>	<i>Rhodobacterales</i>	FR685064.1	<i>Bacteria</i>	96%
0136	<i>Alphaproteobacteria</i>	SAR11	HQ672177.1	<i>Bacteria</i>	<b>97%</b>
0138	<i>Bacteria</i>	<i>Deferribacterales</i>	KU034616.1	<i>Bacteria</i>	<b>97%</b>
0145	<i>Proteobacteria</i>	<i>Rickettsiales</i>	EU802860.1	<i>Bacteria</i>	90%
0146	<i>Proteobacteria</i>	—	JN986470.1	<i>Bacteria</i>	96%
0147	<i>Alphaproteobacteria</i>	SAR11	NR_074224.1	SAR11	95%
0148	<i>Alphaproteobacteria</i>	<i>Rickettsiales</i>	JF488580.1	<i>Alphaproteobacteria</i>	<b>98%</b>
0155	<i>Proteobacteria</i>	<i>Oceanospirillales</i>	HQ163230.1	<i>Actinobacteria</i>	94%
0159	<i>Bacteria</i>	<i>Flavobacteriales</i>	KC336640.1	<i>Bacteria</i>	89%
0165	<i>Proteobacteria</i>	SAR11	NR_074224.1	SAR11	95%
0168	<i>Gammaproteobacteria</i>	<i>Oceanospirillales</i>	JQ269279.1	<i>Bacteria</i>	94%
0170	<i>Bacteria</i>	<i>Gracilibacteria</i>	KU688616.1	<i>Bacteria</i>	<b>97%</b>
0176	<i>Alphaproteobacteria</i>	<i>Rickettsiales</i>	JQ199248.1	<i>Bacteria</i>	93%

<sup>a</sup>Wang's Naïve Bayesian Classifier (WNBC) assignment (80% cutoff).

<sup>b</sup>SPA, straight phylogenetic analysis. Closer taxon according to phylogenetic analysis.

<sup>c</sup>GenBank match ID.

<sup>d</sup>GenBank taxonomy of the matching sequence.

<sup>e</sup>Percent identity. Bold indicates significant homology.

### BLAST analyses revealed similarities to uSEQs from remote regions

The uOTUs were further compared with those in the GenBank database, with the aim of detecting similarities to uSEQs from elsewhere and (or) taxa not represented among our sequences. We performed comparisons with the whole database and with the subsets of sequences corresponding to cultured organisms and type specimens (Federhen 2015). The corresponding results are summarized in Table 1 and detailed in supplemental Tables S4–S7<sup>1</sup>. Eight uOTUs were very similar to sequences from remote places (supplemental Table S7<sup>1</sup>). The uOTU 0022 displayed 100% homology with an eSEQ from California (USA), and it was 99% identical to a sequence from an isolate from the Jiulong River estuary in China. This uOTU also displayed similarities above 97% to many sequences from several places worldwide (supplemental Table S7<sup>1</sup>). The uOTUs 0078 and 0081 were very similar to

sequences amplified from fjords in Norway and Vancouver Island, respectively. In addition, the uOTU 0081 was significantly similar (>97%) to a sequence obtained from the northeast subarctic Pacific Ocean (supplemental Table S7<sup>1</sup>). The uOTUs 0136 and 0138 were 97% homologous with sequences from the North Pacific and Austria, respectively. The uOTUs 0127 and 0148 were very similar to *Alphaproteobacteria* isolates from the Gulf of Maine (USA) and to 65 eSEQs from diverse geographic origins (Table S7<sup>1</sup>). The uOTU 0170 displayed a 97% homology with eSEQs from Italy (Table S7<sup>1</sup>).

As expected, the similarities to type material sequences were rather low, but interestingly, there was quite a good concordance with our SPA assignments. For example, phylogenetic and BLAST comparisons with both cultured and type material suggested a relationship with *Rhodobacterales* for the uOTU 0058 (supplemental

**Fig. 1.** Straight phylogenetic analysis (SPA) of 250 97% operational taxonomic units (OTUs) from the uncharted Engaño Bay. Panel *a* displays the obtained tree, with Wang's Naïve Bayesian Classifier taxonomy mapped on it. Panels *b* and *c* show detailed views of the clades indicated as *b* and *c*, respectively, in panel *a*. Numbers close to branches indicate bootstrap supports ( $n = 100$ , panel *a*) or aBayes local supports (panels *b* and *c*). The scale bars' units in panels *b* and *c* are substitutions per aligned position. [Colour online.]

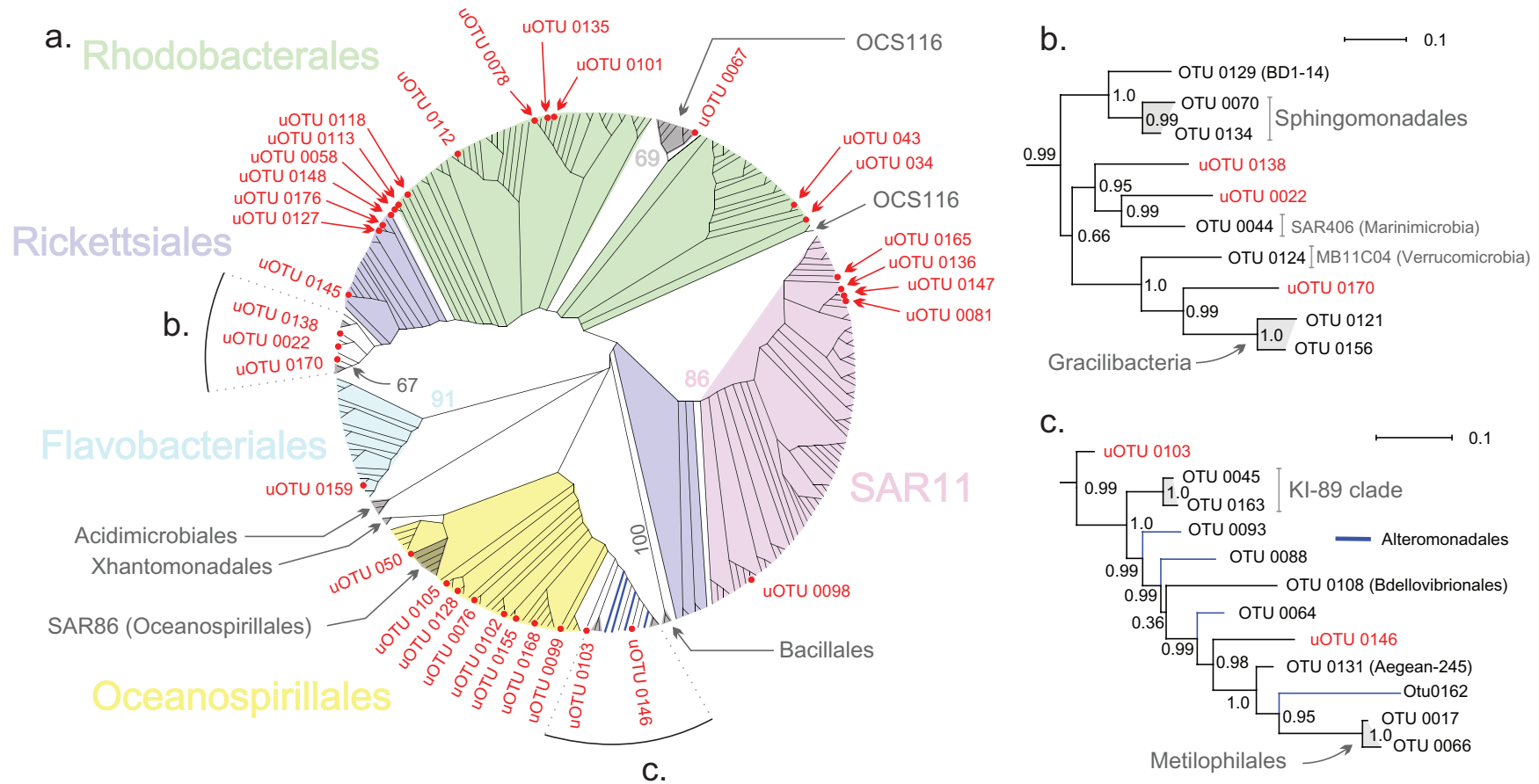


Table S2<sup>1</sup>), and likewise, the uOTUs 0081 and 0136 were monophyletically clustered with SAR11 sequences and more similar to SAR11 type material (supplemental Table S2<sup>1</sup>).

#### The cosmopolitan uSEQs may belong to unknown *Alphaproteobacteria* and *Gracilibacteria* lineages

Our SPA suggested potential relationships between the cosmopolitan uOTUs and several *Alphaproteobacteria* groups and the phyla *Gracilibacteria*, *Verrucomicrobia*, and *Marinimicrobia* (Fig. 1). Thus, we combined the uOTU sequences and the similar sequences from around the world (supplemental Table S7<sup>1</sup>) with 6194 reference sequences from these taxa, and submitted the entire data to phylogenetic analysis, as detailed in section 2 (Extended uOTU analyses) of supplementary Materials and Methods<sup>1</sup>. In agreement with the SPA, the obtained phylogeny showed that the uOTU 0078 and the similar sequences from elsewhere may correspond to a divergent *Rhodobacteriales* lineage and that the uOTUs 0127 and 0148, together with 66 sequences from different places worldwide, may belong to a divergent *Rickettsiales* family (supplemental Fig. S4<sup>1</sup>). In addition, the analysis also corroborated that the uOTUs 0136 and 0081 and 2 unknown sequences from Vancouver Island and the northeast Pacific Ocean were closely related to the SAR11 clade. The uOTU 0138 and 6 uSEQs from elsewhere were clustered in a long branch, close to the Candidatus *Midichloriaceae* (*Rickettsiales*) family. Interestingly, the uOTU 0022 and 159 sequences from diverse places from around the world were included in the *Alphaproteobacteria* clade, but were clearly separated from the 3 previously proposed *Alphaproteobacteria* subclasses (Ferla et al. 2013). The uOTU 0170 and 2 uSEQs from marine epiphytic bacteria obtained in Italy (Mancuso et al. 2016) were clustered forming a divergent *Gracilibacteria* clade.

#### Discussion

As observed in similar studies performed elsewhere, our study region presented a high abundance of SAR11, *Rhodobacteriales*, *Acidimicrobiales*, *Oceanospirillales*, and picoalgae chloroplast sequences (Johnson and Sieburth 1982; Suzuki et al. 2001; Biegala et al. 2003; Lovejoy et al. 2007; Alonso-Gutiérrez et al. 2009; Campbell et al. 2009; Feng et al. 2009; Fortunato et al. 2013; Wu et al. 2014; Liu et al. 2015). Likewise, other taxa, including *Bacillales*, *Xantomonadales*, SAR86, OCS116, *Rhizobiales*, *Rickettsiales*, MB11C04, *Bdellobibrionales*, *Alteromonadales*, *Aegean-245*, *Methylophilales*, KI89A, SAR406, *Sphingomonadales*, and *Gracilibacteria* (also known as BD1-5 and GN02), were detected at low frequencies. About 13% of the OTUs identified here could not be classified. The fact that there was no significant difference between the proportions of Sanger and HTS uSEQs and that there was only a slightly higher proportion of HTS uOTUs, highlight the value of the V1–V3 regions of the 16S gene in taxonomic studies (Liu et al. 2007, 2008). Difficulties in eSEQ classification

can be generally attributed to the sketchy knowledge of marine microbial systematics (Schloss et al. 2016). However, methodological issues can also represent a challenge, such as undetected chimaeras, artificial classifications (i.e., para- or polyphyletic taxa) or phenomena not accounted for by classification algorithms, such as gene transformation and deep coalescence. Notwithstanding that, the identification of previously described uSEQs similar to several uOTUs (supplemental Table S7<sup>1</sup>; supplemental Fig. S4<sup>1</sup>) provided strong evidence that these OTUs actually correspond to real 16S lineages. In addition, this observation supports the concept that locally rare taxa could be relevant at a worldwide scale (Sunagawa et al. 2010; Shade et al. 2014; Choi et al. 2017).

Phylogenetic classification has been used in a limited way in microbial ecology compared with classification algorithms based on sequence similarity and composition (Hiraoka et al. 2016). Here, we show that SPA and (de novo) phylogenetic comparisons with reference sequences can greatly improve the taxonomic assignment of uSEQs and may drive the discovery of new taxa. These findings suggest an untapped potential of phylogenetic evidence in eDNA systematic studies. Some challenges faced by phylogenetic classification could be also foreground. A conspicuous one is that some taxa may not be retrieved by phylogenetic analysis, as observed here for the *Rickettsiales*, *Rhodobacteriales*, and *Rhizobiales* (Fig. 1; supplemental Fig. S4<sup>1</sup>). This can be due to some taxa actually not being monophyletic or to methodological or molecular marker issues. The implementation of scoring statistics is also challenging. The use of branch supports is an apparently straightforward solution. However, depending on the used marker and method, some (actual) groups may not be retrieved as supported clades. Furthermore, it must be considered that the probability of recovering a given clade by chance (even with no or a low support) is extremely low, and depends on the number of sequences from the group of interest and the total number of sequences in the data. This problem becomes more complicated if topological uncertainty (i.e., multiple optimal phylogenies) is taken into account. A good amount of research and analysis of empirical data will be required to approach the subject in a reasonably finished form.

Finally, it is also worth highlighting the observed correspondence between our phylogenetic and BLAST analyses (supplemental Table S2<sup>1</sup>). Interestingly, the correspondence was greater for the BLAST comparisons with sequences from type specimens, which we attribute to a more accurate taxonomic determination of the corresponding exemplars. This highlights the value of sequences from type material in eSEQs classification.

#### Acknowledgements

This work was partially supported by grant PICT-2795 from the National Agency of Scientific and Technological Promotion (ANPCyT) and the Asociación Civil Argentina

Genetics (ArGen). We want to thank the useful comments provided by 2 anonymous reviewers.

## References

- Alonso-Gutiérrez, J., Lekunberri, I., Teira, E., Gasol, J.M., Figueras, A., and Novoa, B. 2009. Bacterioplankton composition of the coastal upwelling system of 'Ría de Vigo', NW Spain. *FEMS Microbiol. Ecol.* **70**(3): 493–505. doi:10.1111/j.1574-6941.2009.00766.x. PMID:19780832.
- Biegala, I.C., Not, F., Vaultot, D., and Simon, N. 2003. Quantitative assessment of picoeukaryotes in the natural environment by using taxon-specific oligonucleotide probes in association with tyramide signal amplification-fluorescence in situ hybridization and flow cytometry. *Appl. Environ. Microbiol.* **69**(9): 5519–5529. doi:10.1128/AEM.69.9.5519-5529.2003. PMID:12957941.
- Campbell, B.J., Yu, L., Straza, T.R.A., and Kirchman, D.L. 2009. Temporal changes in bacterial rRNA and rRNA genes in Delaware (USA) coastal waters. *Aquat. Microb. Ecol.* **57**: 123–135. doi:10.3354/ame01335.
- Choi, C.J., Bachy, C., Jaeger, G.S., Poirier, C., Sudek, L., Sarma, V.V.S.S., et al. 2017. Newly discovered deep-branching marine plastid lineages are numerically rare but globally distributed. *Curr. Biol.* **27**(1): R15–R16. doi:10.1016/j.cub.2016.11.032. PMID:28073013.
- Federhen, S. 2015. Type material in the NCBI Taxonomy Database. *Nucleic Acids Res.* **43**: D1086–D1098. doi:10.1093/nar/gku1127. PMID:25398905.
- Feng, B.W., Li, X.R., Wang, J.H., Hu, Z.Y., Meng, H., Xiang, L.Y., and Quan, Z.X. 2009. Bacterial diversity of water and sediment in the Changjiang estuary and coastal area of the East China Sea. *FEMS Microbiol. Ecol.* **70**(2): 236–248. doi:10.1111/j.1574-6941.2009.00772.x.
- Ferla, M.P., Thrash, J.C., Giovannoni, S.J., and 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**(12): e83383. doi:10.1371/journal.pone.0083383. PMID:24349502.
- Fortunato, C.S., Eiler, A., Herfort, L., Needoba, J.A., Peterson, T.D., and Crump, B.C. 2013. Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. *ISME J.* **7**(10): 1899–1911. doi:10.1038/ismej.2013.79. PMID:23719153.
- Hiraoka, S., Yang, C.C., and Iwasaki, W. 2016. Metagenomics and bioinformatics in microbial ecology: current status and beyond. *Microbes Environ.* **31**(3): 204–212. doi:10.1264/jisme2.ME16024. PMID:27383682.
- Johnson, P.W., and Sieburth, J.M. 1982. In-situ morphology and occurrence of eucaryotic phototrophs of bacterial size in the picoplankton of estuarine and oceanic waters. *J. Phycol.* **18**(3): 318–327. doi:10.1111/j.1529-8817.1982.tb03190.x.
- Liu, J., Fu, B., Yang, H., Zhao, M., He, B., and Zhang, X.H. 2015. Phylogenetic shifts of bacterioplankton community composition along the Pearl Estuary: the potential impact of hypoxia and nutrients. *Front. Microbiol.* **6**: 64. doi:10.3389/fmicb.2015.00064. PMID:25713564.
- Liu, Z., Lozupone, C., Hamady, M., Bushman, F.D., and Knight, R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* **35**(18): e120. doi:10.1093/nar/gkm541. PMID:17881377.
- Liu, Z., DeSantis, T.Z., Andersen, G.L., and Knight, R. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res.* **36**(18): e120. doi:10.1093/nar/gkn491. PMID:18723574.
- Lovejoy, C., Vincent, W.F., Bonilla, S., Roy, S., Martineau, M.-J., Terrado, R., et al. 2007. Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic Seas. *J. Phycol.* **43**(1): 78–89. doi:10.1111/j.1529-8817.2006.00310.x.
- Mancuso, F.P., D'Hondt, S., Willems, A., Airoidi, L., and De Clerck, O. 2016. Diversity and temporal dynamics of the epiphytic bacterial communities associated with the canopy-forming seaweed *Cystoseira compressa* (Esper) Gerloff and Nizamuddin. *Front. Microbiol.* **7**: 476. doi:10.3389/fmicb.2016.00476. PMID:27092130.
- Schloss, P.D., Gevers, D., and Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*, **6**(12): e27310. doi:10.1371/journal.pone.0027310. PMID:22194782.
- Schloss, P.D., Girard, R.A., Martin, T., Edwards, J., and Thrash, J.C. 2016. Status of the archaeal and bacterial census: an update. *mBio*, **7**(3): e00201–16. doi:10.1128/mBio.00201-16. PMID:27190214.
- Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J.A. 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio*, **5**(4): e01371–14. doi:10.1128/mBio.01371-14. PMID:25028427.
- Sunagawa, S., Woodley, C.M., and Medina, M. 2010. Threatened corals provide underexplored microbial habitats. *PLoS One*, **5**(3): e9554. doi:10.1371/journal.pone.0009554. PMID:20221265.
- Suzuki, M.T., Béjà, O., Taylor, L.T., and DeLong, E.F. 2001. Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ. Microbiol.* **3**(5): 323–331. doi:10.1046/j.1462-2920.2001.00198.x. PMID:11422319.
- Wu, W., Huang, B., and Zhong, C. 2014. Photosynthetic picoeukaryote assemblages in the South China Sea from the Pearl River estuary to the SEATS station. *Aquat. Microb. Ecol.* **71**: 271–284. doi:10.3354/ame01681.