



**Universidad Nacional de la Patagonia
San Juan Bosco**

**Facultad de Ciencias Naturales
y Ciencias de la Salud**

Departamento de Biología y ambiente

**Biogeografía de linajes ubicuos y
profusos (LUP) de bacterias del Mar
Argentino: basado en datos de
secuenciación profunda.**

Trabajo de Tesis para optar al Título de Doctor en Ciencias Biológicas

Lic. Giaccardi Laura Inés

Directores

Dr. Jones Leandro R.

Dra. Bioq. Manrique Julieta M.

2023

Agradecimientos

En primer lugar, quiero agradecer a mis directores, Dr. Leandro Jones y Dra. Julieta Manrique, por darme a oportunidad de realizar la tesis doctoral, acompañarme con su dirección en este trayecto del doctorado. Les agradezco por la predisposición, el tiempo y el esfuerzo invertido en mi formación.

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.

También agradecer especialmente a la Secretaría de investigación y posgrado de la Facultad de Ciencias Naturales y Ciencias de la Salud de la Universidad Nacional de la Patagonia San Juan Bosco, por la predisposición.

Y, en segundo lugar, pero no menos importante, a toda mi familia por su apoyo incondicional. Cuando digo familia hablo de mi familia cordobesa, mis viejos, hermanos y abuelos, como así también a mis amigas. Pero por sobre todo agradecer infinitamente a toda la familia madrynense que me acompaña desde 2009, quienes han sido mi apoyo más grande en esta etapa; a quienes no nombró para que no quede nadie afuera.

INDICE GENERAL

AGRADECIMIENTOS	2
INDICE GENERAL.....	3
ÍNDICE FIGURAS.....	6
ÍNDICE TABLAS	7
RESUMEN	8
SUMMARY	12
ABREVIATURAS.....	16
1 INTRODUCCIÓN	17
1.1 Bacteriomas marinos y su rol en el ambiente	17
1.2 El advenimiento de nuevas tecnologías independientes de cultivo ha permitido dilucidar la composición de los microbiomas.....	19
1.3 Biogeografía: distribución espacio-temporal de la diversidad.....	20
1.3.1. Análisis cuantitativos de las distribuciones biológicas	21
1.4 Biogeografía bacteriana.....	23
1.5 Biogeografía bacteriana del Mar Argentino.....	26
1.6 Objetivos	28
Objetivo general.....	28
Objetivos particulares	28
2 MATERIALES Y MÉTODOS.....	29
2.1 Bases de datos utilizadas en este estudio	29
2.1.1 Muestras del Mar Argentino	29
2.1.2 Comunidad Simulada.	31
2.2 Secuencias de alta calidad.....	31
2.2.1 Procesamiento de secuencias: esquema general.....	31
2.2.2 Optimización del control de calidad.....	34

2.2.3 Desarrollos creados para el proceso de control de calidad	34
2.3 Análisis comunitarios	36
2.3.1 Identificación de linajes ubicuos y profusos (LUPs).....	36
2.3.2 Construcción de una base de datos de referencia de LUPs	37
2.3.3 Ecuilización de los esfuerzos de muestreo.....	39
2.3.4 Agrupación en Unidades Taxonómicas Operacionales (OTUs).	40
2.3.5 Generación de archivos complementarios para análisis comunitarios.....	40
2.4 Estimación de relaciones taxa-área (TAR)	41
2.5 Estimación de Decaimientos por Distancia (DD)	43
2.6 Ordenamiento mediante escalado multidimensional no métrico	43
2.7 Determinación de patrones de endemismo	44
2.8 Estimación y análisis espacial de la diversidad	44
3 RESULTADOS	45
3.1 Obtención de secuencias de alta calidad	45
3.1.1 Ajuste mediante una base de datos reducida (comunidad simulada).	45
3.1.2 Obtención de secuencias de alta calidad del Mar Argentino.....	47
3.2 Taxonomía y distribución de grupos bacterianos abundantes	49
3.2.1 Identificación de Linajes Ubicuos y Profusos (LUP).....	49
3.2.2 Distribución de los Linajes Ubicuos y Profusos	51
3.3 Estudio de las relaciones taxa-área (TAR)	53
3.4 Decaimiento por distancia	56
3.5 Escalado multidimensional no métrico	65
3.6 Patrones de endemismo en LUPs del Mar Argentino	75
3.6.1 Patrones de endemismo por linaje.	75
3.6.2 Patrones de endemismo por muestra.....	76
3.7 Distribución espacial de la diversidad	81
4 DISCUSIÓN	90
4.1 Características generales del conjunto de datos (dataset)	90
4.2 Linajes Ubicuos y Profusos	91

4.3 Relaciones taxa-área	93
4.4 Análisis de las curvas de Decaimiento por distancia	94
4.5 Escalado Multidimensional no métrico	96
4.6 Análisis de endemismo	96
4.7 Índices de diversidad	97
5 CONCLUSIONES	98
6 BIBLIOGRAFÍA	99
7 ANEXO	106

Índice de Figuras

1-1 Función de los microorganismos en los ciclos de los principales elementos.	18
1-2 Esquema representando la biogeografía de un conjunto de siete especies hipotéticas.	22
1-3 <i>Kernel</i> de dispersión.	24
1-4 Decaimiento por distancia entre ensamblajes de SAR11 distribuidos globalmente.	25
1-5 1-5 Mapa de la Argentina y Plataforma Continental.	27
2-1 Localización de los sitios de muestreo.	30
2-2 Diagrama de la obtención, agrupamiento y clasificación de las SAC.	33
3-1 Efecto del filtrado por largo de secuencias en la longitud de las secuencias finales.	46
3-2 Porcentajes de SAC de cada LUPs en las muestras del Mar Argentino.	52
3-3 Curvas taxa-área.	54
3-4 Curvas de DD basadas en el índice de disimilitud de Bray- Curtis.	57
3-5 Curvas de DD basadas en el índice de disimilitud de Morisita.	60
3-6 Curvas de DD basadas en el índice de disimilitud de Sorensen.	63
3-7 Ordenamientos del Escalado Multidimensional no Métrico basado en el índice de disimilitud de Bray-Curtis.	67
3-8 Ordenamientos del Escalado Multidimensional no Métrico basado en el índice de disimilitud de Sorensen.	69
2-9 Ordenamientos del Escalado Multidimensional no Métrico basado en el índice de disimilitud de Morisita.	72
3-10 Ordenamientos del Escalado Multidimensional no Métrico basado en el índice de disimilitud de Horn.	74
3-11 Distribución de los valores del índice de Diversidad de Shannon.	82
3-12 Distribución de los valores de índice de Diversidad de Shannon en función de la latitud de las muestras.	83
3-13 Distribución de los valores del índice de Diversidad de Simpson.	85
3-14 Distribución de los valores de índice de Diversidad de Simpson en función de la latitud de las muestras.	86
3-15 Distribución de los valores del índice de Homogeneidad de Shannon	88
3-16 Distribución de los valores de índice de Homogeneidad de Shannon en función de la latitud de las muestras.	89

Índice Tablas

3-1 Efecto del filtrado por largo de secuencia en el largo de las secuencias finales.	45
3-2 Efecto filtrado mediante optimización sobre distintos parámetros.	47
3-3 Distribución de SAC en las muestras y su cobertura.	48
3-4 Proporciones (%) de SAC por muestra alineadas con la base de datos SILVA.	50
3-5 Valores de los coeficientes de correlación ($Rho\ p$), regresión (R^2) y pendientes (z) de la TAR.	55
3-6 Valores de los coeficientes de correlación ($Rho\ p$), regresión (R^2) y pendientes (β) de las curvas DD basados en el índice de disimilitud de Bray- Curtis.	58
3-7 Valores de los coeficientes de correlación ($Rho\ p$), regresión (R^2) y pendientes (β) del análisis DD basados en el índice de disimilitud de Morisita.	61
3-8 Valores de los coeficientes de correlación ($Rho\ p$), regresión (R^2) y pendientes (β) del análisis de DD basados en el índice de disimilitud de Sorensen.	64
3-9 Número de OTUs endémicas de cada LUP.	75
3-10 Distribución de las OTUs endémicas del linaje <i>Amylibacter</i> .	76
3-11 Distribución de las OTUs endémicas del linaje NS5.	77
3-12 Distribución de las OTUs endémicas del linaje OM1.	78
3-13 Distribución de las OTUs endémicas del linaje SAR11.	79
3-14 Distribución de las OTUs endémicas del linaje SAR86.	80

Resumen

Las tecnologías independientes de cultivo han producido un punto de quiebre en la microbiología al demostrar que contamos con conocimientos muy rudimentarios sobre la filogenia, metabolismo y ecología de la vasta mayoría de los microorganismos. Gracias a las nuevas tecnologías de secuenciación, contamos actualmente con un nivel de detalle sin precedente de la composición taxonómica de los microbiomas, el cual ha puesto en evidencia que la misma es en extremo compleja. Por ejemplo, se ha demostrado que, como ocurre con los ecosistemas terrestres, las comunidades bacterianas marinas son dominadas por un reducido conjunto de grupos taxonómicos, los cuales son acompañados por una plétora de grupos minoritarios, aunque en extremo diversos. Dada la elevada biomasa y abundancia de las bacterias en el océano, no es difícil vislumbrar que las mismas influyen significativamente en los ciclos biogeoquímicos, regulando su estado y controlando el destino y la magnitud de la reserva de carbono orgánico. Hay que destacar nuevamente que, a pesar de los avances arriba mencionados, estamos en un punto alejado de contar con un conocimiento acabado de la composición, funcionamiento y distribución de las comunidades bacterianas marinas.

En el presente trabajo se estudió la composición y distribución de comunidades bacterianas de agua superficial de 17 puntos remotos del Mar Argentino, con el fin de identificar grupos abundantes y ubicuos en nuestros mares, y determinar cuantitativamente si la distribución espacial de los mismos es uniforme (concepto tradicional de que todos los microorganismos están potencialmente en todos lados) o heterogénea. La biogeografía bacteriana es un área poco estudiada en la microbiología tradicional. Sin embargo, estudios recientes han demostrado que muchos grupos presentan una biogeografía, hecho que ha reactivado el interés de la comunidad científica, con el consecuente florecimiento de varios trabajos en los últimos años. En esta Tesis, se utilizó una base de datos de aproximadamente tres millones de secuencias pareadas del gen de la subunidad ribosomal pequeña (gen 16S) obtenidas mediante secuenciación Illumina de ADN picoplanctónico. Estas secuencias crudas obtenidas (o lecturas) fueron sometidas a un complejo y estricto control de calidad. El mismo implicó un proceso de optimización en el cual se utilizó una “comunidad simulada”, de tamaño reducido y composición conocida.

Luego del control de calidad, se obtuvieron 302.597 secuencias de alta calidad de las comunidades del Mar Argentino. La clasificación taxonómica de las mismas permitió determinar que las secuencias más abundantes en nuestros mares corresponden al orden *Pelagibacterales* (*Alphaproteobacteria*, *Proteobacteria*), también conocido como SAR11, y al grupo OM1 (actualmente conocido como *Ca. Actinomarina*) del phylum *Actinobacteria*, los cuales en conjunto representaron aproximadamente el 80% de las secuencias. También se identificaron taxa que, aunque menos abundantes que los ya

mencionados, exhibieron abundancias sustancialmente mayores que el resto de los grupos detectados. Estos fueron el orden *Flavobacteriales* (o NS5) del phylum *Bacteroidetes*, el orden *Oceanospirillales* (o SAR86) de la clase *Gammaproteobacteria*, y el orden *Rhodobacterales* (género *Amylibacter*) de la clase *Alphaproteobacteria*. Es decir que se identificaron una total 5 taxa abundantes, a los que de aquí en adelante se hace referencia como Linajes Ubicuos y Profusos (LUPs). En conjunto, los LUPs representaron el 86,31% de las secuencias. El resto de las secuencias fueron asignadas a los grupos *Marinoscillum*, *Formosa*, *NS4*, *Polibacter*, *Synechococcus*, *SAR406*, *Planctomyces*, *Fretibacter*, *PS1*, *Ascidiaceihabitans*, *Planktomarina*, *SAR116*, *Chesapeake*, *OM43*, *Idiomarina*, *OM60*, *KI89A*, *ZD0405* y *JTB255*. Además, un 11,86% de las secuencias resultaron ser no-clasificables.

Para evaluar si los LUPs presentan patrones biogeográficos se estudiaron las correspondientes relaciones taxa-área (TAR), patrones de decaimiento por distancia (DD) y patrones de endemismo, y se realizaron análisis multivariados (NMDS) de cada una de las 17 comunidades de LUPs. Asimismo, se estudió la distribución geográfica de la diversidad de dichas comunidades, la cual se midió mediante índices numéricos de uso común en microecología. Estos estudios se llevaron a cabo a diferentes niveles taxonómicos putativos: especie, género y familia, lo cual se logró agrupando las secuencias en Unidades Taxonómicas Operativas (OTUs) en diferentes niveles de similitud, en base a bibliografía. Además, con el objeto de velar por la consistencia y homogeneidad de la potencia estadística, los estudios fueron repetidos con y sin ecualización. Esto representa un total de: 5 (LUPs) * 3 (niveles taxonómicos putativos) * 2 (ecualizado y no ecualizado) * 14 (TAR, DD con 3 índices, NMDS con 4 índices, endemismo, 5 índices de diversidad) = 420 análisis cuantitativos, los cuales generaron datos originales sobre las comunidades microbianas del Mar Argentino y que además son de relevancia para el conocimiento general de la biogeografía microbiana.

Los análisis de las relaciones taxa-área, en general evidenciaron tasas de reemplazo (valor z de la TAR), o *turnover*, significativos en todos los linajes y con datos ecualizados y sin ecualizar. Dichas tasas fueron similares en los 3 niveles taxonómicos putativos, aunque en general los niveles de agrupamiento menos inclusivos presentaron tasas mayores que los más inclusivos; es decir que los valores más grandes de z se observaron en el nivel putativo de especie y los más pequeños en el de familia.

En concordancia con esto, en general también se observaron patrones de DD en todos los índices de similitud utilizados. Es decir que las comunidades cercanas entre sí resultaron ser composicionalmente más parecidas entre ellas que a comunidades ubicadas en puntos geográficos relativamente alejados. Este patrón pudo observarse en la mayor parte los LUPs, niveles de agrupamientos y para datos ecualizados y no ecualizados, aunque para algunos LUPs y combinaciones de parámetros específicas no se observó el decaimiento, o el mismo no fue estadísticamente significativo.

Los análisis de endemismo indicaron que cada LUP presenta linajes que son característicos de cada localización geográfica en particular, es decir son endémicas en dicha localización. Se identificaron 69, 78 y 86% de OTUs endémicas en los niveles de similitud de 95, 97 y 99%, respectivamente. Los análisis con secuencias ecualizadas y no ecualizadas fueron similares entre sí.

Los ordenamientos realizados mediante NMDS no permitieron evidenciar una relación entre la cercanía en el ordenamiento y la correspondiente cercanía geográfica de manera consistente. Ante estos resultados, se concluye que los ordenamientos mediante NMDS no tienen el poder suficiente para evidenciar por sí solos la existencia de patrones geográficos.

El análisis de la distribución espacial de la diversidad mostró que dos de los índices utilizados, ACE y Chao, estuvieron muy relacionados con los esfuerzos de muestreo, o coberturas, logrados para cada LUP en cada muestra. Por lo tanto, los correspondientes resultados no fueron utilizados en los análisis posteriores. El resto de los índices utilizados (índices de diversidad y homogeneidad de Shannon, e índice de Simpson), resultaron ser más robustos. Su análisis indicó que, si bien existieron variaciones en los distintos puntos de muestreos, las mismas no presentaron relaciones consistentes entre muestras con respecto a su distribución geográfica.

En su conjunto, los resultados arriba expuestos indican que los LUPs pueden presentar patrones de distribución heterogéneos. Este fenómeno podría obedecer a (i) la ocurrencia de procesos aleatorios, en particular aislamiento por distancia, o a (ii) procesos adaptativos (preferencia de nicho). Es ampliamente conocido que la selección puede causar patrones biogeográficos cuando se relevan ambientes contrastantes (por ejemplo, transectas ecuatorial-polares, transectas trans-cordilleranas, transectas trans-estuariales, columna de agua en el océano y lagos profundos, etc.). Sin embargo, esta segunda explicación es menos plausible que la primera en el caso del presente estudio, ya que las muestras aquí estudiadas corresponden a un mismo ambiente (agua superficial). La deriva ecológica es otro proceso que puede asociarse con la existencia de diferencias composicionales entre comunidades de distintas regiones. Sin embargo, no existe mecanismo conocido por el cual la misma pueda generar patrones de DD, por lo cual no resulta plausible que este tipo de deriva pueda, al menos por sí sola, explicar todos los patrones biogeográficos observados en esta Tesis.

Este es el primer estudio en el que se analizó qué linajes del bacterioplancton presentan un mayor aporte a los microbiomas de agua superficial del Mar Argentino. Asimismo, el estudio aporta datos novedosos sobre la existencia de patrones biogeográficos en microbios, tema que, como ya se mencionó, es foco de muchos estudios actuales. En este sentido, se expone la presencia de una estructuración biogeográfica en comunidades de bacterioplancton del Mar argentino, en contraposición al concepto tradicional de que “todo está en todas partes, pero ambiente selecciona”. Si

bien es claro que el ambiente siempre impone restricciones en el proceso de ensamblado comunitario, los resultados aquí presentados demuestran que ensambles de LUPs de un mismo ambiente (agua marina superficial) pero de sitios geográficos remotos presentan variaciones composicionales, las cuales como se explicó en el párrafo previo, pueden atribuirse al fenómeno de aislamiento por distancia. Ahora que se descubrió la existencia de estos patrones en el Mar Argentino, será posible diseñar nuevos estudios y experimentos destinados a avanzar en la comprensión de los mecanismos subyacentes.

Summary

Culture-independent technologies have produced a turning point in microbiology by demonstrating that we have a very rudimentary knowledge of the phylogeny, metabolism, and ecology of most microorganisms. Thanks to the new sequencing technologies, we currently have an unprecedented level of detail on the taxonomic composition of microbiomes, which has turned out to be more complex than we previously thought. For example, it has been shown that, as occurs with terrestrial ecosystems, marine bacterial communities are dominated by a small set of taxonomic groups, which are accompanied by a plethora of minority but highly diverse groups. Given the high biomass and abundance of marine bacteria, they significantly influence the elements' biogeochemical cycles, regulating their state and controlling the fate and magnitude of the organic carbon pool. It should be noted once again that, despite the mentioned advances, we are at a point far from having complete knowledge of the composition, functioning, and distribution of marine bacterial communities.

In the present work, the composition and distribution of bacterial communities of surface water from 17 remote points in the Argentine Sea were studied, to identify abundant and ubiquitous groups in our oceans, and quantitatively determine if their spatial distribution is uniform (a traditional concept that all microorganisms are potentially everywhere) or heterogeneous. Bacterial biogeography is an area little studied in traditional microbiology. However, recent studies have shown that many groups present biogeography, a fact that has reactivated the interest of the scientific community, with the consequent flourishing of various works in recent years. In this Thesis, a database of approximately three million paired sequences of the small ribosomal subunit gene (16S gene), obtained by Illumina sequencing of picoplanktonic DNA, was used. The raw sequences obtained (or reads) were initially subjected to complex and strict quality control. The beginning of this work involved an optimization process, which included the use of a "simulated community" to determine the optimal processing parameters of the sample sequences.

After quality controls, 387,597 high-quality sequences from the communities of the Argentine Sea were obtained. The analysis of the taxonomic composition allowed the determination that the most abundant sequences in our seas correspond to the order *Pelagibacterales* (*Alphaproteobacteria*, *Proteobacteria*), also known as SAR11, and to the OM1 group (currently known as *Ca. Actinomarina*) of the phylum *Actinobacteria*, which together accounted for approximately 80% of the sequences. A preliminary analysis of the abundances indicated that OM1 was more predominant in regions near the coast, while SAR11 was more prevalent in those further away. Taxa were also identified that, although

less abundant than those already mentioned, exhibited substantially higher abundances than the rest of the detected groups. These were the order *Flavobacteriales* (or NS5) of the phylum *Bacteroidetes*, the order *Oceanospirillales* (or SAR86) of the class *Gammaproteobacteria*, and the order *Rhodobacterales* (genus *Amylibacter*) of the class *Alphaproteobacteria*. That is, a total of 5 abundant taxa were identified, which from now on are referred to as Ubiquitous and Profuse Lineages (LUPs). Altogether, LUPs represented 86.31% of the sequences. The rest of the sequences were assigned to the groups *Marinoscillum*, *Formosa*, NS4, *Polibacter*, *Synechococcus*, SAR406, *Planctomyces*, *Fretibacter*, PS1, *Ascidiaceihabitans*, *Planktomarina*, SAR116, *Chesapeake*, OM43, *Idiomarina*, OM60, KI89A, ZD0405, and JTB255. In addition, 11.86% of the sequences turned out to be unclassified.

To assess whether LUPs present biogeographic patterns, the corresponding taxa-area relationships (TAR), patterns of decay by distance (DD), and patterns of endemism were studied, and multivariate analyzes (NMDS) were performed for each of the 17 LUP communities. Likewise, it was evaluated if the diversity of these communities, which was measured by different numerical indices, presents elements compatible with a structured spatial distribution. The studies were carried out at different putative taxonomic levels: species, genus, and family, which was achieved by grouping the sequences into Operative Taxonomic Units (OTUs) at different levels of similarity (99%, 97%, and 95%, respectively), based on the bibliography. In addition, to ensure the consistency and homogeneity of the statistical power, the studies were repeated with and without equalization. This represents a total of 5 (LUPs) * 3 (putative taxonomic levels) * 2 (equalized and non-equalized) * 14 (TAR, DD with 3 indices, NMDS with 4 indices, endemism, 5 diversity indices) = 420 analyses. quantitative, which generated original data on the microbial communities of the Argentine Sea and is also relevant to the general knowledge of microbial biogeography.

The analyzes of the taxa-area relationships, in general, showed geographic patterns in all the lineages, and all the grouping levels, and with equalized and non-equalized data, obtaining values of significant replacement rates (z values), and similar in the 3 putative taxonomic levels analyzed.

Consistent with this, in general, DD patterns were also found with all the dissimilarity indices studied. In general terms, it was observed that as the distance that separates two communities increases, the dissimilarity becomes greater. In other words, nearby communities are compositionally more like each other than communities that are geographically farther away. This pattern could be observed in most LUPs, cluster levels, and for equalized and non-equalized data, although for some LUPs and specific parameter combinations the patterns could not be detected or were not significant.

Endemism analysis indicated that each LUP has lineages that are characteristic of each geographic location, that is, they are endemic to that location. It was found that there

were 69, 78, and 86% of endemic OTUs, when similarity levels of 95, 97, and 99% were used, respectively. The analyzes with equalized and non-equalized sequences were like each other.

The NMDS analyses were less clear than those performed using the other techniques, since for various sampling points there was no evidence of a relationship between the proximity in the obtained ordering and the corresponding geographic proximity. Given these results, it is concluded that the NMDS rankings do not have enough power to demonstrate by themselves the existence of geographic patterns, at least if they are applied in the way they were applied in this study.

The analysis of the spatial distribution of diversity showed that two of the indices used, ACE and Chao, were too related to the sampling efforts achieved for each lineage at each sampling point. Therefore, the corresponding results were not used in subsequent analyses. The rest of the indices (Shannon diversity and homogeneity indices, and Simpson index) turned out to be more robust. Their analysis indicated that, although there were variations between different sampling points, they did not present consistent relationships between samples with respect to their geographic distribution.

Taken together, the above results indicate that LUPs may present heterogeneous distribution patterns. This phenomenon could be attributed to (i) the occurrence of random processes, particularly isolation by distance, or to (ii) adaptative processes (niche preference). It is widely known that selection can cause biogeographic patterns when contrasting environments (e.g., equatorial-polar transects, trans-cordilleran transects, trans-estuarial transects, oceanic and deep lakes' water columns, etc.) are surveyed. However, this second explanation is less plausible than the first one in the case of this study, since the samples studied here correspond to a single environment (surface water). Ecological drift is another process that can be associated with the existence of compositional differences between communities in different regions. However, there is no known mechanism by which it could generate DD patterns, so it is not plausible that this type of drift can, at least by itself, explain the patterns observed in this Thesis.

This is the first study in which it was analyzed which bacterioplankton lineages present a greater contribution to the microbiomes of the Argentine Sea. Likewise, the study provides novel data on the existence of biogeographic patterns in microbes, a topic that, as already mentioned, is the focus of many current studies. The presence of a biogeographic structuring in bacterioplankton communities of the Argentine Sea is exposed, which challenges the traditional concept that "everything is everywhere, and the environment selects". Although the environment always imposes restrictions on the

community assembly process, the results presented here show that assemblages of LUPs from a single environment (surface marine water) but from remote geographic sites can present compositional variations, which, as explained in last paragraph, can be attributed to the phenomenon of isolation by distance. Now that the existence of geographical patterns in the Argentine Sea have been discovered, it will be possible to design new studies and experiments aimed at advancing the understanding of the underlying mechanisms.

Abreviaturas

BOPD: Buque Oceanográfico Puerto Deseado.

DD: Decaimiento por distancia.

ADN: ácido desoxirribonucleico.

LUP: Linajes Ubicuos y Profusos.

N: número.

NMDS: Escalamiento multidimensional no métrico (por su sigla en inglés *Non-metric multidimensional scaling*).

OTUs: Unidad Taxonómica Operacional (por su sigla en inglés *Operational Taxonomic Units*).

ARN: ácido ribonucleico.

rARN: ácido ribonucleico ribosomal.

SAC: Secuencias de Alta calidad.

TAR: Relaciones taxa-área (por su sigla en inglés *taxa-area relationship*).

DOM: materia orgánica disuelta (por su sigla en inglés *dissolved organic matter*).

PCR: reacción en cadena de la polimerasa (por su sigla en inglés *polymerase chain reaction*).

LVGM: Laboratorio de Virología y Genética Molecular.

ACE: por la sigla en inglés *abundance-based coverage estimator*.

1 Introducción

1.1 Bacteriomas marinos y su rol en el ambiente.

Se estima que el número total de bacterias marinas es de 10^{29} , mientras que la abundancia promedio es de 5^4 células por mililitro (Whitman et al. 1998). En consecuencia, se cree que el bacterioplancton supera ampliamente en ubicuidad, abundancia y biomasa al resto de los habitantes de los océanos (Pomeroy et al. 2007; Whitman et al. 1998). Las comunidades de microorganismos tienen un rol central en la descomposición de la materia orgánica en los ecosistemas acuáticos (Pomeroy et al. 2007). Dada su abundancia, además, influyen significativamente en los ciclos biogeoquímicos de varios elementos químicos, regulando el estado metabólico del océano, y controlando el destino y la magnitud del pool de carbono orgánico (Ducklow y Doney 2013; Enke et al. 2018; Falkowsk et al. 2008).

La degradación de partículas de materia orgánica en el océano es un proceso central en el ciclo del carbono, nitrógeno y fósforo en los cuales los microorganismos poseen un rol central; en particular las bacterias (Enke et al. 2018; Pomeroy et al. 2007). En este sentido, el bacterioplancton es responsable de aproximadamente el 50% de la producción primaria de la Tierra, y es un importantísimo actor en dos procesos conocidos como el bucle microbiano (*microbial loop* en inglés) y la bomba microbiana (Guidi et al. 2016; Figura 1-1). El bucle microbiano es el proceso mediante el cual la materia orgánica-disuelta o DOM (siglas en inglés por *Dissolved Organic Matter*), es canalizada hacia niveles tróficos superiores a través de su incorporación a la biomasa bacteriana (Azam et al. 1983). Por otro lado, la bomba microbiana involucra una serie de eventos mediante los cuales el dióxido de carbono es fijado como materia orgánica por la fotosíntesis, y luego se hunde, lo que resulta en el almacenamiento temporal o permanente de carbono (Jiao et al. 2010). A estos dos mecanismos se le suman el cortocircuito viral o *viral shunt* (redireccionamiento de carbono a través de lisis celular, Wilhelm y Suttle 1999) y la bomba biológica (fijación del carbono atmosférico y su exportación por medio de procesos biológicos al fondo del océano; Figura 1-1; Sarmiento 2006). Tomando en cuenta que los océanos cubren el 71% de la superficie del planeta, es evidente que comprender el destino de la materia orgánica en el océano, y los mecanismos de regulación, son fundamentales para entender fenómenos globales como el cambio climático y la productividad global, con el fin de generar pautas para un futuro sustentable (Visbeck et al. 2014).

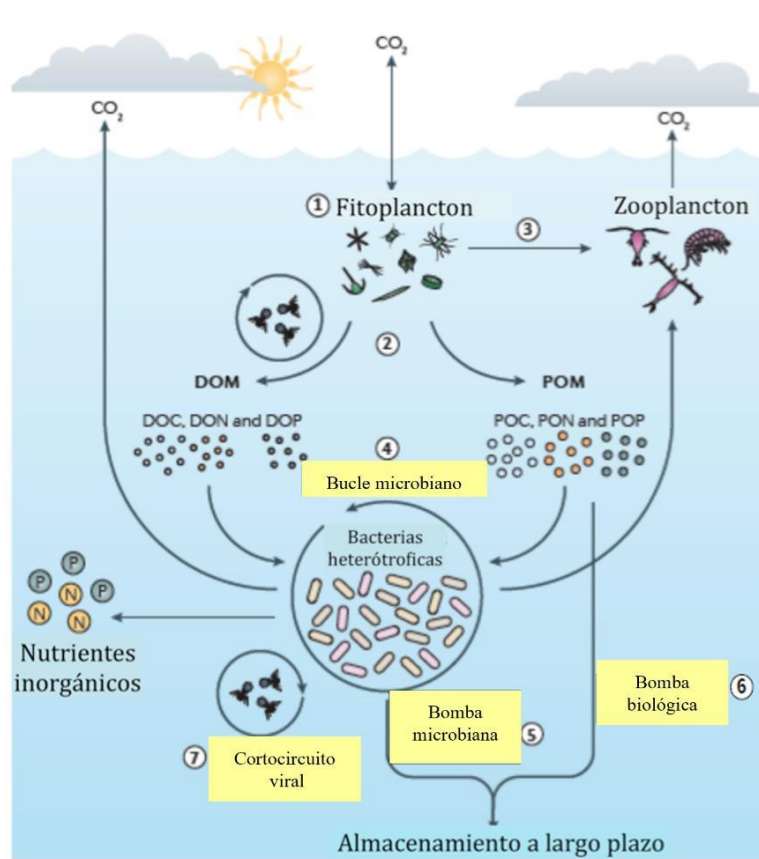


Figura 1-1: Función de los microorganismos en los ciclos de los principales elementos.

Se pueden observar las funciones principales de los microorganismos y su rol en los ciclos de elementos como carbono, nitrógeno y fósforo. Se detalla la fijación de carbono (1); la generación de POM (materia orgánica particulada) y DOM (materia orgánica disuelta) (2), el ingreso a niveles tróficos superiores (3), el loop o bucle microbiano (4), la bomba de carbono microbiana (5); la bomba biológica (6) y el proceso de derivación o cortocircuito viral (7). Las bacterias tienen un rol central en los procesos 4-7. DOC (carbono orgánico disuelto), DON (nitrógeno orgánico disuelto), DOP (fósforo orgánico disuelto), POC (carbono orgánico particulado), PON (nitrógeno orgánico particulado), POP (fósforo orgánico particulado). Esquema adaptado de (Buchan et al. 2014).

Estudiar la dinámica y distribución del bacterioplancton marino es esencial para el desarrollo de modelos para el aprovechamiento correcto y monitoreo de los servicios ecosistémicos marinos, uno de los ecosistemas más explotados del planeta (Grant et al. 2013; Liqueste et al. 2013; Peralta et al. 2014). Los “servicios ecosistémicos”, son los beneficios que obtiene la humanidad a partir de recursos presentes en la naturaleza y que pueden encontrarse afectados por las actividades antropogénicas. En este sentido, uno de los primeros intentos para crear una conciencia pública sobre el hecho de que las acciones antropogénicas conducen a cambios en el funcionamiento de los ecosistemas, e impactan en el bienestar y la salud humana corresponde al trabajo de Westman publicado en el año 1977 (Westman 1977; Liqueste et al. 2013). El mismo destaca la

dependencia que posee la humanidad respecto de los servicios que proporciona la Naturaleza, y el rol de biodiversidad en el mantenimiento de estos. Dado esto, resulta relevante contar con un conocimiento detallado de la biodiversidad que presentan los ecosistemas, para el buen manejo de los servicios ecosistémicos. Y, por lo tanto, es primordial conocer la composición y distribución de las comunidades bacterianas de nuestros mares, para los cuales existen escasos trabajos previos (Grant et al. 2013).

1.2 El advenimiento de nuevas tecnologías independientes de cultivo ha permitido dilucidar la composición de los microbiomas.

La ecología microbiana ha experimentado cambios cualitativos gracias a las tecnologías independientes de cultivo (Cordero y Datta 2016). Especialmente con el desarrollo de las llamadas tecnologías de secuenciación de nueva generación, se ha podido poner en evidencia un nivel de detalle sin precedente de la composición taxonómica de los microbiomas. El continuo desarrollo de las tecnologías de secuenciación de alto rendimiento, la progresiva reducción de su costo, y continuas mejoras en la longitud de lectura y rendimiento, han permitido generar un volumen de conocimiento que ha impactado en forma directa en el campo de la ecología microbiana (D'Amore et al. 2016). En este sentido, se han experimentado avances notables en el conocimiento sobre la composición, funcionamiento y taxonomía de los grandes ecosistemas de la tierra, tales como el suelo, los océanos, entre otros (Grossart et al. 2020).

Dentro de este campo, la determinación de secuencias ambientales del gen que codifica para la subunidad menor del ARN ribosomal pequeña (gen *ARNr-16S*) se ha convertido en una herramienta ampliamente utilizada para describir comunidades microbianas en diversos tipos de hábitat. El gen *ARNr-16S*, contiene regiones hipervariables intercaladas entre regiones altamente conservadas. Las secuencias de éstas últimas han sido utilizadas para el diseño de oligonucleótidos que permiten la amplificación del gen mediante reacción en cadena de la Polimerasa (PCR), utilizando como molde el material genético de muestras ambientales. Esto ha permitido obtener secuencias de una gran cantidad de taxa, muchos de ellos desconocidos para la ciencia (Calvo et al. 2018; Mo et al. 2018). El advenimiento de las técnicas de secuenciación de alto rendimiento del fragmento del gen *ARNr-16S*, actualmente denominada *metabarcoding*, impulsó avances drásticos en la comprensión de la diversidad microbiana (Schloss et al. 2016).

Se descubrió que los ensamblajes microbianos, o microbiomas, son muy complejos en cuanto a su composición taxonómica (Schloss et al. 2016). Asimismo, estos estudios

han demostrado que un reducido conjunto de grupos taxonómicos domina los microbiomas. En este trabajo, se hará referencia a dichos grupos como Linajes Ubicuos y Profusos (LUP).

En el medio marino, uno de los grupos que se encuentra en gran proporción es la clase *Alphaproteobacteria* (Aylward et al. 2015; Brown et al. 2014; Seo et al. 2017). Los grupos de *Alphaproteobacteria* más abundantes son *Pelagibacter* (también llamado SAR11) (Morris et al. 2012; Seo et al. 2017; West et al. 2016), y *Rhodobacterales* (*Amylibacter*) (Laas et al. 2014; Seo et al. 2017). Dentro del phylum *Bacteroidetes*, el orden de *Flavobacteriales* (NS5) es otro de los linajes que suele representar el 10% de las secuencias; siendo abundante en regiones costeras (Díez-Vives et al. 2019; Laas et al. 2014). Otro de los linajes abundantes, que puede representar alrededor de 30-40% de las secuencias es el phylum *Actinobacteria*, particularmente el grupo OM1, el cual también presenta una distribución asociada a ambientes costeros (Laas et al. 2014; Seo et al. 2017). Si bien existen estudios descriptivos y metabólicos de diferentes partes del mundo, hay muy pocos estudios biogeográficos a nivel mundial, regional o local de los LUPs de manera sistemática.

1.3 Biogeografía: distribución espacio-temporal de la diversidad.

La biogeografía es el estudio de la distribución de la biodiversidad en el espacio y el tiempo. Si bien se trata de un campo ampliamente desarrollado para la flora y fauna (Rodríguez-Martínez et al. 2013), la biogeografía de microorganismos ha sido relativamente poco contemplada. Como se detalla más adelante (sección 1.4), debido a los grandes tamaños poblacionales de los microbios, existe una creencia tradicional de que cualquier microorganismo puede probablemente estar en todos lados, aunque lógicamente, la capacidad de sobrevivir en un ambiente determinado depende de las características fisicoquímicas del mismo. Por lo tanto, la selección ambiental sería el único factor que interviene en el ensamblado de las comunidades microbianas. Este concepto, atribuido a MW Beijerinck (Beijerinck 1913) y Lourens Baas Becking (Baas-Becking 1934), ha llevado a que el campo de la biogeografía microbiana haya sido poco estudiado. Sin embargo, estudios más recientes basados en técnicas independientes de cultivo han demostrado que varios grupos de microorganismos pueden exhibir patrones biogeográficos (Dolan 2005; Manrique y Jones 2017; Martiny et al. 2006; Sjöstedt et al. 2014; Zinger et al. 2014). Comprender los mecanismos que generan y mantienen la biodiversidad, como los patrones de distribución, es clave para predecir la evolución de los ecosistemas y el mantenimiento de los ya mencionados servicios ecosistémicos. Resulta particularmente desafiante entender cómo estos procesos impactan en los patrones biogeográficos de los microorganismos, más aún en ambientes como los océanos, en los que no habría barreras físicas aparentes, y que presentan rápidas corrientes (Milici et al. 2016).

1.3.1. Análisis cuantitativos de las distribuciones biológicas.

La diversidad puede medirse en diferentes escalas. Una de ellas es la alfa-diversidad, la cual es la biodiversidad intrínseca de cada comunidad, mientras que la gamma-diversidad es la diversidad a nivel regional (Whitaker et al. 2003). La beta-diversidad es la diferencia entre las diversidades locales (alfa-diversidad) y la diversidad regional total (gamma-diversidad); es decir que mide cuán heterogéneas son dos comunidades (Whitaker et al. 2003). Los análisis de patrones de diversidad se basan en matrices comunitarias que representan la presencia/abundancia de especies a través de las muestras. Cuando se trata de datos de secuencias ambientales obtenidos mediante métodos de alto rendimiento, los datos consisten en cientos o millones de lecturas de secuenciación del ADN molde, que representan muchísimas especies que coexisten. Estos datos se transforman en matrices comunitarias, donde las especies quedan representadas por sus respectivas secuencias, y la abundancia de estas es estimada mediante la cantidad de secuencias de cada clase (Calderón-Sanou et al. 2019). En biogeografía la beta-diversidad se estudia mediante metodologías como el análisis de endemismo, las relaciones taxa-área y el decaimiento por distancia.

El endemismo hace referencia a la existencia de taxa nativas, o simplemente a grupos biológicos con áreas de distribución restringidas. Es un indicador de la estructuración regional de la biodiversidad. Por ejemplo, en la Figura 1-2, la distribución de la especie “a” sugiere la existencia de patrones de endemismo, ya que la misma solamente fue observada en la comunidad “2” (Fig. 1-2, panel inferior izquierdo). La identificación de áreas de endemismo es una tarea fundamental de la biogeografía, ya que se consideran unidades básicas para los estudios biogeográficos históricos (Oliveira et al. 2015). Los patrones de endemismo son ampliamente conocidos para la mayoría de las especies animales y vegetales; pero es un concepto relativamente nuevo para comunidades microbianas.

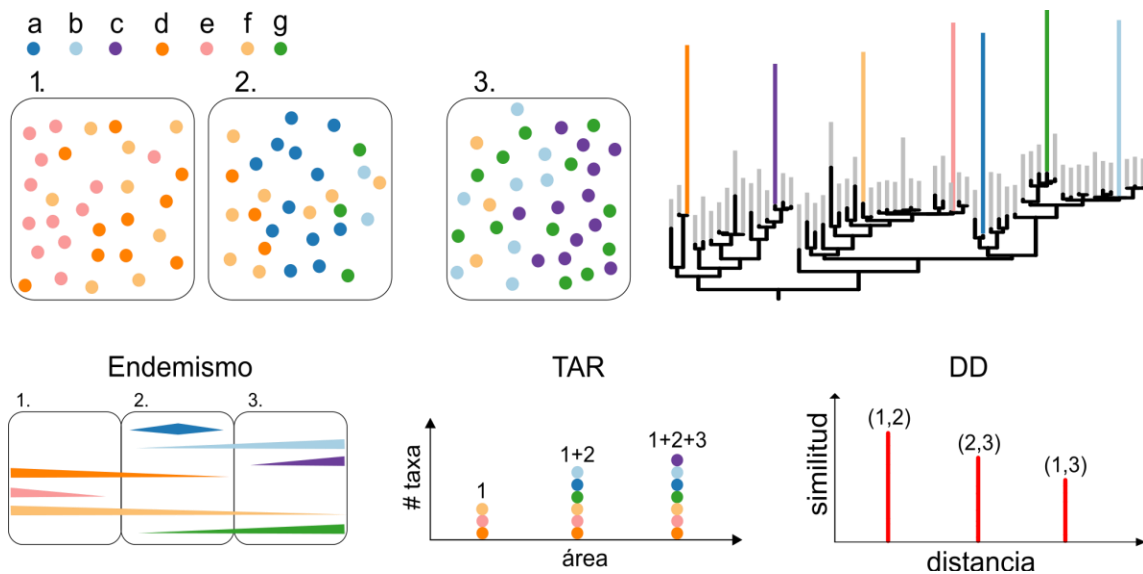


Figura 1-2. Esquema representando la biogeografía de un conjunto de siete especies hipotéticas. En la parte superior se representa la distribución de las siete especies hipotéticas (a-g) en tres regiones (1-3), y su historia evolutiva (árbol filogenético; los terminales en gris corresponden a especies extintas y/o no muestreadas). En la parte inferior, se describen los correspondientes patrones de endemismo (*Endemismo*), relación taxa-área (*TAR*) y patrón de decaimiento por distancia (*DD*). Modificado a partir de (Manrique y Jones 2017).

La relación taxa-área (*TAR*, por su sigla en inglés *taxa-area relationship*) es uno de los patrones generales más consistentes en ecología, y describe el aumento de la riqueza de especies a medida que aumenta el espacio muestreado. En el panel inferior central de Figura 1-2, rotulado como *TAR*, se presenta un patrón de acumulación de especies derivado de un muestreo hipotético de los ensambles 1, 2, y 3. El fenómeno de acumulación de especies en función del área puede representarse mediante la ecuación $S=c A^z$, donde “*S*” es el número de especies, “*A*” es el área muestreada, “*c*” es una constante que se deriva empíricamente del taxón y la ubicación específica estudiada, y el exponente “*z*” representa la tasa de acumulación de especies (Barreto et al. 2014; Drakare et al. 2006). El valor “*z*” es por lo tanto una medida de la tasa de sustitución, o *turn over*, de especies a través del espacio (Horner-Devine et al. 2004). Las relaciones taxa-área pueden ser el resultado de la acumulación de hábitats, y por lo tanto de especies, como también de la dinámica de la población, en lugares donde hay mayores probabilidades de colonización y especiación, con menores tasas de extinción. Por otro lado, la especiación alopatrica en combinación con la limitación de la dispersión también contribuye al desarrollo de relaciones taxa-área. Las tasas de acumulación de especies (valores *z*) son mayores al aumentar el endemismo, independientemente de la causa de este (Zinger et al. 2014).

El decaimiento por distancia (*DD*) es otra característica comúnmente estudiada en las investigaciones biogeográficas (Soininen et al. 2007). La misma cuantifica la relación

entre la similitud composicional de las comunidades y las distancias geográficas que las separan. Cuando existen patrones de decaimiento por distancia, las comunidades cercanas tienden a ser más similares entre sí que las comunidades distantes. El panel inferior derecho de la Figura 1-2, rotulado *DD*, representa un patrón de decaimiento por distancia obtenido a partir de los ensamblajes 1 a 3. Las similitudes (eje *y*), se derivan de las especies compartidas entre cada par de ensamblajes comparados (eje *x*: 1 vs. 2, 2 vs. 3, 1 vs. 3). Al igual que ocurre con las relaciones taxa-área, el decaimiento puede surgir a través de varios procesos interactivos, ecológicos, geográficos, y evolutivos (Clark et al. 2021; Hanson et al. 2012). Además de incorporar la distancia en el análisis, el decaimiento por distancia tiene en cuenta la abundancia relativa de las especies además de su presencia o ausencia. Por ejemplo, la especie hipotética “b” de la Figura 1-2, además de estar ausente en el ensamblaje “1” es mucho más abundante en el ensamblaje “3” que en el “2”. Esto puede responder a procesos como el aislamiento por distancia; la especie, originariamente presente en “3”, se dispersó hacia “2” pero aún no se observa en el ensamblaje “1” dada su lejanía. Este tipo de procesos constituye un importante motor de la diversidad, ya que su accionar es constante y omnipresente a lo largo de la evolución de todos los grupos biológicos, a diferencia de fenómenos como la selección, que actúa de manera espacial y temporalmente acotada.

Otra forma de comparar las composiciones comunitarias es mediante técnicas multivariadas como el escalamiento no métrico multidimensional (del inglés *Non-metric multidimensional scaling*) (Nelson et al. 2013; Pernice et al. 2016). En el caso de las comunidades biológicas, las dimensiones corresponden a los vectores de abundancias de cada una de las unidades taxonómicas operativas presentes en la matriz comunitaria.

1.4 Biogeografía bacteriana

Como ya se ha mencionado, la biogeografía bacteriana ha recibido poca atención en comparación con el relevamiento de su diversidad propiamente, el estudio de los factores ecológicos que la afectan y su metabolismo. Esto se debe principalmente a que hasta hace muy poco tiempo se pensaba que los microorganismos presentaban distribuciones cosmopolitas, gobernadas exclusivamente por la selección ambiental [la hipótesis de que “*Todo está en todas partes, pero el ambiente selecciona*” (De Wit y Bouvier 2006)]. Dicha idea, que data de la década del ‘30, se fundamentó en datos empíricos generados mediante métodos tradicionales (cultivo), los que actualmente sabemos que detectan menos del 1% de la diversidad bacteriana, y en deducciones teóricas derivadas del concepto de *kernel* de dispersión (Whitaker et al. 2003). El *kernel* de dispersión de un individuo es una distribución de probabilidad de colonización en función de la distancia que separa una comunidad fuente de una comunidad receptora

(Oliveira et al. 2015). La probabilidad de dispersión total de una especie es proporcional a la suma de las probabilidades de dispersión de cada uno de los individuos (Figura 1-3). Esto implica que los organismos abundantes, como muchos taxa microbianos, deberían presentar distribuciones cosmopolitas, es decir no presentar patrones biogeográficos.

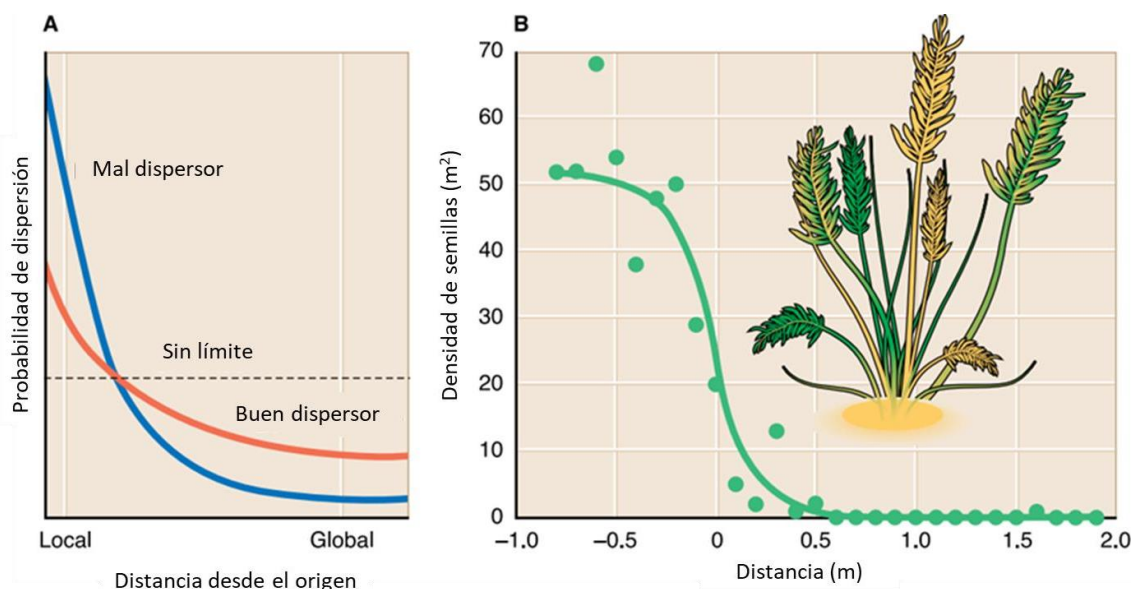


Figura 1-3. Kernel de dispersión.

El panel A ilustra el *kernel* de dispersión de un organismo con dispersión ilimitada (línea punteada), otro con una capacidad elevada de dispersarse (línea roja) y un tercero con una capacidad de dispersión baja (línea azul). En el panel B, el concepto se ejemplifica con la capacidad dispersiva de las semillas de la especie *Schismus barbatus*. Los puntos corresponden a datos empíricos sobre dispersión de las semillas, mientras que la curva corresponde a un modelo ajustado a partir de dichos datos empíricos. Adaptado de (Martiny 2015).

Una de las primeras evidencias contrastantes con este concepto fue un estudio de microorganismos de salinas (*i.e.* un mismo ambiente) de varias partes del mundo. El cual, sorprendentemente, mostró la existencia de patrones de decaimiento por distancia no sólo en escalas regionales y global, sino también a nivel local. Dichos resultados implican que “no todo está en todas partes” y que, además de los ya conocidos factores físicos que afectan el ensamblado comunitario, habría otros factores que influyen en la distribución de los microorganismos (Martiny et al. 2011). Posteriormente se observó que el fenómeno también puede afectar poblaciones globales de microorganismos de todos los océanos del planeta, incluyendo grupos ultra-abundantes como el clado de bacterias marinas SAR11 (Zinger et al. 2014; Hellweger et al. 2014; Manrique y Jones 2017). En la Figura 1-4, se muestran curvas de decaimiento por distancia entre poblaciones de SAR11 de diferentes puntos del planeta, en función de las distancias geográficas y ecológicas que las separan. La distancia latitudinal (panel central) brinda una estimación de las

diferencias ecológicas en cuanto a condiciones lumínicas (por ej. dial-frecuencia, ángulo de incidencia, atenuación atmosférica). La diferencia de temperatura indica qué tan diferentes son cada par de puntos geográficos en cuanto a las temperaturas medias anuales a las que deben subsistir las correspondientes poblaciones. Como puede observarse, si bien la incidencia de la luz y la temperatura presentan cierta relación con las distancias filogenéticas, las distancias geográficas explican mucho mejor el decaimiento. La explicación más plausible de ese patrón es que el aislamiento por distancia es una fuerza evolutiva relevante en este clado bacteriano hiper-abundante (Manrique y Jones 2017).

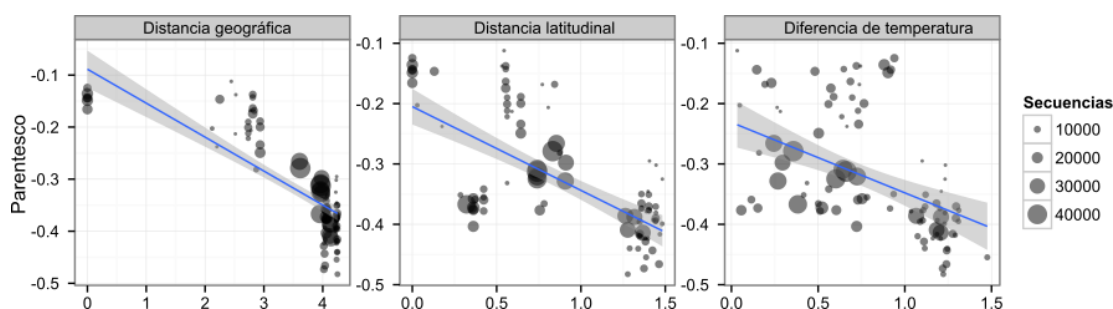


Figura 1-4. Decaimiento por distancia entre ensambles de SAR11 distribuidos globalmente.

Cada punto corresponde a dos comunidades de SAR11 presentes en dos puntos geográficos diferentes. Para estimar el parentesco entre cada par de comunidades (eje y), se realizó la sustracción $1-D_{uni}$, donde D_{uni} es la distancia filogenética entre las 2 comunidades. Es decir que valores más altos corresponden a poblaciones filogenéticamente más relacionadas. El tamaño de los puntos es proporcional a la cantidad de secuencias (Secuencias; escala a la derecha del gráfico) utilizadas en cada comparación. Ambos ejes, x e y, se presentan en escala y valores logarítmicos. La distancia geográfica está expresada en kilómetros. La distancia latitudinal es la diferencia entre los valores absolutos de cada par de latitudes. La diferencia de temperatura es la diferencia entre las temperaturas anuales medias de cada par de puntos geográficos comparados. Modificado a partir de Manrique & Jones (2017).

Los hallazgos descriptos arriba han impactado en el paradigma con el que se interpretaba la biodiversidad microbiana. Es así como, más recientemente, han proliferado numerosos reportes científicos sobre la existencia de patrones biogeográficos. Dichos trabajos se han reportado en variados ambientes como sedimentos marinos de sitios de filtraciones frías (del inglés *cold seep*; Semler et al. 2022), en muestras de agua través de una transecta latitudinal en el Océano Atlántico (Milke et al. 2022), en los frentes de afloramiento de las aguas frías (del inglés *upwelling*; Zhu et al. 2022), en el oeste del Océano Pacífico subtropical (Shao et al. 2022). La potencial consecuencia de este giro conceptual es que los estudios realizados en una locación determinada no podrían extrapolarse directamente a otros sitios, y que los impactos ambientales sobre un sitio particular no afectarían de la misma forma a la diversidad global.

1.5 Biogeografía bacteriana del Mar Argentino

El Mar argentino cuenta con una de las plataformas continentales más grandes del mundo. La misma posee una extensión de unos pocos kilómetros alrededor de los 38°S y luego comienza a ensancharse, llegando hasta unos 850 km alrededor de los 50° - 51°S (Dogliotti et al. 2014) (Figura 1-5). La región es altamente dinámica, caracterizada por la confluencia de dos corrientes: una de ellas la corriente de Brasil y la otra la corriente de Malvinas. A esto se suma la descarga del Río de la Plata que, debido a su gran caudal, extiende su influencia en la dinámica (García et al. 2008).

La corriente de Malvinas, desde su separación de la corriente Circumpolar Antártica, transporta aguas subantárticas hacia el norte. La corriente de Brasil proviene de regiones tropicales, presenta un recorrido de norte a sur, y sus aguas son de mayor temperatura y menor salinidad con respecto a la corriente de Malvinas. Ambas corrientes se encuentran entre los 36° y 38° S sobre el Mar Argentino definiendo la denominada Confluencia Subtropical. Una vez fusionadas se desprenden del borde continental y circulan con dirección sudeste, alejándose del continente americano, para constituir la celda de recirculación anticiclónica del Atlántico Sur. Además de esta dinámica, el mar argentino presenta varios frentes oceanográficos (Martos y Piccolo 1988) y varios frentes de marea que se desarrollan en primavera y verano que definen el límite entre las aguas de la plataforma verticalmente mezcladas y estratificadas (Rivas et al. 2006).

Estos fenómenos se asocian a la alta productividad de esta región, que no solo implica el desarrollo de especies de importancia comercial sino también la producción primaria de fitoplancton (Schloss et al. 2007). En estos campos en los últimos años se han desarrollado muchísimos trabajos, pero los estudios biogeográficos se han focalizado en especies del clado SAR11 de solamente 3 puntos geográficos (Manrique y Jones 2017). No existen estudios sistemáticos para otros grupos marinos abundantes. Llenar

este vacío de conocimiento fue la motivación principal de la realización del presente trabajo.



Figura 1-5: Mapa de la Argentina y Plataforma Continental.

Se detalla sólo la Plataforma Continental Argentina (celeste) de la zona que corresponde al frente de la costa, no se incluyó la región correspondiente a las islas del sur y el territorio de la Antártida. Los límites de la plataforma continental incluyen la ampliación del territorio correspondiente a la ley 27.757 Poder Ejecutivo Nacional. La figura blanca dentro de la Plataforma Continental Argentina corresponde al Territorio de las Islas Malvinas, territorio que se encuentra en reclamo por la República Argentina.

1.6 Objetivos

Objetivo general

Este trabajo de Tesis se enmarca en el objetivo general de “*Avanzar en el conocimiento de la biogeografía de los LUP bacterianos*”. Para ello, se estudió la composición y la presencia de patrones biogeográficos en muestras de agua superficial (~ 1 m), recolectadas en puntos remotos del Mar argentino, utilizando técnicas independientes de cultivo de nueva generación.

Objetivos particulares:

1. *Obtener secuencias de alta calidad (SAC) de muestras del Mar Argentino.* El presente objetivo consistió en generar secuencias de alta calidad a partir de una base de datos de secuencias crudas, correspondiente a 17 muestras de agua superficiales tomadas en puntos remotos del Mar Argentino, obtenidas en el Laboratorio Virología y Genética Molecular (LVGM), de la sede Trelew de la Universidad de la Patagonia San Juan Bosco.
2. *Determinar qué taxa bacterianos son abundantes en los 17 puntos del Mar Argentino estudiados.* El presente objetivo consistió en determinar qué linajes bacterianos son abundantes en cada una de las muestras estudiadas.
3. *Estudiar la TAR en cada uno de los LUPs identificados.*
4. *Determinar si la similitud entre ensamblajes de LUPs se ve afectada por la distancia geográfica.* Este objetivo fue llevado adelante mediante un análisis de decaimiento por distancia, y mediante la técnica no paramétrica *escalado multidimensional no métrico* (NMDS)
5. *Evaluar si existen patrones de endemismo para cada LUP identificado.* Este objetivo consistió en identificar variantes de LUPs que pudieran estar presentes sólo en una o unas pocas muestras. Como ya se explicó más arriba, este tipo de análisis, llamado análisis de endemismo puede poner en evidencia la existencia de especiación alopátrica y posibles limitaciones a la dispersión.
6. *Estudiar la distribución espacial de la diversidad de los LUPs.* Este objetivo consistió en cuantificar, por primera vez, la alfa-diversidad de LUPs en distintos puntos del Mar Argentino. Además, si bien existen estudios de la alfa-diversidad en comunidades completas, la misma no se ha estudiado en linajes individuales, incluyendo los LUPs. Por ejemplo, no se sabe si hay linajes que contribuyan una mayor diversidad al ecosistema que otros.

2 Materiales y Métodos

2.1 Bases de datos utilizadas en este estudio.

2.1.1 Muestras del Mar Argentino.

Los datos utilizados para llevar a cabo el presente estudio corresponden a secuencias obtenidas a partir de muestras recolectadas en diversos puntos del Mar Argentino. Las muestras BH y MDQ fueron tomadas durante la expedición Argentina-Canadiense R / V "Coriolis II", en enero de 2014 (Manrique y Jones 2017). Las muestras BOPD 1-14 fueron colectadas durante la expedición de oceanográfica de verano (enero) del buque ARA "Puerto Deseado" CONICET en 2013. La muestra CHR fue tomada desde una embarcación en condiciones de marea alta a una distancia de aproximadamente cuatro kilómetros desde la desembocadura del Río Chubut en la región de Bahía Engaño, durante enero de 2014 (Giaccardi et al. 2022). Las muestras fueron tomadas en las siguientes puntos geográficos: 51.6°S 68.6°W (BOPD.1), 51.6°S 68.3°W (BOPD.2), 50.4°S 68°W (BOPD.3), 50.25°S 66.17°W (BOPD.4), 47.8°S 65.76°W (BOPD.5), 47.9°S 65.4°W (BOPD.6), 43.4°S 64.9°W(BOPD.7), 43.4°S 64.98°W (BOPD.8), 43.2°S 64.5°W (BOPD.9), 43.01°S 64.1°W (BOPD.10), 41.2°S 62.6°W (BOPD.11), 41.1°S 62.7°W (BOPD.12), 39.7°S 61.8°W (BOPD.13), 38.97°S 60. 07°W (BOPD.14), 39.95°S 55.68°W (MDQ.SURF), 45.93°S 57.7°W (BH.SURF) y 43.44°S 66.11°W (CHR). Los puntos de muestreo pueden también observarse en el mapa de la Figura 2-1.

Cada una de las 17 muestras consistió en aproximadamente 3 litros de agua que se recolectaron a ~ 1 metro de profundidad. Las muestras se pre-filtraron con una malla Nitex de tamaño de poro de 100 µm para eliminar el material particulado grande y el zooplancton. Posteriormente, las muestras se filtraron, secuencialmente, a través de membranas de 20, 10, 5 y 0.22 µm y el ADN picoplanctónico se aisló a partir de las membranas de 0.22 µm. Para ello, las membranas fueron tratadas a 60°C durante 40 minutos con 720 µl de buffer de lisis CTAB (CTAB 2 %, NaCl 1.4M, beta-mercaptoetanol 0.2 %, EDTA 20mM, Tris-Cl pH 8 100mM), luego de lo cual se realizó una primera extracción orgánica mediante cloroformo: alcohol isoamílico (30:1). Posteriormente, se eliminaron enzimáticamente los ARN durante 1 hora a 37°C y el ADN se separó mediante el agregado de isopropanol e incubación de una hora a 4°C seguido por precipitación a 30.000G durante 30 minutos en centrifuga refrigerada (4°C). El remanente de sales del precipitado fue eliminado por tratamiento con etanol 70% y el ADN obtenido se re-suspendió en 20 µl de agua de grado biología molecular. Estos procedimientos han sido validados con anterioridad (Giaccardi et al. 2022). A partir de dicho ADN se procedió, por triplicado, a amplificar 1466 bases del gen de rARN 16S utilizando los cebadores universales para bacterias AF-27 y HR-1492, siguiendo los protocolos descritos en trabajos previos (Giaccardi et al. 2022). Los amplicones obtenidos fueron combinados y se secuenciaron las regiones V1 a V3 del gen utilizando plataforma

Ilumina. La base de datos obtenida consistió en alrededor de 3×10^6 secuencias pareadas que cubren las regiones V1 a V3 del gen rARN16S.

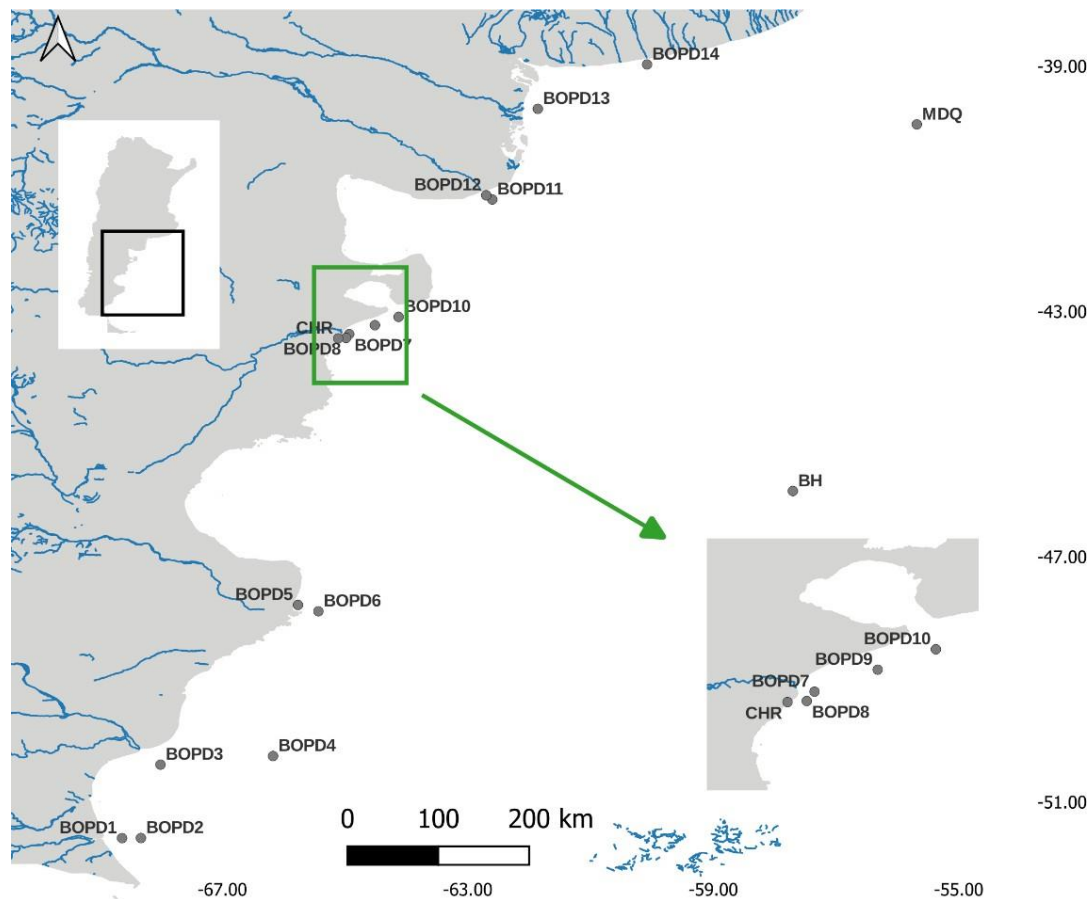


Figura 2-1: Localización de los sitios de muestreo.

Localización geográfica de las comunidades muestreadas en el Mar Argentino. Se detalla la zona donde fueron colectadas y los nombres asignados para su identificación. El recuadro interno corresponde a una ampliación de la región de Bahía Engaño, para una mejor visualización de los puntos muestreados.

2.1.2 Comunidad Simulada.

Una comunidad simulada (o “*mock community*”) es una comunidad que posee una composición conocida, generada en el laboratorio, que se utiliza para optimizar y evaluar los protocolos de curado y análisis de las secuencias crudas (sección 2.2). En este trabajo se usó una comunidad simulada creada a partir del material genético de 5 clones caracterizados por secuenciación Sanger, de una colección de bibliotecas del LVGM generadas a partir de muestras de agua superficiales de la zona de Bahía Engaño. Los clones utilizados para formar la comunidad corresponden a: un clon del linaje OM1, uno de SAR11, uno de *Amylibacter*, y dos clones del linaje SAR116. El material genético (plásmidos purificados) de cada uno de los clones fue cuantificado, combinados y secuenciados en paralelo junto con las muestras.

2.2 Secuencias de alta calidad

Tal como ocurre con todos los datos científicos, los datos de secuencia deben ser sometidos a controles de calidad tendientes a minimizar errores metodológicos (Poulet et al. 2016; Calderón-Sanou et al. 2019). Por tanto, se han desarrollado algoritmos que permiten realizar estrictos controles de calidad a las secuencias crudas, con el fin generar un conjunto de datos de alta calidad. En este trabajo, las secuencias crudas se procesaron siguiendo las pautas descritas por Kozich y colaboradores (Kozich et al. 2013), utilizando el programa Mothur (v. 1.39.5; Schloss et al. 2009). Además, en algunos pasos del procedimiento, se evaluaron diferentes estrategias con el objeto de generar un protocolo propio optimizado que permita obtener el máximo número de secuencias de interés, de la mayor calidad posible.

2.2.1 Procesamiento de secuencias: esquema general

En la Figura 2-2 se muestra un esquema del ensamblado y control de calidad al que fueron sometidas las secuencias crudas en esta Tesis (recuadros azules) y de los subsiguientes análisis comunitarios (recuadros verdes) previos al análisis biogeográfico. Los pasos correspondientes se explican a continuación:

A) Creación de *contigs*. Este procedimiento fue llevado a cabo mediante el comando *make.contigs* con el cual las lecturas pareadas reversas y directas se combinan formando una única secuencia.

B) Selección de secuencias sin ambigüedades ni trectos homopoliméricos mayores que 9 bases, y cuyos largos fueron de al menos 350 bases. Este procedimiento se realizó

mediante la función *screen.seqs*, a la cual se pasaron los siguientes parámetros: *maxambig=0*, *maxhomop=9*, *minlength=350*.

C) Identificación de secuencias únicas. Mediante el comando *unique.seqs*, se identificaron secuencias únicas y se generó un archivo conteniendo las secuencias únicas y la abundancia de cada una de las mismas.

D) Alineamiento de las secuencias. En el presente trabajo, los alineamientos se realizaron con 2 bases de datos de referencia: la base de datos pública y curada SILVA (de rRNA 16s; Pruesse et al. 2007), las flechas rojas en la Figura 2-2 y una base de datos generada en el LVGM como se detalla en el punto 2.3.2 indicada con flechas negras en la Figura 2-2. La función mediante la cual se realizaron los alineamientos fue *align.seqs*; con el argumento *flip=T* (permite evaluar el alineamiento por derecha o reversa). Como se detalla más adelante, para los análisis comunitarios se seleccionó el resultado obtenido con la base de datos que permitió obtener la mayor cantidad de secuencias de los linajes de interés.

E) Eliminación de secuencias que no se alinearon en la zona de interés. Este paso se realizó con el comando de Mothur *filter.seqs*.

F) Evaluación de secuencias con características atípicas con respecto al resto de las secuencias (en inglés “*outliers*”). Cuyo alineamiento se realiza utilizando secuencias de referencia, algunas secuencias pueden alinear en zonas donde no lo hace la mayoría, por lo cual las mismas pueden ser consideradas como atípicas, pudiendo potencialmente contener errores o inducir sesgos en el análisis. Es por ello por lo que las secuencias *outliers* son usualmente eliminadas de los análisis de secuencias ambientales. Una forma de identificar y eliminar estas secuencias es realizar una inspección de los puntos donde comienza y finaliza el alineamiento de cada secuencia del *dataset* con respecto al alineamiento de referencia. En primer lugar, se identifican las posiciones del alineamiento de referencia donde comienzan y finalizan la mayor parte de las secuencias bajo estudio, para luego descartar aquellas que resultaron alineadas en posiciones lejanas a dichas posiciones. En este trabajo, el procedimiento se realizó mediante la función de Mothur *screen.seqs*, implementando los argumentos *optimize=start-end*, *criteria=70* y *minlength=350*.

G) Identificación de secuencias únicas nuevamente, al igual que en el punto C.

H) Fusión de secuencias que presentaron dos bases de diferencia con respecto a secuencias de mayor abundancia. Este procedimiento se basa en la lógica de que en los procesos de secuenciación los errores se producen con mayor probabilidad a partir de

secuencias abundantes. El procedimiento se realizó con la función *precluster* con el argumento *diffs*.

I) Búsqueda de secuencias quiméricas. Las secuencias quiméricas son aquellas que se generaron a partir de dos secuencias diferentes en el proceso de amplificación por PCR. El algoritmo utilizado en esta Tesis para identificar tales secuencias fue V-search, el cual está implementado en la función *chimera.vsearch* de Mothur. Como referencia se usaron las secuencias más abundantes.

J) Eliminación de secuencias quiméricas. Esto fue realizado mediante el comando *remove.seqs*.

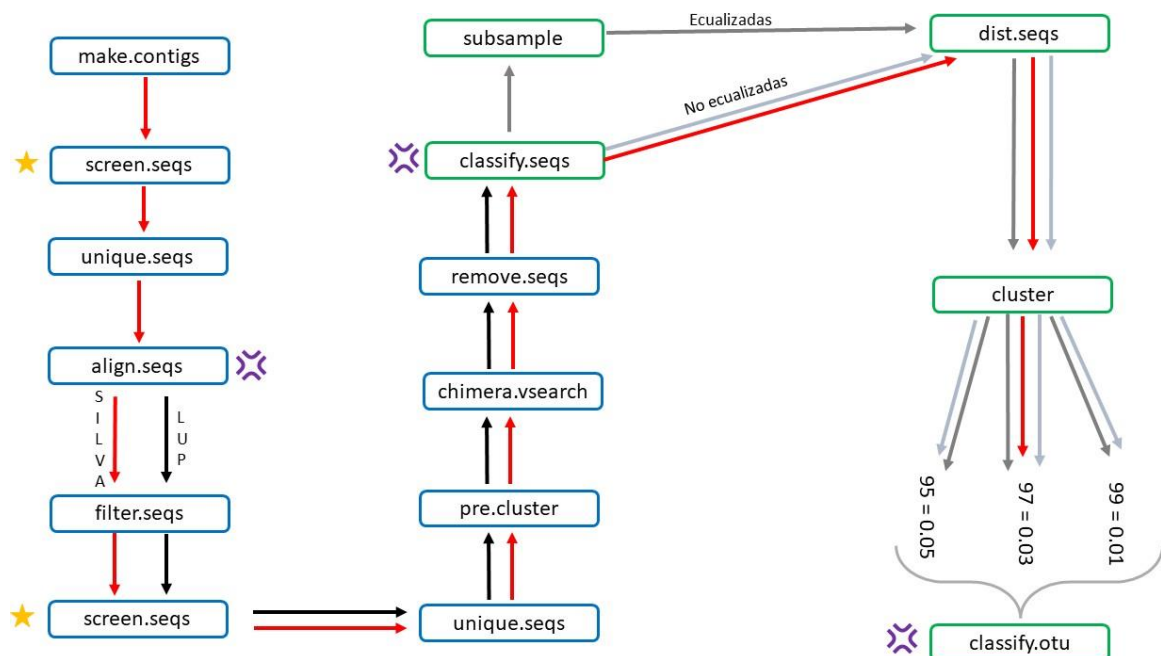


Figura 2-2 Diagrama de la obtención, agrupamiento y clasificación de las SAC.

En azul se detallan los pasos involucrados en el control de calidad, mientras que en verde se detallan aquellos que involucran el agrupamiento y clasificación en OTUs, previo al análisis bio-geográfico propiamente dicho. Las flechas indican el orden de los pasos. Las estrellas amarillas indican los pasos en los que se optimizaron parámetros utilizando la comunidad simulada. Los asteriscos violeta indican los pasos en los cuales se utilizaron diferentes bases de datos de referencia.

2.2.2 Optimización del control de calidad

La comunidad simulada, explicada en el punto 2.1.2, fue utilizada para evaluar y optimizar los análisis delineados en el punto anterior. Los pasos que fueron optimizados se detallan a continuación, y se indican en la Figura 2-2 mediante estrellas amarillas:

A) En el primer paso de selección (punto B de la sección 2.2.1), se optimizó el parámetro *minlength*, para lo cual se realizaron pruebas con dos largos, 250 y 350 bases, y se evaluó cómo cada elección afectó los datos obtenidos.

B) En el paso F (sección 2.2.1) se evaluaron diferentes porcentajes de optimización del argumento *optimize (start-end)*. Los porcentajes evaluados fueron 70, 80 y 90%.

2.2.3 Desarrollos creados para el proceso de control de calidad

Para el proceso de control de calidad descrito en la sección anterior, fue necesario desarrollar una serie de *Scripts* (archivos conteniendo instrucciones informáticas complejas), que permitieron automatizar la estrategia de análisis. A modo demostrativo de las características de dichos desarrollos, a continuación, se muestran dos *scripts* correspondiente al control de calidad. En los mismos ya se encuentran incorporados los resultados de las optimizaciones detalladas en la sección anterior. En color gris se incluyen comentarios que describen los procedimientos llevados a cabo por cada parte del código. Las partes en negro corresponden a las instrucciones pasadas al programa Mothur. El texto en azul corresponde a instrucciones para el editor de texto Vim, las cuales fueron implementadas desde dicho programa.

SCRIPT 1

```
make.contigs(file=argSea.files, processors=2)
#Filtrado de secuencias
screen.seqs(fasta=argSea.trim.contigs.fasta,
group=argSea.contigs.groups, maxhomop=9, minlength=350,
maxambig=0, processors=2)
PROCESAMIENTO CON VIM
#Eliminación de primers/barcodes mediante el editor de texto
Vim.
#remoción de primers y barcodes de secuencias forward:
%s/^\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\
\a\a\a\a\a\a\a\a\a\a//g
# remoción en secuencias reversas:
%s/\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\a\
a\a\a\a\a\a\a\a\a$a$/g
```

SCRIPT 2

```
# identificación de secuencias únicas
unique.seqs(fasta=argSea.trim.contigs.good.SP.fasta)
# alineamiento con secuencias de referencia
align.seqs(fasta=argSea.trim.contigs.good.SP.unique.fasta,
reference=/home/laura/bin/mothur-
1.39.5/References/silva.seed_v128.align, flip=T)
# filtrado de las secuencias que no alinean correctamente
filter.seqs(fasta=argSea.trim.contigs.good.SP.unique.align,
vertical=T)
# Detección de outliers en el alineamiento.
screen.seqs(fasta=argSea.trim.contigs.good.SP.unique.filter.fast
a, group=argSea.contigs.good.groups,
name=argSea.trim.contigs.good.SP.names, optimize=start-end,
criteria=70, processors=2)
#identificación de secuencias únicas
unique.seqs(fasta=argSea.trim.contigs.good.SP.unique.filter.good
.fasta, name=argSea.trim.contigs.good.SP_LUP.good.names)
#pre-agrupameinto de las secuencias
pre.cluster(fasta=argSea.trim.contigs.good.SP.unique.filter.good
.unique.fasta,
name=argSea.trim.contigs.good.SP.unique.filter.good.names,
group=argSea.contigs.good.SP.good.groups, diffs=2)
#identificación de secuencias quiméricas
chimera.vsearch(fasta=argSea.trim.contigs.good.SP.unique.filter.
good.unique.precluster.fasta,name=argSea.trim.contigs.good.SP.un
ique.filter.good.unique.precluster.names,group=argSea.contigs.go
od.good.groups,processors=2)
#eliminación de secuencias quiméricas
remove.seqs(fasta=argSea.trim.contigs.good.SP.unique.filter.good
.unique.precluster.fasta,
name=argSea.trim.contigs.good.SP.unique.filter.good.unique.precl
uster.names, group=argSea.contigs.good.SP.good.groups,
accnos=argSea.trim.contigs.good.SP.unique.filter.good.unique.pre
cluster.denovo.vsearch.accnos)
```

2.3 Análisis comunitarios

2.3.1 Identificación de linajes ubicuos y profusos (LUPs)

2.3.1.1. *Clasificación de las secuencias.* Luego de haber obtenido las secuencias de alta calidad, se clasificaron taxonómicamente utilizando como referencia la base de datos SILVA v128 (Pruesse et al. 2007), mediante los siguientes pasos:

A) En primer lugar, se clasificaron las secuencias de alta calidad (SAC) mediante el clasificador bayesiano naïve de Wang (Wang et al. 2007), comando *classify.seqs* de Mothur, el cual se utilizó con sus parámetros por defecto (*default*).

B) Posteriormente se generaron matrices de distancias pareadas entre las SAC mediante la función *dist.seqs* de Mothur. Los extremos no fueron penalizados (parámetro *countends=F*).

C) En base a las matrices obtenidas según lo descrito en el punto anterior, las SAC fueron agrupadas en unidades taxonómicas operativas (OTUs, por la frase en inglés *Operational Taxonomic Unit*) mediante el algoritmo DGC (del inglés *distance-based greedy clustering*) de VSearch implementado en la función *cluster* Mothur. Esta estrategia funciona asignando una nueva secuencia al más cercano (centroide más similar) cuando existen varios centroides dentro del umbral de similitud dado. El tipo de agrupamiento se especificó mediante el argumento *method= dgc*. El nivel de agrupamiento (parámetro *cutoff*) utilizado para el caso de la asignación taxonómica fue del 97%.

D) La asignación de las OTU a categorías taxonómicas fue realizada mediante la función *classify.otu* de Mothur.

Estos pasos se indican en la Figura 2-2 mediante color verde. El correspondiente flujo de trabajo se indica con las flechas rojas. A continuación, se muestra el *script* que se utilizó para dirigir el mismo:

```
#clasificacion taxonómica de las secuencias, antes de agruparlas
classify.seqs (fasta=argSea.trim.contigs.good.SP_LUP.unique.filter
r.good.unique.precluster.pick.subsample.fasta,reference=/home/laura/ArgSeaData/Taxonomy/References_LUP/LUP.align,taxonomy=/home/laura/ArgSeaData/Taxonomy/References_LUP/taxonomia_LUP.tax)
#calcular la distancia entre pares de secuencias
dist.seqs (fasta=argSea.trim.contigs.good.SP_LUP.unique.filter.go
od.unique.precluster.pick.subsample.fasta,cutoff=0.03,countends= F)
#agrupan las secuencias con diferentes niveles de similitud
cluster (fasta=argSea.trim.contigs.good.SP_LUP.unique.filter.good
.unique.precluster.pick.subsample.fasta,count=argSea.trim.contig
s.good.SP_LUP.unique.filter.good.unique.precluster.pick.subsampl
e.count_table, cutoff=0.03, method=dgc)
```

```
#clasificacion taxonómica de los resultados del agrupamiento
classify.otu(list=argSea.trim.contigs.good.SP_LUP.unique.filter.
good.unique.precluster.pick.subsample.dgc.unique_list.list,count
=argSea.trim.contigs.good.SP_LUP.unique.filter.good.unique.precl
uster.pick.subsample.count_table,taxonomy=argSea.trim.contigs.go
od.SP_LUP.unique.filter.good.unique.precluster.pick.subsample.ta
xonomia_LUP.wang.taxonomy, label=0.03).
```

2.3.1.2. Determinación de las abundancias de los taxa identificados. Una vez clasificadas las OTUs, se procedió a evaluar qué linajes estuvieron representados en forma abundante y ubicua en las 17 comunidades del Mar Argentino. En base a conocimientos previos del grupo de trabajo (Calvo et al. 2018; Manrique y Jones 2017), se utilizó como punto de corte una abundancia de secuencias del 5%, en el total de secuencias en una muestra. Los taxa que superaron dicho umbral, a los que denominamos en este trabajo Linajes Ubicuos y Profusos (LUPs), fueron utilizados en los análisis comunitarios.

2.3.2 Construcción de una base de datos de referencia de LUPs

La base de datos SILVA, que se utilizó para realizar el alineamiento y clasificación de las secuencias, como se mencionó en la sección anterior, es una base de datos muy amplia desde el punto de vista taxonómico. En consecuencia, muchos taxa no se encuentran representados por todas las variantes conocidas de cada uno de los mismos, lo cual obedece a la ventaja de mantener relativamente reducido el tamaño de la base de datos. Sin embargo, esto tiene la consecuencia negativa de que, al no incluir toda la diversidad de cada grupo, el diccionario de k-meros utilizados por la función *classify-seqs* puede no incluir entradas para muchas de las secuencias problema, resultando en que las mismas pueden potencialmente quedar como “no clasificables” debido a que la base de datos no tiene representantes lo suficientemente cercanos. Por esta razón, en este trabajo se generó una base de datos específica para los LUPs identificados (punto 2.3.1.2.). Esta base de datos “*in-house*” se utilizó en paralelo con la base de datos SILVA y, para los análisis biogeográficos, se eligió el resultado que resultó en una mayor cantidad de secuencias de LUPs (lo que interpretamos como esfuerzos de muestreo superiores). Para crear la base de datos de referencia se utilizaron secuencias completas del gen 16S presentes en

diferentes bases de datos públicas, todas las cuales contaron con una descripción y clasificación respaldada por trabajos publicados. Dichas secuencias se detallan a continuación. Se describe el linaje, clados específicos encontrados para cada linaje y números de acceso mediante el cual las secuencias pueden ser identificadas en la colección anotada y curada de secuencias de ADN disponibles públicamente, Genbank (Ferster 1973):

Linaje SAR11

Clado 1a: AF245616, AF268220-1446, AF327027, AJ400350, AY033306, AY033320, L10935, U13159, U75253 y X52280.

Clado 1b: AY033312, U75649, X52172, AY0332199 y AY033230.

Clado 2: AY033322, U75254, U75256 y U75257.

Clado 3: AF418965, AY145598 y Z99997.

Clado 4: AF353229, AF353226 y U70686.

Linaje SAR86

Clado1: AY033304, AY033311, AY033314, AF16650, 233932639 AF, AF353229, U70686, AJ400345, AY033317 y U70693.

Clado 3a: AF173975, AF241653, AJ240912 y AY033328.

Clado 3b: AF001651, AF268217, AF354613, AJ400344, AJ400356 y AY458632.

Linaje OM1

MBY015C0326-A00 y U707105.

Linaje NS5

JX524915.1.1437, FJ352248.1.858, EU733897.1.730, GO349983.1.1368, EF572510.1.1322, EU010174.1.1471, JQ195035.1.1360, DQ656151.1.614, HM037.1455.16.868, AY274871.1.1213, JQ516707.1515, JX525253.1.1434, JX526605.1.1431, FJ744847.1.1374, KP887226.1.1157, LC018891.1.917, JQ1990 66.1.1360, EF572758.1.1494, JQ197656.1.1360, HM023186.1.984 y KP887518.1.1164.

Linaje *Amylibacter*

AF245635, JX016850, KX934278, U70680 y U78942.

Estas secuencias de referencia fueron alineadas mediante el programa MAFFT (Kato y Toh 2010) con los argumentos de la estrategia LINSI. El alineamiento obtenido fue manualmente formateado según los requerimientos del programa Mothur, utilizando el editor de texto Vim. Las filiações de las secuencias de referencia de la base de datos “*in-house*” fue corroborada mediante análisis filogenéticos con programa fastTree (Jones et al. 1992). El modelo evolutivo (JC69) se infirió mediante el programa Mr.AIC (Nylyer 2004). Para generar los árboles con el programa FastTree se utilizaron los parámetros por defecto de este, excepto que se utilizó una única categoría de tasa en el modelo evolutivo. Bajo estas condiciones, el programa realiza hasta $4 \times \log_2(N)$ rondas de evolución mínima con reordenamiento por intercambio de vecinos (en inglés *Nearest Neighbor Interchange, NNI*), 2 rondas de reordenamiento por poda y re-injerto de ramas (en inglés *Subtree Pruning Regrafting, SPR*), y hasta $2 \times \log_2(N)$ rondas de máxima verosimilitud con *NNI*, donde N es el número de secuencias únicas en el alineamiento. Este paso permitió corroborar la monofilia de los taxa incluidos, y que ninguna de las secuencias incluidas presenta secuencias divergentes (ausencia de ramas atípicamente largas).

2.3.3 Ecuación de los esfuerzos de muestreo.

Existen distintas opiniones sobre la utilidad/necesidad de ecualizar los datos de secuenciación masiva. McMurdie y Holmes 2014 (McMurdie y Holmes 2014) sostienen que las muestras ecualizadas son ineficientes en sentido estadístico, ya que generan observaciones adicionales de especies raras llevando a un aumento en las estimaciones de diversidad alfa y beta, además el tamaño de las bibliotecas queda definido de manera arbitraria. La ecualización está lejos de ser el método óptimo, este enfoque es inapropiado para la detección de especies con abundancia diferencial. Por el contrario, Weiss y colaboradores (Weiss et al. 2017), afirman que la mayoría de los métodos de ecualización permiten la agrupación exitosa de muestras según el origen biológico cuando los grupos difieren sustancialmente en su composición microbiana general. La ecualización utilizada de manera adecuada puede potencialmente eliminar los sesgos y las variaciones introducidas por el muestreo; y los resultados del proceso de secuenciación, de tal modo que los datos ecualizados reflejan la biología subyacente. Los enfoques no ecualizados/normalizados y de alta proporción muestral son propensos a la generación de grupos de artefactos basados en la profundidad de secuenciación en el análisis de diversidad beta. Para estudios que presentan muestras con grandes diferencias en el tamaño medio de la biblioteca entre grupos, la ecualización ayuda a disminuir la tasa de falso descubrimiento.

Como no se encuentra un consenso en cuanto a la utilidad de ecualizar o no los datos, tal como se mencionó anteriormente, todos los análisis de las comunidades de esta Tesis fueron realizados a partir de datos ecualizados y no ecualizados.

Para realizar la ecualización se utilizó el programa Mothur (Schloss et al. 2009), con

el comando *sub.sample*, que permite realizar re-muestreos de las secuencias de cada muestra hasta obtener un número determinado de las mismas que sea homogéneo entre muestras. En microecología, es común ecualizar de forma tal que todas las muestras tengan el mismo número de secuencias que la muestra que presentó el menor número de secuencias, estrategia utilizada en esta Tesis utilizando el argumento *persample=T*. El paso de ecualización se representa en la Figura 2-2 con flechas en diferentes tonos de grises, especificando el camino que se toma en cada uno.

2.3.4 Agrupación en Unidades Taxonómicas Operacionales (OTUs).

Como se indica en los pasos que se esbozaron en la sección 2.3.1 (puntos B y C) y Figura 2-2, se utilizaron 3 niveles de agrupamiento: 95, 97, 99%. Esto se realizó mediante los comandos *dist.seqs* y *cluster* de Mothur, especificando el nivel en el argumento mediante el parámetro *cutoff*. Las secuencias agrupadas en OTUs con 95, 97 y 99% de similitud, corresponden a rangos taxonómicos putativos de familia, género y especie, respectivamente (Yarza et al. 2014).

2.3.5 Generación de archivos complementarios para análisis comunitarios.

Una vez que las secuencias fueron agrupadas en OTUs y se clasificaron taxonómicamente, se utilizó Mothur para reordenar la información para realizar los análisis comunitarios. Las funciones que se utilizaron fueron:

count.seqs: Esta función permite contar el número de secuencias agrupadas en cada OTU.

make.shared: *make-shared* permite inferir conteos parciales, por muestra, de la abundancia de cada OTU.

A continuación, se muestra cómo se implementaron estas funciones:

```
#conteo de las secuencias y distribución de estas
count.seqs(name=argSea.trim.contigs.good.SP_LUP.unique.filter.goo
d.unique.precluster.pick.subsample.names,group=argSea.contigs.goo
d.LUP.good.pick.subsample.groups)
#establecer las diferentes OTUs presentes en cada muestra y la
cantidad de secuencias
make.shared(list=argSea.trim.contigs.good.SP_LUP.unique.filter.goo
od.unique.precluster.pick.subsample.dgc.unique_list.list,
count=argSea.trim.contigs.good.SP_LUP.unique.filter.good.unique.p
recluster.pick.subsample.count_table)
```

Una vez realizados estos procedimientos para todos los niveles de agrupamiento con datos ecualizados y datos no ecualizados, se identificaron y separaron las OTUs de

cada linaje utilizando los archivos obtenidos y funciones base del lenguaje de programación estadística R.

2.4 Estimación de relaciones taxa-área (TAR).

El primer paso para realizar las curvas taxa-área fue calcular la distancia geográfica que existe entre cada par de comunidades. Se calcularon todas las distancias de a pares entre las comunidades mediante la página <https://www.movabletype.co.uk/scripts/latlong.html>, la cual permite extraer la distancia en kilómetros que separa dos localizaciones utilizando las coordenadas geográficas de las mismas. Una vez obtenidas estas distancias, se procedió a ordenar cada par de comunidades dependiendo de la distancia que las separa, desde las que presentaron menor distancia, hasta las que presentaron mayor distancia. A partir de aquí, se realizó el procedimiento descrito por Manrique y Jones 2017. Brevemente, utilizando un método tipo Monte Carlo se tomaron series de cascos semiesféricos, de área creciente, a partir de cada una de las 17 locaciones estudiadas. En cada una de estas series, se incluyó en orden de cercanía cada una de las demás locaciones y cada vez que se incluía una nueva locación se contabilizó la cantidad de OTUs conjuntas. La cuantificación de las OTUs se realizó mediante la función `get.otus` del programa Mothur, indicando las muestras a tener en cuenta en cada caso mediante el argumento `groups`. A continuación, se muestra el *script* generado para realizar dichos conteos.

```
#a modo de ejemplo se muestra el script correspondiente al linaje
Amylibacter, agrupado a 95% de similitud, con datos no ecualizados y
tomando como punto de partida la muestra MDQ. Esto se repitió a
partir de cada una de las muestras, colapsando en una única
observación todo par de muestras que pudiera verse duplicado. Se
realizaron análisis equivalentes para cada LUP y nivel de similitud,
con y sin ecualización.
```

```
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13)
```

```
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11)
```

```
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12)
```

```
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF)
```

```
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10)
```

```
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9)
```

```

get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8-
SUMMER.12)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8-
SUMMER.12-BOPD.6)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8-
SUMMER.12-BOPD.6-BOPD.5)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8-
SUMMER.12-BOPD.6-BOPD.5-BOPD.4)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8-
SUMMER.12-BOPD.6-BOPD.5-BOPD.4-BOPD.3)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8-
SUMMER.12-BOPD.6-BOPD.5-BOPD.4-BOPD.3-BOPD.2)
get.groups(shared=Amylibacter_95.shared, groups=MDQ.SURF-BOPD.14-
BOPD.13-BOPD.11-BOPD.12-BH.SURF-BOPD.10-BOPD.9-BOPD.7-BOPD.8-
SUMMER.12-BOPD.6-BOPD.5-BOPD.4-BOPD.3-BOPD.2-BOPD.1)

```

Para calcular el área abarcada de cada casco semiesférico se utilizó las distancias geográficas calculadas anteriormente y el radio medio de la Tierra (6378) con la fórmula:

$$2*\pi()*6378*(6378*(1-COS((2*Distancia)/(2*6378))))$$

Este proceso fue realizado para todos los niveles de agrupamiento de OTUs, 95, 97 y 99%, y con datos ecualizados y datos no ecualizados.

Una vez obtenidos estos datos, se procedió a graficarlos utilizando el entorno de programación R. Se utilizaron escalas logarítmicas en ambos ejes (Legendre y Legendre 1998) y se realizaron análisis de regresión y correlación entre las áreas y los correspondientes números de OTUs. Las estimaciones de la pendiente y coeficiente de regresión (R^2) y valor de probabilidad asociado, se obtuvieron mediante la función *lm* que se aplicó al logaritmo de ambas variables. El análisis de la correlación (Rho ρ) y su valor

de probabilidad asociado, se realizó mediante el método de Pearson, con la función *cor.test* de R.

2.5 Estimación de Decaimientos por Distancia (DD).

Para realizar los estudios de DD, se utilizaron 3 índices de disimilitud diferentes. El índice de disimilitud de Sorensen, el cual mide las diferencias del ensamble en base a la presencia/ausencia de las diferentes OTUs, el índice de disimilitud de Bray-Curtis, que además tiene en cuenta las abundancias de las OTUs, y el índice de disimilitud de Morisita, el cual está basado en la dispersión, y se cree que posee la virtud de ser más robusto al tamaño muestral (Wolda 1981).

Las disimilitudes se calcularon en el entorno de programación R, a partir de matrices comunitarias obtenidas con Mothur, mediante la biblioteca "Vegan". Las matrices obtenidas, fueron manualmente volcadas en tablas y se incluyó la distancia geográfica entre pares de muestras. Una vez confeccionadas las mismas, se procedió a realizar análisis gráficos y a evaluar las potenciales relaciones lineales a través de un análisis de regresión y correlación, además de calcular la pendiente. Dichos análisis se realizaron mediante procedimientos equivalentes a los utilizados en el punto 2.4. Los métodos de correlación usados fueron Spearman y Kendall.

2.6 Ordenamiento mediante escalado multidimensional no métrico.

El escalado multidimensional no métrico (NMDS; del inglés *Non-metric multidimensional scaling*) es una técnica multivariada cuyo objetivo, al igual que en otros métodos de análisis multivariado, es representar en un espacio geométrico reducido (pocas dimensiones) las relaciones existentes entre conjuntos de datos. Una de las desventajas a nivel técnico de utilizar este método, es que puede tener dificultades en encontrar una solución estable única (Van Hanne et al. 1999). Esto se soluciona optimizando los valores de una función de stress, aceptando por convención aquellos ordenamientos con valores menores a 0.2 (Dexter et al. 2018). Las distancias entre las muestras pareadas se calcularon utilizando cuatro índices: Bray-Curtis, Sorensen, Morisita y Horn. El índice Horn es una variante del índice Morisita que permite trabajar con datos de abundancia; los demás índices están descritos en la sección 2.5.

El cálculo de los ordenamientos, así como el de las distancias, se realizaron en el entorno de programación R utilizando la biblioteca "Vegan" mediante las funciones *vegdist* y *metaMDS*. Los argumentos que se fijaron en esta función fueron *try=20* y *trymax=100*.

2.7 Determinación de patrones de endemismo

Para determinar la existencia de endemismo se partió de archivos *shared* (matrices comunitarias) generados por Mothur. Como toda matriz comunitaria, los mismos especifican numéricamente cómo se distribuyen las OTUs y secuencias en cada una de las muestras de estudio. La identificación de OTUs endémicas se realizó mediante funciones base del entorno de programación de R. La búsqueda consistió en identificar OTUs que solo se detectaron en una de las muestras. Esta búsqueda se realizó mediante un *loop* para ingresar de forma particular en cada una de las OTUs, combinado con la función *which*, la cual fue utilizada para buscar las muestras que presentan secuencias de cada OTU. Este procedimiento devuelve un vector que contiene el nombre de las muestras que tienen secuencias de la OTU clasificada en ellas. Luego se realizó una selección de aquellos vectores que tengan largo igual a 1; es decir que sólo presentaron secuencias en una de las muestras.

Estos análisis se complementaron con cálculos de la cobertura en cada muestra (para cada linaje en cada nivel de agrupamiento con datos ecualizados y sin ecualizar), con el objeto de evaluar si la ausencia de OTUs en alguna/s muestra/s presentaba relación con la correspondiente cobertura. Los valores de cobertura se obtuvieron mediante la función *summary.single* de Mothur, especificando el valor *coverage* para el argumento *calc*.

2.8 Estimación y análisis espacial de la diversidad

Para estudiar la alfa-diversidad de los LUP en cada una de las localidades muestreadas se utilizaron los índices de ACE, CHAO, Shannon, Simpson, y el índice de homogeneidad de Shannon.

Estos análisis se realizaron mediante el programa Mothur; utilizando el comando *summary.single*. Los índices se escogieron utilizando el argumento *calc* mediante el cual se especificaba el índice de diversidad a calcular.

Los resultados obtenidos se volcaron manualmente en una tabla, la cual se utilizó para realizar análisis gráficos mediante funciones base del entorno de programación estadística R.

Al igual que para todos los análisis realizados en esta Tesis, el análisis de la diversidad con cada uno de los cinco índices se realizó por linaje individual, para cada uno de los tres niveles de agrupamiento (95, 97 y 99%), y con datos ecualizados y no ecualizados.

3 Resultados

3.1 Obtención de secuencias de alta calidad

3.1.1 Ajuste mediante una base de datos reducida (comunidad simulada).

Las secuencias crudas de la comunidad simulada (99678 secuencias pareadas) fueron sometidas a los controles de calidad que se detallan en los Materiales y Métodos. Como se detalla en la subsección 2.2.2, además de efectuar los controles sugeridos en la literatura, se evaluó el efecto de filtrar los datos crudos utilizando dos umbrales de largos de secuencias diferentes, 250 bases (menos astringente y menor información por secuencia) y 350 bases (más astringente y más información por secuencia). Los resultados correspondientes se observan en la Tabla 3-1.

Tabla 3-1: Efecto del filtrado por largo de secuencia en el largo de las secuencias finales

	Total	250 bases largo	350 bases largo
N° de contigs/secuencias	99678	69697	64035
Media del largo de secuencias	446	462	475
Mediana del largo de secuencias	476	476	479

Como puede observarse en dicha tabla, al realizar los filtrados por largos de secuencias en ambos casos disminuyó considerablemente el número de secuencias con respecto al número de secuencias totales.

En la Figura 3-1 se puede observar que la dispersión disminuye sustancialmente con respecto a los datos sin procesar, lo que indica que se obtienen conjuntos de datos más homogéneos. A su vez, se observa que, al filtrar a 350 bases, la dispersión resultó ser menor que al hacerlo por 250 bases.

A la luz de los resultados de la Figura 3-1 y la Tabla 3-1, se decidió seleccionar el filtrado de largo mínimo de secuencias de 350 bases, ya que permitió obtener secuencias de mayores largos, y con una menor dispersión de los largos de secuencias. A través de este paso se excluyeron 35643 secuencias.

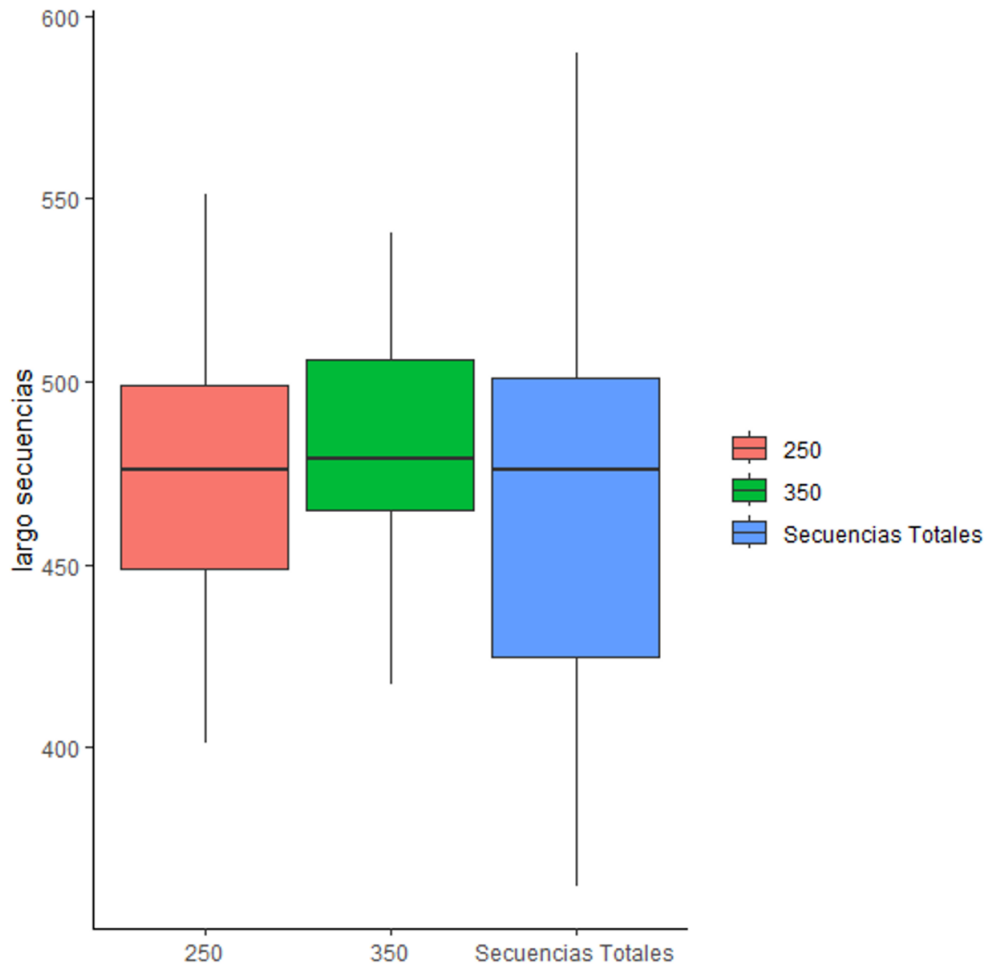


Figura 3-1: Efecto del filtrado por largo de secuencias en el largo de las secuencias finales.

Posteriormente se optimizó el paso destinado a identificar secuencias que presentaron comportamientos anómalos durante el proceso de alineamiento con la base de datos de referencia. Esto se realizó mediante una optimización por criterio de mayoría donde se observa el porcentaje de secuencias que inician y finalizan en cada posición (puntos F y B de las secciones 2.2.1 y 2.2.2, respectivamente). En la Tabla 3-2 se observa que a medida que se permitió un porcentaje mayor, es decir menor astringencia, las posiciones de inicio y finalización fueron menos definidas y aumentó el número de OTUs. Por el contrario, al disminuir el porcentaje, los valores de inicio y finalización fueron mucho más homogéneos. Esto indica que la zona del alineamiento fue mucho más acotada, permitiendo seleccionar una mayor proporción de secuencias con información genética homogénea (*i.e.* alinean en una región más acotada del gen blanco), y cuyos largos fueron más homogéneos.

Tabla 3-2: Efecto filtrado mediante optimización sobre distintos parámetros.

Porcentaje de prueba	Base inicial del alineamiento	Base final del alineamiento	Largo promedio de las secuencias	N° total de secuencias obtenidas	N° OTUs totales obtenidas
70	2	1139	460	36470	938
80	2	1129;1435	459	41268	1047
90	1; 2; 8	1100; 1115; 1139; 1722	457	48072	1301

En base a estos resultados, se eligió el criterio de selección, donde del 70% de las secuencias comienzan y terminan en la misma posición; ya que, si bien se pierde un mayor número de secuencias, estas presentan mayor homogeneidad entre sí. Este paso implicó la exclusión de 27565 secuencias.

Luego del control de calidad, se obtuvo un total de 24959 secuencias de alta calidad, habiéndose excluido un total de 74719 secuencias que no cumplieron con los requisitos impuestos por criterios de calidad establecidos en este trabajo.

La clasificación taxonómica de las secuencias de alta calidad así obtenidas se correspondió con la composición de la comunidad simulada generada para este fin.

3.1.2 Obtención de secuencias de alta calidad del Mar Argentino.

Las secuencias pareadas del Mar Argentino utilizadas en este trabajo consistieron en 1.647.614 lecturas directas y 1.523.564 reversas. En el proceso de formación de contigs se generaron 965.262 secuencias pareadas, quedando sin aparearse 2.205.916 lecturas. Luego de someter estas 965.262 secuencias al control de calidad (Fig. 2-2), se obtuvieron 302.597 secuencias de alta calidad (SAC), excluyéndose de esta forma 662.665 secuencias que podrían contener errores de secuenciación.

La distribución de las SAC entre los puntos muestreados se detalla en la Tabla 3-3. Si bien existieron variaciones entre las cantidades de secuencias obtenidas por muestra, en todos los casos se observó una alta cobertura; indicando que las muestras resultaron ser representativas de las correspondientes comunidades bacterianas.

Tabla 3-3: Distribución de SAC en las muestras y su cobertura.

Muestra	N° secuencias	Cobertura
BH	24405	0.94
BOPD1	18073	0.92
BOPD2	17956	0.90
BOPD3	9459	0.85
BOPD4	13065	0.85
BOPD5	12207	0.90
BOPD6	12350	0.88
BOPD7	15872	0.91
BOPD8	16306	0.90
BOPD9	15144	0.92
BOPD10	14248	0.91
BOPD11	12570	0.89
BOPD12	16711	0.88
BOPD13	11911	0.88
BOPD14	16159	0.88
MDQ	22807	0.95
CHR	53354	0.97

3.2 Taxonomía y distribución de grupos bacterianos abundantes

3.2.1 Identificación de Linajes Ubicuos y Profusos (LUP).

Luego de clasificar las SAC utilizando la base de datos de SILVA, tal como se detalla en Materiales y Métodos, se investigó cuáles fueron los linajes más ubicuos y abundantes, siguiendo los criterios explicados en la sección 2.3.1.2.

Se encontró que sólo cinco linajes superaron el umbral de abundancia de referencia. Estos linajes, a los cuales se denomina en este trabajo como Linajes Ubicuos y Profusos (LUP), fueron: SAR11, OM1, SAR86, NS5 y *Amylibacter*. En conjunto, los LUP constituyeron el 86,31% del total de las SAC. Dentro de los grupos minoritarios, se encontraron los linajes *Marinoscillum*, *Formosa*, NS4, *Polibacter*, *Synechococcus*, SAR406, *Planctomyces*, *Fretibacter*, PS1, *Ascidiaceihabitans*, *Planktomarina*, SAR116, *Chesapeake*, OM43, *Idiomarina*, OM60, K189A, ZD0405 y JTB255, los cuales en conjunto representaron el 1,82% del total de las SAC. Las secuencias no clasificables, comunes en todo estudio basado en ADN ambiental, representaron un 11,87% del total de las secuencias.

Los resultados obtenidos se resumen en la Tabla 3-4, la cual describe resultados correspondientes a OTUs que presentaron abundancias mayores a 10 secuencias.

Puede observarse que los cinco LUPs identificados en este trabajo no se distribuyeron en forma homogénea entre las diferentes comunidades del Mar Argentino, sino que algunos de ellos presentaron abundancias inusualmente altas en algunas de las comunidades muestreadas. En la siguiente sección se presenta un análisis más detallado de la distribución de los LUPs.

Tabla 3-4: Linajes Ubicuos y Profusos (LUP) y sus abundancias (% de SAC de cada taxón) en cada muestra.

	<i>Amylibacter</i>	NS5	OM1	SAR11	SAR86	Otras*	No clasificables**
BH	0.60	2.85	1.98	65.37	8.37	6.10	14.73
BOPD.1	0.60	11.35	44.76	32.23	0.49	0.23	10.34
BOPD.10	0.37	6.15	56.16	15.02	0.25	0.77	21.29
BOPD.11	0.06	5.63	5.96	47.65	2.64	0.77	37.28
BOPD.12	0.03	4.75	29.41	57.41	1.14	0.84	6.42
BOPD.13	0.03	2.49	35.87	52.99	1.08	2.80	4.75
BOPD.14	0.02	1.32	54.62	39.24	0.45	0.83	3.51
BOPD.2	1.06	2.49	62.43	28.68	0.53	0.49	4.30
BOPD.3	15.56	4.92	46.63	19.28	0.73	2.24	10.63
BOPD.4	3.23	2.18	59.02	28.64	0.94	0.78	5.20
BOPD.5	1.47	1.28	66.86	23.67	1.44	1.69	3.61
BOPD.6	1.35	3.31	57.98	29.88	1.03	0.49	5.96
BOPD.7	0.99	5.10	34.45	22.58	1.30	0.81	34.77
BOPD.8	0.86	2.07	38.57	45.42	1.23	1.18	10.67
BOPD.9	0.09	2.57	56.55	31.01	0.58	0.93	8.28
MDQ	10.14	0.26	0.41	64.08	12.44	6.08	6.59
CHR	0.34	7.96	33.16	44.50	0.41	1.13	12.50

* Taxa de baja abundancia (<5%): *Marinoscillum*, *Formosa*, *NS4*, *Polibacter*, *Synechococcus*, *SAR406*, *Planctomyces*, *Fretibacter*, *PS1*, *Asciaceihabitans*, *Planktomarina*, *SAR116*, *Chesapeake*, *OM43*, *Idiomarina*, *OM60*, *KI89A*, *ZD0405*, *JTB255*.

** Secuencias no clasificables.

En resumen, los cinco linajes ubicuos y profusos, correspondieron a los siguientes taxa según la clasificación SILVA:

- Amylibacter*: *Proteobacteria*, *Alphaproteobacteria*, *Rhodobacterales*, *Rhodobacteraceae*
- NS5: *Bacteroidetes*, *Flavobacteriia*, *Flavobacteriales*, *Flavobacteriaceae*
- OM1: *Actinobacteria*, *Acidimicrobiia*, *Acidimicrobiales*, OM1, *Candidatus _actinomarina*
- SAR11: *Proteobacteria*, *Alphaproteobacteria*, SAR11, *Surface_1*, *Candidatus_pelagibacter* y/o *Proteobacteria*, *Alphaproteobacteria*, SAR11, *Surface_1*, *Surface_1_ge*.
- SAR86: *Proteobacteria*, *Gammaproteobacteria*, *Oceanospirillales*, SAR86_clade, SAR86_clade_ge.

En base a estos resultados, las SAC fueron reprocesadas con la base de datos de referencia “*in house*” desarrollada en este estudio, como se explicó en la subsección 2.3.2 de Materiales y Métodos. En la Tabla A-01 del Anexo se indican las proporciones de LUPs observadas mediante este procedimiento. Como cabría esperar, los mismos fueron levemente superiores, aunque comparables, a los obtenidos mediante la base de datos Silva (Tabla 3-4).

3.2.2 Distribución de los Linajes Ubicuos y Profusos.

Como se detalla en la sección 2.3.2, para el reclutamiento final de SACs de LUPs se utilizó una base de datos “*in house*”, lo cual permitió obtener cantidades de secuencias levemente mayores con respecto a las obtenidas utilizando la base de datos Silva. Las cantidades de OTUs obtenidas fueron 8.391, 15.198 y 42.744 para los agrupamientos, 95, 97 y 99%, respectivamente, con datos ecualizados, mientras que para los datos no ecualizados se obtuvieron 12.552, 23.349 y 71.159 OTUs.

Como se mencionó en la sección anterior, los diferentes linajes no se distribuyeron homogéneamente en las diferentes comunidades. En la Figura 3-2 se puede observar la distribución de taxa en las diferentes comunidades del Mar Argentino estudiadas en este trabajo. En la categoría “otras”, se incluyeron las secuencias que no pertenecen a ninguno de los LUPs. Los linajes más abundantes fueron SAR11 y OM1. En las muestras tomadas más cerca de la costa se observó que el linaje que presentó mayor cantidad de secuencias fue el linaje OM1, mientras que en las muestras que fueron tomadas más lejos de la costa, el mayor aporte de secuencias correspondió al linaje SAR11. Los linajes SAR86, NS5 y *Amylibacter* resultaron ser relativamente muy abundantes en algunas de las muestras (por ejemplo, *Amylibacter* en BOPD3 y MDQ), aunque no fueron mayoritarias en ninguna

de las mismas. La distribución de las secuencias varió levemente en los diferentes niveles de agrupamiento, y no se observaron cambios sustanciales en las abundancias de cada LUP en función del uso de datos ecualizados o sin ecualizar. En el Anexo se puede observar la Figura A-02, donde se muestra las cantidades de SAC obtenidas por LUP y por muestra.

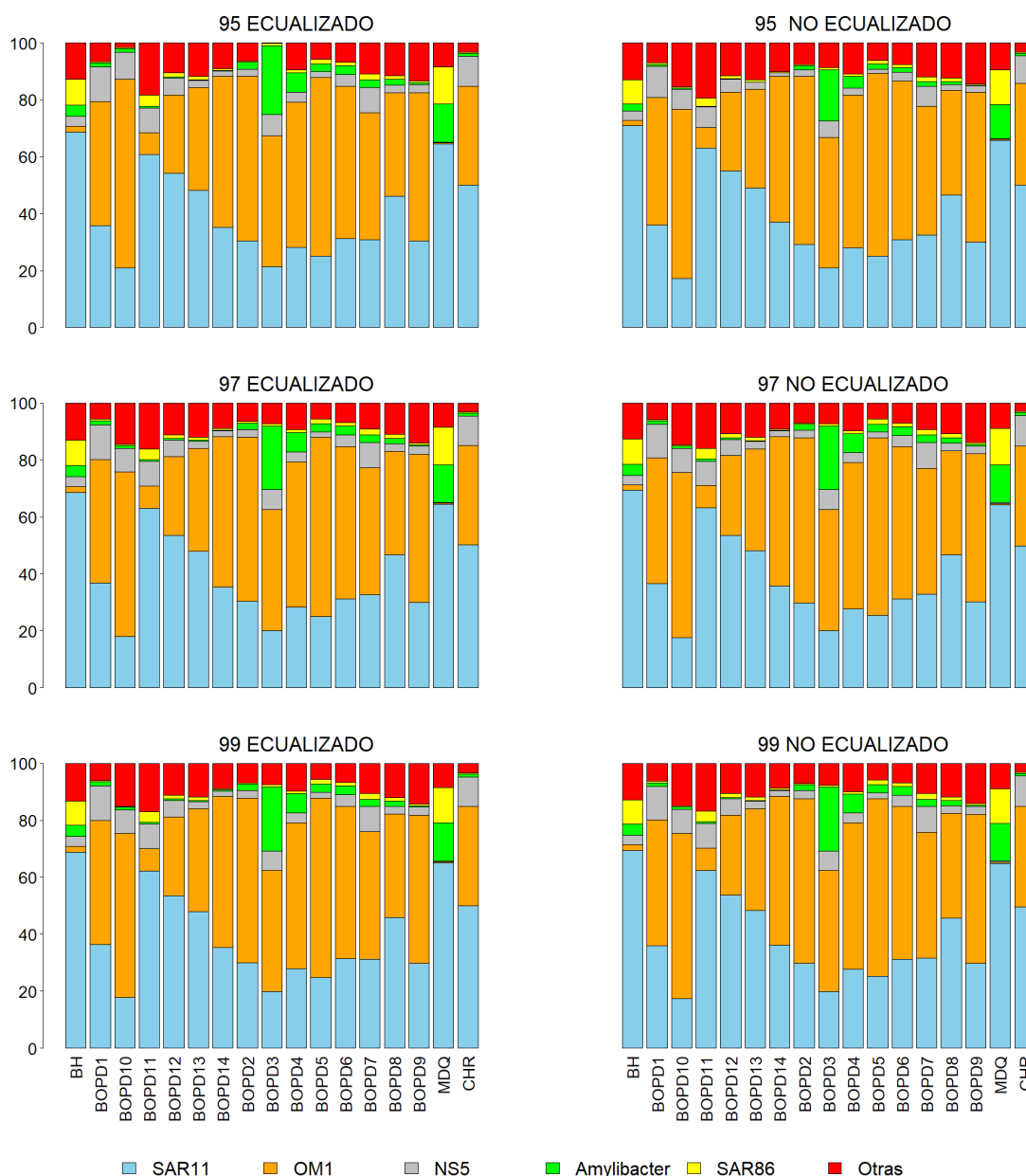


Figura 3-2: Porcentajes de SAC de cada LUPs en las muestras del Mar Argentino.

Cada barra corresponde a la proporción de SAC encontradas en cada una de las muestras discriminadas por el nivel de similitud agrupadas; ecualizados o no ecualizadas. En el eje “y” se encuentra la escala de porcentajes. En la parte inferior del grafico se especifica la identidad de cada muestra.

3.3 Estudio de las relaciones taxa-área (TAR)

Las curvas taxa-área correspondientes a todos los LUPs se muestran en la Figura 3-3. En cada panel se pueden observar 3 análisis, los cuales representan los niveles de agrupamiento por similitud. La columna de la derecha corresponde a las curvas taxa-área obtenidas a partir de datos ecualizados y la de la izquierda a datos no ecualizados.

Se observó que a medida que el área de muestreo se incrementó, el número de OTUs identificadas por cada LUP también aumentó. En otras palabras, se observó un patrón positivo para todos los análisis de relaciones taxa-área realizados. Las curvas que presentaron cambios más pronunciados corresponden a los agrupamientos de 99% de similitud, seguido por el agrupamiento de 97% y luego por 95%, como era de esperarse en relación con los correspondientes niveles taxonómicos putativos. Los resultados obtenidos a partir de datos ecualizados no presentaron diferencias relevantes con respecto a las curvas obtenidas a partir de datos sin ecualizar.

En la tabla 3-5 se presentan los resultados numéricos de las regresiones y correlaciones correspondientes a cada uno de los LUPs, en sus diferentes niveles de agrupamiento, como así también utilizando datos ecualizados y no ecualizados. Todos los parámetros resultaron estadísticamente significativos ($p < 0.05$). Para ρ (rho), parámetro de correlación, se encontraron valores por encima de 0.7; indicando que ambas variables se relacionan de manera positiva. También se observó que a medida que se aumentó el nivel de similitud en el agrupamiento, los valores de ρ disminuyeron o se mantuvieron iguales. El parámetro de regresión lineal R^2 presentó valores cercanos a 0,5. Estos valores de R^2 indican que la relación existente entre las variables puede ser explicado moderadamente por un modelo lineal.

En cuanto a las tasas de reemplazo o pendiente z , los datos sin ecualizar generaron valores levemente más bajos que los ecualizados (Tabla 3-5). Además, como pudo deducirse a partir del análisis gráfico exploratorio (Fig. 3-3), a medida que aumentó el nivel taxonómico putativo, se observó un aumento de los valores de la pendiente z .

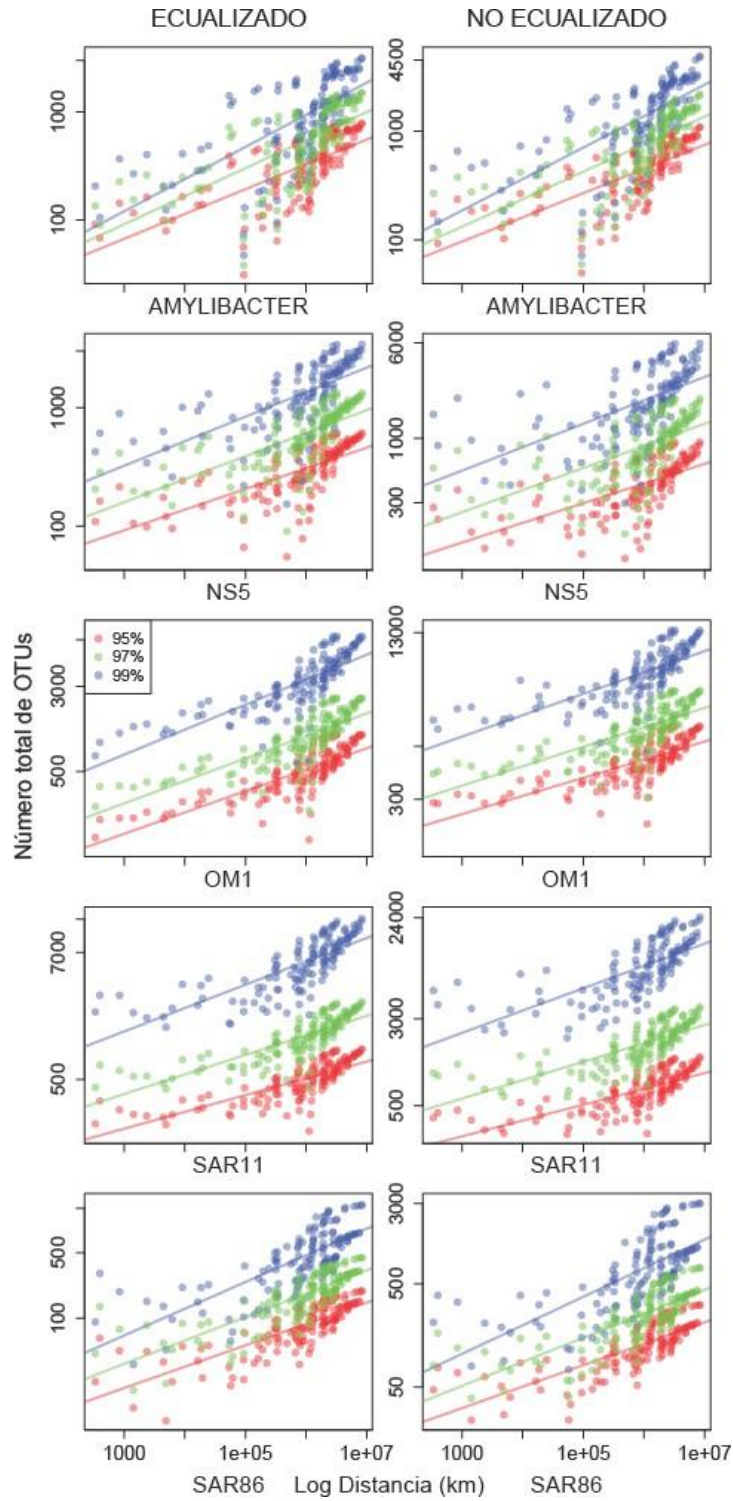


Figura 3-3: Curvas taxa-área.

Los colores indican los niveles de agrupamiento utilizados. Cada punto corresponde a la cantidad de OTUs (eje y) identificadas en una determinada área (eje x).

Tabla 3-5: Valores coeficientes de correlación (Rho ρ), regresión (R^2) y pendientes (z) correspondientes a las relaciones taxa-área (ver también Figura 3-3).

Linaje	% similitud	Ecuilizado			No ecualizado		
		ρ	R^2	z	ρ	R^2	z
<i>Amylibacter</i>	95	0.79	0.47	0.23	0.83	0.5	0.22
	97	0.78	0.46	0.26	0.82	0.49	0.25
	99	0.77	0.44	0.3	0.81	0.46	0.29
NS5	95	0.82	0.46	0.17	0.77	0.4	0.16
	97	0.82	0.46	0.19	0.76	0.38	0.18
	99	0.79	0.42	0.21	0.72	0.33	0.19
OM1	95	0.85	0.62	0.19	0.83	0.55	0.18
	97	0.85	0.61	0.2	0.82	0.53	0.19
	99	0.86	0.58	0.23	0.84	0.52	0.21
SAR11	95	0.84	0.59	0.15	0.81	0.52	0.14
	97	0.86	0.61	0.18	0.83	0.54	0.16
	99	0.84	0.58	0.21	0.8	0.48	0.2
SAR86	95	0.83	0.62	0.23	0.81	0.59	0.21
	97	0.83	0.63	0.25	0.78	0.56	0.23
	99	0.77	0.57	0.29	0.73	0.48	0.28

3.4 Decaimiento por distancia.

Como se detalló en los Materiales y Métodos, en este trabajo se estudiaron las curvas de Decaimiento por Distancia (DD) utilizando los índices de disimilitud de Bray-Curtis, Sorensen y Morisita. Los resultados obtenidos se presentan en las Figuras 3-4, 3-5 y 3-6, respectivamente.

En las curvas de DD basadas en el índice de Bray-Curtis (Figura 3-4) se observó que a medida que aumentó la distancia entre las comunidades estudiadas, los valores de disimilitud tendieron a incrementarse. Las curvas con mayores pendientes fueron observadas cuando el nivel de agrupamiento fue del 95% de similitud, seguidas por las de 97%, y luego las de 99%.

En la Tabla 3-6 pueden observarse los valores de los parámetros estadísticos de las curvas de DD correspondientes a la Figura 3-4. Las pendientes, β , se ordenaron según los niveles de agrupamiento utilizados, en concordancia con lo observado en el correspondiente gráfico. Particularmente para el linaje *Amylibacter*, con un nivel de agrupamiento del 95%, se pudieron observar valores de β mayores, contrariamente a lo que ocurrió en la mayoría de los otros linajes, en los cuales los mayores valores de β se observaron en las curvas de 97 y 99%. Los valores de β sin ecualizar fueron ligeramente mayores que los ecualizados.

Los valores de rho (ρ) que se obtuvieron fueron positivos, indicando una variación directamente proporcional de las variables. Los valores obtenidos para los linajes *Amylibacter* y SAR11 presentaron, al realizar el agrupamiento 99% de similitud, diferencias con respecto a los demás agrupamientos. En *Amylibacter* el valor de ρ fue cercano a cero mientras que en SAR11 el valor de ρ se incrementó. En el resto de los linajes y agrupamientos de los linajes antes mencionados, los valores provenientes de datos no ecualizados fueron mayores a los provenientes de datos ecualizados. Los valores de ρ y R^2 resaltados (negrita) corresponden a valores cuyas probabilidades asociadas son no significativas ($p > 0,05$). Los valores de R^2 obtenidos se dividieron en dos rangos de valores; por un lado, los valores de los linajes NS5, OM1 en todos los niveles de agrupamiento y SAR11 al agrupar a 99% donde los valores están por encima de 0,3. Por otro lado, los valores correspondientes a *Amylibacter* y SAR86 en todos los niveles de agrupamiento, y SAR11 al agrupar al 95 y 97% de similitud, fueron valores bajos, muy cercanos a cero. A pesar de esto, las probabilidades asociadas fueron significativas ($p < 0,05$), salvo en *Amylibacter* al utilizar 97% de similitud y partiendo de datos ecualizados, *Amylibacter* agrupado al 99% y con datos ecualizados y no ecualizados, SAR11 en los agrupamientos del 95 y 97% sin ecualizar.

Esto implica que, si bien los análisis no paramétricos (rho) evidencian una correlación, el modelo lineal explica poco la variabilidad de los datos en torno a su media.

En la Tabla A- 03 del Anexo se presentan análisis de correlación realizados con el

coeficiente de Kendall, los cuales resultaron ser concordantes con los arriba descritos.

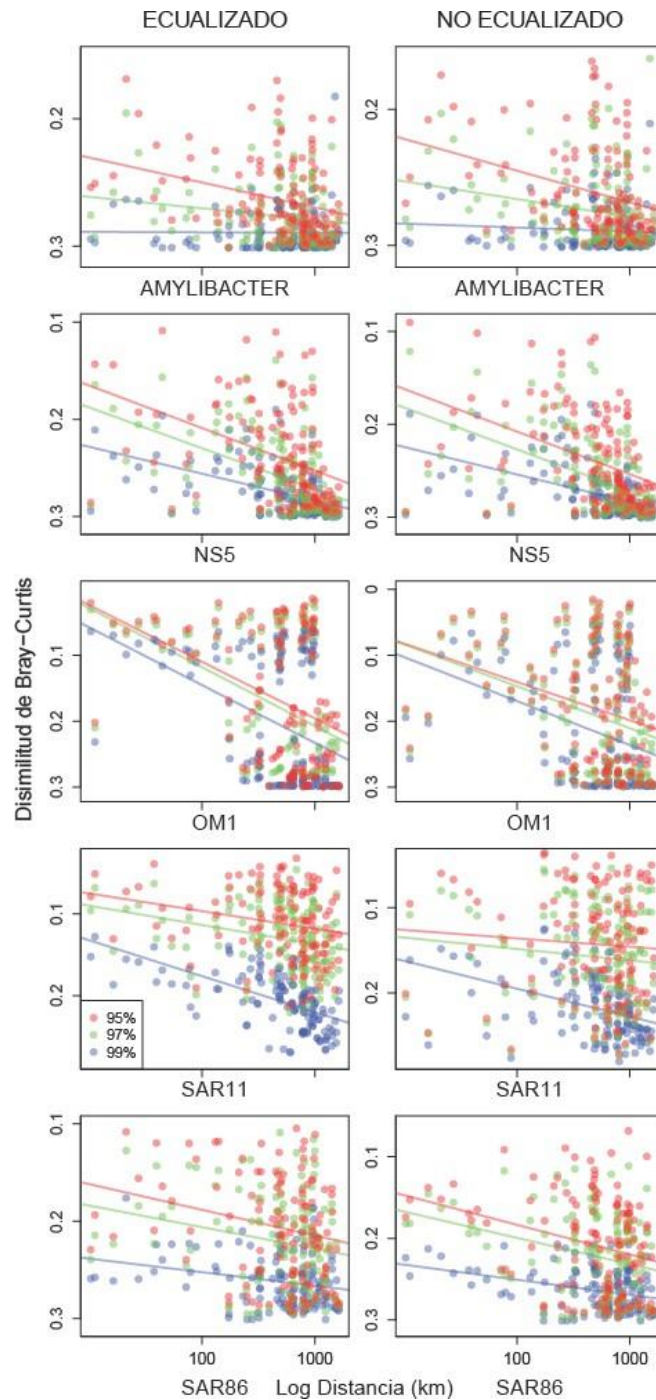


Figura 3-4: Curvas de DD basadas en el índice de disimilitud de Bray- Curtis.

Discriminados por LUP. Cada color de líneas y puntos corresponde a un nivel de agrupamiento (95, 97 y 99%), como se encuentra detallado en el gráfico. Ecuilizado del lado izquierdo y del lado derecho no Ecuilizado.

Tabla 3-6: Valores de los coeficientes de correlación (Rho ρ), regresión (R^2) y pendientes (β) de las curvas DD basados en el índice de disimilitud de Bray- Curtis (ver también Figura 3-4).

Disimilitud de Bray- Curtis		Ecuilizado			No ecualizado		
		β	ρ	R^2	β	ρ	R^2
<i>Amylibacter</i>	95	2.00E-02	0.35	0.07	2.00E-02	0.37	0.08
	97	9.00E-03	0.23	0.02	1.00E-02	0.31	0.04
	99	4.00E-03	0.03	-0.01	3.00E-03	0.13	-0.001
NS5	95	4.40E-02	0.51	0.16	4.60E-02	0.49	0.17
	97	4.20E-02	0.55	0.23	4.50E-02	0.53	0.23
	99	3.00E-02	0.54	0.23	3.00E-02	0.52	0.23
OM1	95	8.60E-02	0.35	0.14	5.00 E-02	0.32	0.08
	97	8.90E-02	0.36	0.15	6.40E-02	0.34	0.09
	99	8.80E-02	0.42	0.16	6.70E-02	0.4	0.11
SAR11	95	2.20E-02	0.28	0.04	1.00E-02	0.11	-0.002
	97	2.40E-02	0.32	0.06	1.30E-02	0.15	0.005
	99	4.00E-02	0.65	0.34	3.30E-02	0.49	0.16
SAR86	95	2.60E-02	0.21	0.04	3.60E-02	0.28	0.07
	97	2.20E-02	0.19	0.03	3.20E-02	0.26	0.07
	99	1.00E-02	0.19	0.05	2.00E-02	0.25	0.1

Los valores remarcados son aquellos con valores de probabilidad mayores a 0.05 ($p > 0.05$).

Las curvas de DD basadas en el índice de disimilitud de Morisita se observan en la Figura 3-5. En general en todos los LUPs se observaron patrones de decaimiento por distancia. Pero particularmente en *Amylibacter*, en todos los niveles de agrupamiento con datos ecualizados y no ecualizados, se observó que a medida que aumentó la distancia, las curvas de cada nivel de agrupamiento se invirtieron y entrecruzaron. Además, para el linaje SAR11 las curvas correspondientes a los agrupamientos al 95 y 97% con datos ecualizados y no ecualizados, presentaron pendiente cercanas a cero.

En la Tabla 3-7 se presentan los valores de los parámetros de regresión y correlación correspondientes a los gráficos de la Figura 3-5. En la primera columna se observan los valores de β , la pendiente. Estos valores aumentaron a medida que se aumentó el nivel taxonómico putativo, tendencia que se observó en los linajes SAR11, SAR86 y OM1. Particularmente para el linaje NS5 las pendientes presentaron una pequeña variación entre los niveles de agrupamiento con datos ecualizados y sin ecualizar. Tal como se observa en el análisis gráfico exploratorio, el linaje *Amylibacter* presentó un patrón atípico con respecto al resto de los LUPs en el cual se observó una tendencia a que los valores de β disminuyan conforme aumentó el nivel de similitud. A excepción del linaje NS5, en los restantes linajes la diferencia entre los valores de las curvas obtenidas a partir de datos ecualizados y no ecualizados fue muy pequeña.

En las siguientes columnas de la Tabla 3-7 se observan los valores de los parámetros de regresión y correlación. Los valores que no resultaron ser significativos se indican en negrita. Los valores de p se presentaron en dos rangos numéricos: unos muy bajos, cercanos al 0 (cero) en los linajes *Amylibacter*, SAR86 y SAR11 al agrupar a 95 y 97% todos con datos ecualizados y sin ecualizar. Estos bajos valores de p indican que no hay relación entre las variables. Por otro lado, los linajes NS5 donde los valores de p fueron mayores a 0.4; OM1 con valores superiores a 0.2 y SAR11 al agrupar a 99%, datos con y sin ecualizar, que presentó valores de alrededor de 0.73: Estos valores indicaron la existencia de una relación entre las variables, creciente y positiva.

Los valores de R^2 obtenidos fueron bajos en casi todos los linajes, niveles de agrupamiento, con datos ecualizados y sin ecualizar. Lo que implica que en aquellos linajes que, si sugieren una relación entre las variables, se ajusta pobremente al modelo lineal.

En la Tabla A- 04 del Anexo se encuentran los valores de correlación obtenidos con el coeficiente de Kendall, cuyos resultados concuerdan con los descritos anteriormente.

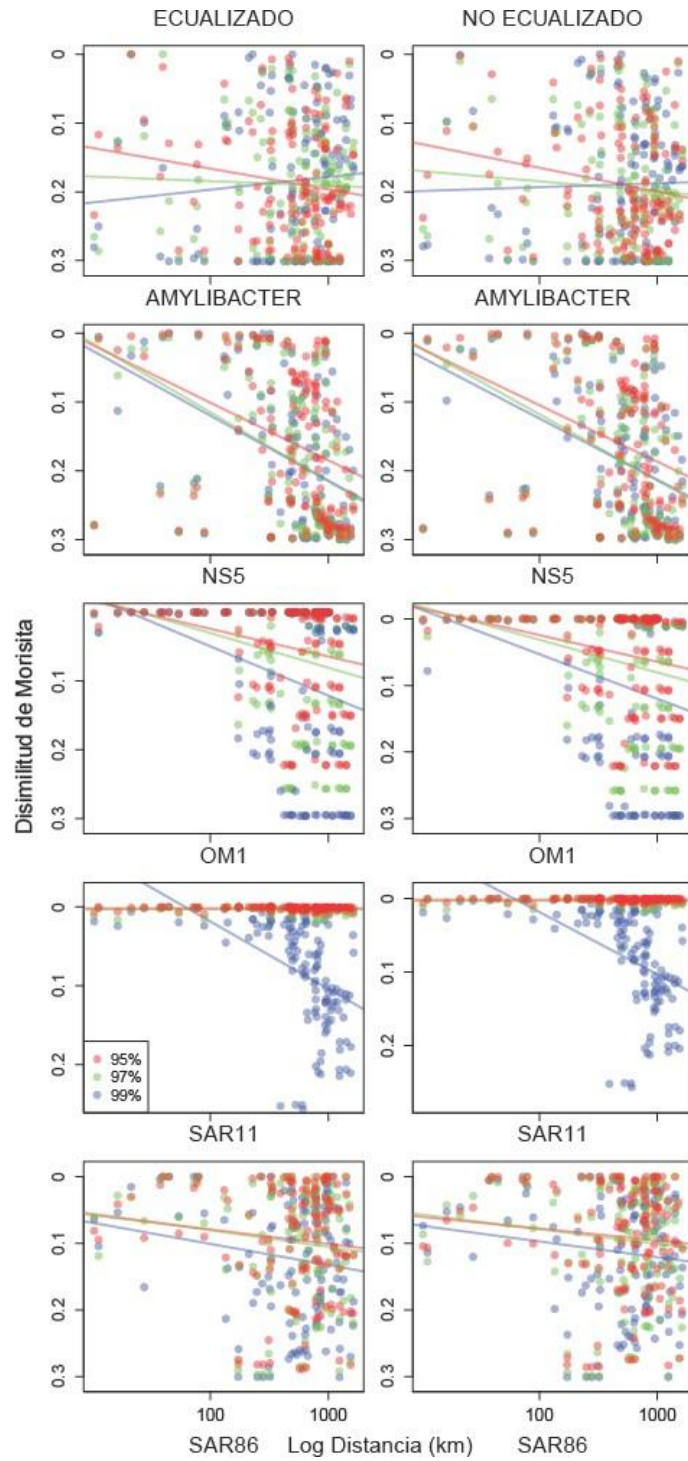


Figura 3-5: Curvas de DD basadas en el índice de disimilitud de Morisita. Discriminados por LUP. Cada color de líneas y puntos corresponde a un nivel de agrupamiento (95, 97 y 99%).

Tabla 3-7: Valores de los coeficientes de correlación (Rho ρ), regresión (R^2) y pendientes (β) del análisis DD basados en el índice de disimilitud de Morisita (ver también Figura 3-5).

Disimilitud de Morisita		Ecuilizado			No ecualizado		
		β	ρ	R^2	β	ρ	R^2
<i>Amylibacter</i>	95	3.00E-02	0.11	0.02	3.40E-02	0.17	0.03
	97	7.00E-03	0.02	-0.01	1.50E-02	0.08	-0.001
	99	-2.00E-02	-0.082	-0.001	-6.00E-03	0.0004	-0.007
NS5	95	8.40E-02	0.47	0.12	8.20E-02	0.46	0.12
	97	1.00E-01	0.49	0.18	9.60E-02	0.48	0.17
	99	9.00E-02	0.46	0.17	8.90E-02	0.45	0.15
OM1	95	4.00E-02	0.23	0.06	4.00E-02	0.22	0.05
	97	5.00E-02	0.21	0.06	4.90E-02	0.2	0.05
	99	7.00E-02	0.24	0.07	6.70E-02	0.21	0.06
SAR11	95	-1.00E-05	0.22	-0.007	3.00E-06	0.25	-0.01
	97	1.00E-04	0.28	-0.007	1.00E-04	0.31	-0.01
	99	9.00E-02	0.73	0.32	8.50E-02	0.73	0.31
SAR86	95	2,00E-02	0.078	0.006	1.80E-02	0.1	0.002
	97	3,00E-02	0.09	0.01	2.20E-02	0.12	0.01
	99	3.20E-02	0.08	0.02	2.40E-02	0.06	0.01

Los valores remarcados son aquellos con valores de probabilidad mayores a 0.05 ($p > 0.05$).

En la Figura 3-6 se pueden observar las curvas de DD basadas en el índice de disimilitud de Sorensen. Para la mayor parte de los linajes se pudo observar patrones de decaimiento con la distancia geográfica. Particularmente las curvas de los linajes *Amylibacter* en todos los niveles de agrupamiento y SAR11 al agrupar al 95 y 97%, con datos ecualizados y no ecualizados, presentaron pendientes con cambios sutiles, mientras que las pendientes correspondientes a los linajes NS5 y OM1 y SAR11 en el nivel de agrupamiento al 99% presentaron curvas con mayor pendiente en aquellas curvas generadas a partir de datos no ecualizados. Para el linaje SAR86 las pendientes fueron próximas a cero.

En la Tabla 3-8 se muestran los valores de las pendiente, regresión y correlación correspondientes a los datos de los gráficos de la Figura 3-6 (índice de Sorensen). Los valores de β tendieron a aumentar a medida que aumentó el nivel de ordenamiento. Los valores de β que se obtuvieron a partir de datos ecualizados fueron ligeramente mayores que los obtenidos a partir de datos no ecualizados. Sin embargo, en el linaje SAR86 se observó lo contrario, es decir que los valores de la pendiente disminuyeron al aumentar la similitud intra-OTU.

En cuanto a los valores de p , se observó que en el linaje *Amylibacter* al agrupar al 99% con datos ecualizado y en el linaje SAR86 en los niveles de agrupamiento del 95 y 97% obtenidos a partir de datos ecualizados y no ecualizados; fueron bajos y no-significativos (marcados en negrita). Mientras que, en el resto de los agrupamientos de los linajes antes mencionados, y en los linajes NS5, OM1 y SAR11 en todos los niveles de agrupamiento con datos ecualizados y no ecualizados; se observaron valores de p mayores a 0.25. Infiriendo la existencia de una relación creciente positiva entre las variables. Los valores de p aumentaron a medida que aumentó el nivel de similitud de los agrupamientos. Los valores R^2 fueron en su mayoría muy bajos, presentando valores más grandes las curvas provenientes de datos no ecualizados y sus valores aumentaron a medida que el nivel de agrupamiento creció. Estos valores evidenciaron que el modelo lineal no es el adecuado para explicar la variabilidad de los datos respuesta con respecto a la media.

En la Tabla A-05 del Anexo se encuentran los valores de correlación obtenidos con el coeficiente de Kendall, cuyos resultados concuerdan con los descriptos anteriormente.

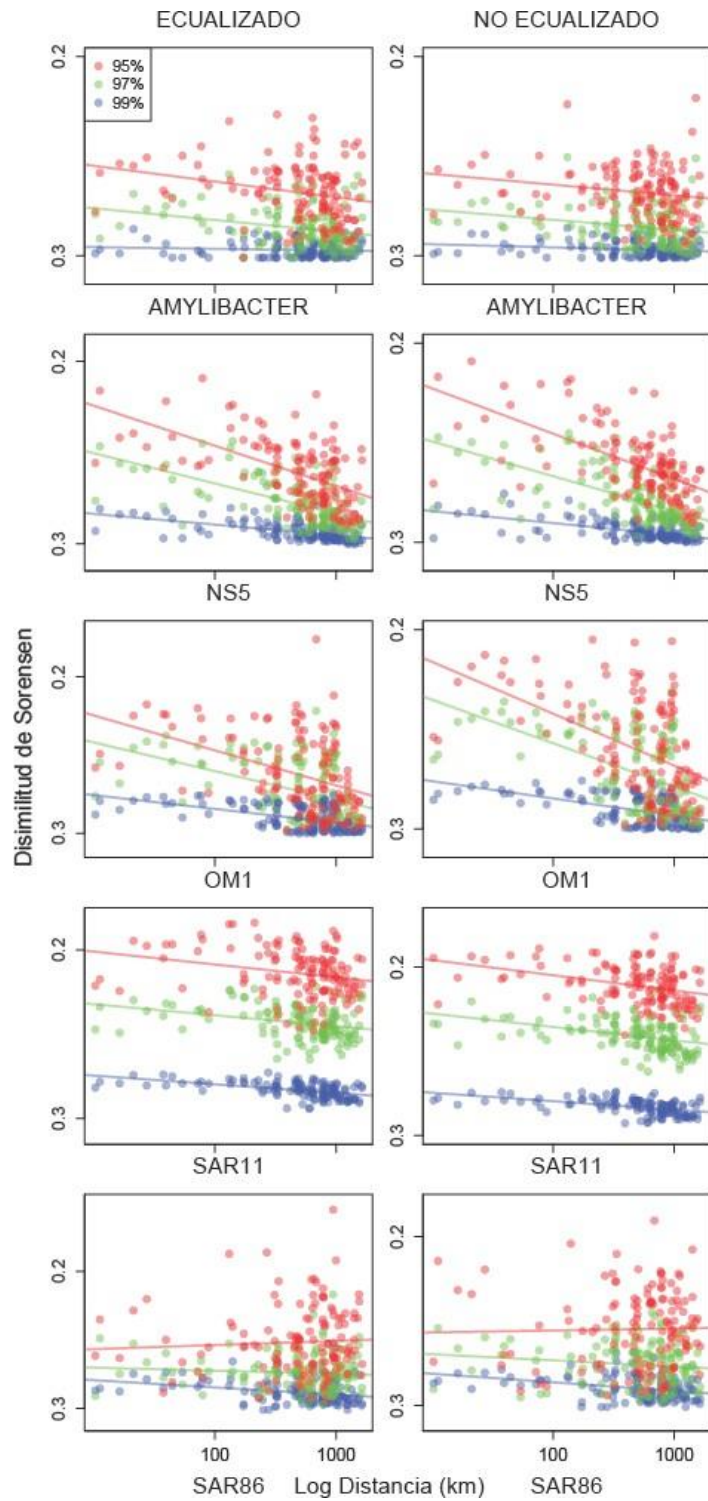


Figura 3-6: Curvas de DD basadas en el índice de disimilitud de Sorensen.

Discriminados por LUP. Cada color de líneas y puntos corresponde a un nivel de agrupamiento (95, 97 y 99%).

Tabla 3-8: Valores de los coeficientes de correlación (Rho ρ), regresión (R^2) y pendientes (β) del análisis de DD basados en el índice de disimilitud de Sorensen (ver también Figura 3-6).

Disimilitud de Sorensen		Ecuado			No ecualizado		
		β	ρ	R^2	β	ρ	R^2
Amylibacter	95	8.00E-03	0.31	0.05	5.40E-03	0.26	0.03
	97	6.00E-03	0.31	0.07	5.00E-03	0.3	0.01
	99	1.00E-03	0.07^(*)	0.002	2.00E-02	0.2	0.03
NS5	95	2.20E-02	0.58	0.32	2.30E-02	0.6	0.37
	97	1.70E-02	0.62	0.34	1.70E-02	0.67	0.39
	99	6.00E-03	0.57	0.31	6.00E-03	0.53	0.34
OM1	95	2.20E-02	0.45	0.16	2.60E-02	0.46	0.22
	97	1.80E-02	0.44	0.18	2.20E-02	0.52	0.25
	99	9.00E-03	0.48	0.24	9.00E-03	0.55	0.28
SAR11	95	8.00E-03	0.28	0.05	9.00E-03	0.39	0.12
	97	7.00E-03	0.36	0.09	8.00E-03	0.46	0.14
	99	5.00E-03	0.56	0.26	5.00E-03	0.57	0.25
SAR86	95	-3.00E-03	-0.08	-0.004	-1.00E-03	-0.06	-0.01
	97	2.00E-03	0.08	-0.002	4.00E-03	0.17	0.01
	99	5.00E-02	0.29	0.08	5.00E-03	0.34	0.1

(*) Los valores destacados en negrita se asociaron con valores p mayores que 0.05.

3.5 Escalado multidimensional no métrico.

En las Figuras 3-7, 3-8, 3-9 y 3-10 se presentan los ordenamientos multidimensionales no métricos (NMDS) de las muestras en base a diferentes índices de disimilitud.

En la Figura 3-7 se presentan los ordenamientos correspondientes al índice de disimilitud de Bray- Curtis. La primera columna, de izquierda a derecha, corresponde al linaje *Amylibacter*, al nivel de agrupación del 95%, con datos ecualizados, en los ordenamientos se observó que las muestras denominadas oceánicas estaban juntas y MDQ se encontró cerca de las muestras BOPD3 y BOPD4 a pesar de la distancia geográfica que las separa. Las muestras BOPD13 y BOPD14, se presentaron separadas en el ordenamiento a pesar de su cercanía geográfica. El resto de las muestras se encontraron agrupadas en forma consistente con su cercanía geográfica. Al analizar los agrupamientos al 95% sin ecualización y al 97% con ecualización se observó que las muestras en general se agruparon en forma consistente con las distancias geográficas que las separa. Al agrupar a 97% con datos sin ecualizar se observó que las muestras de Río Gallegos, Bahía Engaño, Río Negro, Río Colorado, Puerto Deseado y las denominadas Oceánicas ubicaron cerca unas de otras en el ordenamiento a pesar de la distancia que las separa. Al agrupar al 99%, tanto con datos ecualizados como con datos no ecualizados, los ordenamientos presentaron un patrón similar al observado al agrupar a 97% con datos sin ecualización. Los valores de *stress* que se obtuvieron para los ordenamientos de *Amylibacter* fueron menores a 0.2.

En la columna 2 de la Figura 3-7 se observan los ordenamientos obtenidos con respecto al linaje NS5. Al agrupar a 95% de similitud, con datos ecualizados, las muestras tendieron a ordenarse según la distancia geográfica que las separa. La muestra CHR se encontró alejada de las muestras con las que comparte cercanía física. A 95% similitud, datos sin ecualización, se observó que las muestras se ordenaron según la distancia geográfica e incluso las muestras oceánicas se encontraron separadas. Al agrupar a 97 y 99%, los ordenamientos coincidieron con la distancia geográfica que los separa, salvo en el caso de las muestras oceánicas, que a pesar de ser distantes quedaron cercanas en el ordenamiento. Los valores de *stress* que se obtuvieron para los ordenamientos fueron menores 0.2.

Para el linaje OM1 se observó, en todos los niveles de agrupamiento con datos ecualizados y no ecualizados, que las muestras Río Gallegos, Bahía Engaño y Puerto Deseado siempre se encontraron juntas, salvo al agrupar 99% sin ecualización, donde la muestra CHR se separó de las demás. Las muestras correspondientes a Río Negro, Río Colorado y las oceánicas se ordenaron coincidentemente con la distancia geográfica que

las separa. Los valores de *stress* que se obtuvieron para los ordenamientos de OM1 en todos los niveles fueron menores a 0.2.

En la fila número 4 de la Figura 3-7 se observan los ordenamientos correspondientes al linaje SAR11. Al agrupar a 95% y 97%, con datos ecualizados y sin ecualización, el componente que permitió una mayor discriminación entre las muestras es NMDS2, el cual puede observarse en el eje de las ordenadas en la figura. Las diferencias más importantes se observaron para las muestras que pertenecen a Bahía Engaño y Río Gallegos. A partir de los agrupamientos de 99%, con datos ecualizados o no, la componente NMDS1 resultó relevante, y las muestras se agruparon según su proximidad geográfica. Los valores de *stress* obtenidos para los ordenamientos fueron menores a 0.2.

Los ordenamientos del linaje SAR86 se observan en la última columna de izquierda a derecha de la Figura 3-7. En todos los niveles de agrupamiento (95%, 97% y 99%) con datos ecualizados o no ecualizados los ordenamientos observados fueron iguales; las muestras oceánicas y las de Río Negro se presentaron separadas de las muestras Bahía Engaño, Río Gallegos, Puerto Deseado y Río Colorado, las cuales se ordenaron cercanamente entre ellas, a pesar de la distancia geográfica que las separa. Cabe mencionar que las muestras que permanecieron juntas corresponden a muestras más cercanas a la costa en comparación a las primeras mencionadas. Los valores de *stress* de los ordenamientos de SAR86 fueron menores a 0.2.

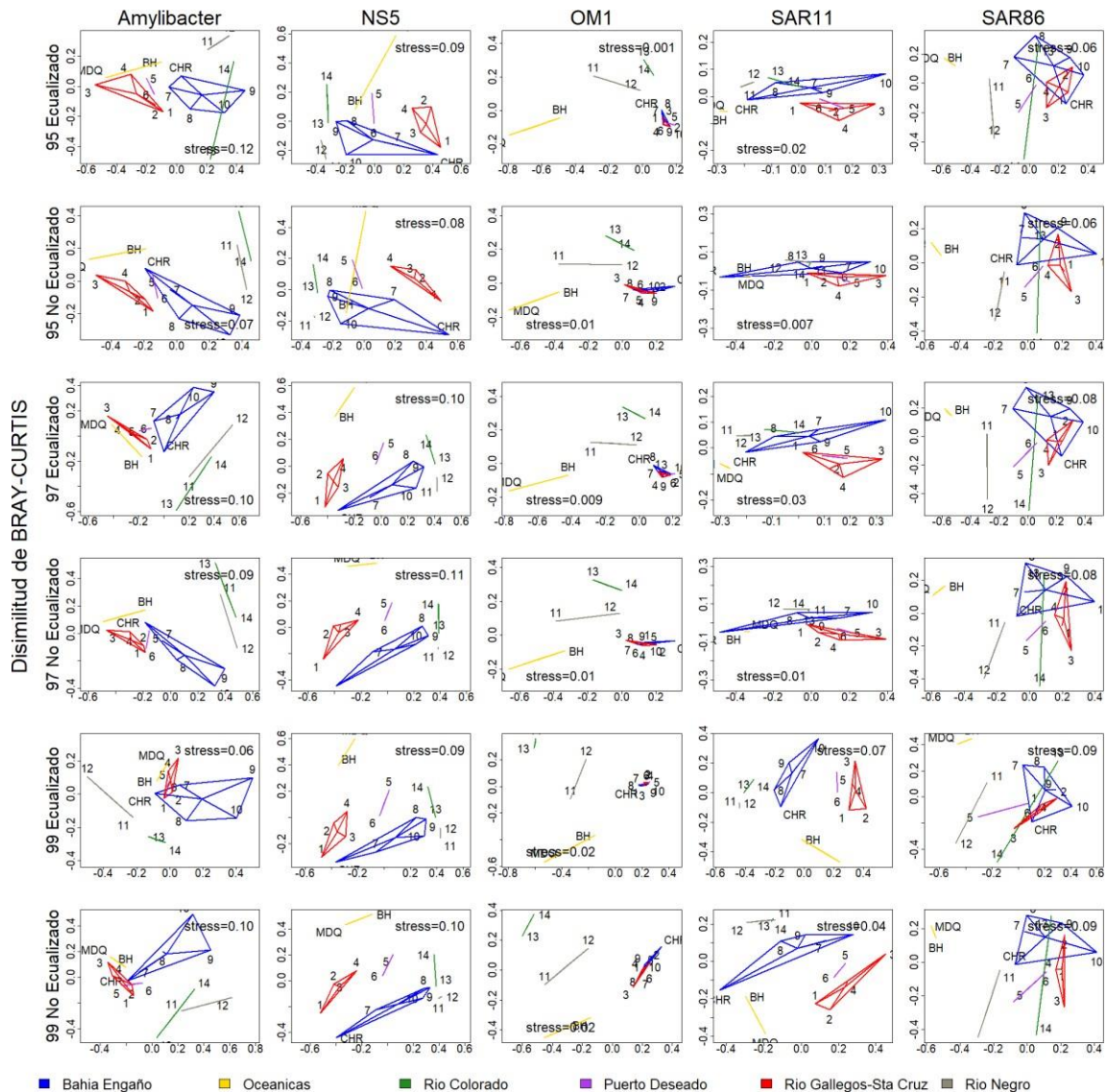


Figura 3-7: Ordenamientos obtenidos mediante Escalado Multidimensional no Métrico basado en el índice de disimilitud de Bray-Curtis.

Discriminado por linaje y según se haya aplicado o no equalización. Los números (1-14) en los ordenamientos corresponden a las muestras BOPD 1-14. Los polígonos de color se agregaron manualmente para indicar proximidad entre las muestras abarcadas por cada uno de ellos (ver también Figura 2-1).

En la Figura 3-8 se observan los ordenamientos basados en el índice de disimilitud de Sorensen. Para el linaje *Amylibacter*, en la primera columna, las muestras se ordenaron en concordancia con la distancia geográfica que los separa. Sin embargo, al agrupar a 95% con datos sin ecualizar y 97% de similitud con y sin ecualización, se observó que las muestras de Bahía Engaño tendieron a presentarse de manera separada en los ordenamientos a pesar de que son muestras cercanas geográficamente. Los valores de *stress* obtenidos para los ordenamientos antes mencionados fueron valores menores a 0.2.

En columna 2 se observan los ordenamientos del linaje NS5. Al agrupar 95%, 97% y 99% con datos ecualizados y no ecualizados, los ordenamientos mostraron una correspondencia conspicua con las distancias geográficas que separan las muestras. Los valores de *stress* para linaje NS5 fueron inferiores a 0.2.

En los ordenamientos correspondientes al linaje OM1; las muestras se corresponden con la distancia geográfica. Particularmente las muestras correspondientes a Bahía Engaño, Río Gallegos y Puerto Deseado pueden observarse en forma cercana, a pesar de la distancia geográfica que las separa, generando un bloque de muestras australes. Estas observaciones se presentan cuando se agrupa en 95%, 97% y 99%, tanto con datos ecualizados, como sin ecualizar. En este sentido, se observa que, a medida que el nivel de similitud aumentaba, las muestras más oceánicas tendieron a separarse entre ellas. Los valores de *stress* obtenidos para los ordenamientos fueron valores menores a 0.2.

En la columna 4 de la Figura 3-8 se observan los ordenamientos pertenecientes al linaje SAR11. Al agrupar al 95 y 97% con datos sin ecualización los ordenamientos concordaron con la distancia geográfica que los separa. En los agrupamientos a 95% y 97% con datos ecualizados la muestra CHR se separó de aquellas que están geográficamente cerca. En los ordenamientos de 99% similitud con datos ecualizados y sin ecualizar las muestras oceánicas se presentaron cercanas a las muestras de Río Gallegos, a pesar de la distancia geográfica que las separa. Los valores de *stress* obtenidos para los ordenamientos fueron menores a 0.2.

En la última columna se presentan los ordenamientos correspondientes al linaje SAR86. En los agrupamientos a 95% y 97% de similitud, con datos ecualizados y sin ecualizar, los ordenamientos no se correspondieron con la distancia geográfica, en los ordenamientos de 99% similitud se observó que los ordenamientos fueron de acuerdo con la distancia geográfica. Los valores de *stress* obtenidos fueron inferiores a 0.2.

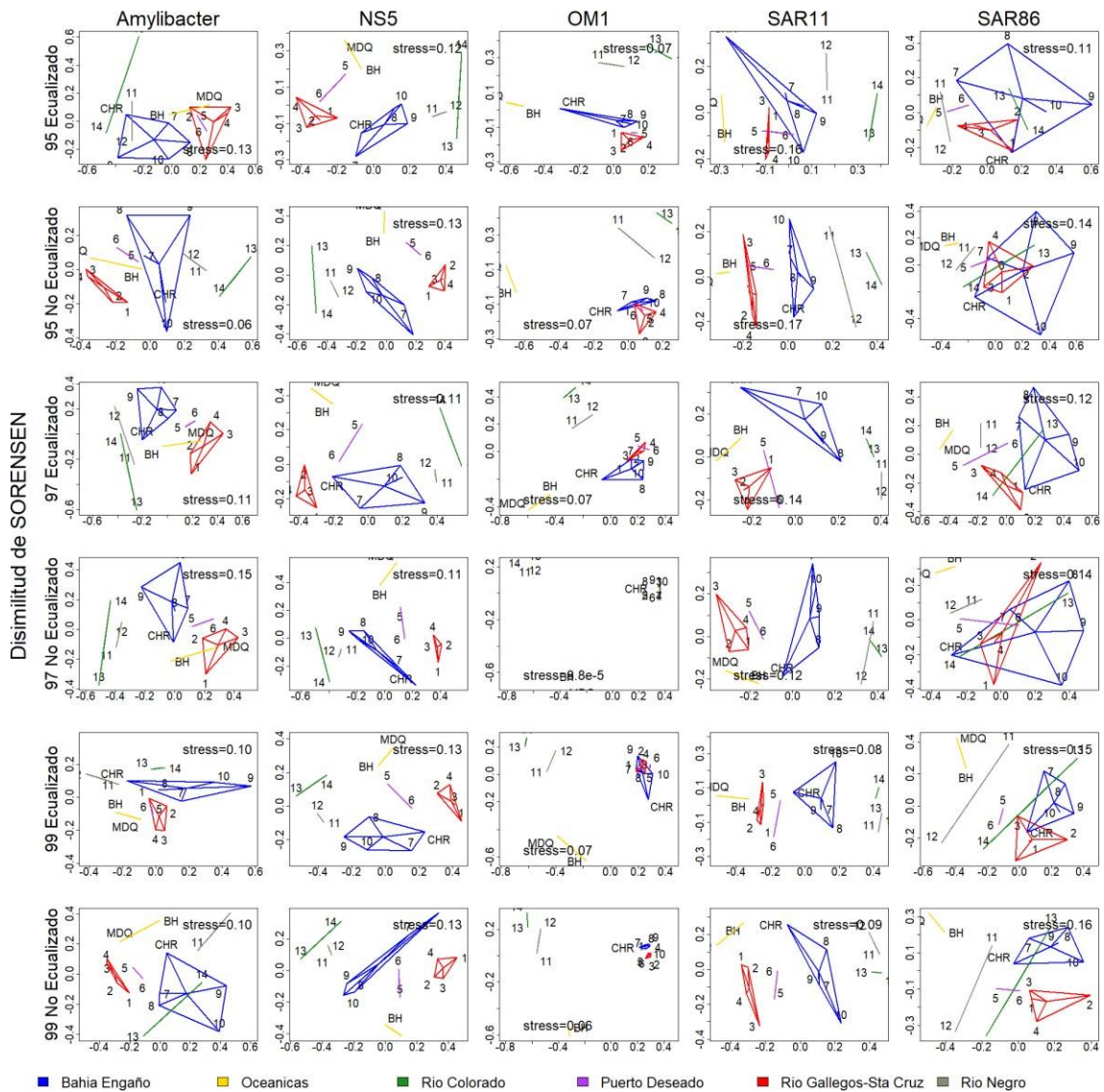


Figura 3-8: Ordenamientos obtenidos mediante Escalado Multidimensional no Métrico basado en el índice de disimilitud de Sorensen.

Discriminado por linaje y según se haya aplicado o no ecuialización. Los números (1-14) en los ordenamientos corresponden a las muestras BOPD 1-14. Los polígonos de color se agregaron manualmente para indicar proximidad entre las muestras abarcadas por cada uno de ellos (ver también Figura 2-1).

En la Figura 3-9 se observan los ordenamientos basados en el índice de disimilitud de Morisita. Los ordenamientos de la primera columna pertenecen al linaje *Amylibacter*; se observó al agrupar al 95% con datos ecualizados y sin ecualizar que las muestras se ordenaron según su distancia geográfica, excepto en las muestras de Río Colorado que tendieron a separarse a pesar de su cercanía geográfica. En los agrupamientos al 97% y 99% con datos ecualizados y no ecualizados las muestras también se ordenaron según la distancia geográfica, salvo las muestras de Bahía Engaño y Río Negro, las cuales se no se agruparon en el ordenamiento, pesar de su cercanía geográfica. El valor de *stress* que se obtuvo para los ordenamientos fue inferior a 0.2.

En la siguiente columna de la Figura 3-9 se observan los ordenamientos del linaje NS5. En todos los niveles de agrupamiento con datos ecualizados y sin ecualizar se presentó un patrón repetido, el cual constó en ordenamientos que se correspondieron con la distancia geográfica. Observamos sin embargo que particularmente las muestras correspondientes a la cercanías de Bahía engaño no se agrupan según la distancia geográfica y estas se separaron en 2 grupos; por un lado, BOPD7 y CHR las cuales se ordenaron cerca de las muestras de Río Gallegos, Puerto Deseado y por otra lado las muestras BOPD8, BOPD9 y BOPD10 que se presentaron cercanas a las muestras de Río Negro y Río Colorado, ordenándose las muestras de Bahía Engaño cerca de muestras con las que no comparten cercanía geográfica. Los valores de *stress* obtenidos de los ordenamientos de NS5 fueron menores a 0.2.

En la tercera columna se observan los ordenamientos pertenecientes al linaje OM1. Los ordenamientos concordaron con la distancia geográfica, particularmente las muestras Río Gallegos, Puerto Deseado y Bahía Engaño formaron un bloque, ordenándose juntas a pesar de la distancia geográfica que las separa, cabe mencionar que las muestras son aquellas de distribución más austral. También se observó que conforme se aumentó el nivel de similitud de agrupamiento las muestras oceánicas tendieron a separarse entre sí; tal como la distancia geográfica que las separa. Los valores de *stress* obtenidos para estos ordenamientos fueron muy pequeños, cercanos a cero, por ende, menores a 0.2.

En la columna cuatro de la Figura 3-9 se presentan los ordenamientos correspondientes al linaje SAR11, en estos las muestras se ordenaron siguiendo la distancia geográfica que las separa, patrón que se repitió en todos los niveles de agrupamiento con datos ecualizados y sin ecualizar. Particularmente en los ordenamientos de los agrupamientos de 95 y 97% con datos ecualizados y sin ecualizar se observó que la muestra CHR tendió a separarse de las muestras del grupo de Bahía Engaño, con las cuales comparte proximidad geográfica. El valor de *stress* de los ordenamientos fue menor a 0.2.

En la última columna (Figura 3-9) los ordenamientos pertenecen al linaje SAR86. Al agrupar a 95 y 97% con datos ecualizados y sin ecualizar los ordenamientos no se presentaron de acuerdo con la distancia geográfica, los grupos de muestras se encontraron cercanos entre si a pesar de la distancia geográfica que los separa y en algunos casos (Río Negro y Río Colorado) se presentaron distantes a pesar de su cercanía geográfica. En los ordenamientos de los agrupamientos de 99% (ecualizado y no ecualizado) se observó el mismo patrón antes mencionado, particularmente en estos las muestras oceánicas se encontraron separadas de las demás. Los valores de *stress* obtenidos para los ordenamientos fueron menores a 0.2.

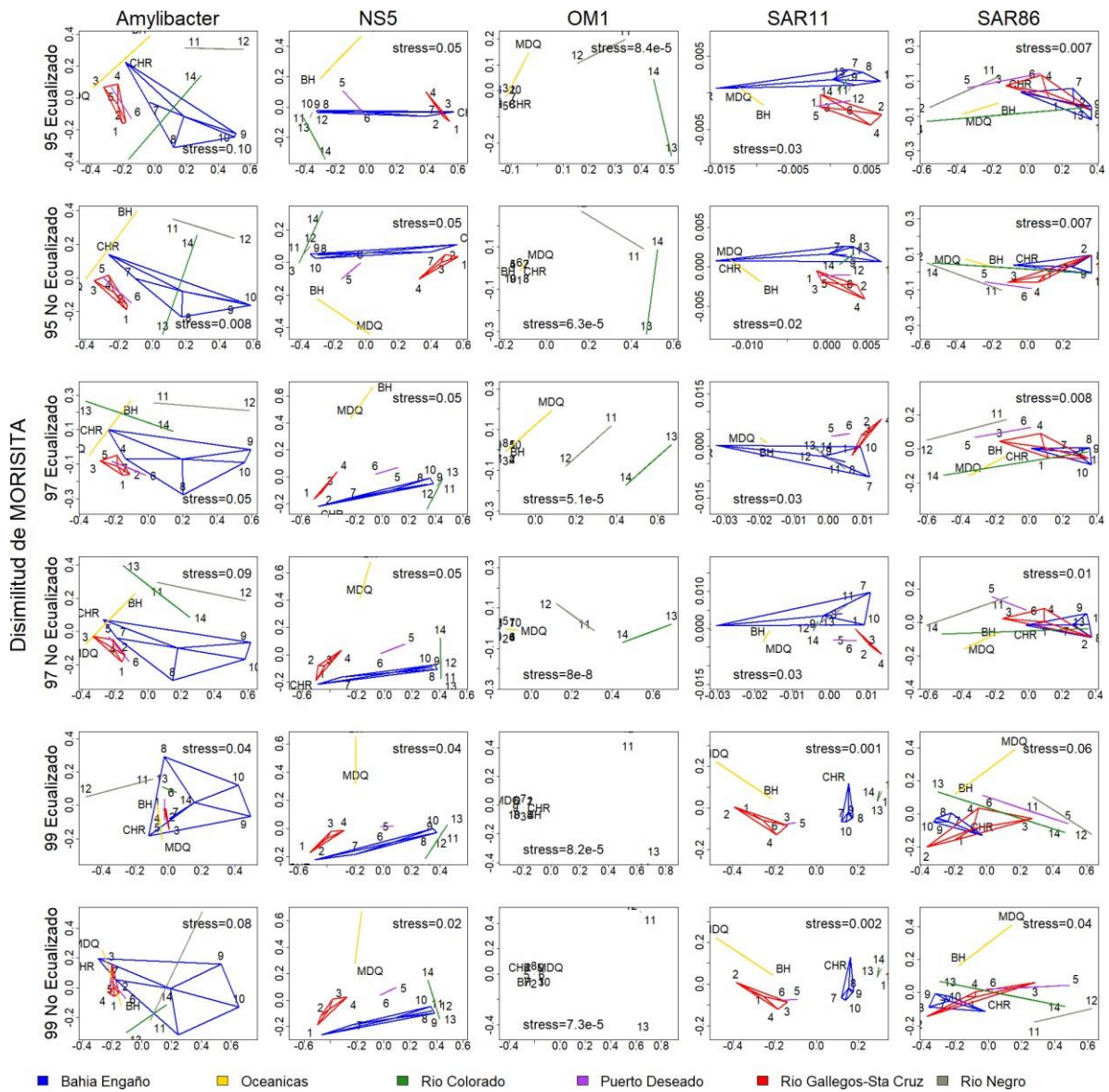


Figura 3-9: Ordenamientos obtenidos mediante Escalado Multidimensional no Métrico basado en el índice de disimilitud de Morisita.

Discriminado por linaje y según se haya aplicado o no ebalización. Los números (1-14) en los ordenamientos corresponden a las muestras BOPD 1-14. Los polígonos de color se agregaron manualmente para indicar proximidad entre las muestras abarcadas por cada uno de ellos (ver también Figura 2-1).

Por último, en la Figura 3-10 se presentan los ordenamientos que corresponden al índice de disimilitud de Horn. En la primera columna se presentan los ordenamientos correspondientes al linaje *Amylibacter*. Al agrupar 95% con datos ecualizados y sin ecualizar, las muestras de Río Colorado se ordenaron distanciadas entre sí a pesar de su cercanía geográfica, lo mismo ocurrió con la muestra CHR de Bahía Engaño y la muestra MDQ la cual se ordenó cerca de las muestras de Río Gallegos y Puerto Deseado a pesar de la distancia geográfica que las separa. En los ordenamientos de los agrupamientos al 97% ecualizado y no ecualizado, las muestras de Bahía Engaño se ordenaron muy separadas entre sí y el resto de las muestras se ordenaron de acuerdo con la distancia geográfica. Al agrupar a 99% con datos ecualizados y sin ecualizar se observó que las muestras de Bahía Engaño se ordenaron muy dispersas entre sí y el resto de las muestras se ordenaron juntas a pesar de la distancia geográfica que las separa. Los valores de *stress* de los ordenamientos fueron menores a 0.2.

En la segunda columna se muestran los ordenamientos del linaje NS5, donde en todos los niveles de agrupamiento (95, 97, 99%), con datos ecualizados y no ecualizados las muestras se ordenaron de acuerdo con la distancia geográfica que las separa. Los valores de *stress* para los ordenamientos fueron menores a 0.2.

En la tercera columna se encuentran los ordenamientos que corresponden al linaje OM1. Las muestras se ordenaron por la distancia geográfica que las separa, particularmente las muestras más australes (Río Gallegos, Puerto Deseado y Bahía Engaño) quedaron juntas. Esto se repitió en los diferentes niveles de agrupamiento (95, 97 y 99%) con datos ecualizados y sin ecualizar. Los valores de *stress* obtenidos para los ordenamientos fueron pequeños, cercanos a cero y menores a 0.2.

En la cuarta columna de la Figura 3-10 se observan los ordenamientos correspondientes al linaje SAR11. Los ordenamientos se correspondieron con la distancia geográfica que los separa, particularmente la muestra CHR del grupo Bahía Engaño, siempre se presentó alejada del resto de las muestras del grupo. Estos patrones se presentaron en todos los niveles de agrupamiento 95, 97 y 99% con datos ecualizados y no ecualizados. Los valores de *stress* obtenidos para los ordenamientos fueron menores a 0.2.

Para concluir en la última columna de la Figura 3-10, se encuentran los ordenamientos correspondientes al linaje SAR86. Se observó al agrupar a 95 y 97% con datos ecualizados y no ecualizados que las muestras de Río Gallegos y Bahía Engaño se ordenaron cercanamente a pesar de estar geográficamente distantes, mientras que las muestras de Río Negro y Río Colorado se ordenaron separadas pese a estar geográficamente cerca una de otra. En los agrupamientos a 99% con y sin ecualizar las muestras se ordenaron correspondiéndose con la distancia geográfica; las muestras

oceánicas se ordenaron bien alejadas del resto de las muestras y separadas entre sí. Los valores de *stress* para los ordenamientos fueron menores a 0.2.

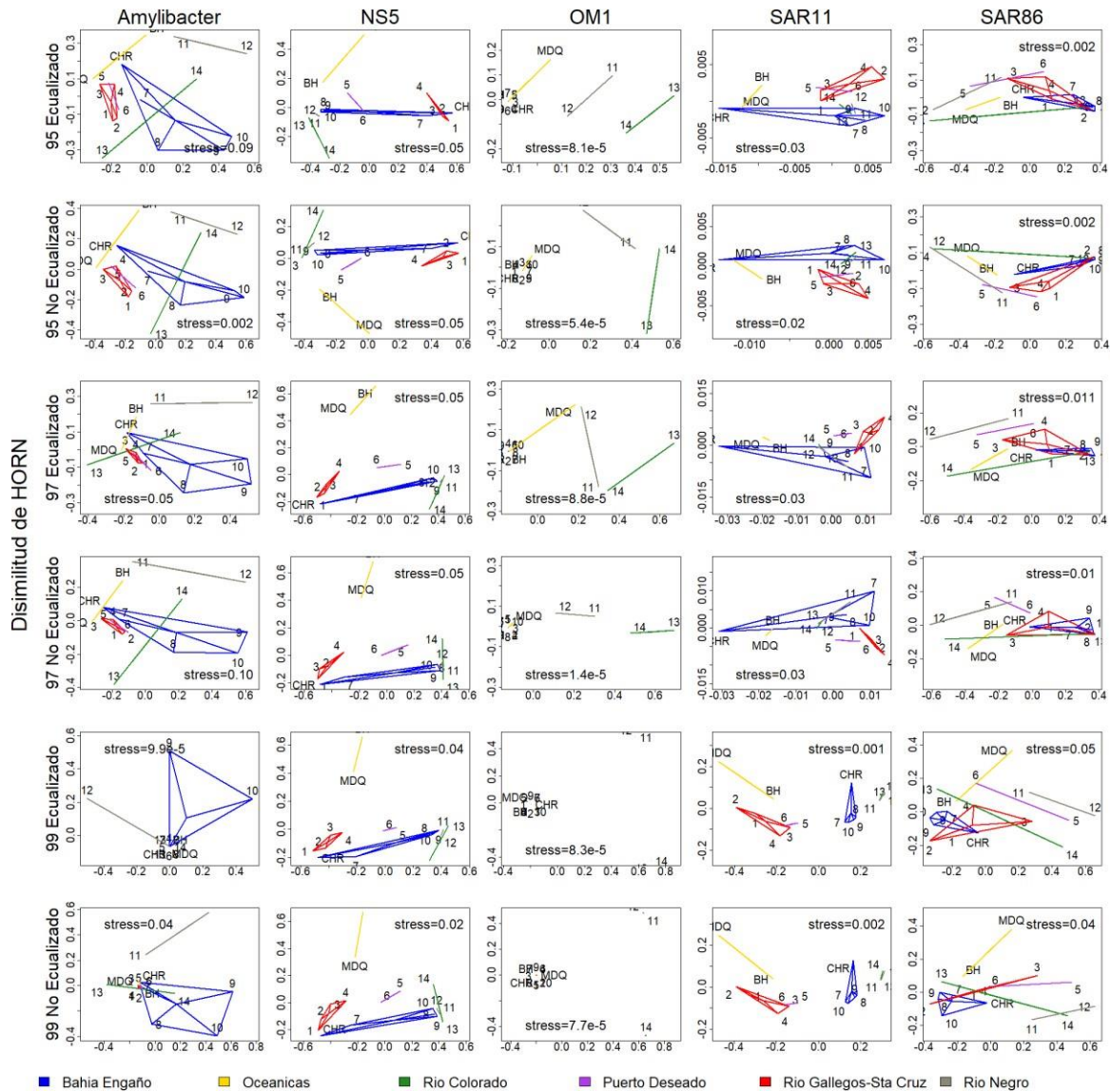


Figura 3-10: Ordenamientos obtenidos mediante Escalado Multidimensional no Métrico basado en el índice de disimilitud de Horn.

Discriminado por linaje y según se haya aplicado o no eculización. Los números (1-14) en los ordenamientos corresponden a las muestras BOPD 1-14. Los polígonos de color se agregaron manualmente para indicar proximidad entre las muestras abarcadas por cada uno de ellos (ver también Figura 2-1).

3.6 Patrones de endemismo en LUPs del Mar Argentino

3.6.1 Patrones de endemismo por linaje.

En la Tabla 3-9 se indica la cantidad de OTUs endémicas, es decir OTUs que sólo se detectaron en 1 muestra, identificadas para cada LUP en los distintos niveles taxonómicos putativos, a partir de datos ecualizados y no ecualizados. Todos los LUPs presentaron OTUs endémicas, y a medida que se aumentó nivel de agrupamiento se observó un mayor grado de endemismo. Tal como se esperaba, el número de OTUs endémicas fue mayor cuando se utilizaron datos no-ecualizados, lo cual es atribuible a la mayor cantidad de secuencias. El linaje que presentó mayor número de OTUs endémicas en los niveles de agrupamiento de 95 y 97% con datos ecualizados y no ecualizados fue OM1, mientras que en el nivel de agrupamiento de 99% con datos ecualizados y no ecualizados el mayor grado de endemismo fue observado en el clado SAR11.

Tabla 3-9: Número de OTUs endémicas de los LUPs, discriminados por nivel de agrupamiento; ecualizados y no ecualizados.

	Número de OTUs Endémicas					
	Ecuallizado			No Ecuallizado		
	95	97	99	95	97	99
Amylibacter	587	1268	3018	782	1772	4602
NS5	408	1070	3178	622	1716	5538
OM1	767	1803	7858	1111	2727	12591
SAR11	601	1771	12514	867	2658	21461
SAR86	148	357	1506	238	590	2780

3.6.2 Patrones de endemismo por muestra.

En la Tabla 3-10 se puede observar la distribución geográfica de las OTUs endémicas del género *Amylibacter*. La mayor cantidad de OTUs endémicas se observó en las muestras más australes, detectándose una aparente disminución en el endemismo en dirección sur a norte, siendo la muestra BOPD14, la muestra más boreal y costera, la que presentó el menor número de OTUs endémicas en todos los niveles de similitud, tanto en el caso de datos ecualizados y como en el caso de datos sin ecualizar. En la Tabla 3-10 también se proporcionan valores de cobertura, los cuales no mostraron relación con la cantidad de OTUs endémicas detectadas.

Tabla 3-10: Distribución de las OTUs endémicas del linaje *Amylibacter*

Muestras	<i>Amylibacter</i>											
	Ecuallizado						No Ecuallizado					
	E 95	C 95	E 97	C 97	E 99	C 99	E 95	C 95	E 97	C 97	E 99	C 99
BH	29	0.84	98	0.72	256	0.28	60	0.9	211	0.76	621	0.36
BOPD.1	19	0.65	38	0.54	83	0.2	28	0.7	62	0.54	145	0.24
BOPD.10	23	0.58	47	0.27	77	0.09	27	0.63	60	0.36	103	0.12
BOPD.11	18	0.5	30	0.36	50	0.08	16	0.5	27	0.35	50	0.1
BOPD.12	16	0.33	35	0.81	47	0.04	29	0.45	55	0.22	84	0.03
BOPD.13	10	0.48	17	0.37	23	0.2	13	0.42	20	0.37	26	0.17
BOPD.14	5	0.6	11	0.4	20	0.16	12	0.53	23	0.33	35	0.09
BOPD.2	29	0.72	54	0.6	133	0.28	54	0.75	105	0.61	231	0.31
BOPD.3	149	0.91	352	0.8	922	0.54	128	0.9	321	0.79	887	0.54
BOPD.4	88	0.79	177	0.61	386	0.3	116	0.79	234	0.64	544	0.34
BOPD.5	28	0.76	62	0.63	147	0.39	30	0.78	74	0.63	174	0.39
BOPD.6	51	0.71	97	0.53	193	0.26	53	0.72	104	0.57	242	0.26
BOPD.7	29	0.76	62	0.6	140	0.34	37	0.76	75	0.65	176	0.37
BOPD.8	33	0.67	64	0.45	147	0.06	50	0.72	99	0.55	234	0.14
BOPD.9	14	0.49	21	0.34	41	0.04	14	0.6	35	0.39	71	0.03
MDQ	38	0.93	89	0.88	321	0.71	87	0.95	208	0.9	827	0.71
CHR	8	0.82	14	0.7	32	0.57	28	0.88	59	0.79	152	0.62

E = OTUs endémicas presentes en la muestra; C= cobertura de la muestra

En la Tabla 3-11 se observa la distribución de las OTUs endémicas del linaje NS5 en todos los niveles de agrupamiento analizados, así como las correspondientes coberturas. Se muestran los resultados obtenidos a partir de datos ecualizados y a partir de conteos directos, sin ecualizar. No se observaron tendencias geográficas como se observó en el caso del linaje *Amylibacter*. La muestra que presentó el menor nivel de endemismo fue MDQ y las que presentaron los mayores números de OTUs endémicas fueron BOPD7 y CHR, las cuales se encuentran relativamente cercanas entre sí en la zona de Bahía Engaño (ver Figura 2-1 de la sección Materiales y Métodos). Los valores de cobertura logrados para cada muestra no se relacionaron con los correspondientes números de OTUs endémicas.

Tabla 3-11: Distribución de las OTUs endémicas del linaje NS5

Muestras	NS5											
	Ecualizado						No Ecualizado					
	E 95	C 95	E 97	C 97	E 99	C 99	E 95	C 95	E 97	C 97	E 99	C 99
BH	10	0.93	27	0.89	80	0.73	21	0.95	52	0.93	190	0.76
BOPD.1	25	0.95	76	0.9	350	0.67	40	0.95	128	0.91	647	0.67
BOPD.10	31	0.91	97	0.82	287	0.6	38	0.92	128	0.83	397	0.62
BOPD.11	32	0.92	94	0.84	305	0.58	34	0.91	91	0.83	324	0.58
BOPD.12	43	0.86	90	0.76	235	0.53	63	0.88	160	0.79	444	0.53
BOPD.13	5	0.91	23	0.84	71	0.63	4	0.93	27	0.86	98	0.65
BOPD.14	20	0.83	51	0.65	114	0.31	39	0.78	101	0.64	219	0.3
BOPD.2	12	0.86	33	0.75	96	0.53	28	0.88	60	0.79	188	0.54
BOPD.3	32	0.89	86	0.8	292	0.51	26	0.88	83	0.8	284	0.49
BOPD.4	36	0.79	85	0.65	188	0.33	47	0.81	117	0.67	273	0.37
BOPD.5	21	0.78	48	0.63	107	0.33	25	0.79	60	0.66	144	0.35
BOPD.6	19	0.89	57	0.73	165	0.47	22	0.89	74	0.74	223	0.48
BOPD.7	62	0.87	155	0.76	434	0.43	61	0.88	184	0.76	520	0.46
BOPD.8	18	0.85	41	0.71	101	0.48	24	0.87	66	0.77	175	0.52
BOPD.9	12	0.85	31	0.78	92	0.56	26	0.87	54	0.79	145	0.6
MDQ	3	0.78	8	0.75	14	0.56	3	0.86	11	0.82	29	0.61
CHR	27	0.95	68	0.91	247	0.74	121	0.97	320	0.93	1238	0.77

E = OTUs endémicas presentes en la muestra; C= cobertura de la muestra

En la Tabla 3-12, se presenta la distribución de las OTUs endémicas identificadas en el linaje OM1 y las correspondientes coberturas. Las muestras con menor número de OTUs endémicas fueron MDQ y BH, es decir las muestras que fueron colectadas más lejos de la costa. En las muestras restantes se observó el número de OTUs endémicas fue notablemente mayor en las muestras más boreales (BOPD13 y BOPD14). Los valores de cobertura no presentaron relación con el número de OTUs endémicas detectadas. Incluso se puede observar (tabla 3-12) que muestras como BH y MDQ presentaron grandes valores de cobertura y el número de OTUs endémicas encontrado fue relativamente bajo.

Tabla 3-12: Distribución de las OTUs endémicas del linaje OM1.

Muestras	OM1											
	Ecualizado						No Ecualizado					
	E 95	C 95	E 97	C 97	E 99	C 99	E 95	C 95	E 97	C 97	E 99	C 99
BH	2	0.97	2	0.93	23	0.8	5	0.98	6	0.98	48	0.84
BOPD.1	21	0.98	61	0.97	413	0.87	48	0.98	117	0.97	760	0.88
BOPD.10	41	0.98	92	0.96	455	0.89	59	0.98	132	0.96	620	0.9
BOPD.11	14	0.92	47	0.86	172	0.66	10	0.93	37	0.86	172	0.66
BOPD.12	56	0.95	146	0.91	555	0.74	91	0.96	243	0.92	1002	0.76
BOPD.13	103	0.95	216	0.91	837	0.72	121	0.95	263	0.92	1071	0.73
BOPD.14	161	0.96	377	0.91	1255	0.72	242	0.96	593	0.93	2183	0.74
BOPD.2	48	0.98	117	0.96	603	0.86	78	0.98	218	0.96	1214	0.86
BOPD.3	40	0.98	73	0.96	414	0.86	33	0.98	69	0.96	394	0.76
BOPD.4	82	0.97	167	0.94	629	0.83	92	0.97	216	0.95	930	0.84
BOPD.5	38	0.98	97	0.96	559	0.88	46	0.98	117	0.97	703	0.88
BOPD.6	42	0.98	100	0.96	524	0.87	48	0.98	127	0.96	681	0.87
BOPD.7	27	0.98	71	0.96	355	0.89	31	0.98	82	0.96	431	0.89
BOPD.8	47	0.97	110	0.94	383	0.85	84	0.97	181	0.95	671	0.86
BOPD.9	43	0.98	114	0.96	527	0.87	70	0.98	191	0.96	870	0.88
MDQ	0	1,00	0	0.96	0	0.82	1	0.95	3	0.9	6	0.83
CHR	2	0.99	13	0.98	154	0.93	52	0.99	132	0.98	835	0.94

E = OTUs endémicas presentes en la muestra; C= cobertura de la muestra

En la Tabla 3-13 se presentan la distribución de las OTUs endémicas correspondientes al linaje SAR11, el cual presentó el mayor grado de endemismo y las mayores coberturas en todos los niveles de agrupamiento. Las muestras más oceánicas presentaron el mayor grado de endemismo a 99% similitud, pero no en los niveles del 95 y 97%. En el resto de las muestras se observó una tendencia al aumento del número de OTUs endémicas en sentido sur a norte (muestras desde BOPD1 a BOPD14; Figura 2-1 de la sección Materiales y Métodos). Al igual que en los otros linajes, con la salvedad del caso de SAR11 no se observó relación entre la cobertura y el número de OTUs endémicas identificadas.

Tabla 3-13: Distribución de las OTUs endémicas del linaje SAR11

Muestras	SAR11											
	Ecualizado						No Ecualizado					
	E 95	C 95	E 97	C 97	E 99	C 99	E 95	C 95	E 97	C 97	E 99	C 99
BH	21	0.99	75	0.97	945	0.8	44	0.99	191	0.97	2357	0.82
BOPD.1	24	0.97	94	0.93	791	0.69	45	0.97	170	0.94	1420	0.71
BOPD.10	20	0.95	66	0.9	391	0.65	27	0.96	81	0.9	519	0.66
BOPD.11	49	0.97	165	0.94	1210	0.74	50	0.97	147	0.94	1232	0.74
BOPD.12	66	0.97	172	0.93	1016	0.74	92	0.97	273	0.94	1799	0.76
BOPD.13	69	0.97	151	0.94	893	0.74	81	0.97	201	0.94	1185	0.75
BOPD.14	52	0.96	122	0.93	659	0.74	85	0.97	204	0.93	1170	0.76
BOPD.2	40	0.96	115	0.91	671	0.67	73	0.96	186	0.93	1243	0.69
BOPD.3	21	0.95	64	0.91	420	0.66	18	0.95	57	0.9	382	0.67
BOPD.4	45	0.95	118	0.89	708	0.62	48	0.96	145	0.91	977	0.65
BOPD.5	27	0.96	85	0.92	529	0.67	29	0.96	82	0.92	671	0.68
BOPD.6	35	0.96	128	0.9	772	0.65	42	0.96	143	0.9	926	0.67
BOPD.7	31	0.96	98	0.93	667	0.68	33	0.97	110	0.93	807	0.69
BOPD.8	36	0.97	111	0.94	957	0.7	49	0.97	179	0.94	1555	0.72
BOPD.9	26	0.96	87	0.92	616	0.68	41	0.97	127	0.94	990	0.72
MDQ	22	0.98	72	0.97	800	0.83	37	0.99	144	0.98	1857	0.85
CHR	17	0.98	48	0.97	469	0.85	73	0.99	218	0.98	2371	0.88

E = OTUs endémicas presentes en la muestra; C= cobertura de la muestra

En la Tabla 3-14 se observa la distribución de OTUs endémicas del linaje SAR86 y las correspondientes coberturas por muestra. Este linaje presentó la menor cantidad de OTUs endémicas en todas las muestras, tanto con datos ecualizados como sin ecualizar. Además, fue el linaje que presentó la menor cantidad de secuencias por muestras, aunque, no obstante, no se correspondieron con valores bajos de cobertura, salvo al 99% similitud, donde los valores de cobertura encontrados fueron menores que 0.5. Las muestras tomadas más cerca del talud presentaron mayor número de OTUs endémicas, particularmente la muestra MDQ, la cual además es la muestra con mayor número de secuencias del linaje. El sitio con el menor número de OTUs endémicas fue BOPD10. En este linaje, a diferencia del resto, se encontró una relación entre los valores de cobertura y el número de OTUs endémicas, donde los valores más altos de cobertura se asociaron con los mayores valores de endemismo. Esto puede atribuirse al relativamente bajo número de secuencias que presente este linaje en todas las muestras con respecto al resto de los LUPs.

Tabla 3-14: Distribución de las OTUs endémicas del linaje SAR86

Muestras	SAR86											
	Ecuilizado						No Ecuilizado					
	E 95	C 95	E 97	C 97	E 99	C 99	E 95	C 95	E 97	C 97	E 99	C 99
BH	14	0.96	46	0.91	269	0.61	37	0.97	115	0.92	647	0.64
BOPD.1	1	0.89	3	0.87	13	0.64	3	0.86	9	0.8	27	0.51
BOPD.10	1	0.93	1	0.8	15	0.4	3	0.82	4	0.75	18	0.4
BOPD.11	21	0.91	55	0.76	191	0.38	18	0.9	54	0.78	193	0.4
BOPD.12	13	0.84	25	0.68	70	0.34	19	0.88	46	0.7	129	0.38
BOPD.13	6	0.9	15	0.78	56	0.42	8	0.88	22	0.77	69	0.44
BOPD.14	6	0.83	10	0.68	23	0.37	12	0.75	20	0.67	45	0.37
BOPD.2	2	0.88	6	0.76	11	0.64	3	0.92	11	0.8	29	0.58
BOPD.3	5	0.85	10	0.72	28	0.44	4	0.83	7	0.71	26	0.46
BOPD.4	8	0.79	19	0.66	45	0.38	11	0.81	20	0.72	64	0.38
BOPD.5	14	0.82	26	0.69	65	0.43	15	0.84	33	0.7	93	0.42
BOPD.6	4	0.88	18	0.72	53	0.36	12	0.84	25	0.72	65	0.43
BOPD.7	14	0.87	23	0.81	88	0.45	14	0.88	21	0.8	105	0.47
BOPD.8	10	0.88	21	0.75	59	0.43	14	0.91	27	0.83	93	0.53
BOPD.9	3	0.92	8	0.75	19	0.45	13	0.84	21	0.72	50	0.43
MDQ	24	0.97	68	0.93	491	0.55	39	0.98	134	0.94	1052	0.62
CHR	2	0.89	3	0.86	10	0.63	13	0.91	21	0.87	75	0.49

E = OTUs endémicas presentes en la muestra; C= cobertura de la muestra

3.7 Distribución espacial de la diversidad

A diferencia del índice de Simpson y de los índices de diversidad y homogeneidad de Shannon, los índices ACE y Chao estuvieron influenciados por las cantidades de secuencias de cada LUP en las muestras (Tablas A-06 y A-07 del ANEXO). Esto sugiere una menor robustez estadística con respecto al esfuerzo de muestreo, por lo cual se decidió no utilizarlos.

En la Figura 3-11 y la Tabla A-08 del anexo se observan los valores obtenidos para el índice de diversidad de Shannon para todos los LUPs en los diferentes niveles de agrupamiento con datos ecualizados y no ecualizados. Se pudo observar que los linajes *Amylibacter* y NS5 fueron los más diversos en todos los niveles de agrupamiento con datos ecualizados y no ecualizados. En SAR86 los valores de diversidad fueron intermedios, mientras que los linajes SAR11 y OM1 presentaron diversidades relativamente bajas. Por otra parte, se observó que la diversidad de Shannon aumentó a medida que aumentó el nivel de similitud de agrupamiento. Esto indica que, como es de esperar, los rangos taxonómicos más bajos presentan una mayor cantidad de grupos con respecto a los rangos más inclusivos. En *Amylibacter*, los valores correspondientes a los puntos de muestreo ubicados más cerca de la costa fueron relativamente pequeños mientras que en los puntos más alejados dichos valores tendieron a incrementarse, en particular en las muestras tomadas hacia el sur. En las muestras más boreales, estas diferencias fueron menores. Para el Linaje OM1 los mayores valores del índice se observaron en las locaciones BOPD11, BOPD12, BOPD13 y BOPD14, las cuales poseen una distribución boreal en relación con las demás muestras (ver Figura 2-1 de los Materiales y Métodos). Sin embargo, la muestra MDQ, la cual proviene de una latitud similar a las anteriores, presentó una diversidad relativamente menor. El resto de los puntos muestreados presentaron diversidades relativamente similares entre sí. Para el linaje SAR11, se observaron diferencias entre las muestras más oceánicas, BH y MDQ, y el resto de las muestras. Sin embargo, ninguna de estas tendencias pudo observarse en los linajes SAR86 y NS5. Los valores obtenidos a partir de datos ecualizados y no ecualizados, en todos los niveles de agrupamiento, no presentaron diferencias sustanciales entre sí.

En la Figura 3-12 se observan cómo se distribuyó la diversidad de Shannon con respecto a la latitud. Si bien NS5, *Amylibacter*, SAR11 y SAR86 no presentaron tendencias detectables, en el grupo OM1 se pudo observar que a latitudes mayores los valores del índice tendieron a disminuir en los 3 niveles de agrupamiento y con datos ecualizados y no ecualizados.

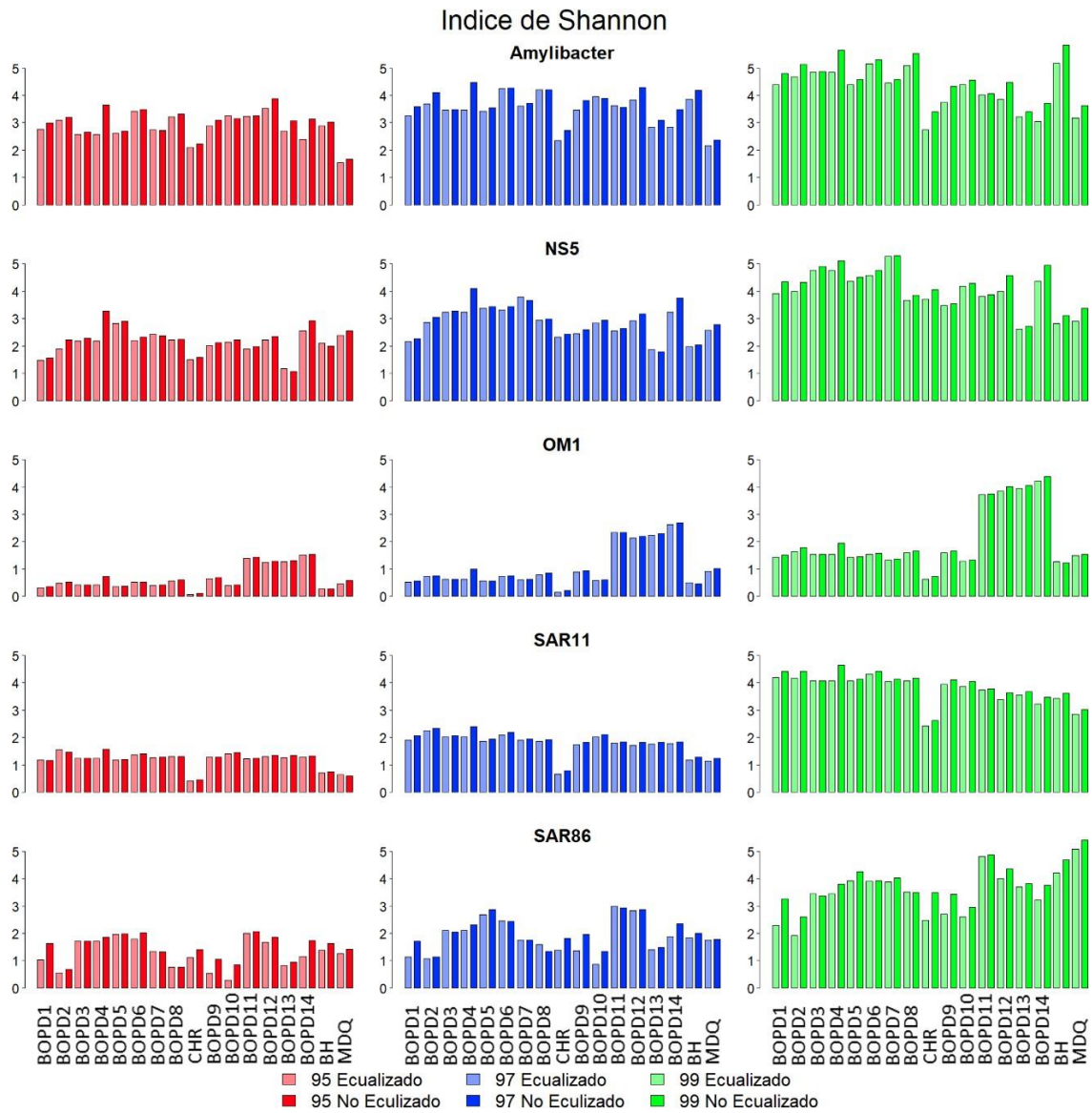


Figura 3-11: Distribución de los valores del índice de Diversidad de Shannon.

Las barras corresponden al valor del índice en cada una de las muestras. El color de cada columna (rojo, azul y verde) indica el nivel de agrupamiento, mientras que la intensidad de cada uno (claro u oscuro) indica si se trata de datos ecualizados o no ecualizados.

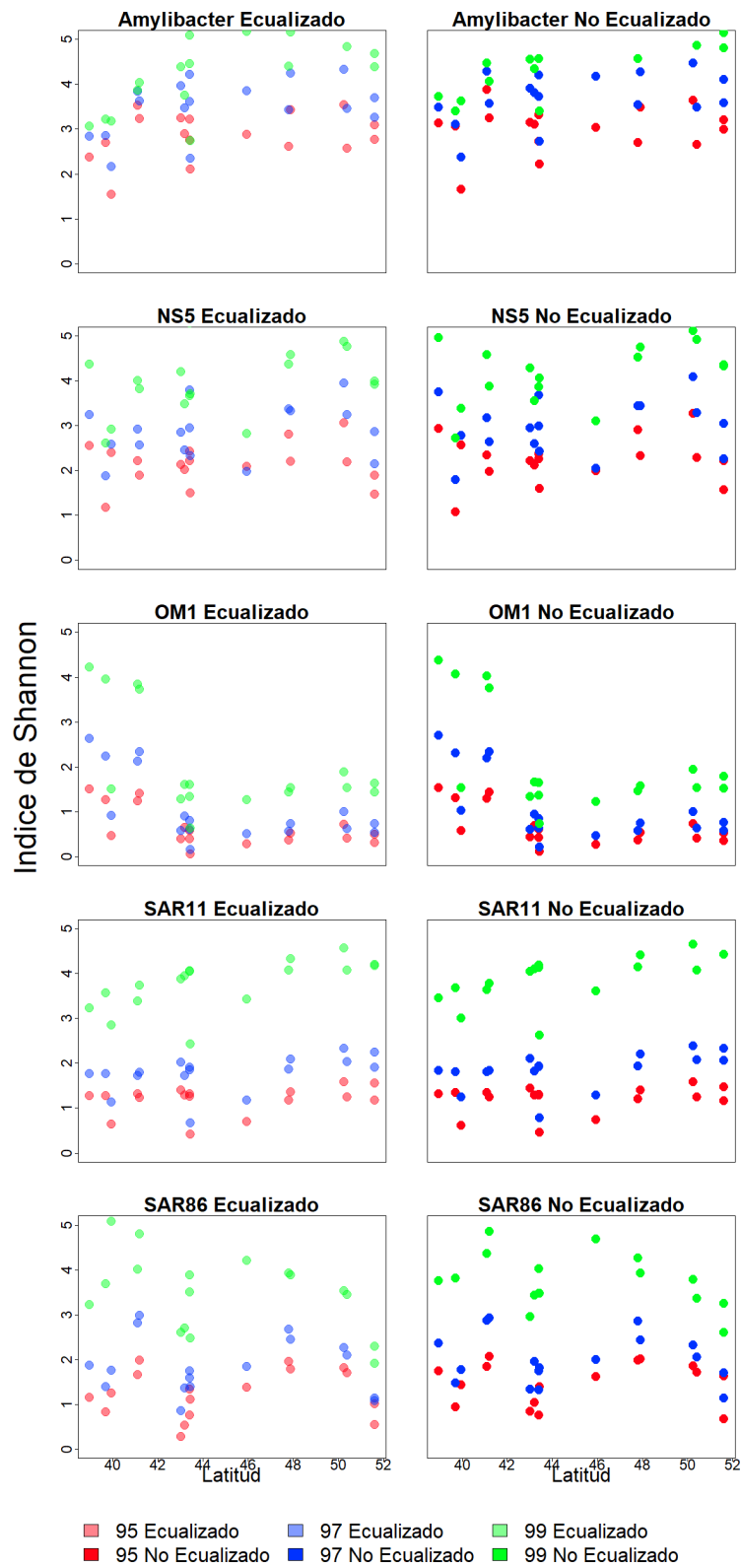


Figura 3-12: Distribución de los valores del índice de Diversidad de Shannon en función de la latitud de las muestras.

En la Figura 3-13 y la Tabla A-09 del Anexo se observan las estimaciones de la diversidad obtenidas mediante el índice de Simpson, para todos los LUPs, en todos los niveles de agrupamiento, con datos ecualizados y sin ecualizar. Se observó que a medida que los niveles de similitud de agrupamiento aumentaron, los valores del índice disminuyeron, lo que indicaría una mayor diversidad a nivel de especie, que de familia. Los valores obtenidos con datos ecualizados y no ecualizados no presentaron diferencias importantes. El linaje que presentó la menor diversidad fue *Amylibacter*, seguido por NS5, mientras que SAR11 y SAR86 presentaron valores intermedios. OM1 fue el linaje que presentó las mayores diversidades.

El análisis de la diversidad de Simpson tampoco evidenció patrones geográficos. Sin embargo, en algunos de los linajes pudieron observarse tendencias sutiles. En el linaje OM1 se observó que la diversidad fue mayor en las localidades con distribución más boreal, a excepción de la muestra MDQ, que, si bien encuentra en latitudes próximas a las demás, presentó valores similares a las muestras más australes. En el linaje *Amylibacter* los valores del índice aumentaron en las muestras más oceánicas (MDQ y BH).

En la Figura 3-14 se observa la distribución de los valores del índice de Simpson con respecto a la latitud. En los linajes *Amylibacter*, NS5, SAR11 y SAR86 no se observan tendencias. Sin embargo, la diversidad del linaje OM1 presentó una tendencia a aumentar a medida que aumentó la latitud. Esto se encuentra en concordancia con los datos de la Figura 3-13.

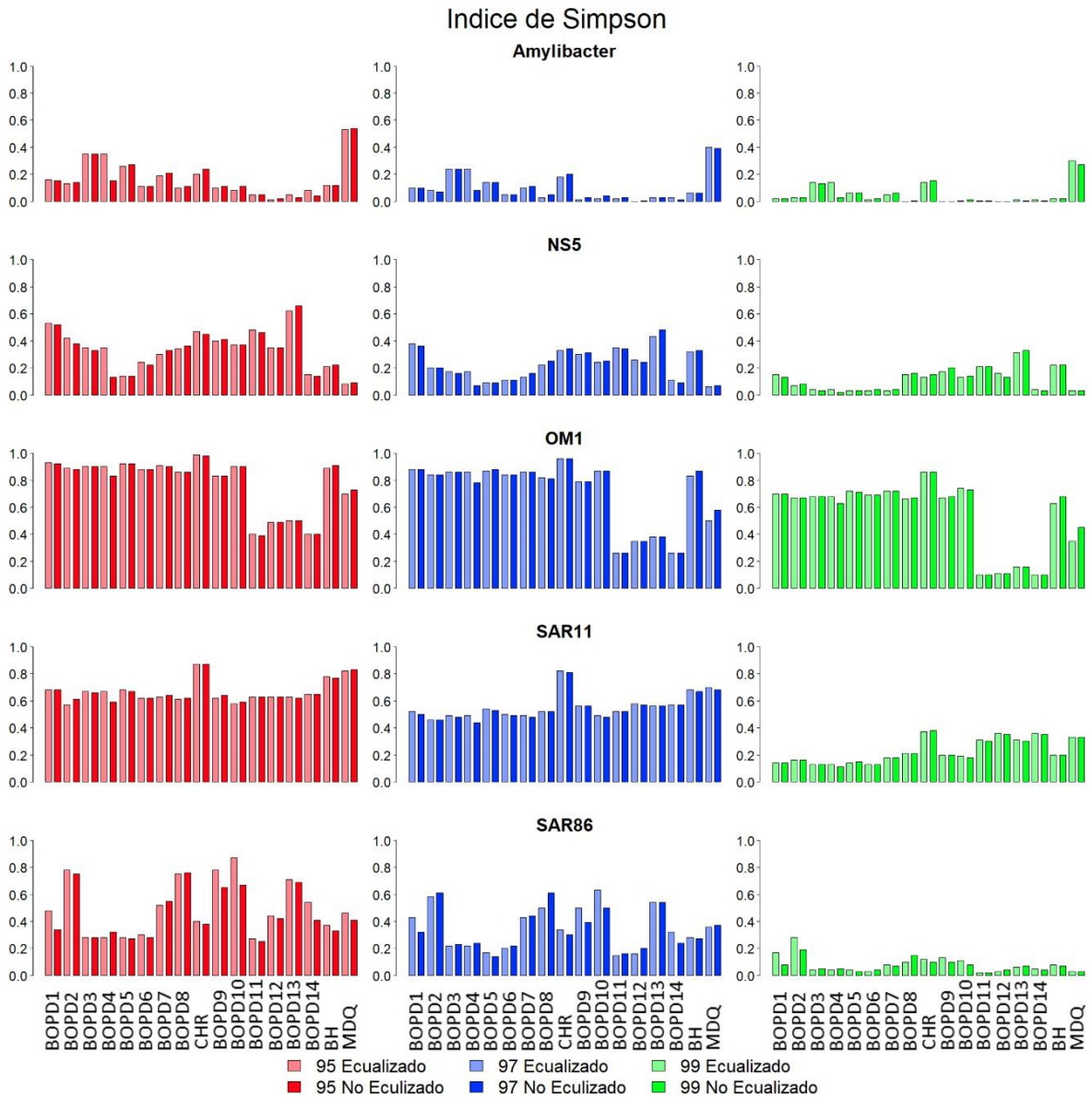


Figura 3-13: Distribución de los valores del índice de Diversidad de Simpson.

Las barras corresponden al valor del índice en cada una de las muestras. El color de cada columna (rojo, azul y verde) indica el nivel de agrupamiento, mientras que la intensidad de cada uno (claro u oscuro) indica si se trata de datos ecualizados o no ecualizados.

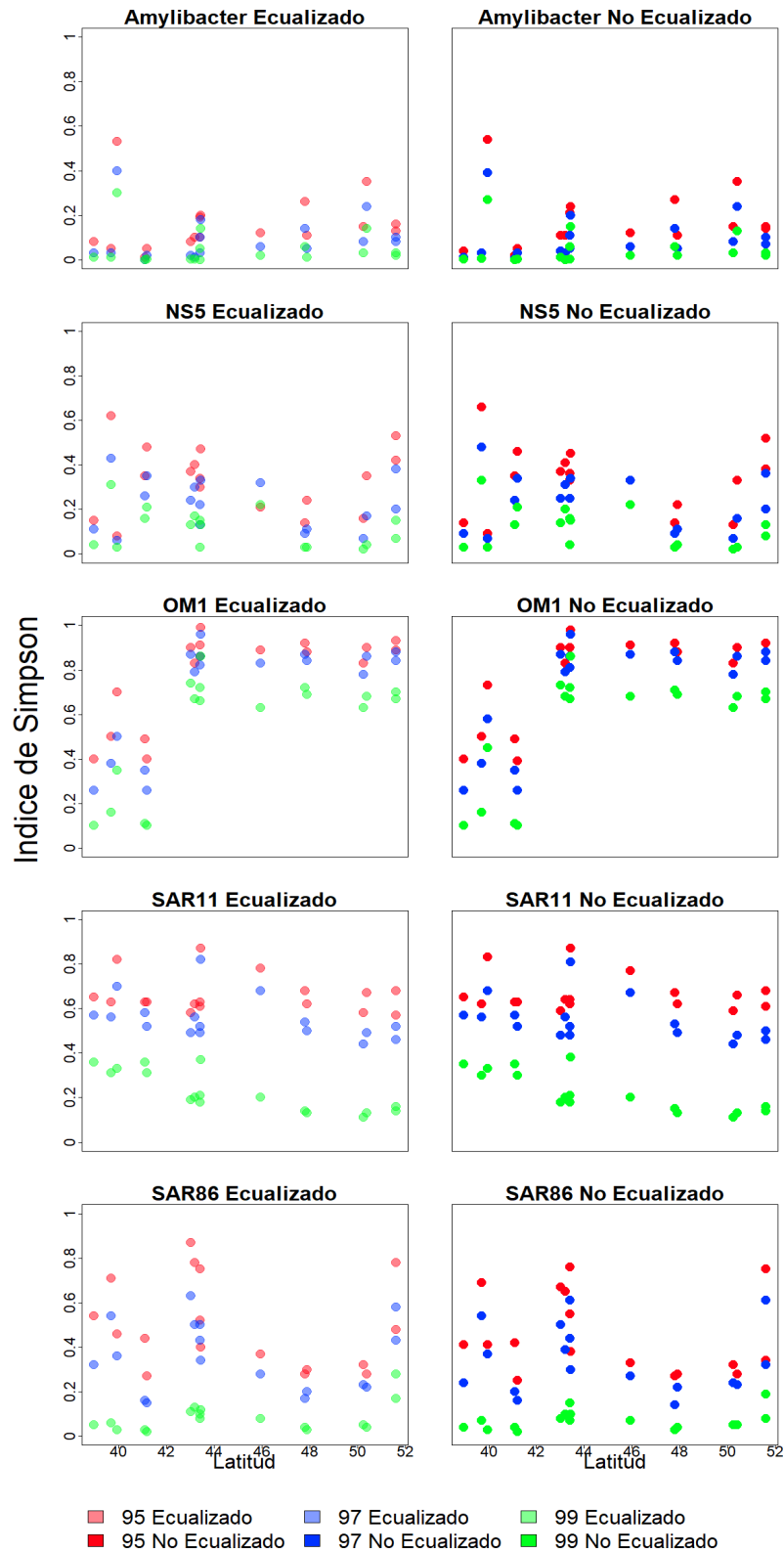


Figura 3-14: Distribución de los valores del índice de Diversidad de Simpson en función de la latitud de las muestras.

En la Figura 3-15 y Tabla A-10 del Anexo se presentan los valores obtenidos para el índice de homogeneidad de Shannon. Se observa que, a medida que aumentó el porcentaje de similitud de agrupamiento, aumentó la equitatividad; es decir que a medida que se agrupó a niveles taxonómicos putativos más bajos, las distribuciones relativas de abundancias de OTUs tendieron a ser más homogéneas. Los valores obtenidos con los datos ecualizados y no ecualizados no presentaron diferencias sustanciales, aunque los valores obtenidos para los datos no ecualizados fueron levemente superiores. El linaje que presentó los valores de homogeneidad más altos fue *Amylibacter*, seguido por NS5 y SAR86, mientras que OM1 y SAR11 resultaron ser los más inequitativos.

En general no se observaron patrones en la distribución geográfica de la homogeneidad, a excepción del linaje OM1, para el cual se observó que los puntos de muestreo más boreales presentaron valores mayores, independientemente de su cercanía a la costa.

En la Figura 3-16 se observa la variación del índice de homogeneidad de Shannon con respecto a la latitud. Como se mencionó en el párrafo previo, las homogeneidades de los linajes SAR11, SAR86, NS55 y *Amylibacter* resultaron ser similares a lo largo del rango latitudinal, mientras que para el linaje OM1 se observó una disminución en la homogeneidad a medida que la latitud aumentó.

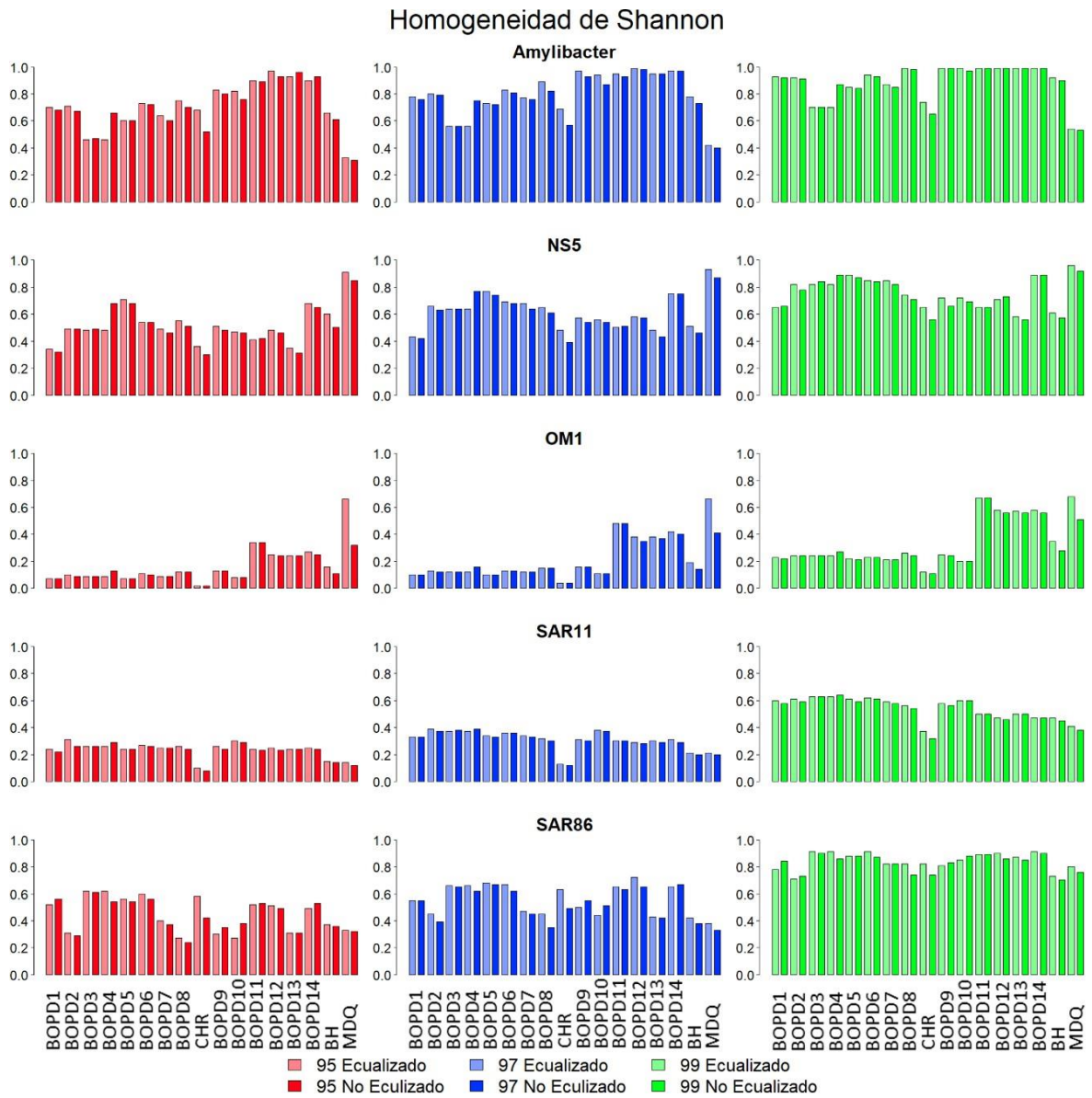


Figura 3-15: Distribución de los valores del índice de Homogeneidad de Shannon.

Las barras corresponden al valor del índice en cada una de las muestras. El color de cada columna (rojo, azul y verde) indica el nivel de agrupamiento, mientras que la intensidad de cada uno (claro u oscuro) indica si se trata de datos ecualizados o no ecualizados.

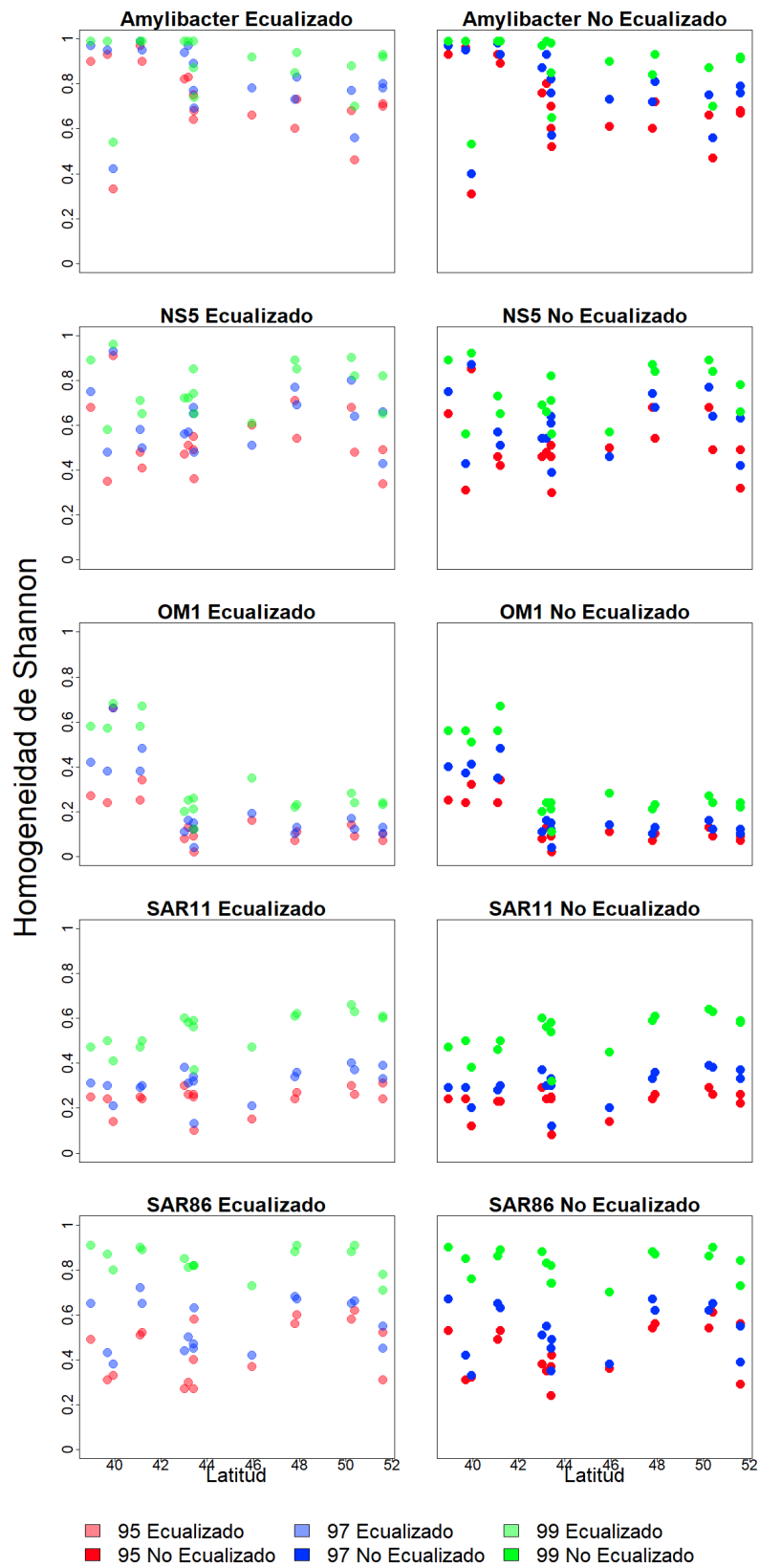


Figura 3-16: Distribución de los valores del índice de Homogeneidad de Shannon en función de la latitud de las muestras.

4 Discusión

4.1 Características generales del conjunto de datos (dataset).

El conjunto de datos de las muestras del mar argentino presentó casi tres millones de secuencias pareadas, a partir de las cuales pudieron obtenerse 302.597 secuencias de alta calidad (SAC). Estas secuencias se obtuvieron luego de optimizar los procedimientos y parámetros del control de calidad con la ayuda de una comunidad simulada de tamaño reducido y composición conocida; sumado al uso de una base de datos de referencia “*in-house*” diseñada y desarrollada específicamente para el presente estudio.

Las mencionadas SAC brindaron coberturas superiores a 0.8 (llegando en ocasiones a 0.98) en todas las muestras, lo que indica que las secuencias obtenidas son un reflejo de la composición de las comunidades presentes en cada punto de muestreo. Generalmente, en muestreos de grandes superficies geográficas tales como el ambiente marino, es difícil lograr muestreos representativos, donde las comunidades bacterianas no estén sub-muestreadas (Brown et al. 2014; Shade et al. 2014), sin embargo, para este estudio, los valores de cobertura revelaron esfuerzos de muestreo adecuados a nivel de las comunidades locales.

Por otro lado, los resultados obtenidos mostraron que la cantidad de secuencias que pueden obtenerse para los grupos más abundantes dentro de los LUPs (en este estudio SAR11 y OM1) pueden ser mucho mayores que las obtenidas para otros LUPs (como fue el caso de *Amylibacter*, NS5 y SAR86). Además, las abundancias de algunos LUPs pueden presentar fuertes variaciones entre muestras, como fue el caso del linaje *Amylibacter* en las muestras BOPD2 y BOPD3, y SAR86 en las muestras MDQ y CHR, entre varios otros casos (Fig.3-2). Ambas situaciones, es decir la presencia de grupos hiperabundantes como SAR11 y OM1, y de linajes de abundancia muy variables como lo fueron SAR86 y *Amylibacter*, podría potencialmente generar distorsiones en los análisis debidos a diferencias en las coberturas logradas por linaje. Si bien en este estudio no se observaron problemas relacionados con estos fenómenos, el dato seguramente resultará de utilidad para estudios futuros focalizados en el estudio de la biología de los LUPs.

4.2 Linajes Ubicuos y Profusos.

En este trabajo se identificaron 5 linajes ubicuos y profusos a los que hemos denominado LUPs. Dos de dichos linajes contribuyeron con la mayor cantidad de secuencias en todas las muestras, OM1 (*Ca. Actinomarina*) (Ghai et al. 2013) y SAR11, por lo que, a partir de los resultados de este estudio, y según ya se adelantó en la subsección previa, pueden considerarse como grupos hiperabundantes. El resto de los LUPs, *Amylibacter*, NS5 y SAR86, fueron muy abundantes en algunos de los puntos muestreados mientras que en otros presentaron abundancias relativamente bajas, aunque siempre cualitativamente mayores que las presentadas por el resto de los linajes presentes en la meta comunidad (*Marinoscillum*, *Formosa*, NS4, *Polibacter*, *Synechococcus*, SAR406, *Planctomyces*, *Fretibacter*, PS1, *Asciidiaceihabitans*, *Planktomarina*, SAR116, *Chesapeake*, OM43, *Idiomarina*, OM60, KI89A, ZD0405, JTB255, y la mayoría de las OTUs no clasificables).

Los resultados obtenidos, en particular la predominancia de SAR11 y OM1, (con un aporte levemente superior al 80% entre ambos), se encuentran en concordancia con trabajos anteriores realizados en el Mar Argentino y en otras regiones (Calvo et al. 2018; Laas et al. 2014; Seo et al. 2017). La mayor abundancia de estos grupos indica que ambos grupos son componentes claves de la comunidad bacterio-planctónica del Mar Argentino. Ambos grupos constituyen ejemplos de bacterias marinas caracterizadas por su amplia distribución y abundancia principalmente en la zona fótica, que poseen pequeño tamaño y un genoma adaptativamente reducido o racionado (fenómeno denominado *streamlining* en inglés) con bajo contenido en GC (Giovannoni 2017; Mizuno et al. 2015). Se ha documentado que ambos grupos, *Pelagibacterales* (SAR11) y OM1, pueden encontrarse en forma conjunta en los mismos hábitats, exponiendo una distribución global similar (Giaccardi et al. 2022; Giovannoni 2017). No obstante, los representantes de OM1 suelen encontrarse en menor abundancia que los de SAR11. En este trabajo, las proporciones de ambos linajes fueron relativamente similares en la mayoría de las muestras, a excepción de aquellas colectadas en puntos de muestreos más alejados de la costa (MDQ y BH), donde OM1 presentó una menor proporción de secuencias que SAR11, lo cual deberá verificarse en futuros estudios.

El clado SAR11 es un grupo ampliamente estudiado, cuyo predominio en cuanto a abundancia y distribución a nivel mundial es bien conocida, principalmente en la zona fótica de los ambientes marinos (Laas et al. 2014; Manrique y Jones 2017; Morris et al. 2012; Seo et al. 2017; West et al. 2016). Los miembros de SAR11 pertenecen a la clase *Alphaproteobacteria*, orden *Pelagibacterales*, los cuales, como ya se mencionó, son

bacterias pequeñas, aeróbicas, quimio-heterótrofas, de vida libre, con genomas pequeños, que respiran oxígeno y oxidan carbono (Giovannoni 2017).

El Linaje OM1 (*Ca. Actinomarina*), perteneciente al phylum *Actinobacteria*, orden *Acidimicrobiales* que son exclusivamente marinas, de reducido genoma y de vida libre (López-Pérez et al. 2020). Este linaje presenta una amplia distribución a nivel mundial, presentando algunos grupos distantes dentro del orden que se encuentran asociados ambientes de agua dulce (Ghai et al. 2013), como también regiones marinas de alta salinidad (Mehrshad et al. 2016). La amplia distribución de los miembros de *Actinobacteria* es testigo de su gran plasticidad metabólica (Morris et al. 2012).

Entre los LUPs que se detectaron en esta Tesis, *Amylibacter* es un género del phylum *Proteobacteria* que como ya se destacó fue particularmente abundante en las muestras más oceánicas, MDQ y BH. A medida que las muestras presentaban una geolocalización más cercana la costa, se observó una disminución en el número de secuencias presentes de este linaje (Figura 3.2, Tabla 3-4). En concordancia con esto, un trabajo de J. Liu y colaboradores, llevado a cabo en regiones estuariales, demostró la prevalencia del grupo en aguas marinas superficiales (Liu et al. 2015). Por otro lado, el linaje ha sido caracterizado como un rápido colonizador de superficies oceánicas, junto con el orden *Oceanospirillales* (Dang et al. 2008). Dentro del orden *Oceanospirillales*, del phylum *Proteobacteria*, clase *Gammaproteobacterias*, se encuentra el clado SAR86, otro de los LUPs observados en este trabajo. Este taxón se encuentra constituido por bacterias heterotróficas ubicuas, que se presentan como un clado diverso, con por lo menos tres grupos, que se encuentran asociadas a variables ambientales (Hoarfrost et al. 2020; Suzuki et al. 2001). En concordancia con lo reportado por otros grupos (Brown et al. 2014; Dang et al. 2008), en este trabajo, el linaje presentó mayor abundancia en los puntos de muestreo relativamente más alejados de la costa, MDQ y BH (Brown et al. 2014; Dang et al. 2008).

El quinto linaje fue el grupo NS5. Este linaje, perteneciente al phylum *Bacteroidetes*, clase *Flavobacteriia*, presentó mayores abundancias en los ambientes más cercanos a las desembocaduras de ríos. El phylum *Bacteroidetes*, puede asociarse a fuentes de agua dulce; sin embargo, la clase *Flavobacteriia* en particular se ha reportado en aguas superficiales de océanos (Liu et al. 2015), generalmente de distribución costera y no suele ser un grupo muy abundante, aunque en realidad podría ser un linaje que ha sido subestimado (Dang et al. 2008), o que presenta grandes variaciones en la abundancia como sugieren los datos de la Figura 3-2 (por ejemplo, obsérvese la gran diferencia que se observa entre las muestras BOPD1 y BOPD14, o MDQ y CHR, entre otras). Esto coincide con datos de trabajos realizados

por otros autores, en los cuales las abundancias de NS5 variaron entre 4% y 22% (Aylward et al. 2015; Díez-Vives et al. 2019).

4.3 Relaciones taxa-área.

Algunas de las primeras descripciones de relaciones taxa-área (o especie-área) en bacterias corresponden a Bell et al. 2005 y Horner-Devine et al. 2004. Los hallazgos presentados en dichos trabajos sentaron sólidos precedentes dado que hasta ese momento la idea de que “Todo está en todas partes, pero el ambiente selecciona” era prácticamente dogmática en microbiología. Como se detalló en la sección 1.4, dicha teoría asume que los taxa bacterianos, especialmente aquellos muy abundantes, no deberían presentar relaciones taxa-área. A partir del año 2004, se han publicado otros trabajos (Barreto et al. 2014; Bell et al. 2005; Horner-Devine et al. 2004; Martiny et al. 2011; Woodcock et al. 2006) en los que se describen dichas relaciones en diferentes ambientes. Los mismos evidencian que las comunidades bacterianas pueden presentar patrones biogeográficos, incluso en ambientes homogéneos como la superficie del océano (Zinger et al. 2014). En el presente trabajo, se realizó un estudio de las relaciones taxa-área (TAR) para cada uno de los LUPs identificados en el Mar Argentino (Figura 3-3). Como resultado, pudieron descubrirse patrones taxa-área en todos ellos, en todos los niveles de agrupamientos, y tanto a partir de datos ecualizados como a partir de datos sin ecualizar.

A grandes rasgos el presente estudio indica que las curvas taxa-área presentan patrones similares en todos los LUPs. Los valores de las correspondientes pendientes (z) oscilaron entre 0.15 y 0.3 (Tabla 3-5), pudiéndose observar que a igual resolución taxonómica los LUPs menos abundantes presentaron valores de pendientes ligeramente más altos. Esto podría deberse, sin embargo, a una mayor dispersión o varianza asociada con la presencia de “n” menores (menos OTUs), indicando que algunos LUPs requerirán profundidades de secuenciación mayores que las utilizadas comúnmente. De manera similar, los valores de las pendientes z fueron levemente menores con datos no ecualizados con respecto a los obtenidos utilizando datos ecualizados. Esto puede atribuirse a que la ecualización tiene el efecto negativo de sacrificar parte de la variabilidad presente en las muestras. Este no es el primer estudio en el cual se observa que la ecualización, procedimiento muy difundido entre los estudios independientes de cultivo, puede requerir operaciones más complejas que las usualmente implementadas y a veces tener propiedades no deseables desde el punto de vista estadístico (McMurdie y Holmes 2014, Weiss et al 2017).

Según era esperable a partir de trabajos publicados con anterioridad (Horner-Devine et al. 2004, Manrique y Jones 2017), los valores de las pendientes se ordenaron según su resolución taxonómica utilizada en cada experimento. Es decir que los

agrupamientos de 99% presentaron mayores valores de pendiente que los agrupamientos al 97 y 95% del mismo linaje. Esto sostiene el concepto de que al comparar los valores obtenidos de los diferentes estudios realizados sobre las TAR es importante considerar el o los rangos taxonómicos utilizados.

Los valores de las pendientes (z) obtenidas en este trabajo fueron ligeramente mayores que los observados en estudios previos realizados en ambientes marinos (Zinger et al. 2014), aunque se encuentran en concordancia con los obtenidos en otros trabajos ([0.002 - 0.26]; Green et al. 2004; Horner-Devine et al. 2004; Woodcock et al. 2006). Sin embargo, debe tenerse en cuenta que los trabajos citados corresponden a comunidades bacterianas completas, es decir no a linajes individuales cual fue el objetivo de esta Tesis. Son escasos los trabajos donde se hayan analizado los TAR de linajes individuales, por lo que la información que se proporciona en esta Tesis resulta de particular valor, ya que demuestran que el patrón se verifica en linajes individuales, cada uno presumiblemente adaptado a un determinado conjunto de factores ambientales.

La obtención de valores significativamente diferentes de 0 para las pendientes (z), las que constituyen el parámetro más relevante de este tipo de estudios (MacArthur y Wilson 1967), corrobora la existencia de patrones biogeográficos para los grupos estudiados.

4.4 Análisis de las curvas de Decaimiento por distancia

El decaimiento por distancia evalúa cambios en la similitud entre comunidades en función de la distancia que los separa. Esta relación, a diferencia de los análisis de la relación taxa-área, no solo se basa en datos presencia/ausencia, sino que además contiene información de la abundancia relativa de las OTUs en las diferentes distancias contempladas (Barreto et al. 2014). La existencia de patrones de distribución puede atribuirse además de a efectos ambientales, a limitaciones a la dispersión, las cuales hacen que las comunidades cercanas en el espacio compartirán más especies que las que se encuentran más separadas (Clark et al. 2021; Oliveira et al. 2015).

En el presente trabajo, se observó que los LUPs presentaron patrones de decaimiento por distancia (Figuras 3-4, 3-5 y 3-6). En líneas generales, puede inferirse que a medida que aumenta la distancia entre ensamblajes de LUPs, sus disimilitudes se incrementan. Este patrón pudo evidenciarse en todos los LUPs y en los 3 niveles de agrupamientos, tanto en el caso de análisis realizados a partir de datos ecualizados como en el caso de aquellos realizados con datos no ecualizados. De los 3 índices de disimilitud utilizados para realizar las curvas, puede destacarse que las curvas realizadas con el índice de Morisita presentaron diferencias con respecto a los otros índices utilizados. En

particular, los LUPs que presentaron curvas con diferencias más conspicuas fueron *Amylibacter* y el clado SAR11. Este último presentó pendientes de las curvas de 95 y 97% muy cercanas a cero, lo cual podría atribuirse a la capacidad resolutoria del índice, ya que las curvas basadas en otros índices no presentaron el patrón. En el caso del linaje *Amylibacter* se obtuvieron curvas en los análisis basados tanto en el índice de Morisita como en las basadas en el índice de Bray-Curtis. Es posible que esto se deba no sólo a la capacidad de resolución de dichos índices, sino también a la abundancia de linaje, la cual al ser muy baja resulta en la obtención de una menor cantidad de observaciones.

Los valores de β (beta), la pendiente de las curvas DD, representan la tasa de recambio de especies o composición y también permite predecir el número de taxa no compartidos entre áreas de diferentes tamaños (Zinger et al. 2014). Los valores de las pendientes β para las curvas de Bray-Curtis variaron desde 0.01 a 0.09, mientras que para el índice de Morisita los valores fueron de -0.00005 a 0.07 y para el índice de Sorensen fueron desde 0 a 0.05. Estos valores de pendiente β son mayores con respecto a los valores observados en otros trabajos (Zinger et al. 2014; tablas 3-6, 3-7 y 3-8). Este incremento puede deberse a que en el trabajo antes mencionado se focalizaron en la comunidad completa y se retiraron del estudio algunos taxa de baja frecuencia, lo que puede introducir variabilidad (Youssef et al. 2010; Dolan y Stoeck 2011). Otra causa posible es que los diferentes valores observados para la pendiente β reflejen diferencias en las tasas de dispersión. Vale decir que los valores de β de cada LUP, y por ende las tasas de dispersión, podrían ser diferentes de los de la comunidad completa, posibilidad que abre futuras vías de investigación.

En el trabajo de Meyer y colaboradores de 2018 se realizaron los análisis con diferentes niveles de agrupamiento, 95, 97, 99 y 100% de similitud (Meyer et al. 2018). Dichos autores, observaron que las pendientes se relacionaron con el nivel de similitud de agrupamiento de la misma forma que lo hicieron en el presente trabajo. Este patrón indica que las mayores tasas de recambio ocurren a nivel de especie, ya que el mismo fue el nivel taxonómico putativo que presentó los mayores valores de pendientes, seguidas por las tasas en los niveles putativos de género (97%) y familia (95%). Estos resultados son consistentes con lo que cabía esperar a-priori, ya que a nivel putativo de especies es esperable encontrar una cantidad de OTUs que en niveles superiores, y por ende una mayor diferencia composicional cuyo se trabaja a dicho nivel.

Los patrones de Decaimiento por distancia pueden reflejar distintos mecanismos de ensamblado, o bien combinaciones de estos. Entre estos podemos mencionar: 1) disminución de la similitud en el medio ambiente (Nekola y White 1999), 2) la tasa de dispersión de los organismos entre los sitios, en un ecosistema con gries barreras de dispersión (Soininen et al. 2007) y 3) por teoría neutral, donde la similitud no se ve

afectada por heterogeneidad ambiental, sino más bien por deriva ecológica, dispersión y especiación aleatorias (Hancock 2010). Al analizar las curvas encontradas en este trabajo, puede hipotetizarse que la distancia geográfica puede actuar como barrera para la dispersión, lo cual concuerda también con otros trabajos ya mencionados (Dolan 2005; Hellweger et al. 2014; Manrique y Jones 2017; Martiny et al. 2006; Sjöstedt et al. 2014; Zinger et al. 2014, entre otros).

4.5 Escalado Multidimensional no métrico

Como ya se explicó, el escalado multidimensional no métrico es una técnica que permite simplificar las relaciones entre comunidades a partir de las similitudes (o disimilitudes) entre las mismas (Mead 1992). Si bien los análisis realizados en esta Tesis guardaron ciertas relaciones con el resto de los análisis biogeográficos clásicos (*i.e.* TAR, DD y endemismo), no pudieron por sí solos revelar la existencia de patrones de aislamiento por distancia. Cabe destacar que los valores de *stress* fueron inferiores a 0.2 en todos los análisis realizados. Esto indica que los ordenamientos obtenidos fueron estables, y que el ajuste entre de las distancias originales y las distancias proyectadas en el diagrama de ordenación fue metodológicamente aceptable (Dexter et al. 2018). Por lo tanto, puede inferirse que, al menos bajo el diseño del presente estudio, el análisis mediante NMDS por sí solo no posee la capacidad de evidenciar patrones biogeográficos, destacando la importancia de aplicar métodos biogeográficos tradicionales como los análisis de TAR, DD y endemismo.

4.6 Análisis de endemismo

En los últimos años, ya se han realizado algunos estudios que indicaron la existencia de diferentes grados de endemividad (Bell 2010; Martiny et al. 2006). Por definición, un taxón endémico está restringido a un lugar particular, región o tipo de hábitat y por lo tanto no se distribuye uniformemente. Por esta razón, el endemismo constituye una de las pruebas más simples de la existencia de una biogeografía microbiana (Hanson et al. 2012). De manera similar a lo que se conoce en los organismos superiores, algunos microorganismos parecen ser endémicos de distintas regiones geográficas (Whitaker et al. 2003).

En el presente trabajo se observó que al agrupar al 95% el 69% de las OTUs fueron endémicas, al agrupar al 97% de similitud el 78% de las OTUs se encontraron en una muestra, mientras que al agrupar a 99% las OTUs endémicas representaron el 86%. Valores aproximados fueron observados para todos los linajes con datos ecualizados y no ecualizados. En concordancia con esto, en un estudio de bacterioplancton marino realizado por Pommier y colaboradores (Pommier et al. 2007), que utilizó bibliotecas del gen 16S de diversos lugares, se observó que al agrupar a 97% de similitud, el 69% de las OTUs se detectaban sólo en un único lugar (*i.e.* eran endémicas), 17% en dos lugares,

6% en tres lugares, y sólo el 0,4% eran cosmopolitas.

Un resultado que llamó la atención es que, para el linaje SAR11 se observó un incremento de la cantidad de OTUs endémicas al nivel taxonómico putativo de especie, agrupamiento del 99%, tanto con datos ecualizados como no ecualizados, con respecto a los niveles taxonómicos putativos más inclusivos (Tabla 3-9). El mismo efecto puede observarse con el número de OTUs en la meta comunidad; es decir que, al agrupar a 99% de similitud, el linaje SAR11 presentó el doble de OTUs que el linaje OM1, mientras que la proporción de OTUs clasificadas al agrupar a 95 y 97% fue de 1 a 1 aproximadamente (Tabla A-11). Esto no observó para los demás linajes, ya que al comprar *Amylibacter* y NS5, que son similares en abundancia, la proporción de OTUs endémicas y totales en todos los niveles de agrupamiento se mantuvo igual. Este resultado podría ser un reflejo de las características del linaje SAR11, como podría ser una capacidad diferente de respuesta ante cambios en el ambiente o bien a características intrínsecas del linaje referentes al tiempo y modo de su evolución. Se requerirán otros estudios para establecer el mecanismo específico que subyace a la aparente hiper-diversificación de SAR11 con respecto al resto de los LUP en el nivel putativo de especie evidenciada por los estudios realizados en esta Tesis.

4.7 Índices de diversidad

Los análisis de los índices de riqueza de CHAO y ACE (Tablas Anexo A-06 y A-07), mostraron que los mismos estuvieron muy influenciados por el número de secuencias determinado para cada muestra, lo cual podría deberse a que ambos índices dan un peso elevado a los linajes raros (Kim et al. 2017), o bien a que son menos robustos al esfuerzo de muestreo que el resto de los índices utilizados. Estas posibilidades deberán verificarse mediante diseños experimentales posteriores creados para tal fin. Sin embargo, ante la posibilidad de incurrir en errores estadísticos en esta Tesis se decidió no tener en cuenta los análisis correspondientes.

En cuanto a los índices de diversidad de Shannon y Simpson, y el índice de homogeneidad de Shannon (Figuras 3-11,3-13 y 3-15), se observó que no presentaron patrones geográficos equivalentes a los revelados por los análisis de TAR, DD y endemismo. Sin embargo, resultó interesante que la diversidad en el linaje OM1 podría variar con la latitud, lo cual abre hipótesis para realizar nuevas investigaciones.

5 Conclusiones

❖ A partir de 17 comunidades de agua superficiales se determinó la existencia de cinco linajes ubicuos y abundantes (Linajes Ubicuos y Profusos, LUP). Esto son:

•*Amylibacter*: *Proteobacteria*, *Alphaproteobacteria*, *Rhodobacterales*, *Rhodobacteraceae*

•NS5: *Bacteroidetes*, *Flavobacteriia*, *Flavobacteriales*, *Flavobacteriaceae*

•OM1: *Actinobacteria*, *Acidimicrobiia*, *Acidimicrobiales*, OM1, *Candidatus_actinomarina*

•SAR11: *Proteobacteria*, *Alphaproteobacteria*, SAR11, *Surface_1*, *Candidatus_pelagibacter y/o Proteobacteria*, *Alphaproteobacteria*, SAR11, *Surface_1*, *Surface_1_ge*.

•SAR86: *Proteobacteria*, *Gammaproteobacteria*, *Oceanospirillales*, SAR86_clade, SAR86_clade_ge.

❖ Al igual que se ha observado en otras regiones del mundo, los LUPs se encuentran acompañados por una plétora de taxa minoritarios y por una cantidad importante de grupos no clasificables. Los linajes OM1 y SAR11, representan casi el 80% de las secuencias totales; mientras que los otros tres LUPs presentaron abundancias menores, pero mucho mayores que las observadas entre los taxa minoritarios.

❖ En los 5 LUPs es posible observar patrones biogeográficos, es decir distribuciones heterogéneas. Dichos patrones pueden evidenciarse mediante análisis de las relaciones taxa-área, decaimiento por distancia, patrones de endemismo y, en menor medida, mediante análisis multivariado (NMDS).

❖ Los resultados obtenidos entre datos ecualizados y no ecualizados no han presentado mayores diferencias, por lo que los resultados de esta Tesis sugieren que sería indistinto trabajar con uno u otro enfoque.

❖ Los índices de alfa diversidad por sí solos no aportan datos relevantes en el estudio de patrones biogeográficos, sino que deben ir acompañados de otros tipos de análisis para comprender en forma cabal la diversidad microbiana.

❖ Este es el primer estudio en el que se analizó qué linajes del bacterioplancton realizan un mayor aporte a los microbiomas del Mar Argentino. El estudio aporta datos novedosos sobre la existencia de patrones biogeográficos en microorganismos, tema que es foco de muchos estudios actuales.

❖ La detección de patrones biogeográficos en los microbiomas del Mar Argentino abre las puertas a la planificación de nuevos estudios y experimentos destinados a avanzar en la comprensión de los mecanismos subyacentes a los patrones evidenciados en esta Tesis.

6 Bibliografía

- Azam, F et al. 1983. "The Ecological Role of Water-Column Microbes in the Sea." *Marine Ecology Progress Series* 10(3): 257–63. <http://www.jstor.org/stable/24814647>.
- Baas-Becking, L. 1934. *Geobiologie; of Inleiding Tot de Milieukunde*. WP Van Stockum & Zoon NV.
- Barreto, D. et al. 2014. "Distance-Decay y Taxa-Area Relationships for Bacteria, Archaea y Methanogenic Archaea in a Tropical Lake Sediment." *PLoS ONE* 9(10).
- Beijerinck, M. 1913. "De Infusies En de Ontdekking Der Bacteriën, Jaarboek van de Koninklijke Akademie V." *Wetenschappen. Müller, Amsterdam, The Netherlys*.
- Bell, T. et al. 2005. "Larger Islys House More Bacterial Taxa." *Science* 308(5730): 1884.
- Bell, T. 2010. "Experimental Tests of the Bacterial Distance-Decay Relationship." *ISME Journal* 4(11): 1357–65. <http://dx.doi.org/10.1038/ismej.2010.77>.
- Brown, V.; Ostrowski, M.; Grzymski, J. y Lauro F. 2014. "A Trait Based Perspective on the Biogeography of Common y Abundant Marine Bacterioplankton Clades." *Marine Genomics* 15: 17–28. <http://dx.doi.org/10.1016/j.margen.2014.03.002>.
- Buchan, A.; LeCleir, G.; Gulvik, C. y González, J. 2014. "Master Recyclers: Features y Functions of Bacteria Associated with Phytoplankton Blooms." *Nature Reviews Microbiology* 12(10): 686–98. <https://doi.org/10.1038/nrmicro3326>.
- Calderón-Sanou, I. et al. 2019. "From Environmental DNA Sequences to Ecological Conclusions: How Strong Is the Influence of Methodological Choices?" *Journal of Biogeography* (January 2019): 193–206.
- Calvo, A.; Manrique J. y Jones, L. 2018. "Rare Unclassified 16S RRNA Operational Taxonomic Units from the Uncharted Engaño Bay (Argentinean Patagonia)." *Canadian Journal of Microbiology* 64(1): 91–96.
- Clark, D.; Graham, J.; Underwood, T.; McGenity, J. y Dumbrell, A. 2021. "What Drives Study-Dependent Differences in Distance–Decay Relationships of Microbial Communities?" *Global Ecology y Biogeography* 30(4): 811–25.
- Cordero, O. y Manoshi S. 2016. "Microbial Interactions y Community Assembly at Microscales." *Current Opinion in Microbiology* 31(Figure 1): 227–34. <http://dx.doi.org/10.1016/j.mib.2016.03.015>.
- D'Amore, R. et al. 2016. "A Comprehensive Benchmarking Study of Protocols y Sequencing Platforms for 16S RRNA Community Profiling." *BMC Genomics* 17(1). <http://dx.doi.org/10.1186/s12864-015-2194-9>.
- Dang, H.; Tiegang, L.; Mingna, C. y Guiqiao H. 2008. "Cross-Ocean Distribution of Rhodobacterales Bacteria as Primary Surface Colonizers in Temperate Coastal Marine Waters." *Applied y Environmental Microbiology* 74(1): 52–60.

- Dexter, E.; Rollwagen-Bollens G., y Bollens, C. 2018. "The Trouble with Stress: A Flexible Method for the Evaluation of Nonmetric Multidimensional Scaling." *Limnology y Oceanography: Methods* 16(7): 434–43.
- Díez-Vives, C. et al. 2019. "Delineation of Ecologically Distinct Units of Marine Bacteroidetes in the Northwestern Mediterranean Sea." *Molecular Ecology* 28(11): 2846–59.
- Dogliotti, A., Lutz V., y Segura; M. 2014. "Estimation of Primary Production in the Southern Argentine Continental Shelf y Shelf-Break Regions Using Field y Remote Sensing Data." *Remote Sensing of Environment* 140: 497–508. <http://dx.doi.org/10.1016/j.rse.2013.09.021>.
- Dolan, J. 2005. "An Introduction to the Biogeography of Aquatic Microbes." *Aquatic Microbial Ecology* 41(1): 39–48.
- Drakare, S.; Lennon, J. y Hillebry, H. 2006. "The Imprint of the Geographical, Evolutionary Ecological Context on Species-Area Relationships." *Ecology Letters* 9(2): 215–27.
- Ducklow, H. y Doney, S. 2013. "What Is the Metabolic State of the Oligotrophic Ocean? A Debate." *Annual Review of Marine Science* 5: 525–33.
- Enke, T. et al. 2018. "Microscale Ecology Regulates Particulate Organic Matter Turnover in Model Marine Microbial Communities." *Nature Communications* 9(1). <http://dx.doi.org/10.1038/s41467-018-05159-8>.
- Falkowski, P.; Fenchel, T. y Delong, E. 2008. "The Microbial Engines That Drive Earth's Biogeochemical Cycles." *Science* 320(5879): 1034–39.
- Ferster, C. 1973. "A Functional Analysis of Depression." *American psychologist* 28(10): 857.
- García, V. et al. 2008. "Environmental Factors Controlling the Phytoplankton Blooms at the Patagonia Shelf-Break in Spring." *Deep-Sea Research Part I: Oceanographic Research Papers* 55(9): 1150–66.
- Ghai, R. et al. 2013. "Metagenomics Uncovers a New Group of Low GC y Ultra-Small Marine Actinobacteria." *Scientific Reports* 3: 1–8.
- Giaccardi, L.; Badenas, M.; Jones, L. y Manrique, J. 2022. "Abundant Microbes of Surface Sea Waters of the Uncharted Engaño Bay at the Atlantic Patagonian Coast: Relevance of Bacteria-Sized Photosynthetic Eukaryotes." *Aquatic Ecology*. <https://doi.org/10.1007/s10452-022-09962-w>.
- Giovannoni, S. 2017. "SAR11 Bacteria: The Most Abundant Plankton in the Oceans." *Annual Review of Marine Science* 9(1): 231–55.
- Grant, S.; Hill, S.; Trathan, P. y Murphy, E. 2013. "Ecosystem Services of the Southern Ocean: Trade-Offs in Decision-Making." *Antarctic Science* 25(5): 603–17.
- Green, J. et al. 2004. "Spatial Scaling of Microbial Eukaryote Diversity." *Nature* 432(7018):

747–50.

- Grossart, H.; Massana, R.; McMahon, K. y Walsh, D. 2020. “Linking Metagenomics to Aquatic Microbial Ecology y Biogeochemical Cycles.” *Limnology y Oceanography* 65(S1): S2–20.
- Guidi, L. et al. 2016. “Plankton Networks Driving Carbon Export in the Oligotrophic Ocean.” *Nature* 532(7600): 465–70. <http://dx.doi.org/10.1038/nature16942>.
- Hancock, L. 2010. “Book Review: Book Review.” *Criminology & Criminal Justice* 10(4): 419–20.
- Van Hannen, E. et al. 1999. “Changes in Bacterial y Eukaryotic Community Structure after Mass Lysis of Filamentous Cyanobacteria Associated with Viruses.” *Applied y Environmental Microbiology* 65(2): 795–801.
- Hanson, C.; Fuhrman, J.; Horner-Devine, M. y Martiny, J. 2012. “Beyond Biogeographic Patterns: Processes Shaping the Microbial Lyscape.” *Nature Reviews Microbiology* 10(7): 497–506. <http://dx.doi.org/10.1038/nrmicro2795>.
- Hellweger, F.; Van Sebille, E. y Fredrick, N. 2014. “Biogeographic Patterns in Ocean Microbes Emerge in a Neutral Agent-Based Model.” *Science* 345(6202): 1346–49.
- Hoarfrost, A. et al. 2020. “Global Ecotypes in the Ubiquitous Marine Clade SAR86.” *ISME Journal* 14(1): 178–88. <http://dx.doi.org/10.1038/s41396-019-0516-7>.
- Horner-Devine, M.; Lage, M.; Hughes, J. y Bohannan, B. 2004. “A Taxa–Area Relationship for Bacteria.” *Nature* 432: 750–53.
- Jiao, N. et al. 2010. “Microbial Production of Recalcitrant Dissolved Organic Matter: Long-Term Carbon Storage in the Global Ocean.” *Nature Reviews Microbiology* 8(8): 593–99. <http://dx.doi.org/10.1038/nrmicro2386>.
- Jones, D.; Taylor, W. y Thornton, J. 1992. “The Rapid Generation of Mutation Data Matrices from Protein Sequences.” *Computer applications in the biosciences: CABIOS* 8(3): 275–82.
- Katoh, K. y Toh, H. 2010. “Parallelization of the MAFFT Multiple Sequence Alignment Program.” *Bioinformatics* 26(15): 1899–1900.
- Kim, B. et al. 2017. “Deciphering Diversity Indices for a Better Understying of Microbial Communities.” *Journal of Microbiology y Biotechnology* 27(12): 2089–93.
- Kozich, J. et al. 2013. “Development of a Dual-Index Sequencing Strategy y Curation Pipeline for Analyzing Amplicon Sequence Data on the Miseq Illumina Sequencing Platform.” *Applied y Environmental Microbiology* 79(17): 5112–20.
- Laas, P.; Simm, J.; Lips, I. y Metsis, M. 2014. “Spatial Variability of Winter Bacterioplankton Community Composition in the Gulf of Finly (the Baltic Sea).” *Journal of Marine Systems* 129: 127–34. <http://dx.doi.org/10.1016/j.jmarsys.2013.07.016>.
- Legendre, P. y Legendre, L. 1998. *Numerical Ecology*. 2nd Englis. Amsterdam: Elsevier

science.

- Liquete, C. et al. 2013. "Current Status y Future Prospects for the Assessment of Marine y Coastal Ecosystem Services: A Systematic Review." *PLoS ONE* 8(7).
- Liu, J. et al. 2015. "Phylogenetic Shifts of Bacterioplankton Community Composition along the Pearl Estuary: The Potential Impact of Hypoxia y Nutrients." *Frontiers in Microbiology* 6(FEB): 1–13.
- López-Pérez, M.; Haro-Moreno, J.; Iranzo, J. y Rodríguez-Valera, F. 2020. " Genomes of the ' Cyidatus Actinomarinales' Order: Highly Streamlined Marine Epipelagic Actinobacteria." *mSystems* 5(6).
- MacArthur, R. y Wilson, E. 1967. "The Theory of Isly BiogeographyNo Title." In *The Theory of Isly Biogeography*, Princeton University Press.
- Manrique, J. y Jones, L. 2017. "Are Ocean Currents too Slow to Counteract SAR11 Evolution? A next-Generation Sequencing, Phylogeographic Analysis." *Molecular Phylogenetics y Evolution* 107: 324–37. <http://dx.doi.org/10.1016/j.ympev.2016.11.015>.
- Martiny, J. 2015. "Dispersal y the Microbiome: Learning How Fast y How Far Microorganisms Move Will Help Us Better Understy the Diversity of Microbial Communities." *Microbe Magazine* 10: 191–96.
- Martiny, J. et al. 2006. "Microbial Biogeography: Putting Microorganisms on the Map." *Nature Reviews Microbiology* 4(2): 102–12.
- Martos, P & Piccolo, M. 1988. "Hydrography of the Argentine Continental Shelf between 38S° y 42S°." *Continental Shelf Research* 8(9): 1043–56.
- McMurdie, P. y Holmes, S. 2014. "Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible." *PLoS Computational Biology* 10(4).
- Mead, A. 1992. "Review of the Development of Multidimensional Scaling Methods." *The Statistician* 41(1): 27.
- Mehrshad, M. et al. 2016. "Genome Reconstruction from Metagenomic Data Sets Reveals Novel Microbes in the Brackish Waters of the Caspian Sea." *Applied y Environmental Microbiology* 82(5): 1599–1612.
- Meyer, K. et al. 2018. "Why Do Microbes Exhibit Weak Biogeographic Patterns?" *ISME Journal* 12(6): 1404–13. <http://dx.doi.org/10.1038/s41396-018-0103-3>.
- Milici, M. et al. 2016. "Bacterioplankton Biogeography of the Atlantic Ocean: A Case Study of the Distance-Decay Relationship." *Frontiers in Microbiology* 7(APR): 1–15.
- Milke, F. et al. 2022. "Composition y Biogeography of Planktonic Pro- y Eukaryotic Communities in the Atlantic Ocean: Primer Choice Matters." *Frontiers in Microbiology* 13(June): 1–17.
- Mizuno, C.; Rodríguez-Valera, F. y Ghai, R. 2015. "Genomes of Planktonic."

- Mo, Y. et al. 2018. "Biogeographic Patterns of Abundant y Rare Bacterioplankton in Three Subtropical Bays Resulting from Selective y Neutral Processes." *ISME Journal* 12(9): 2198–2210. <http://dx.doi.org/10.1038/s41396-018-0153-6>.
- Morris, R.; Frazar, C. y Carlson, C. 2012. "Basin-Scale Patterns in the Abundance of SAR11 Subclades, Marine Actinobacteria (OM1), Members of the Roseobacter Clade y OCS116 in the South Atlantic." *Environmental Microbiology* 14(5): 1133–44.
- Nekola, J. y White, P. 1999. "The Distance Decay of Similarity in Biogeography y Ecology." *Journal of Biogeography* 26(4): 867–78.
- Nelson, T.; Rogers, T. y Brown, M. 2013. "The Gut Bacterial Community of Mammals from Marine y Terrestrial Habitats." *PLoS ONE* 8(12): 1–8.
- Nylyer, J. 2004. "MrAIC. Pl." *Program distributed by the author* 1(3).
- Oliveira, U.; Brescovit, A. y Santos, A. 2015. "Delimiting Areas of Endemism through Kernel Interpolation." *PLoS ONE* 10(1): 1–18.
- Peralta, A.; Stuart, D.; Kent, A. y Lennon, J. 2014. "A Social-Ecological Framework for 'Micromanaging' Microbial Services." *Frontiers in Ecology y the Environment* 12(9): 524–31.
- Pernice, M. et al. 2016. "Large Variability of Bathypelagic Microbial Eukaryotic Communities across the World's Oceans." *ISME Journal* 10(4): 945–58.
- Pomeroy, L.; Le Williams, P.; Azam, F. y Hobbie, J. 2007. "The Microbial Loop." *Oceanography* 20(SPL.ISS. 2): 28–33.
- Pommier, T. et al. 2007. "Global Patterns of Diversity y Community Structure in Marine Bacterioplankton." *Molecular Ecology* 16(4): 867–80.
- Poulet, A. et al. 2016. "Improved Efficiency y Reliability of NGS Amplicon Sequencing Data Analysis for Genetic Diagnostic Procedures Using AGSA Software." *BioMed Research International* 2016.
- Pruesse, E. et al. 2007. "SILVA: A Comprehensive Online Resource for Quality Checked y Aligned Ribosomal RNA Sequence Data Compatible with ARB." *Nucleic Acids Research* 35(21): 7188–96.
- Rivas, A.; Dogliotti, A. y Gagliardini, D. 2006. "Seasonal Variability in Satellite-Measured Surface Chlorophyll in the Patagonian Shelf." *Continental Shelf Research* 26(6): 703–20. <https://www.sciencedirect.com/science/article/pii/S0278434306000343>.
- Rodríguez-Martínez, R.; Rocap, G.; Salazar, G. y Massana, R. 2013. "Biogeography of the Uncultured Marine Picoeukaryote MAST-4: Temperature-Driven Distribution Patterns." *ISME Journal* 7(8): 1531–43.
- Sarmiento, J. 2006. *Ocean Biogeochemical Dynamics*. Princeton University Press. <https://doi.org/10.1515/9781400849079>.
- Schloss, I. et al. 2007. "Role of Plankton Communities in Sea-Air Variations in PCO₂ in

- the SW Atlantic Ocean." *Marine Ecology Progress Series* 332: 93–106.
- Schloss, P. et al. 2009. "Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing y Comparing Microbial Communities." *Applied y Environmental Microbiology* 75(23): 7537–41.
- Schloss, P. et al. 2016. "Status of the Archaeal y Bacterial Census: An Update." *mBio* 7(e00201–16).
- Semler, A.; Fortney, J.; Fulweiler, R. y Dekas, A. 2022. "Cold Seeps on the Passive Northern U.S. Atlantic Margin Host Globally Representative Members of the Seep Microbiome with Locally Dominant Strains of Archaea." *Applied y Environmental Microbiology* 88(11): e00468-22. <https://doi.org/10.1128/aem.00468-22>.
- Seo, J.; Kang, I.; Yang, S. y Cho, J. 2017. "Characterization of Spatial Distribution of the Bacterial Community in the South Sea of Korea." *PLoS ONE* 12(3): 1–18.
- Shade, A. et al. 2014. "Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity." *mBio* 5(4): 1–9.
- Shao, Q. et al. 2022. "Biodiversity y Biogeography of Abundant y Rare Microbial Assemblages in the Western Subtropical Pacific Ocean." *Frontiers in Microbiology* 13(March): 1–14.
- Sjöstedt, J.; Martiny, J.; Munk, P. y Riemann, L. 2014. "Abundance of Broad Bacterial Taxa in the Sargasso Sea Explained by Environmental Conditions but Not Water Mass." *Applied y Environmental Microbiology* 80(9): 2786–95.
- Soininen, J.; McDonald, R. y Hillebry, H. 2007. "The Distance Decay of Similarity in Ecological Communities." *Ecography* 30(1): 3–12.
- Suzuki, M.; Béjà, O.; Taylor, L. y DeLong, E. 2001. "Phylogenetic Analysis of Ribosomal RNA Operons from Uncultivated Coastal Marine Bacterioplankton." *Environmental Microbiology* 3(5): 323–31. <https://doi.org/10.1046/j.1462-2920.2001.00198.x>.
- Visbeck, M. et al. 2014. "A Sustainable Development Goal for the Ocean y Coasts: Global Ocean Challenges Benefit from Regional Initiatives Supporting Globally Coordinated Solutions." *Marine Policy* 49: 87–89. <https://www.sciencedirect.com/science/article/pii/S0308597X14000487>.
- Wang, Q.; Garrity, G.; Tiedje, J. y Cole, J. 2007. "Naïve Bayesian Classifier for Rapid Assignment of RRNA Sequences into the New Bacterial Taxonomy." *Applied y Environmental Microbiology* 73(16): 5261–67.
- Weiss, S. et al. 2017. "Normalization y Microbial Differential Abundance Strategies Depend upon Data Characteristics." *Microbiome* 5(1): 1–18.
- West, N. et al. 2016. "Distinct Spatial Patterns of SAR11, SAR86, y Actinobacteria Diversity along a Transect in the Ultra-Oligotrophic South Pacific Ocean." *Frontiers in Microbiology* 7(MAR): 1–16.

- Westman, W. 1977. "How Much Are Nature's Services Worth?" *Science* 197(4307): 960–64. <https://doi.org/10.1126/science.197.4307.960>.
- Whitaker, R.; Grogan, D. y Taylor, J. 2003. "Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea." *Science* 301(5635): 976–78.
- Whitman, W.; Coleman, D. y Wiebe, W. 1998. "Prokaryotes: The Unseen Majority." *Proceedings of the National Academy of Sciences of the United States of America* 95(12): 6578–83.
- Wilhelm, S. y Suttle, C. 1999. "Viruses y Nutrient Cycles in the Sea Aquatic Food Webs." *BioScience* 49(October): 781–88. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Viruses+y+Nutrient+Cycles+in+the+Sea#0>.
- De Wit, R. y Bouvier T. 2006. "'Everything Is Everywhere, but, the Environment Selects'; What Did Baas Becking y Beijerinck Really Say?" *Environmental Microbiology* 8(4): 755–58.
- Wolda, H. 1981. "Similarity Indices, Sample Size y Diversity." *Oecologia* 50(3): 296–302. <https://doi.org/10.1007/BF00344966>.
- Woodcock, S. et al. 2006. "Taxa-Area Relationships for Microbes: The Unsampled y the Unseen." *Ecology Letters* 9(7): 805–12.
- Yarza, P. et al. 2014. "Uniting the Classification of Cultured y Uncultured Bacteria y Archaea Using 16S rRNA Gene Sequences." *Nature Reviews Microbiology* 12(9): 635–45. <http://dx.doi.org/10.1038/nrmicro3330>.
- Youssef, N.; Couger, N. y Elshahed, M. 2010. "Fine-Scale Bacterial Beta Diversity within a Complex Ecosystem (Zodletone Spring, OK, USA): The Role of the Rare Biosphere." *PLoS ONE* 5(8).
- Zhang, L. et al. 2020. "Temporal y Spatial Variations of Bacterial Community Compositions in Two Estuaries of Chaohu Lake." *Journal of Oceanology y Limnology* 38(3): 745–58.
- Zhu, W. et al. 2022. "Different Responses of Bacteria y Microeukaryote to Assembly Processes y Co-Occurrence Pattern in the Coastal Upwelling." *Microbial Ecology*. <https://doi.org/10.1007/s00248-022-02093-7>.
- Zinger, L., Boetius, A. y Ramette, A. 2014. "Bacterial Taxa-Area y Distance-Decay Relationships in Marine Environments." *Molecular Ecology* 23(4): 954–64.

7 Anexo

Tabla A-01: Linajes Ubicuos y Profusos (LUP) y sus abundancias (% de SAC de cada taxón) en cada muestra, al 97% de similitud, con datos no ecualizados y procesadas con la base de datos "in-house".

Taxon	SAR11	OM1	NS5	SAR86	Amylibacter	No clasifi- cadas
BH	72.50	2.01	3.26	2.62	8.59	11.02
BOPD.1	37.68	47.07	11.51	0.73	0.48	2.53
BOPD.10	18.22	62.81	7.53	0.43	0.29	10.72
BOPD.11	70.92	8.12	8.07	0.30	3.26	9.33
BOPD.12	59.12	29.65	4.96	0.12	1.09	5.06
BOPD.13	52.45	36.97	2.57	0.07	1.01	6.94
BOPD.14	39.74	54.90	1.42	0.09	0.42	3.42
BOPD.2	30.78	62.41	2.37	1.38	0.43	2.63
BOPD.3	22.14	48.77	6.18	19.12	0.70	3.09
BOPD.4	30.74	58.87	2.56	4.61	0.84	2.38
BOPD.5	26.13	67.03	1.47	1.96	1.34	2.08
BOPD.6	32.63	58.73	3.28	1.74	1.02	2.61
BOPD.7	35.10	48.84	7.48	1.73	1.83	5.02
BOPD.8	50.46	39.95	2.23	1.10	1.27	4.99
BOPD.9	31.38	55.19	2.46	0.22	0.48	10.28
MDQ	66.16	0.40	0.26	12.09	12.20	8.90
CHR	51.03	36.42	10.00	0.71	0.46	1.38

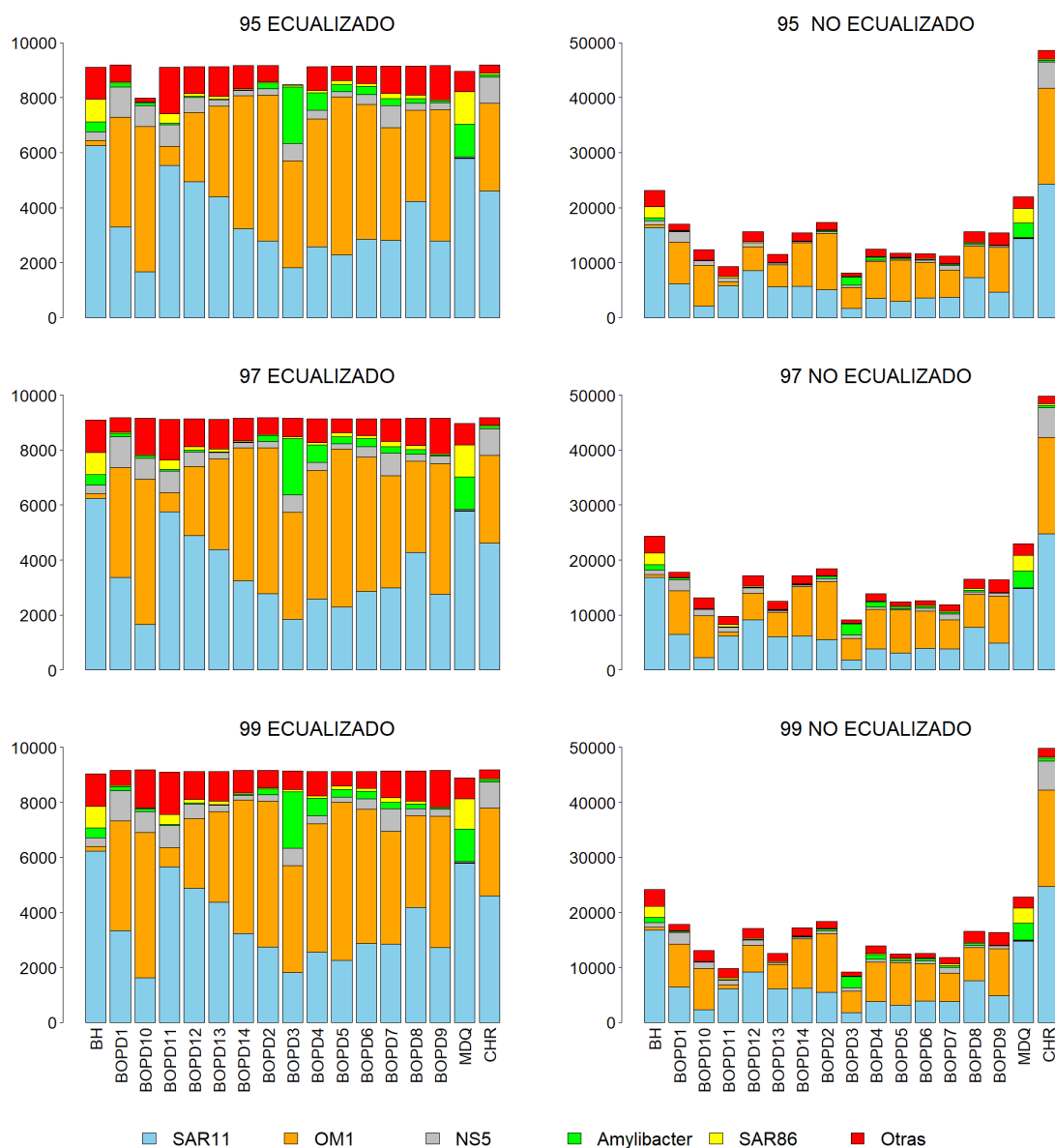


Figura A-02: Número de SAC de cada LUPs en las muestras del Mar Argentino.

Cada barra corresponde al total de SAC encontradas en cada una de las muestras discriminadas por el nivel de similitud agrupadas; eualizados o no eualizadas. En la parte inferior del grafico se especifica la identidad de cada muestra

Tabla A-03: Valores de la correlación de Kendall para las curvas de Decaimiento por distancia basadas en el índice de disimilitud de Bray-Curtis

Linaje	Agrupamiento	Ecualizado	No ecualizado
	95	0.23	0.25
Amylibacter	97	0.15	0.21
	99	0.012	0.087
	95	0.36	0.35
NS5	97	0.39	0.38
	99	0.38	0.37
	95	0.22	0.21
OM1	97	0.28	0.22
	99	0.24	0.26
	95	0.18	0.07
SAR11	97	0.21	0.1
	99	0.47	0.35
	95	0.15	0.19
SAR86	97	0.13	0.17
	99	0.12	0.16

Los valores remarcados son aquellos con valores de probabilidad mayores a 0.05 ($p > 0.05$).

Tabla A-04: Valores de la correlación de Kendall para las curvas de Decaimiento por distancia basadas en el índice de disimilitud de Morisita

Linaje	Agrupamiento	Ecuilizado	No ecualizado
	95	0.070	0.11
Amylibacter	97	0.015	0.056
	99	-0.057	-0.006
	95	0.36	0.33
NS5	97	0.35	0.34
	99	0.33	0.32
	95	0.15	0.15
OM1	97	0.16	0.13
	99	0.14	0.14
	95	0.14	0.16
SAR11	97	0.19	0.21
	99	0.53	0.53
	95	0.05	0.07
SAR86	97	0.06	0.08
	99	0.04	0.03

Los valores remarcados son aquellos con valores de probabilidad mayores a 0.05 ($p > 0.05$).

Tabla A-05: Valores de la correlación de Kendall para las curvas de Decaimiento por distancia basadas en el índice de disimilitud de Sorensen

Linaje	Agrupamiento	Ecuilizado	No ecualizado
	95	0.21	0.17
Amylibacter	97	0.22	0.21
	99	0.04	0.13
	95	0.42	0.44
NS5	97	0.45	0.49
	99	0.41	0.38
	95	0.30	0.31
OM1	97	0.32	0.35
	99	0.29	0.37
	95	0.20	0.27
SAR11	97	0.25	0.33
	99	0.39	0.40
	95	-0.054	-0.036
SAR86	97	0.05	0.11
	99	0.21	0.21

Los valores remarcados son aquellos con valores de probabilidad mayores a 0.05 ($p > 0.05$).

Tabla A-06: Valores del índice ACE para todos los LUPs, niveles de agrupamiento, ecualizados y no ecualizados, discriminados por muestras.

ACE	BH	BOPD									MDQ	CHR							
		1	10	11	12	13	14	2	3	4			5	6	7	8	9		
Amylibacter	Ecuallizado	95	464	739	603	648	164	94	61	646	946	755	1341	860	640	322	822	328	162
		97	679	1938	5353	554	576	154	58	1034	4682	3468	1434	1819	1887	1853	147	2752	836
		99	7313	2073	2450	1081	1176	126	132	1589	26395	26991	1296	2095	1467	11822	946	11133	3040
	No Ecuallizado	95	513	758	701	849	419	85	168	1350	942	1516	1017	1074	692	1386	570	904	440
		97	2079	3616	2514	1163	505	229	146	2766	5494	4303	2798	1696	1719	2010	528	3089	504
		99	21220	2193	3081	736	10414	176	462	9646	24159	25592	1694	8345	1695	19711	7983	20098	3857
NS5	Ecuallizado	95	71	327	567	381	456	205	364	274	529	657	360	214	741	197	296	22	275
		97	211	546	2432	1403	1541	331	599	897	1054	896	359	983	1486	660	377	25	670
		99	2558	17828	6638	7648	4195	4161	10635	1884	12246	12984	5980	5840	11070	5875	1318	42	6103
	No Ecuallizado	95	279	772	605	549	1048	193	816	412	520	550	430	315	711	392	493	51	921
		97	488	1569	2832	1818	2070	457	1621	1415	1193	1640	460	1673	4147	995	962	61	3203
		99	6972	16560	6968	8162	7313	3373	11227	7075	14516	18705	28332	4905	21601	11029	3421	206	43912
OM1	Ecuallizado	95	48	341	469	495	882	832	1222	462	500	726	801	566	174	557	640	2	72
		97	314	1192	1526	1118	2755	2599	4105	1863	1288	2685	1964	2023	1507	3221	1248	5	730
		99	704	13978	22646	2600	28324	26827	39633	17231	24688	24193	15723	18043	14499	14798	17087	15	9045
	No Ecuallizado	95	246	636	744	242	2251	766	1952	882	389	871	1121	706	363	1156	994	60	781
		97	413	1750	2893	1108	3869	3383	4906	2938	1361	3202	2016	3641	1506	3159	2573	157	2920
		99	3367	31830	34699	2464	29652	20003	67851	57109	24791	36293	28892	27802	18885	29530	26651	109	43607
SAR11	Ecuallizado	95	366	390	255	712	1253	870	608	765	505	740	423	548	475	521	474	298	376
		97	1178	1813	1498	4315	4558	2246	1931	2282	1334	2845	1908	3029	2643	2016	2118	1027	1473
		99	37222	28090	34366	37289	46058	27679	18281	19098	22378	27093	19150	27320	25794	36226	21734	19948	15784
	No Ecuallizado	95	535	930	507	713	1409	758	1037	1275	494	763	870	710	525	957	989	536	1104
		97	2578	2560	2600	3067	4439	3072	3865	3254	1996	3058	2938	5341	2845	2864	2151	1729	3962
		99	72644	49586	41692	41788	58580	37011	33575	39407	19396	31382	27155	28275	28533	48794	33748	61893	59689
SAR86	Ecuallizado	95	519	0	0	227	149	36	34	0	45	243	741	42	118	121	16	254	0
		97	924	0	0	716	202	203	202	0	134	498	447	134	182	860	106	747	57
		99	9377	159	211	23133	1152	1653	1332	57	692	3778	2404	2182	2151	1463	2040	11839	350
	No Ecuallizado	95	356	154	0	201	208	61	656	28	47	359	489	170	129	1087	101	391	189
		97	3537	503	92	696	1368	638	687	172	233	600	3622	322	359	673	144	1276	163
		99	34446	750	604	12878	8019	1556	2755	928	596	3759	9386	884	4120	2849	494	15790	3203

Tabla A-07: Valores del índice de CHAO para todos los LUPs, niveles de agrupamiento, ecualizados y no ecualizados, discriminados por muestras.

CHAO	BH	BOPD															MDQ	CHR	
		1	10	11	12	13	14	2	3	4	5	6	7	8	9				
Amylibacter	Ecualizado	95	307	260	211	192	162	64	37	371	636	509	336	495	265	271	158	307	82
		97	434	951	2084	292	377	156	54	500	2082	1380	609	900	822	568	145	892	355
		99	4017	1529	1663	828	589	89	85	1475	11620	7315	1223	2056	1194	4568	474	3600	706
	No Ecualizado	95	326	259	408	215	253	120	106	473	653	844	319	645	439	444	148	599	205
		97	860	971	1706	456	481	142	123	1003	2268	1919	1048	1099	519	646	242	1342	503
		99	5607	2088	2037	592	3917	126	302	2866	11253	13111	1591	3069	1738	6018	3004	7728	1979
NS5	Ecualizado	95	66	178	327	229	399	61	97	179	232	266	131	128	421	123	148	19	178
		97	129	446	783	498	570	139	320	417	546	663	326	769	798	282	250	22	451
		99	679	4358	2226	2753	1928	962	1447	835	3717	2914	4007	2212	5411	1188	693	39	2350
	No Ecualizado	95	115	358	267	363	517	66	355	192	254	391	224	179	577	189	439	31	482
		97	244	820	979	640	937	226	784	519	510	801	370	770	1555	387	517	44	1569
		99	3341	7424	3119	3109	3197	945	4529	2865	3974	4010	4795	2430	7983	3454	1318	142	15839
OM1	Ecualizado	95	12	236	411	381	496	556	751	355	310	437	566	319	176	388	503	2	48
		97	80	491	778	474	1398	1141	2052	1004	638	1236	1011	778	592	1270	637	4	194
		99	353	7119	6869	3057	8594	8628	12975	10097	6154	9714	5705	6574	5107	4578	6191	12	4028
	No Ecualizado	95	30	372	453	240	898	801	1219	627	269	533	756	377	239	639	604	12	442
		97	130	906	1276	491	1662	1587	2897	1533	692	1435	1020	1501	666	1118	1191	48	1580
		99	1354	11455	9381	3467	11215	10398	22046	18337	7467	15465	8158	9751	7113	9499	8223	125	14440
SAR11	Ecualizado	95	217	385	221	410	648	538	425	481	260	540	272	337	349	306	301	207	144
		97	679	826	712	1578	1837	1103	933	1124	677	1242	721	1218	940	1180	793	648	608
		99	12268	10877	7200	13736	15596	9918	9057	8219	7252	9172	6672	11208	10036	13309	9435	7951	5670
	No Ecualizado	95	367	623	288	452	837	578	675	590	263	469	514	460	440	667	608	370	755
		97	1478	1457	1074	1415	2332	1574	2100	1537	971	1458	1075	2050	1241	1674	1211	1139	2108
		99	26651	17135	11375	15048	19763	12620	14336	14511	6736	13079	9026	10812	10426	19079	13402	19660	25465
SAR86	Ecualizado	95	150	17	4	132	112	29	25	16	44	194	150	37	113	62	9	202	17
		97	933	23	22	452	173	153	71	56	87	251	312	125	141	267	54	311	24
		99	2772	72	107	2968	1167	558	179	37	186	1382	1628	488	1121	810	353	4311	51
	No Ecualizado	95	181	57	37	162	78	45	103	31	34	194	164	82	122	118	65	230	132
		97	1154	193	47	427	428	499	169	87	118	239	559	276	299	176	123	700	139
		99	9001	242	192	5146	2395	603	462	268	161	1506	2389	755	1637	1794	477	6889	759

Tabla A-08: Valores del índice de Diversidad de Shannon para todos los LUPs, niveles de agrupamiento, ecualizados y no ecualizados, discriminados por muestras.

SHANNON		1	2	3	4	5	6	7	8	CHR	9	10	11	12	13	14	BH	MDQ	
Amylibacter	E	95	2.77	3.09	2.57	3.54	2.61	3.43	2.75	3.22	2.1	2.89	3.25	3.23	3.53	2.7	2.38	2.88	1.54
		97	3.26	3.69	3.46	4.33	3.43	4.24	3.61	4.21	2.35	3.47	3.96	3.62	3.84	2.85	2.84	3.85	2.17
		99	4.39	4.68	4.84	5.39	4.4	5.16	4.46	5.09	2.74	3.75	4.39	4.03	3.86	3.22	3.06	5.17	3.18
	NE	95	2.99	3.2	2.65	3.64	2.7	3.48	2.72	3.31	2.22	3.1	3.15	3.25	3.88	3.07	3.13	3.04	1.66
		97	3.58	4.1	3.49	4.47	3.54	4.27	3.72	4.2	2.72	3.81	3.9	3.57	4.29	3.1	3.49	4.18	2.37
		99	4.81	5.14	4.86	5.65	4.57	5.31	4.57	5.52	3.4	4.34	4.56	4.06	4.47	3.4	3.72	5.84	3.62
NS5	E	95	1.47	1.89	2.18	3.06	2.81	2.2	2.42	2.22	1.5	2.01	2.13	1.89	2.21	1.17	2.55	2.09	2.39
		97	2.15	2.86	3.24	3.94	3.37	3.32	3.79	2.95	2.33	2.45	2.84	2.56	2.92	1.87	3.24	1.97	2.58
		99	3.92	3.99	4.76	4.87	4.36	4.58	5.28	3.66	3.7	3.48	4.19	3.81	4	2.61	4.36	2.82	2.91
	NE	95	1.57	2.21	2.28	3.27	2.9	2.33	2.37	2.25	1.59	2.12	2.21	1.97	2.34	1.07	2.93	1.99	2.56
		97	2.26	3.05	3.28	4.09	3.44	3.44	3.67	2.98	2.42	2.59	2.94	2.64	3.17	1.79	3.75	2.04	2.78
		99	4.35	4.32	4.91	5.11	4.52	4.75	5.29	3.86	4.05	3.55	4.28	3.87	4.57	2.72	4.95	3.1	3.38
OM1	E	95	0.31	0.49	0.41	0.72	0.36	0.52	0.4	0.57	0.06	0.65	0.4	1.4	1.24	1.27	1.51	0.28	0.46
		97	0.53	0.73	0.62	1	0.56	0.73	0.6	0.8	0.15	0.9	0.58	2.33	2.13	2.23	2.63	0.51	0.91
		99	1.44	1.63	1.53	1.88	1.43	1.53	1.33	1.6	0.63	1.6	1.28	3.73	3.84	3.95	4.22	1.26	1.5
	NE	95	0.35	0.52	0.41	0.73	0.37	0.53	0.42	0.61	0.11	0.69	0.43	1.44	1.29	1.31	1.53	0.27	0.58
		97	0.57	0.76	0.63	1	0.57	0.74	0.63	0.85	0.21	0.94	0.6	2.33	2.19	2.3	2.7	0.46	1.02
		99	1.52	1.78	1.53	1.94	1.46	1.57	1.37	1.65	0.73	1.66	1.33	3.75	4.02	4.06	4.38	1.22	1.53
SAR11	E	95	1.17	1.56	1.25	1.58	1.17	1.36	1.26	1.31	0.42	1.29	1.4	1.23	1.31	1.27	1.28	0.7	0.64
		97	1.9	2.25	2.03	2.33	1.87	2.09	1.9	1.85	0.67	1.73	2.02	1.8	1.72	1.76	1.77	1.17	1.13
		99	4.2	4.17	4.07	4.56	4.07	4.32	4.05	4.06	2.42	3.94	3.87	3.73	3.38	3.56	3.23	3.42	2.85
	NE	95	1.16	1.47	1.24	1.58	1.21	1.4	1.29	1.3	0.46	1.29	1.44	1.25	1.34	1.34	1.32	0.74	0.61
		97	2.06	2.33	2.07	2.39	1.94	2.2	1.94	1.92	0.78	1.82	2.11	1.83	1.81	1.81	1.84	1.29	1.24
		99	4.43	4.42	4.07	4.65	4.14	4.41	4.13	4.18	2.62	4.1	4.04	3.77	3.63	3.68	3.46	3.61	3.01
SAR86	E	95	1.02	0.55	1.71	1.81	1.96	1.79	1.34	0.76	1.12	0.54	0.29	1.99	1.66	0.83	1.16	1.38	1.26
		97	1.14	1.08	2.1	2.27	2.68	2.45	1.75	1.59	1.39	1.36	0.86	2.99	2.82	1.4	1.87	1.84	1.76
		99	2.29	1.91	3.45	3.53	3.92	3.89	3.88	3.51	2.48	2.7	2.6	4.8	4.01	3.69	3.22	4.21	5.08
	NE	95	1.63	0.68	1.72	1.86	1.98	2.02	1.32	0.76	1.4	1.05	0.84	2.07	1.85	0.94	1.74	1.62	1.43
		97	1.71	1.14	2.05	2.32	2.86	2.43	1.75	1.33	1.82	1.96	1.34	2.93	2.87	1.48	2.36	2	1.77
		99	3.25	2.6	3.36	3.79	4.26	3.92	4.03	3.48	3.48	3.43	2.95	4.86	4.36	3.82	3.75	4.68	5.41

E: Ecualizado, NE: No ecualizado. Los números corresponden a la serie de muestras BOPD, el prefijo se retiró para una mejor comprensión de la tabla

Tabla A-09: Valores del índice de Diversidad de Simpson para todos los LUPs, niveles de agrupamiento, ecualizados y no ecualizados, discriminados por muestras

SIMPSON		1	2	3	4	5	6	7	8	CHR	9	10	11	12	13	14	BH	MDQ	
Amylibacter	E	95	0.16	0.13	0.35	0.15	0.26	0.11	0.19	0.1	0.2	0.1	0.08	0.05	0.01	0.05	0.08	0.12	0.53
		97	0.1	0.08	0.24	0.08	0.14	0.05	0.1	0.03	0.18	0.01	0.02	0.02	0.0017	0.03	0.03	0.06	0.4
		99	0.02	0.03	0.14	0.03	0.06	0.01	0.05	0.0009	0.14	0.001	0.002	0.002	0.008	0.01	0.01	0.02	0.3
	NE	95	0.15	0.14	0.35	0.15	0.27	0.11	0.21	0.11	0.24	0.11	0.11	0.05	0.02	0.03	0.04	0.12	0.54
		97	0.1	0.07	0.24	0.08	0.14	0.05	0.11	0.05	0.2	0.03	0.04	0.03	0.0047	0.03	0.01	0.06	0.39
		99	0.02	0.03	0.13	0.03	0.06	0.02	0.06	0.0018	0.15	0.0009	0.01	0.002	0.0007	0.005	0.002	0.02	0.27
NS5	E	95	0.53	0.42	0.35	0.16	0.14	0.24	0.3	0.34	0.47	0.4	0.37	0.48	0.35	0.62	0.15	0.21	0.08
		97	0.38	0.2	0.17	0.07	0.09	0.11	0.13	0.22	0.33	0.3	0.24	0.35	0.26	0.43	0.11	0.32	0.06
		99	0.15	0.07	0.04	0.02	0.03	0.03	0.03	0.15	0.13	0.17	0.13	0.21	0.16	0.31	0.04	0.22	0.03
	NE	95	0.52	0.38	0.33	0.13	0.14	0.22	0.33	0.36	0.45	0.41	0.37	0.46	0.35	0.66	0.14	0.22	0.09
		97	0.36	0.2	0.16	0.07	0.09	0.11	0.16	0.25	0.34	0.31	0.25	0.34	0.24	0.48	0.09	0.33	0.07
		99	0.13	0.08	0.03	0.02	0.03	0.04	0.04	0.16	0.15	0.2	0.14	0.21	0.13	0.33	0.03	0.22	0.03
OM1	E	95	0.93	0.89	0.9	0.83	0.92	0.88	0.91	0.86	0.99	0.83	0.9	0.4	0.49	0.5	0.4	0.89	0.7
		97	0.88	0.84	0.86	0.78	0.87	0.84	0.86	0.82	0.96	0.79	0.87	0.26	0.35	0.38	0.26	0.83	0.5
		99	0.7	0.67	0.68	0.63	0.72	0.69	0.72	0.66	0.86	0.67	0.74	0.1	0.11	0.16	0.1	0.63	0.35
	NE	95	0.92	0.88	0.9	0.83	0.92	0.88	0.9	0.86	0.98	0.83	0.9	0.39	0.49	0.5	0.4	0.91	0.73
		97	0.88	0.84	0.86	0.78	0.88	0.84	0.86	0.81	0.96	0.79	0.87	0.26	0.35	0.38	0.26	0.87	0.58
		99	0.7	0.67	0.68	0.63	0.71	0.69	0.72	0.67	0.86	0.68	0.73	0.1	0.11	0.16	0.1	0.68	0.45
SAR11	E	95	0.68	0.57	0.67	0.58	0.68	0.62	0.63	0.61	0.87	0.62	0.58	0.63	0.63	0.63	0.65	0.78	0.82
		97	0.52	0.46	0.49	0.44	0.54	0.5	0.49	0.52	0.82	0.56	0.49	0.52	0.58	0.56	0.57	0.68	0.7
		99	0.14	0.16	0.13	0.11	0.14	0.13	0.18	0.21	0.37	0.2	0.19	0.31	0.36	0.31	0.36	0.2	0.33
	NE	95	0.68	0.61	0.66	0.59	0.67	0.62	0.64	0.62	0.87	0.64	0.59	0.63	0.63	0.62	0.65	0.77	0.83
		97	0.5	0.46	0.48	0.44	0.53	0.49	0.48	0.52	0.81	0.56	0.48	0.52	0.57	0.56	0.57	0.67	0.68
		99	0.14	0.16	0.13	0.11	0.15	0.13	0.18	0.21	0.38	0.2	0.18	0.3	0.35	0.3	0.35	0.2	0.33
SAR86	E	95	0.48	0.78	0.28	0.32	0.28	0.3	0.52	0.75	0.4	0.78	0.87	0.27	0.44	0.71	0.54	0.37	0.46
		97	0.43	0.58	0.22	0.23	0.17	0.2	0.43	0.5	0.34	0.5	0.63	0.15	0.16	0.54	0.32	0.28	0.36
		99	0.17	0.28	0.04	0.05	0.04	0.03	0.08	0.1	0.12	0.13	0.11	0.02	0.03	0.06	0.05	0.08	0.03
	NE	95	0.34	0.75	0.28	0.32	0.27	0.28	0.55	0.76	0.38	0.65	0.67	0.25	0.42	0.69	0.41	0.33	0.41
		97	0.32	0.61	0.23	0.24	0.14	0.22	0.44	0.61	0.3	0.39	0.5	0.16	0.2	0.54	0.24	0.27	0.37
		99	0.08	0.19	0.05	0.05	0.03	0.04	0.07	0.15	0.1	0.1	0.08	0.02	0.04	0.07	0.04	0.07	0.03

E: Ecualizado, NE: No ecualizado. Los números corresponden a la serie de muestras BOPD, el prefijo se retiró para una mejor comprensión de la tabla

Tabla A-10: Valores del índice de Homogeneidad de Shannon para todos los LUPs, niveles de agrupamiento, ecualizados y no ecualizados, discriminados por muestras

HOMOGENEIDAD SHANNON		1	2	3	4	5	6	7	8	CHR	9	10	11	12	13	14	BH	MDQ	
Amylibacter	E	95	0.7	0.71	0.46	0.68	0.6	0.73	0.64	0.75	0.68	0.83	0.82	0.9	0.97	0.93	0.9	0.66	0.33
		97	0.78	0.8	0.56	0.77	0.73	0.83	0.77	0.89	0.69	0.97	0.94	0.95	0.996	0.95	0.97	0.78	0.42
		99	0.93	0.92	0.7	0.88	0.85	0.94	0.87	0.99	0.74	0.997	0.99	0.99	0.998	0.99	0.99	0.92	0.54
	NE	95	0.68	0.67	0.47	0.66	0.6	0.72	0.6	0.7	0.52	0.8	0.76	0.89	0.93	0.96	0.93	0.61	0.31
		97	0.76	0.79	0.56	0.75	0.72	0.81	0.76	0.82	0.57	0.93	0.87	0.93	0.98	0.95	0.97	0.73	0.4
		99	0.92	0.91	0.7	0.87	0.84	0.93	0.85	0.98	0.65	0.996	0.97	0.99	0.996	0.99	0.995	0.9	0.53
NS5	E	95	0.34	0.49	0.48	0.68	0.71	0.54	0.49	0.55	0.36	0.51	0.47	0.41	0.48	0.35	0.68	0.6	0.91
		97	0.43	0.66	0.64	0.8	0.77	0.69	0.68	0.65	0.48	0.57	0.56	0.5	0.58	0.48	0.75	0.51	0.93
		99	0.65	0.82	0.82	0.9	0.89	0.85	0.85	0.74	0.65	0.72	0.72	0.65	0.71	0.58	0.89	0.61	0.96
	NE	95	0.32	0.49	0.49	0.68	0.68	0.54	0.46	0.51	0.3	0.48	0.46	0.42	0.46	0.31	0.65	0.5	0.85
		97	0.42	0.63	0.64	0.77	0.74	0.68	0.64	0.61	0.39	0.54	0.54	0.51	0.57	0.43	0.75	0.46	0.87
		99	0.66	0.78	0.84	0.89	0.87	0.84	0.82	0.71	0.56	0.66	0.69	0.65	0.73	0.56	0.89	0.57	0.92
OM1	E	95	0.07	0.1	0.09	0.14	0.07	0.11	0.09	0.12	0.02	0.13	0.08	0.34	0.25	0.24	0.27	0.16	0.66
		97	0.1	0.13	0.12	0.17	0.1	0.13	0.12	0.15	0.04	0.16	0.11	0.48	0.38	0.38	0.42	0.19	0.66
		99	0.23	0.24	0.24	0.28	0.22	0.23	0.21	0.26	0.12	0.25	0.2	0.67	0.58	0.57	0.58	0.35	0.68
	NE	95	0.07	0.09	0.09	0.13	0.07	0.1	0.09	0.12	0.02	0.13	0.08	0.34	0.24	0.24	0.25	0.11	0.32
		97	0.1	0.12	0.12	0.16	0.1	0.13	0.12	0.15	0.04	0.16	0.11	0.48	0.35	0.37	0.4	0.14	0.41
		99	0.22	0.24	0.24	0.27	0.21	0.23	0.21	0.24	0.11	0.24	0.2	0.67	0.56	0.56	0.56	0.28	0.51
SAR11	E	95	0.24	0.31	0.26	0.3	0.24	0.27	0.25	0.26	0.1	0.26	0.3	0.24	0.25	0.24	0.25	0.15	0.14
		97	0.33	0.39	0.37	0.4	0.34	0.36	0.34	0.32	0.13	0.31	0.38	0.3	0.29	0.3	0.31	0.21	0.21
		99	0.6	0.61	0.63	0.66	0.61	0.62	0.59	0.56	0.37	0.58	0.6	0.5	0.47	0.5	0.47	0.47	0.41
	NE	95	0.22	0.26	0.26	0.29	0.24	0.26	0.25	0.24	0.08	0.24	0.29	0.23	0.23	0.24	0.24	0.14	0.12
		97	0.33	0.37	0.38	0.39	0.33	0.36	0.33	0.3	0.12	0.3	0.37	0.3	0.28	0.29	0.29	0.2	0.2
		99	0.58	0.59	0.63	0.64	0.59	0.61	0.58	0.54	0.32	0.56	0.6	0.5	0.46	0.5	0.47	0.45	0.38
SAR86	E	95	0.52	0.31	0.62	0.58	0.56	0.6	0.4	0.27	0.58	0.3	0.27	0.52	0.51	0.31	0.49	0.37	0.33
		97	0.55	0.45	0.66	0.65	0.68	0.67	0.47	0.45	0.63	0.5	0.44	0.65	0.72	0.43	0.65	0.42	0.38
		99	0.78	0.71	0.91	0.88	0.88	0.91	0.82	0.82	0.82	0.81	0.85	0.89	0.9	0.87	0.91	0.73	0.8
	NE	95	0.56	0.29	0.61	0.54	0.54	0.56	0.37	0.24	0.42	0.35	0.38	0.53	0.49	0.31	0.53	0.36	0.32
		97	0.55	0.39	0.65	0.62	0.67	0.62	0.45	0.35	0.49	0.55	0.51	0.63	0.65	0.42	0.67	0.38	0.33
		99	0.84	0.73	0.9	0.86	0.88	0.87	0.82	0.74	0.74	0.83	0.88	0.89	0.86	0.85	0.9	0.7	0.76

E: Ecualizado, NE: No ecualizado. Los números corresponden a la serie de muestras BOPD, el prefijo se retiró para una mejor comprensión de la tabla

Tabla A-11: Número de OTUs totales de cada LUP, especificada por nivel de agrupamiento, con datos ecualizados y sin ecualizar.

	Número de OTUs Totales					
	Ecualizado			No Ecualizado		
	95	97	99	95	97	99
Amylibacter	789	1523	3175	1097	2176	4886
NS5	613	1334	3414	931	2110	5917
OM1	1087	2344	8585	1579	3516	13757
SAR11	944	2516	14238	1363	3833	24420
SAR86	195	445	1618	314	719	3010