



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Efectos de la diversidad del plancton sobre el funcionamiento del ecosistema marino

Tesis presentada para optar al título de Magíster en Explotación de
Datos y Descubrimiento del Conocimiento

Osvaldo Burastero
Lic. Ciencias Biológicas

Directores

Federico Ibarbalz y Pedro Flombaum

Fecha de defensa

Diciembre 2022

Esta página se ha dejado en blanco intencionalmente

0.1. Resumen

Efectos de la diversidad del plancton sobre el funcionamiento del ecosistema marino

Comprender la relación entre la biodiversidad y la productividad de un ecosistema representa un desafío vigente. Experimentos controlados en ecosistemas terrestres indican que, en general, la productividad primaria y la diversidad de especies vegetales se asocian de forma positiva y que ambas se ven afectadas por la disponibilidad de recursos y las condiciones ambientales. En ecosistemas oceánicos, la biodiversidad y la productividad son determinadas por las comunidades microscópicas del plancton. La estructura y dinámica del ecosistema oceánico sumadas a las altas tasas reproductivas y de dispersión de sus microorganismos podrían alterar la relación entre diversidad y productividad. De acuerdo a la teoría ecológica, los mecanismos por los cuales la productividad de un ecosistema podría aumentar a medida que aumenta la biodiversidad son tres: complementariedad de nicho (uso diferencial de recursos), las interacciones positivas entre especies, y el efecto de muestreo (si hay más especies, es más probable encontrar una que sea muy productiva). En este trabajo de Tesis se evaluó la relación biodiversidad-productividad por medio de aproximaciones observacionales, así como los posibles mecanismos subyacentes. Para ello se utilizaron diferentes datos de la expedición Tara Oceans (2009-2013), la cual recolectó datos metagenómicos y fisicoquímicos en los océanos del planeta. Nuestros resultados presentan evidencia preliminar, para el grupo de los autótrofos, de una asociación significativa entre la abundancia de organismos y diversidad, y de interacciones positivas en base a la fijación de nitrógeno. A su vez, nuestros análisis basados en diversidad funcional proveen por primera vez evidencia preliminar de complementariedad de nicho en el grupo de las bacterias heterótrofas marinas.

Palabras clave: Biodiversidad, Productividad, Ecosistemas, Plancton, Efecto de muestreo, Complementariedad de Nicho, Interacciones positivas, Tara Oceans.

0.2. Abstract

Effects of plankton diversity on the functioning of the marine ecosystem

Understanding the relationship between biodiversity and productivity is still a major challenge. Controlled experiments in terrestrial ecosystems have shown that the primary productivity and the diversity of plants is positively associated, and that both variables depend on the resources availability and environmental conditions. In ocean ecosystems, the biodiversity and productivity are determined by the plankton microscopic communities. The structure and dynamics of the ocean ecosystems, together with the high reproductive and dispersion rates of microorganisms could alter the biodiversity-productivity relationship. According to the theory of ecology, there are three mechanisms by which productivity and biodiversity could be positively correlated: niche complementarity (differential use of resources), positive interactions between species, and sampling effect (if there are more species, it is more likely to find one that is very productive). In this work, we evaluated the biodiversity-productivity relationship and underlying mechanisms using observational data. For this purpose, we used metagenomic and environmental data from the Tara Oceans expeditions (2009-2013). Our results present preliminary evidence, in the case of the autotrophs, of a positive association between abundance of organisms and biodiversity, and of positive interaction based on Nitrogen fixation. Moreover, our analysis based on functional diversity provides, for the very first time, preliminary evidence of niche complementarity in the group of marine heterotrophic bacteria.

Keywords: Biodiversity, Productivity, Ecosystems, Plankton, Sampling effect, Niche Complementarity, Positive interactions, Tara Oceans.

Glosario

Capa / zona fótica	Capa superficial de un cuerpo de agua donde penetra suficiente luz para permitir que se produzca fotosíntesis
DAE	Distribución de abundancia de especies
Filo	Categoría taxonómica que se ubica por encima de la clase y por debajo del reino.
Genoma	ADN de una célula de una especie (incluyendo los genes y las partes no codificantes)
Genotipo	Constitución genética completa de un individuo
Metagenoma	Conjunto de genes presentes en un sistema determinado
Metatranscriptoma	Conjunto de genes transcritos en un sistema determinado
MAG	<i>Metagenome-assembled genomes</i>
mOTUs	<i>Phylogenetic marker gene based operational taxonomic units</i>
Productividad	Producción de biomasa por unidad de tiempo y área
Productividad primaria	Tasa de conversión de la energía solar en materia orgánica mediante fotosíntesis

0.1. Abstract

0.2. Resumen

Capítulo 1 - Introducción

- 1.1. El ecosistema marino y su comunidad de microorganismos
- 1.2. Productividad de un ecosistema
- 1.3. Ensamble de una comunidad
- 1.4. Abundancia de las especies en una comunidad
- 1.5. Expedición Tara Oceans
- 1.6. Objetivos de la tesis
- 1.7. Organización de la tesis

Capítulo 2 - Métodos

- 2.1. Descripción general de los datos
 - 2.1.1. Metagenome-Assembled Genomes (MAGs)
 - 2.1.2. Metagenomic operational taxonomic units (mOTUs)
 - 2.1.3. Conteo celular por citometría de flujo
 - 2.1.4. Variables fisicoquímicas
- 2.2. Análisis de asociación entre variables
 - 2.2.1. Variables fisicoquímicas
 - 2.2.1. Variables fisicoquímicas y productividad
- 2.3. Biodiversidad
 - 2.3.1. Esfuerzo de muestreo
 - 2.3.2. Índices de biodiversidad
- 2.4. Expresión génica
- 2.5. Similitud funcional
- 2.6. Ensamble de una comunidad planctónica

2.7. Software.

Capítulo 3 - Comprensión de las comunidades microbianas marinas

- 3.1. Variables contextuales
- 3.2. Variables contextuales y productividad
- 3.3. Biodiversidad
- 3.4. Ensamble de una comunidad planctónica
- 3.5. Distribución de abundancia de especies

Capítulo 4 - Funcionamiento de los ecosistemas microbianos marinos

- 4.1. Biodiversidad y productividad
- 4.2. Mecanismo de facilitación
- 4.3. Complementariedad de nicho
- 4.4. Efecto de muestreo

Capítulo 5- Conclusiones

- 5.1. Conclusiones específicas
- 5.2. Conclusiones generales

Agradecimientos

Índice de Figuras

Índice de Tablas

Esta página se ha dejado en blanco intencionalmente

Capítulo 1

Introducción

1.1. El ecosistema marino y su comunidad de microorganismos

El océano representa aproximadamente el 70 % de la superficie terrestre y contiene el 97 % del agua de nuestro planeta. Allí, la vida es principalmente microscópica, o microbiana. Enormes y diversas poblaciones de organismos unicelulares, virus y animales diminutos, colectivamente como plancton, se trasladan pasivamente. De acuerdo a su fuente de energía, los microorganismos se diferencian entre heterótrofos, también llamados quimiótrofos por depender de la energía química, y fotoautótrofos, que necesitan de la energía solar. Los metabolismos representados en la comunidad microbiana global impactan directamente sobre el clima a través de su contribución a los ciclos biogeoquímicos, es decir, a los movimientos de materia y energía entre los distintos componentes del sistema Tierra. Los microorganismos del plancton son responsables de cerca de la mitad de la producción primaria neta a nivel mundial, y de exportar una parte de este carbono fijado por fotosíntesis al océano profundo. Así, además, se convierten en la base de la cadena alimenticia en el océano.¹⁻⁴

Los factores abióticos juegan un rol crucial en la estructuración del ecosistema marino. Entre los factores más importantes se encuentran los niveles de luz solar, la temperatura, la salinidad, las concentraciones de oxígeno y nutrientes disueltos, y la proximidad con el continente. Por ejemplo, según la cantidad de luz recibida, en la columna de agua se diferencian tres zonas desde la superficie hasta el fondo: la zona fótica, que concentra a la mayor parte de la vida marina y se extiende hasta los 200 metros desde la superficie; la zona disfótica, que va desde los 200 hasta los 1000 metros; y la zona afótica, sin luz solar.

1.2. Productividad de un ecosistema y su relación con la biodiversidad

La productividad (o tasa de generación de biomasa por unidad de superficie y tiempo) es una manifestación central del funcionamiento de un ecosistema. La misma se encuentra íntimamente ligada al ciclo de carbono, uno de los ciclos biogeoquímicos más importantes por la magnitud de los flujos entre sus componentes.⁵ Dado que la biota juega un rol principal en el ciclado de materia y el flujo de energía, se espera que cambios en la biodiversidad alteren la productividad en particular y el funcionamiento de los ecosistemas en general. En tiempos en que las actividades humanas están produciendo cambios drásticos sobre el planeta, resulta esencial comprender el rol de la biodiversidad sobre la productividad de los ecosistemas.⁶

A veces, no se cuenta con mediciones precisas de la productividad (sobre todo en ecosistemas marinos) y esta debe ser reemplazada por aproximaciones. Una de las alternativas más comunes es utilizar la biomasa total presente una muestra, o en su defecto, la abundancia total de organismos. Si se conoce la biomasa de cada organismo, es posible inferir la biomasa total.

Un caso especial de productividad es la llamada productividad primaria, es decir, la velocidad con la cuál la energía lumínica es convertida a sustancias orgánicas por organismos fotosintéticos. En ecosistemas marinos, su medición requiere incubaciones con radioisótopos a bordo de las embarcaciones. Se ha propuesto también que la productividad primaria podría ser aproximada mediante la concentración de clorofila, una variable que representa de alguna manera la biomasa del fitoplancton. Debe tenerse en cuenta, sin embargo, que la concentración de clorofila varía con la intensidad de luz solar.

Una de las hipótesis centrales en ecología que vincula la biodiversidad y el funcionamiento de ecosistemas postula que a medida que aumenta la biodiversidad, aumenta la tasa de funcionamiento de los ecosistemas (Figura 1.1.).⁷ Los mecanismos biológicos para esta hipótesis son tres: (i) la complementariedad de nicho, (ii) las interacciones positivas entre especies y (iii) el efecto de muestreo. La complementariedad de nicho se refiere al uso diferencial de recursos por parte de

las especies o genotipos, por lo que en ecosistemas ricos en especies se utilizan más recursos que en los ecosistemas pobres en especies, lo que resulta en mayores tasas de funcionamiento.⁸ Las interacciones positivas como la simbiosis y la facilitación aumentan la capacidad de producción de la/las especie/s involucrada/s. Es decir, las especies que interactúan positivamente producen más cuando coexisten que cuando están separadas.^{9,10} Un ejemplo clásico de facilitación es la presencia de especies fijadoras de nitrógeno. Por último, el efecto de muestreo está asociado con la existencia de especies con características que permiten hacer el mayor uso de los recursos. En este caso, el aumento de la tasa de funcionamiento se debe a que a medida que aumenta la diversidad de especies en un ecosistema es más probable encontrar dicha especie.¹¹

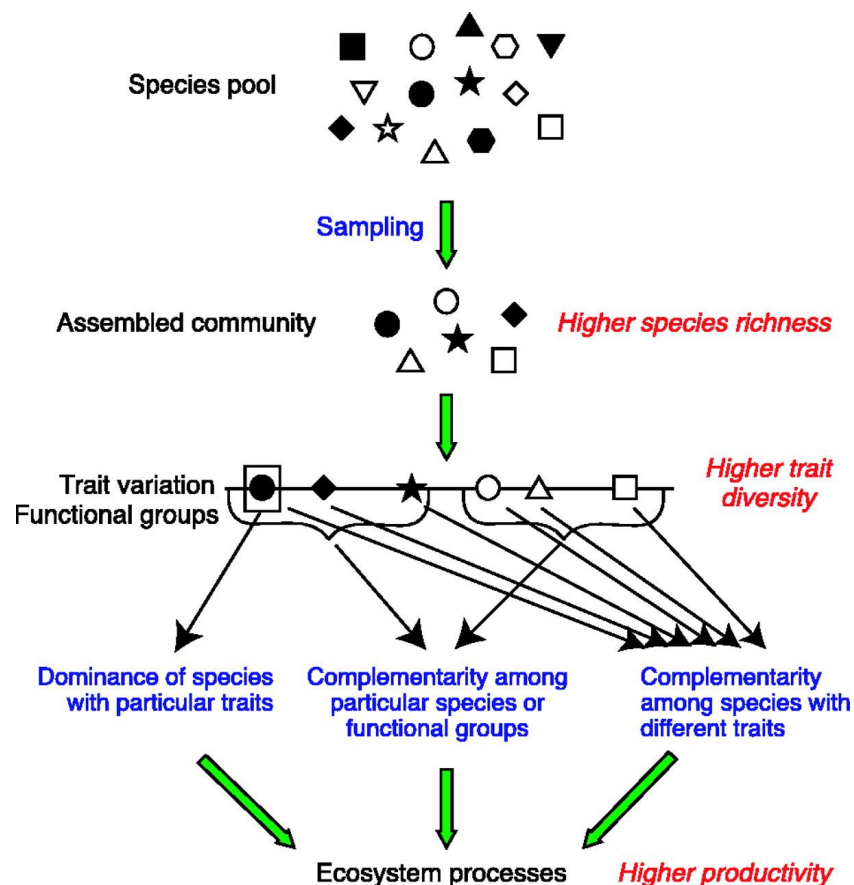


Figura 1.1. Tomada de Loreau et al., 2001.¹² Mecanismos hipotéticos que correlacionan positivamente biodiversidad con productividad. El efecto de muestreo está involucrado en el ensamble de la comunidad, de modo que las comunidades con más especies tienen mayor probabilidad de tener una especie muy productiva. La complementariedad se puede dar por dos razones, debido al uso diferencial de recursos, o debido a la interacción positiva.

Los mecanismos antes mencionados se pueden agrupar en dos clases. Por un lado, la facilitación y diferenciación de nicho son procesos locales determinísticos que incrementan la productividad de comunidades sobre lo esperado conforme a la productividad de cada especie. Por otro lado, el efecto de muestreo involucra un proceso estocástico en el cual la dominancia de especies altamente productivas conlleva a un incremento en la productividad. Un sistema con más especies, tiene más probabilidad de contener una especie altamente productiva, y por eso mayor biodiversidad implica mayor productividad.^{11,13,14}

Para poner a prueba estas hipótesis y los mecanismos subyacentes se han realizado experimentos controlados—mayormente en comunidades vegetales de ecosistemas terrestres—en los que se varía el número y la composición de especies y se mantienen constante el resto de las variables. Un proyecto único de este estilo fue BIODDEPTH (*Bio-Diversity and Ecological Processes in Terrestrial Herbaceous ecosystems: experimental manipulations of plant communities*), en el cual, a lo largo de dos transectas que cruzan Europa, se crearon ocho comunidades artificiales de plantas nativas con variaciones en la diversidad.¹⁵ El objetivo de la metodología era generar un gradiente de especies que iba desde el monocultivo hasta el número de especies esperado para un pastizal cercano sin manipular de 4 m². Un posterior análisis global multisitio reveló que las relaciones entre diversidad y productividad variaban a de acuerdo al sitio, pero que existía un patrón general descrito por una relación lineal entre la riqueza de especies (en escala logarítmica con base 2) y la biomasa.¹⁶ Además, hallaron efectos significativos de complementariedad pero no del efecto de muestreo (presencia de especies dominantes).¹⁶

Otro experimento controlado que provee evidencia de complementariedad de nicho y/o facilitación fue publicado por Jasper van Ruijven *et al.*, 2005.¹⁷ Los autores analizaron la relación biomasa-diversidad mediante el seguimiento de parcelas con combinaciones de ocho plantas distintas (gramíneas y dicotiledóneas) a lo largo de cuatro años. Como resultado, hallaron una asociación positiva entre ambas variables a partir del segundo año y evidencia de “sobre rendimiento” (*overyielding*),

es decir, de una mayor productividad de los cultivos mixtos respecto a la suma de los monocultivos. A su vez, descubrieron para cinco especies una asociación positiva entre la cantidad de biomasa por gramo de nitrógeno y el número de especies totales, lo cual indica un uso más eficiente de recursos a mayor diversidad.

Recientemente, se publicó una revisión de trabajos (*review*) que propone a la facilitación como un factor crucial en el funcionamiento de los ecosistemas y, al mismo tiempo, expone su extrema complejidad.¹⁸ Entre los trabajos que se mencionan, se encuentra el de Mulder et al. (2001), donde testean explícitamente la hipótesis de gradiente de estrés mediante el ensamblaje de comunidades de musgo de Nueva Zelanda en condiciones de humedad limitante y exposición solar alta (caso A), o humedad alta y baja exposición solar (caso B).^{19,20} En el primer caso (caso A) sí encontraron relaciones entre la diversidad y la productividad, pero no en el segundo (caso B). Otro trabajo con conclusiones similares es el publicado por Wright et al. (2021), que demuestra que las especies de pastizales sensibles a sequía están protegidas de los impactos negativos de la sequía cuando crecen en comunidades de mayor diversidad.²¹

También existen trabajos observacionales para investigar los mecanismos de interés. Dichos trabajos en este contexto estudian la productividad en gradientes naturales de biodiversidad, y observan otras variables asociadas que frecuentemente co-varían con las primeras dos. El aporte de este tipo de experimentos es que agregan realidad, aunque no es simple aislar efectos o mecanismos. Un meta-análisis reciente sugirió que los estudios observacionales también dan cuenta del efecto de la biodiversidad sobre la productividad en una multiplicidad de ambientes que incluye pastizales, bosques y cuerpos de agua dulce.²² Los estudios observacionales también permiten entender cuáles son los principales mecanismos. En 2017, Flombaum et al. publicaron un artículo que sugiere la existencia del efecto de muestreo en ecosistemas invadidos.²³ Los autores analizaron la relación entre productividad y diversidad en los bosques subtropicales del noreste argentino que se encuentra invadido por la especie *Ligustrum lucidum* (proveniente de Asia). Hallaron que los sistemas invadidos

triplicaron su biomasa, sin importar el número de especies, y que los monocultivos de *L. lucidum* eran más productivos que cualquiera de las comunidades vírgenes ricas en especies.

Es importante notar que estos mecanismos interaccionan entre sí, y que la productividad puede ser afectada a su vez por variables fisicoquímicas. En este sentido, la productividad y biodiversidad son variables respuesta fuertemente dependientes de las condiciones climáticas y toda relación correlacional se debe suponer en principio no causal.²⁴⁻²⁸ Dicho de otra manera, los factores abióticos suelen ser los principales impulsores de variaciones en procesos de los ecosistemas a lo largo de gradientes ambientales.¹²

En ecosistemas marinos/oceánicos, en donde la productividad primaria está determinada por el fitoplancton (microalgas unicelulares), la evidencia es muy limitada.²² A pesar de que existen numerosos experimentos que ponen a prueba las hipótesis mencionadas, los mismos pertenecen casi exclusivamente a ecosistemas terrestres. Comprender la relación entre la biodiversidad y la productividad en ecosistemas marinos es importante porque nos permite comprender si los principales mecanismos observados en ecosistemas terrestres se repiten en el ecosistema más grande del planeta, el océano, y con organismos unicelulares.

1.3. Ensamble de una comunidad

En ecología de comunidades, una de las preguntas fundamentales es qué factores determinan la biodiversidad a nivel local. En definitiva, se trata de factores que afectan el ensamble de la comunidad a diferentes escalas temporales. Se han hipotetizado numerosos mecanismos al respecto, siendo todos agrupables en cuatro procesos distintos, la selección ambiental, la deriva ecológica, la diversificación *in situ* y la dispersión.²⁹

El concepto de selección aplicado a especies en comunidades requiere dos cambios de referencia respecto al marco tradicional. Primero, se define como población a la

comunidad que contiene individuos de muchas especies diferentes, y segundo, el fenotipo bajo selección corresponde a la identidad de las especies. Así como la selección puede favorecer el alelo "A" sobre el alelo "a" dentro de una población de especies, la selección podría favorecer la especie "X" sobre la especie "Y" en una comunidad. Existen tres formas relevantes de selección: constante, dependiente de densidad, y temporal o espacialmente variable. La selección constante es simple, si la aptitud relativa es constante en el tiempo e independiente de la densidad de las especies, la especie con mayor aptitud predominará. La selección dependiente de densidad implica que la aptitud de una dada especie depende parcialmente de la densidad de dicha especie, así como también de la densidad de otras especies. Este tipo de selección se relaciona con las posibles interacciones entre especies, que incluyen la predación, competencia y mutualismo. Por último, dado que la aptitud relativa de las especies depende de los factores abióticos, la selección puede variar a lo largo del espacio o tiempo.

La deriva ecológica es un componente estocástico causado por el hecho de que una comunidad posee un número finito de individuos, y el nacimiento y muerte de nuevos integrantes son procesos no determinísticos. En una comunidad cerrada, si todos los individuos fueran idénticos en aptitud, la deriva sería el único impulsor de la dinámica comunitaria. La diversificación *in situ*, o especiación, se refiere a la aparición de nuevos tipos de individuos. En el caso de los microbios, este proceso requiere de escalas temporales cortas debido a sus altas tasas de crecimiento.³⁰ Luego, la dispersión involucra el movimiento de individuos a lo largo del espacio y posee un efecto que depende del tamaño y la composición de las comunidades desde donde inmigran y hacia donde migran los individuos. Respecto a los tamaños relativos de las comunidades fuente y receptora se distinguen dos extremos de dispersión. Los modelos de islas continentales asumen dispersión direccional desde una comunidad de tamaño infinito (continente) a una comunidad local (las islas) más pequeña con una dinámica relativamente más rápida. Por el contrario, los modelos de islas suponen una red de pequeñas comunidades vinculadas por la dispersión entre ellas. Esta red de comunidades, denominada metacomunidad, es probablemente el caso de las comunidades planctónicas.

1.4. Distribución de la abundancia de las especies en una comunidad

Uno de los patrones más documentados de diversidad microbiana en los últimos años es la “biosfera rara”, que describe como la gran mayoría de los taxones en una muestra ambiental están representados por pocos individuos.^{31,32} Este fenómeno parece ser un patrón ecológico fundamental y se lo estudia dentro del marco de la distribución de abundancia de especies (DAS , o SAD del inglés *species abundance distribution*) .³³

La DAE, uno de los descriptores más básicos de una comunidad ecológica, describe la abundancia (número de individuos observados) para cada especie distinta dentro de una comunidad (Figura 1.2.). Típicamente, si se grafica un histograma del número de especies en el eje “y” contra la abundancia en el eje “x”, se observa una curva hiperbólica, indicando unas pocas especies abundantes y muchas especies raras (Figura 1.2. a). Esta ley parece ser universal, y no se conocen comunidades que la contradigan, desde comunidades bentónicas marinas hasta de la selva amazónica. Cuando se grafica la abundancia transformada con logaritmo, aparece mayor variabilidad en la forma (Figura 1.2. b). A menudo, la DAE se presenta como un diagrama de ranking-abundancia, con lo cuál se evita perder información debido al agrupamiento en clases (Figura 1.2. c, e). Una característica importante de la DAE es que las especies no tienen clases, con lo cual se pueden comparar comunidades sin especies en común.

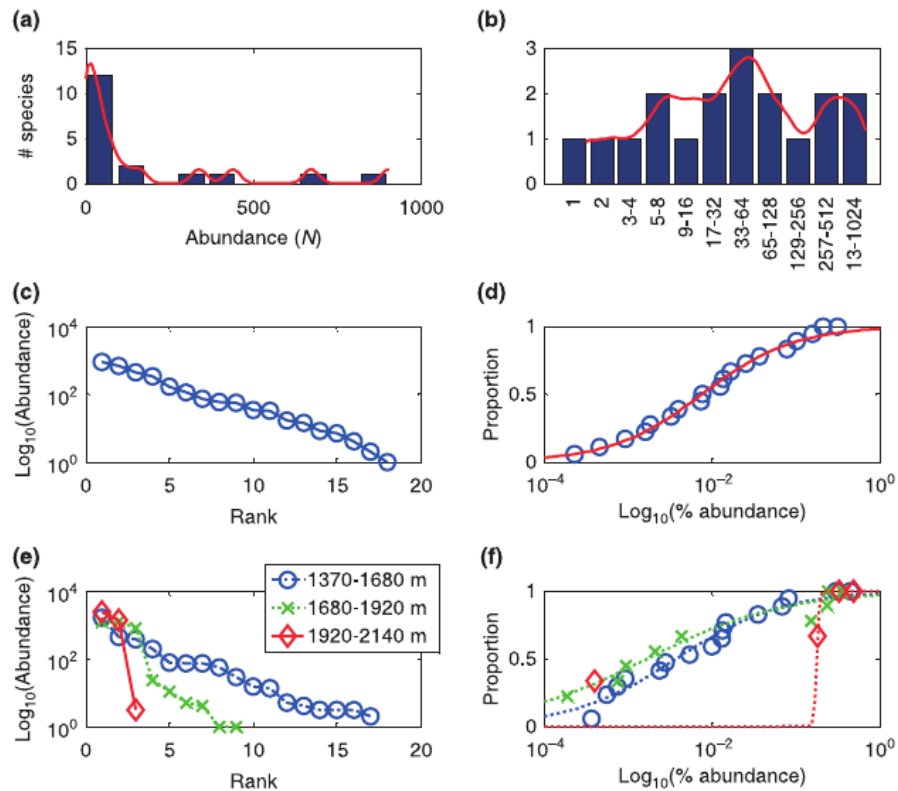


Figura 1.2. Tomada de McGill *et al.*, 2007.³³ Distintas maneras de graficar SADs. Los datos de abundancia de árboles recolectados por Whittaker en las montañas Siskiyou son graficados en tres formatos distintos. (a) Histograma del número de especies versus abundancia en una escala aritmética. (b) Histograma con la abundancia en escala logarítmica (de base 2). (c) Un gráfico de ranking-abundancia (RAD, *rank-abundance distribution*) en escala logarítmica (base 10). (d) Distribución empírica acumulada (ECDF, *empiric cumulative distribution function*) ajustada por regresión logística no lineal. (e) Un gráfico de RAD para distintas elevaciones mostrando diferentes patrones. (f) Las mismas tres elevaciones graficadas como ECDF.

La DAE es central para las teorías de biodiversidad y macroecología, que buscan comprender los patrones en abundancia y distribución a lo largo de distintas escalas espaciales y temporales.³³ En efecto, algunas de estas teorías predicen patrones intensivamente estudiados tales como la curva de especies-área que relaciona el área de un hábitat con el número de especies presentes, o la relación distancia-decaimiento, que implica que la similaridad biológica decrece con la distancia.³⁴

Se han desarrollado más de 20 modelos para predecir la DAE, siendo algunos puramente estadísticos y otros basados en principios de diversas teorías.^{33,35-37} De todos los modelos, los más exitosos poseen distribuciones lognormales (o de series logarítmicas) y sirven como el estándar contra el cuál se prueban otros modelos. Dicha distribución se caracteriza por una distribución de frecuencias sesgada a la derecha que se vuelve normal cuando se aplica la transformación con logaritmo. Se ha sugerido que la forma lognormal de la DAE resulta de procesos ecológicos multiplicativos tales como el crecimiento y las interacciones estocásticas.³⁸

A pesar de que la curva lognormal es uno de los modelos más exitosos, no es predicho por ninguna teoría general de biodiversidad y raramente se usa en estudios microbianos.³⁹⁻⁴³ Por otra parte, la forma de series logarítmicas sí es predicha por uno de las teorías más recientes de biodiversidad, llamada la “Teoría ecológica de máxima entropía” (*maximum entropy theory of ecology*, METE).³⁷ La METE establece que el patrón ecológico esperado es aquel que puede ocurrir en el mayor número de formas para un conjunto dado de restricciones y se basa en el principio de máxima entropía.⁴⁴ La METE usa solamente el número de especies y el número total de individuos para predecir la DAE. Se ha demostrado, con datos a escala global de plantas y animales, que la METE puede superar a la lognormal y explicar la variación en abundancia entre y dentro de comunidades.^{45,46}

Por otro lado, recientemente se evaluó la capacidad explicativa del modelo lognormal, de las series logarítmicas y dos modelos bien conocidos: uno que predice una forma muy desigual (distribución de Zipf), y uno que predice una forma muy uniforme.⁴⁷ Para ellos se utilizó una compilación de ARN ribosomal 16S de 20.000 ubicaciones geográficas únicas. En contraste con estudios anteriores, se halló que la lognormal explicaba hasta el 94 % de la variación entre y dentro de las DAE microbianas, comparado con un 91 % usando la distribución de Zipf y un 64 % usando series lognormales predichas por la METE. Los autores propusieron como explicación la manera en que se muestrean las comunidades de micro y macroorganismos, dado que en el caso de los microbiomas, la escala espacial

utilizada excede largamente la escala de las interacciones y tiende a juntar taxones ecológicamente muy distintos que no utilizan los mismos recursos o ocupan los mismos microhabitats.

1.5. Expedición Tara Oceans

En el ecosistema oceánico, la mayor parte de la biomasa, diversidad y abundancia (número de individuos) se encuentra en el plancton. A la detección remota por satélite y a las innumerables observaciones ópticas realizadas en cientos de expediciones (e.g. microscopía, citometría de flujo), en la última década se ha consolidado el estudio del plancton por medio de métodos basados en secuenciación masiva de ADN. Aquellos denominados “metagenómicos” consisten en recuperar el ADN total de una comunidad microbiana (es decir, el conjunto de genomas, el metagenoma), fragmentarlo en pequeñas moléculas y secuenciarlas en simultáneo o en paralelo. Este tipo de análisis está contribuyendo a la comprensión de la diversidad y el funcionamiento de los diferentes componentes del plancton, su rol ecológico dentro de la comunidad, y su posible devenir en ecosistemas cambiantes.

Recientemente, un consorcio público-privado se propuso recolectar muestras metagenómicas de forma estandarizada a una escala sin precedentes.⁴⁸ Entre 2009 y 2013, la goleta TARA surcó las aguas de todas las cuencas oceánicas (Figura 1.3.). Además de realizar una caracterización exhaustiva del hábitat a lo largo de la expedición, la estrategia consistió en separar la comunidad por tamaño mediante un sistema de filtros. Es así que se lograron establecer fracciones enriquecidas en virus, procariotas, y eucariotas unicelulares y multicelulares. El estudio de la fracción enriquecida en procariotas de vida libre fue uno de los primeros en ser procesado y ha resultado en la generación de un catálogo genético inmenso.^{49,50} Estos nuevos recursos confirmaron en repetidas ocasiones como la temperatura es capaz de estructurar esta comunidad.⁴⁹⁻⁵¹ Asimismo, el análisis focalizado de genes vinculados a la fotosíntesis permitió esclarecer la diversidad y biogeografía de cianobacterias y asociarlos a patrones de luz y nutrientes.^{52,53} Por último, en el

contexto de fenómenos oceanográficos singulares como la mezcla de aguas del Índico con las del Atlántico, o en procesos de fertilización natural alrededor de islas del Pacífico, este muestreo aclaró vías por las cuales las comunidades podrían estar respondiendo ante variaciones físicas y/o biogeoquímicas .^{54,55}

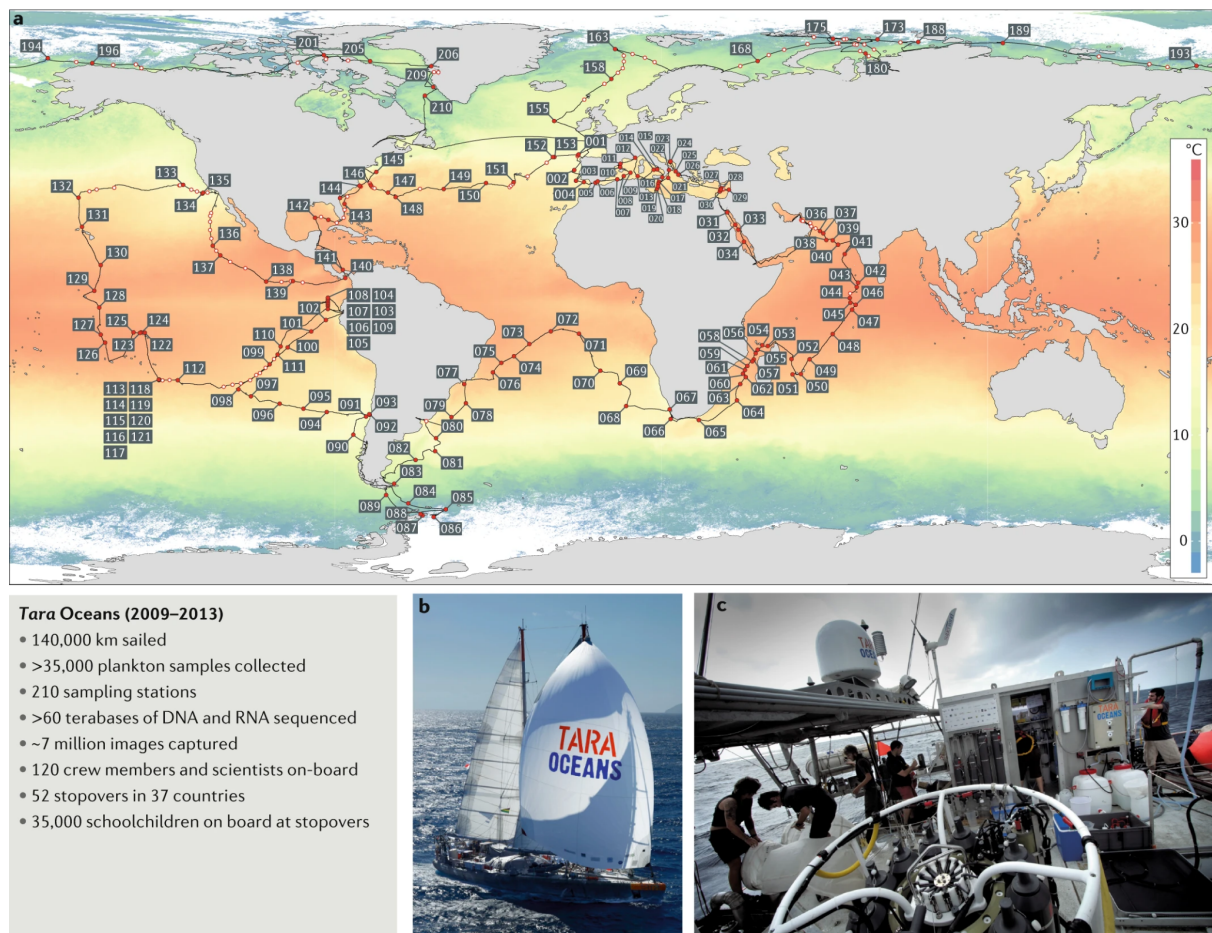


Figura 1.3. Tomada de Sunagawa *et al.* 2020.⁴⁸ a) Ruta de la goleta TARA en la expedición Tara Oceans desde 2009 hasta 2013, incluyendo 210 estaciones de muestreo. En total, más de 35000 muestras fueron recolectadas y archivadas. b) La goleta TARA de 36 m con casco de aluminio. c) Laboratorio montado dentro de la goleta para el muestreo del plancton.

1.6. Objetivos de la tesis

El objetivo del presente trabajo es estudiar, de forma observacional, el ensamble de comunidades marinas, los patrones de biodiversidad (DAS) y la relación entre la biodiversidad microbiana y el funcionamiento de ecosistemas en el ambiente oceánico. La hipótesis principal de trabajo es que a medida que aumenta la diversidad, aumentan los procesos del ecosistema, en particular para este trabajo, la

abundancia o el stock de biomasa de microorganismos. Los mecanismos detrás de este aumento serían la complementariedad de nicho, la facilitación, y el efecto de muestreo. Para poner a prueba esta hipótesis se propone utilizar datos metagenómicos, de citometría de flujo, y contextuales asociados a procesos ecosistémicos a lo largo de una transecta realizada en el océano global.

1.7. Organización de la Tesis

El muestreo global de Tara Oceans no fue diseñado para poner a prueba hipótesis de diversidad de las comunidades y la productividad del ecosistema. Es por ello que la Tesis presenta, en primera instancia, un análisis general de los datos en donde 1) se estudia la asociación entre las variables contextuales (físicoquímicas) y la variable “objetivo” que consiste en la abundancia de microorganismos, 2) se valida la convergencia en el muestreo de biodiversidad, y se remueven algunas estaciones del análisis principal (ver siguiente párrafo), 3) se investigan los factores que determinan el ensamblado de una comunidad, y 4) se evalúa la distribución de abundancia de especies en cada muestra. Todo esto se desarrolla en el Capítulo 3.

En segunda instancia (Capítulo 4), se trabaja sobre la relación entre la biodiversidad y el funcionamiento de los ecosistemas. 1) Se describe la relación entre la biodiversidad y productividad para los autótrofos y heterótrofos, 2) se presenta evidencia preliminar del mecanismo de facilitación mediante el análisis de genes fijadores de nitrógeno, 3) se investiga la complementariedad de nicho usando un índice de similaridad genómica, y 4) se busca comprender si el efecto de muestreo se encuentra presente.

Esta página se ha dejado en blanco intencionalmente

Capítulo 2

Métodos

2.1. Descripción general de los datos

Este trabajo hizo foco en muestras tomadas del océano por el consorcio Tara Oceans.⁴⁹ Para estudiar el gradiente de biodiversidad del plancton marino se utilizaron genomas ensamblados a partir de metagenomas (MAGs, del inglés *metagenome-assembled genomes*) y unidades taxonómicas operacionales definidas por medio de genes marcadores (mOTUs, del inglés *phylogenetic marker gene (MG)-based operational taxonomic units*), también obtenidas de metagenomas.^{56,57} Por otro lado, el funcionamiento de los ecosistemas, que estaría representado por la productividad primaria (cantidad de carbono fijado / hora, por ejemplo), fue reemplazado por el conteo de células por citometría de flujo, considerado como una aproximación. Además, el análisis del contenido genético de los MAGs y de datos suplementarios provenientes de metatranscriptomas y variables contextuales de cada lugar de muestreo permitió analizar cuál de los mecanismos domina en el ecosistema marino. Nos centramos en las muestras de la superficie, aunque algunas de las estaciones cuentan también con muestras más profundas a la altura del valor máximo de concentración de clorofila. Ambas pertenecen a la capa fótica del mar.

2.1.1. Genomas ensamblados a partir de metagenomas (*Metagenome-Assembled Genomes*, MAGs)

El dataset consiste en 957 MAGs, de los cuales 820 son bacterianos, 72 MAGs de picoeucariotas y 65 MAGs de arqueas. Los MAGs junto a su respectivo valor de abundancia relativa en 93 muestras metagenómicas (61 a nivel superficial y 32 a un nivel de profundidad mayor) fueron obtenidos por Delmont et al., 2018.⁵⁶ La clasificación taxonómica de los MAGs alcanza el nivel de Filo en un 93 % de los casos, de Clase en un 78 %, de Orden en un 57 %, de Familia en un 33 %, y de género

en un 10 %. El 75 % de los MAGs tienen una completitud (estimada mediante el software Anvio)⁵⁸ mayor al 70 %.

2.1.2. Unidades taxonómicas operativas metagenómicas (*Metagenomic operational taxonomic units*, mOTUs)

Para 81 muestras metagenómicas, se posee información de la abundancia relativa de aproximadamente 4000 unidades operacionales taxonómicas. Las mismas fueron obtenidas a partir de una base de datos de genes codificantes de tipo *housekeeping* de copia única (Milanese et al., 2019).

2.1.3. Conteo celular por citometría de flujo

Las cuentas de abundancia celular por unidad de volumen están asociadas a tres grupos de organismos distintos: heterótrofos (bacterias y arqueas), bacterias autótrofas (fotoautótrofas, principalmente cianobacterias de los géneros *Prochlorococcus* y *Synechococcus*) y picoeucariotas fotosintéticos. Por cada estación de muestreo se usó solamente una fracción de tamaño (0.22-1.6 μm o 0.22-3 μm). Además, se cuenta, para algunas estaciones, con las abundancias de las bacterias autótrofas separadas por grupo (*Prochlorococcus* y *Synechococcus*).⁵¹

2.1.4. Variables fisicoquímicas

Cada metagenoma está asociado a variables contextuales del punto de muestreo que incluyen la latitud, longitud, temperatura, salinidad, concentración de clorofila, nitrito/nitrato, oxígeno, entre otros (Tabla 2.1.).

Tabla 2.1. Descripción de las variables fisicoquímicas asociadas a cada muestra.

Parámetro	Detalle
Temperatura	En °C.

Salinidad	Propiedad basada en la conductividad del agua. Medida en PSU (del inglés , <i>Practical Salinity Unit</i>).
Oxígeno (O ₂)	Concentración de O ₂ disuelto en µmol/kg.
Nitratos y nitritos (NO ₂ ²⁻ y NO ₃ ³⁻)	Concentración de NO ₂ ²⁻ y NO ₃ ³⁻ disuelto en µmol/L.
Fosfato (PO ₄)	Concentración de PO ₄ disuelto en µmol/L.
Silica (Si)	Concentración de Si disuelta en µmol/L.
Radiación fotosintética disponible (PAR, <i>Photosynthetically Available Radiation</i>)	En Einsteins*metros ⁻² *días ⁻¹ . Cantidad de luz con longitud de onda en el rango de los 400-700 nm.
Lyapunov	En días ⁻¹ . Relacionado con la dinámica del agua. ⁵⁹
Okubo-Weiss	En segundos ⁻² . Relacionado con la capacidad de mezclado y turbulencia. ⁶⁰
Grad SST adv	En °C/100 km. Gradiente de temperatura de la superficie del mar.
Eddy retention	En días. Provee información acerca de la permanencia de una partícula dentro de un área de interés. ⁶¹
MLD (<i>Mean Depth MLD Sigma</i>)	En metros. Profundidad de la capa mixta.
Fluorescence	En metros. Profundidad máxima en la cual se detecta fluorescencia por clorofila.
BV (<i>Mean Depth of the Maximum of Brünt-Väisälä Frequency</i>)	En metros. Relacionado con la estabilidad de la columna de agua.
Max O ₂	En metros. Profundidad con la concentración máxima de oxígeno disuelto.
Min O ₂	En metros. Profundidad con la concentración mínima de oxígeno disuelto.
Nitraclina (<i>Nitracline</i>)	En metros. Relacionado con el gradiente vertical de nitrógeno. El suministro ascendente de nitratos a la zona fótica es un limitante de la productividad primaria en una columna de agua estratificada. ⁶²
ChlorophyllA	Concentración de clorofila

2.2. Análisis de asociación entre variables

2.2.1. Variables fisicoquímicas

Para investigar la relación entre las variables abióticas, utilizamos el coeficiente Tau de correlación de Kendall. Brevemente, este parámetro toma valores entre -1 y 1 y valores positivos significan que el ranking de las variables se incrementa de modo conjunto. La hipótesis nula es que no existe asociación entre las variables bajo estudio.

Para estudiar relaciones no-monotónicas, calculamos la información mutua entre los pares de variable que, para variables continuas, se calcula como

$$mi(X, Y) = \int \int_{Y \ X} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$$

dónde $p(x, y)$ es la función de densidad de probabilidad conjunta de X e Y y $p(x)$, $p(y)$ son las funciones de densidad de probabilidad marginal de X e Y, respectivamente. De modo intuitivo, la información mutua mide en cuánto el conocimiento de una variable reduce nuestra incertidumbre sobre la otra y permite captar tanto relaciones lineales como no lineales. Dado que en la práctica no se conocen las distribuciones antes mencionadas, la información mutua debe ser estimada.

Para evaluar la significancia del valor estimado de información mutua, empleamos un test no paramétrico de permutación basado en que un p -valor es simplemente la probabilidad de obtener cierta métrica (o un valor más extremo) asumiendo que la hipótesis nula es verdadera. En este caso particular, el p -valor se calcula como la proporción de veces en las cuales la muestra aleatorizada tuvo un valor de información mutua mayor al estimado (una variable se deja fija, la otra variable se re muestrea sin reposición). En una primera instancia, realizamos 200 repeticiones. Luego, para aquellos resultados con ningún valor (o solo uno) de información mutua

de una muestra aleatorizada mayor o igual al de la muestra real nos extendemos hasta 5000 repeticiones (de modo que el mínimo valor distinto de 0 sea 2×10^{-4}). Los resultados con p -valores iguales o menores a 2×10^{-4} fueron explorados visualmente.

2.2.1. Variables fisicoquímicas y productividad

Con el fin de investigar la relación entre la abundancia de microorganismos determinada por citometría de flujo (aquí nuestra variable *proxy* para la productividad), y las variables fisicoquímicas, estimamos la información mutua entre cada variable fisicoquímica y de abundancia. Luego, mediante el test de permutación (600 repeticiones) calculamos el p -valor. Las posibles relaciones significativas (p -valor < 0.1 , previa corrección por *False Discovery Rate*) fueron inspeccionadas visualmente.

2.3. Biodiversidad

2.3.1. Esfuerzo de muestreo

Para construir las curvas de acumulación de nuevos tipos de individuos versus proporción de individuos muestreados, usando el dataset de mOTUs, seguimos el siguiente protocolo.

- 1) Construir un vector *mOTUvec* donde cada mOTU aparece tantas veces como fue detectado. *mOTUvec* tiene largo n .
- 2) Cambiar al azar el orden de los elementos dentro de *mOTUvec*.
- 3) Construir dos vectores vacíos, *mOTUvecDiff* y *mOTUsObsProp*.
- 4) Arrancar un contador en 0.
- 5) Desde $i = 1 \dots n$, tomar el elemento i de *mOTUvec* y si no está presente en *mOTUvecDiff*, agregarlo y sumar uno al contador. Si el elemento no está presente en *mOTUvecDiff*, el contador no cambia. Luego, en la posición i de *mOTUsObsProp* se completa con el valor del contador.
- 6) Dividir *mOTUsObsProp* por n .

Construir el vector con la proporción de individuos muestreados es trivial. Este procedimiento se puede repetir varias veces para tener un error. En nuestro caso, lo repetimos 15 veces para construir las curvas de la Figura 3.4, donde se denota que el error es casi despreciable.

Para caracterizar analíticamente el esfuerzo de muestreo, ajustamos cada una de las curvas de acumulación a una función del tipo

$$\hat{y} = \frac{ax}{x+b}$$

donde el parámetro b es el valor de x en el cual y equivale a la mitad de su valor máximo, y a es el valor de la asíntota. En otras palabras, en qué proporción de individuos muestreados ya no aparecen individuos de nuevos tipos. Entonces, mientras más cercano el valor de b a 0, mejor fue el muestreo.

2.3.2. Índices de biodiversidad

Existen varias maneras de estimar la diversidad de una muestra.⁶³ Entre ellas encontramos a la riqueza, que es el número de especies distintas, el coeficiente de Hill de orden 1 (exponente de la entropía de Shannon, donde la abundancia relativa de cada especie es tomada en cuenta) o número efectivo de especies, el índice de Pielou y el índice de dominancia de Berger-Parker.

Tabla 2.2. Índices de biodiversidad y fórmulas asociadas.

Nombre	Fórmula
Riqueza	Número de especies distintas
Hill (Número efectivo de especies distintas)	$\exp(-\sum_{i=1}^R p_i \ln(p_i))$ donde p_i es la frecuencia relativa del tipo de especie i y R es el número total de especies distintas (Riqueza)

$$\left(\sum_{i=1}^R p_i = 1\right).$$

Pielou
(Equitatividad)

$$Shannon * \ln(riqueza)^{-1} \text{ donde}$$

$$Shannon = \ln(Hill) = - \sum_{i=1}^R p_i \ln(p_i).$$

Berger-Parker
(Dominancia)

N_{max}/N , donde N_{max} es el número de individuos de la especie más abundante y N el número total de individuos.

2.4. Modelos de distribución de abundancia de especies

Los siguientes modelos fueron ajustados a los datos mOTUS para investigar si podían predecir las DAE microbianas observadas. Los tres modelos están basados en la publicación de Shoemaker *et al.*, 2017.⁴⁷

Lognormal

Para evitar abundancias fraccionales y tener en cuenta el error de muestreo, se utilizó un modelo de Poisson basado en un muestreo lognormal (Poisson lognormal).⁶⁴ La ecuación que da la probabilidad de que una especie elegida al azar contenga n individuos está dada por $p(n)$, donde

$$p(n) = \int_0^{\infty} \lambda^n e^{-\lambda} n^{-1} p_{LN(\lambda)} d\lambda$$

donde λ es un parámetro con una media (μ) y desvío estándar (σ) que se ajusta mediante un estimador de máxima verosimilitud, y P_{LN} es la probabilidad lognormal.⁶⁵

Teoría de máxima entropía ecológica

El modelo METE utiliza dos parámetros empíricos como input para predecir la SAD: la riqueza de especies S y la abundancia total de individuos N . Para predecir la SAD,

el modelo asume que la forma esperada es aquella que puede ocurrir en el mayor número de maneras. La probabilidad de que la abundancia de una dada especie sea n está dada por

$$\Phi(n|S, N) = - e^{-\beta n} / (n \log(\beta))$$

donde β es un parámetro ajustado que representa un “coeficiente metabólico” y se deriva de la optimización numérica de

$$N / S = (\sum_{n=1}^N e^{-\beta n}) / (\sum_{n=1}^N e^{-\beta n} / n)$$

Es decir, β se obtiene minimizando $f(\beta)$:

$$f(\beta) = | (\sum_{n=1}^N e^{-\beta n}) / (\sum_{n=1}^N e^{-\beta n} / n) - N/S |$$

Distribución de Zipf

La distribución de Zipf es una ley de potencias que predice una de las formas de DAE con abundancias más desiguales. Está caracterizada por el parámetro γ , donde la frecuencia de un elemento de rango k está dada por

$$f(k; \gamma, N) = (1 / k^\gamma) / (\sum_{n=1}^N (1/n^\gamma))$$

El parámetro γ se obtiene usando un estimador de máxima verosimilitud.⁶⁵

Comparación de modelos

Para evaluar el ajuste de los tres modelos propuestos utilizamos el criterio de información de Akaike dado por

$$AIC = 2k - 2LL$$

donde LL es la probabilidad logarítmica (*log-likelihood*) de haber observado dichos datos (abundancias absolutas) dado los parámetros ajustados, y k es la cantidad de parámetros. En el caso de los modelos Poisson lognormal y Zipf, el valor de LL sale de la optimización. Para el modelo METE, dada una muestra particular, LL se obtiene usando

$$LL = \sum_{i=1}^S \log(\Phi(n_i))$$

donde S es el número total de especies, y n_i es la abundancia observada para la especie i . Si el valor de AIC de un modelo es dos o más unidades menor a otro, dicho modelo se considera significativamente mejor.

2.4. Expresión génica

Para 42 muestras se cuenta con datos metatranscriptómicos, es decir, con información sobre la identidad y el nivel de expresión de genes de la comunidad microbiana.⁵⁰ En particular, utilizamos los datos provistos por Salazar *et al.*, (2019) que incluyen la abundancia genómica normalizada ($Gnorm$), o número de copias de un gen por célula, y la abundancia transcriptómica, o número de transcritos de un gen por célula ($Tnorm$) de aproximadamente 9000 genes. Ambos valores fueron preprocesados mediante transformación a cuentas, estabilización de varianza y aplicación de logaritmo en base 2.⁵⁰

2.5. Similitud funcional

Para poder comparar la similitud funcional entre pares de MAGs utilizamos el coeficiente de Jaccard que tiene en cuenta presencia y ausencia de genes, y solamente aquellos MAGs con completitud mayor al 70 % (*anvio completeness*). La validación de este enfoque se realizó mediante la comparación entre clusters obtenidos por similitud genómica y la filogenia. Primero, del dataset de MAGs se seleccionaron todos aquellos correspondientes a filos (*phyla*) bacterianos con al menos 4 MAGs, y se calcularon todas las distancias “genómicas” entre pares de MAGs.

$$Distancia(A, B) = 1 - Jaccard(A, B) = 1 - \frac{A \cap B}{A \cup B}$$

donde A y B son los dos MAGs del par, la intersección representa los genes en común, y la unión los genes no compartidos más los genes compartidos. A partir de la matriz simétrica de distancias de 618*618, se construyó un árbol jerárquico con el método aglomerativo de *ward.D2*. La cantidad de clusters se seleccionó mediante validación interna optimizando el índice de Dunn, que resultó en 11 clusters.^{66,67} Posteriormente verificamos que los clusters obtenidos de modo no supervisado contengan elementos que concuerden con lo esperado de acuerdo a la taxonomía.

2.6. Ensamble de una comunidad planctónica

La similitud entre estaciones de muestreo se calculó mediante el índice de Jaccard, pero esta vez con el dataset de mOTUs. Para evaluar la relación entre las abundancias celulares de autótrofos y la composición (Figura 3.7), la matriz de similitud fue transformada en una de distancias ($1 - similitud$) y posteriormente reducida a 4 dimensiones usando escalamiento multidimensional métrico. El coeficiente de correlación R^2 de las distancias proyectadas usando las nuevas dimensiones y las distancias originales fue de 0.81. Posteriormente, mediante validación interna optimizando el índice de Dunn, cinco grupos fueron detectados usando el algoritmo *k-means*. Finalmente, la distribuciones de abundancia de los

diferentes grupos fueron comparadas usando ANOVA de Welch seguida del test post-hoc Dunnett T3 para comparaciones múltiples.

2.7. Software

Los análisis y gráficos de este trabajo de Tesis fueron realizados en lenguaje de programación R y por medio de los paquetes tidyverse, ggplot2, corrplot, varrank, xlsx, reshape2, mpmi, poilog, ggpubr, broom, sads y minpack.lm.^{65,68-78}

2.8 Disponibilidad de los datos

Diversas fuentes de datos fueron integradas. La descripción de las mismas se presenta a continuación. Es importante tener en cuenta que en cada caso distintas muestras fueron analizadas, con lo cual se pierde información al unir datasets.

Tabla 2.3 Fuentes de datos asociadas a la expedición Tara Oceans empleadas en este trabajo.

Link	Datos utilizados
http://ocean-microbiome.embl.de/data/OM.CompanionTables.xlsx	Descripción de las muestras de Tara Oceans. Variables fisicoquímicas asociadas. Abundancias celulares de heterótrofos, picoeucariotas y autótrofos. ⁴⁹
https://www.ebi.ac.uk/biostudies/files/S-BSST297/Files/OM-RGC_v2_functional_profile_KEGG.tar.gz	Abundancia funcional del gen nifH. ⁵⁰
https://prod-dcd-datasets-cache-zipfiles.s3.eu-west-1.amazonaws.com/p9r9wttjkm-2.zip	Abundancias celulares de <i>Synechococcus</i> y <i>Prochlorococcus</i> . ⁵¹
https://static-content.springer.com/esm/art%3A10.1038%2Fs41564-018-0176-9/MediaObjects/41564_2018_176_MOESM5_ESM.xlsx	Distribución, completitud, y clasificación taxonómica de los 957 MAGs no redundantes. ⁵⁶
https://static-content.springer.com/esm/art%3A10.1038%2Fs41564-018-0176-9	Anotación funcional para los 1077 MAGs redundantes. ⁵⁶

/MediaObjects/41564_2018_176_MOESM13_ESM.xlsx	
http://ocean-microbiome.embl.de/companion.html	Abundancia y clasificación de los mOTUs.

Esta página se ha dejado en blanco intencionalmente

Capítulo 3

Comprensión de las comunidades microbianas marinas

3.1. Variables contextuales

Las muestras de comunidades planctónicas fueron recolectadas en diferentes condiciones oceanográficas y ambientales que fueron capturadas por una serie de variables contextuales de origen físico y químico. El registro de dichos parámetros se fundamenta en el efecto que tienen sobre el desarrollo de los microorganismos (por ejemplo, la disponibilidad de nutrientes y luz es crucial para los microorganismos fotosintéticos; y la temperatura afecta el metabolismo de los microorganismos en general). Aquí se evaluaron las posibles relaciones entre estas variables así llamadas 'explicativas'. En la Figura 3.1. presentamos el resultado de calcular la dependencia monotónica entre ellas. En concordancia con los hallazgos presentados por Sunagawa *et al.* (2015) para muestras epi pelágicas (es decir, de la capa superficial del océano con mezcla vertical y que recibe luz solar), se descubrieron asociaciones positivas entre la concentración de clorofila y la salinidad, la concentración de fosfato (PO_4) disuelto y la suma de nitratos y nitritos ($\text{NO}_2 + \text{NO}_3$) o el sílice (SI); la concentración de SI y $\text{NO}_2 + \text{NO}_3$; la profundidad de la capa de mezcla (MLD, *Mixed Layer Depth*; determina, entre otras cosas, la cantidad de luz promedio recibida por los microorganismos marinos) y la profundidad de frecuencia Brunt-Väisälä máxima (BV, *Mean Depth of the Maximum of Brunt-Väisälä Frequency*; una medida de la estabilidad de la columna de agua); entre la concentración de oxígeno disuelto y la profundidad con la máxima concentración de oxígeno disuelto; entre otras.⁴⁹ Por otra parte, existen asociaciones negativas entre la temperatura y la concentración de oxígeno disuelto, la salinidad y la cantidad de PO_4 , o SI; el tiempo de retención en remolinos y el parámetro de Okubo-Weiss, vinculado con la capacidad de mezclado y turbulencia; entre otras.

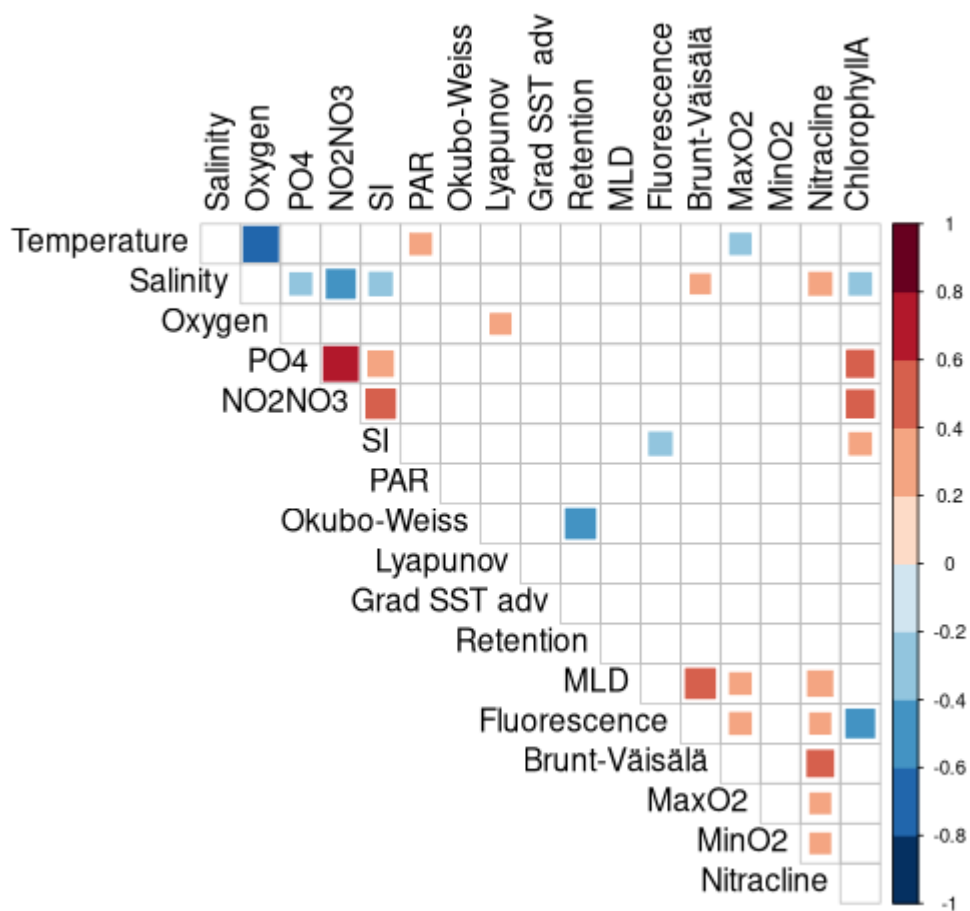


Figura 3.1. Matriz de correlación usando el método de Kendall para las variables fisicoquímicas. Solamente se muestran valores significativos (p -valor < 0.05 corregido por *False Discovery Rate*).

Luego, dado que las variables podrían estar asociadas de un modo no monotónico, para todas los pares de relaciones no significativas se estimó el valor de información mutua. Brevemente, este método se basa en calcular el grado de dependencia entre variables aleatorias usando las funciones de densidad de probabilidad conjunta y marginal y sirve para cualquier tipo de relación. Las únicas tres posibles asociaciones halladas no presentan patrones claros (Figura 3.2).

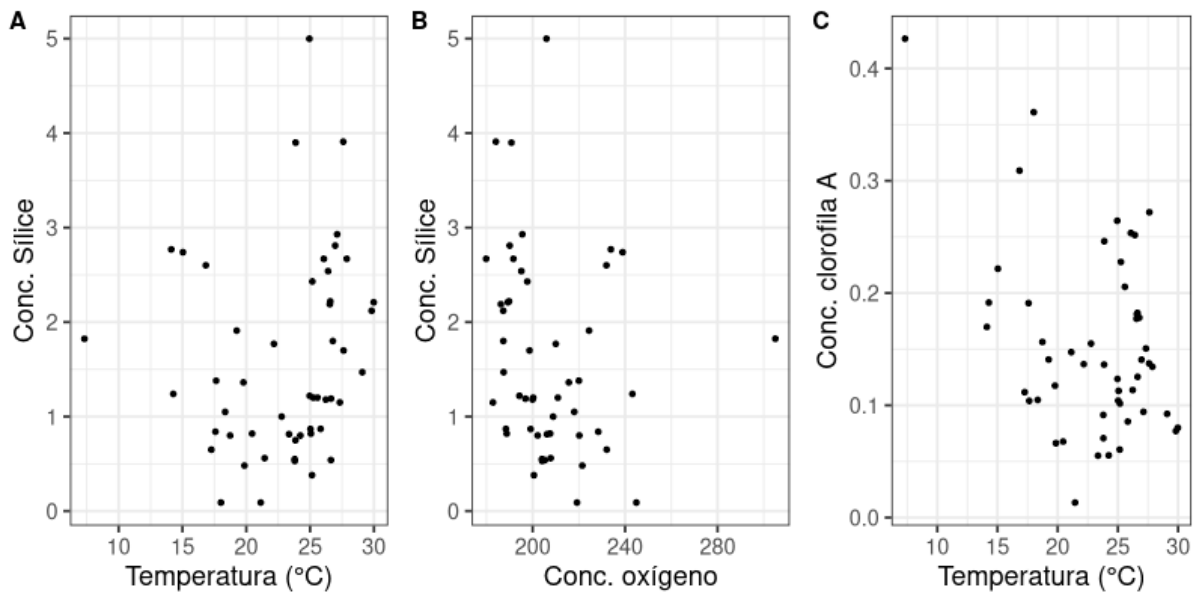


Figura 3.2. Pares de variables con valores de información mutua significativos. A) Silica versus temperatura. B) Silica versus oxígeno. C) Concentración de clorofila predicha versus temperatura.

3.2. Variables contextuales y abundancia

Como se dijo anteriormente, aquí consideramos a la abundancia de microorganismos como un indicador de la productividad de un ecosistema. En la expedición de *Tara Oceans* se diferenciaron tres grupos: (i) los autótrofos, compuesto principalmente por bacterias del género *Prochlorococcus* y *Synechococcus*, (ii) los heterótrofos, y (iii) los picoeucariotas. En la Figura 3.3.A se observa la distribución de la variable de abundancia para cada grupo. Se destaca, como era de esperar, que el número de bacterias autótrofas y heterótrofas es mayor respecto a los picoeucariotas, y que la abundancia de heterótrofos es mayor a la de los autótrofos. Además, la abundancia de heterótrofos está más concentrada y posee menos varianza que la abundancia de autótrofos. Respecto al tamaño de filtro utilizado, parece haber un marcado efecto sobre los picoeucariotas (lo cual afecta los resultados de los análisis genómicos), detectándose mayor abundancia con un filtro de 0.22-3 μm . Esto coincide con el hecho de que los mismos son microorganismos de mayor tamaño. Por último, existe un valor de abundancia de heterótrofos alrededor de 1×10^4 que será considerado como un outlier.

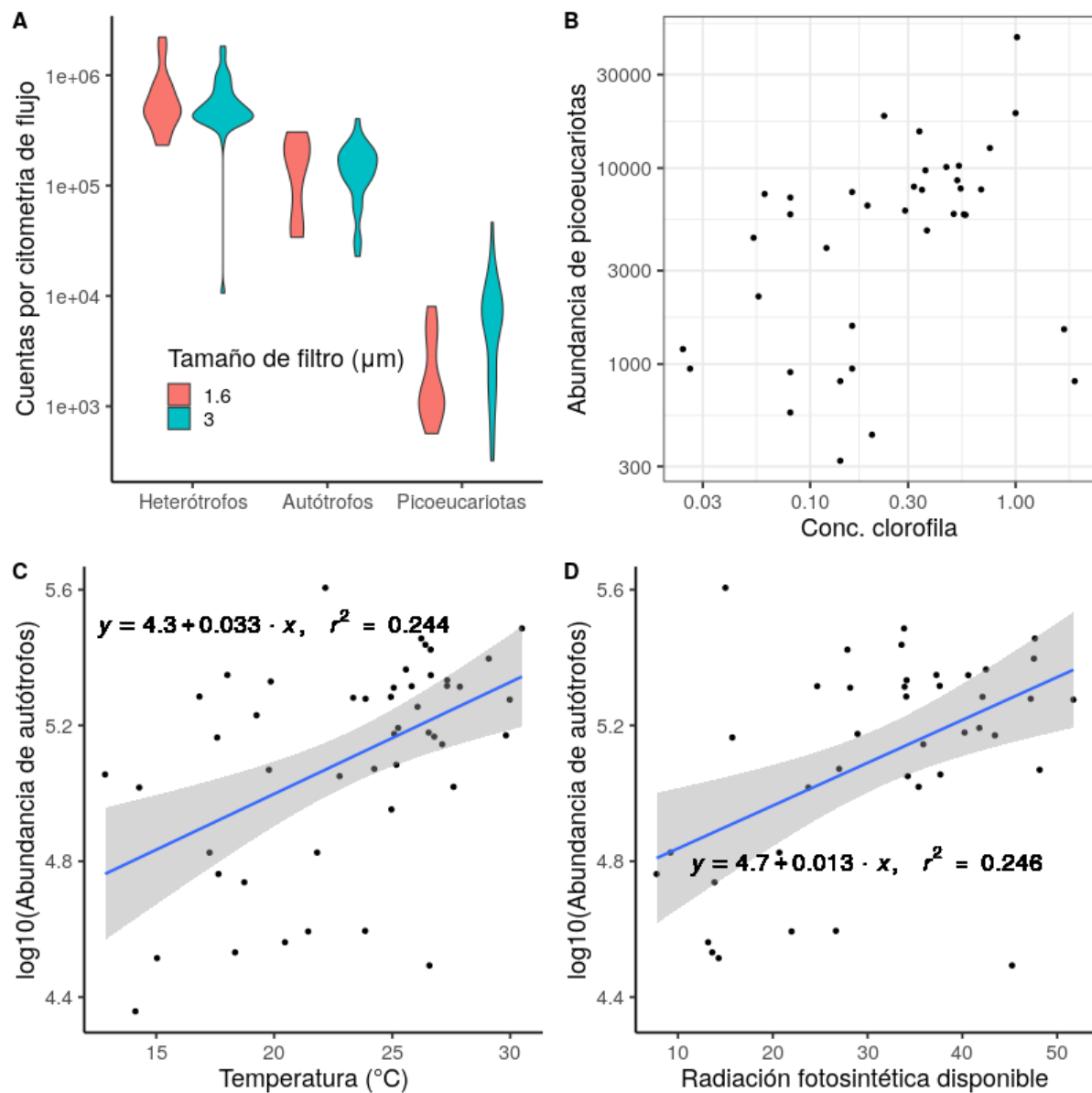


Figura 3.3. A) Distribución de la variable cuentas de citometría de flujo separado por grupos de organismos. B) Par de variables con valor de información mutua significativo. Abundancia de picoeucariotas versus nitratos y nitritos (NO_2NO_3). C) y D) Asociación positiva entre la abundancia de autótrofos, y radiación fotosintética disponible, respectivamente.

Para evaluar la relación entre las variables fisicoquímicas y la variable objetivo aplicamos la metodología basada en la información mutua. Se priorizo esta técnica respecto a la correlación de Kendall ya que en trabajos previos se mostró que el número de especies distintas podría seguir una dependencia lineal (*índice de*

Shannon) pero también un de tipo “parabólica” (*riqueza*) respecto a la temperatura.^{49,51} En la Figura 3.3, panel B, se muestra una de las dos asociaciones significativas halladas. Según estos resultados, existe una clara relación entre la cantidad de microorganismos picoeucariotas y la concentración de clorofila. Las otras dos variables asociadas eran la abundancia de picoeucariotas y la concentración de fosfato (PO_4). Esto último era esperable ya que la concentración de clorofila y la concentración de fosfato son dos variables correlacionadas entre sí (Figura 3.1).

Es necesario mencionar que la productividad es una propiedad emergente del ecosistema de causas multifactoriales. Los datos de Tara Oceans poseen muchas variables contextuales que fluctúan en simultáneo, con lo cual, podría ocurrir que al analizar la relación entre productividad y la diversidad no se observen asociaciones significativas, que si existiría en caso de realizar un experimento controlado. Por lo tanto, se considera en principio que la ausencia de evidencia de efecto no implica evidencia de ausencia. Siguiendo esta línea argumental, la relación entre temperatura y abundancia de autótrofos no mostró una relación significativa usando información mutua, no obstante, tenemos indicios de que existe una asociación positiva puesto que ajustar la ecuación de una recta a la curva $\log_{10}(\text{Abundancia de autótrofos})$ versus Temperatura, da un ajuste con una pendiente significativamente distinta a 0 y positiva (0.033, p-valor de 0.0004 con la hipótesis nula de que el valor es 0, R^2 ajustado de 0.22) (Figura 3.3. C). De la misma manera, la radiación fotosintética disponible (PAR) también parece estar relacionada con la abundancia de autótrofos (pendiente igual a 0.013, p-valor 0.001) (Figura 3.3. D). Estos resultados son consistentes con los principales controles de la productividad en el océano. PAR, nutrientes y temperatura juegan un rol clave en la abundancia de autótrofos, en particular en el rango de temperatura evaluado.^{79–81} Por lo tanto, el patrón observado aquí es el esperado.

3.3. Biodiversidad

Antes de comenzar a investigar la relación productividad-diversidad, es necesario averiguar cuánta información acerca de la presencia de especies existe en los metagenomas. Con este fin, se construyeron curvas de acumulación del número de “especies” distintas (dataset mOTUs) versus el esfuerzo de muestreo, es decir, número de individuos muestreados. Dado que todas las comunidades poseen un número finito de especies, estas curvas deberían llegar eventualmente a una asíntota. Mientras más cóncava hacia abajo es la curva, mejor muestreada está la comunidad.⁸² En la Figura 3.4. se muestran las curvas de acumulación para dos muestras distintas, una en la cual el número de mOTUs distintos presentes esta bien estimado ya que luego de haber analizado el 30 % de los individuos, no aparecen nuevos mOTUs (panel A), y otra curva donde pareciera que si seguimos muestreando individuos (en este caso, más secuenciación), deberían aparecer nuevos mOTUs (panel B).

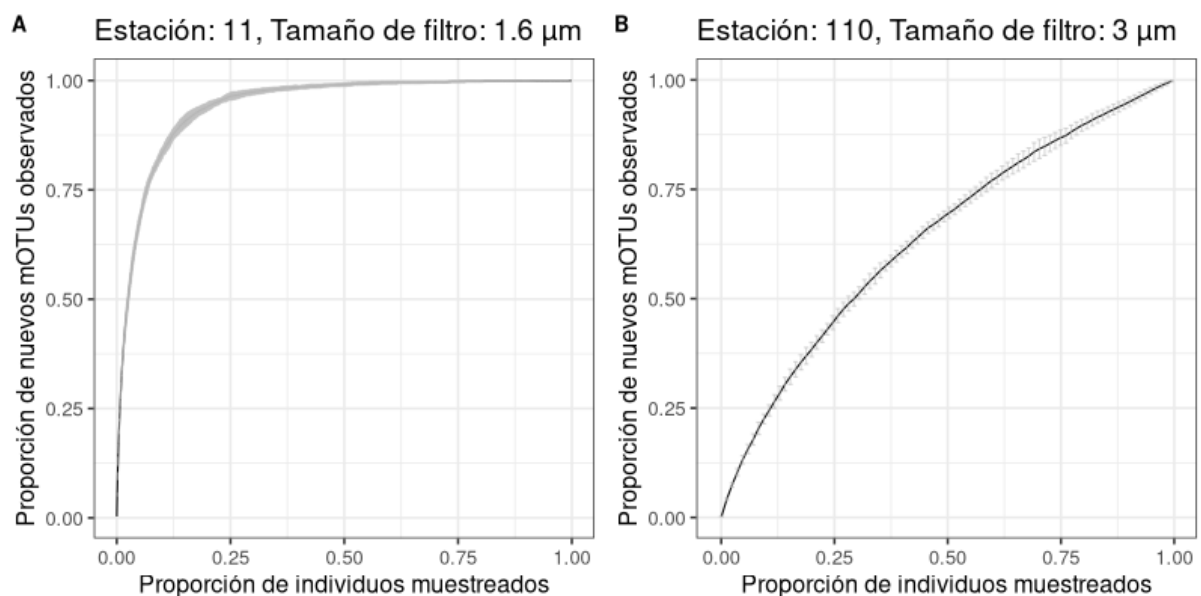


Figura 3.4. Curvas de acumulación de nuevos tipos de individuos versus proporción de individuos muestreados (mOTUs). A) Estación 11, tamaño de fracción 0.22-1.6 µm. B) Estación 110, tamaño de fracción 0.22-3 µm. Ajustando las curvas a una función de tipo michaeliana se observa que, en el panel A, la mitad de mOTUs distinto se consigue luego de contar el 2-3 % de los individuos, mientras que en el panel B la mitad de mOTUs distinto se consigue recién con un 67 % de los individuos.

Para cuantificar la esperanza de no hallar nuevos tipos de individuos al muestrear (secuenciar) más, se ajustó cada curva de acumulación a una función hiperbólica y se obtuvo un parámetro de ajuste que describe el valor de la proporción de individuos contados en la cual se observaron la mitad de nuevos tipos de individuos. La distribución de dicho parámetro se presenta en la Figura 3.5., de la cual se concluye que para las futuras estimaciones de biodiversidad es mejor descartar a las estaciones 128 y 110.

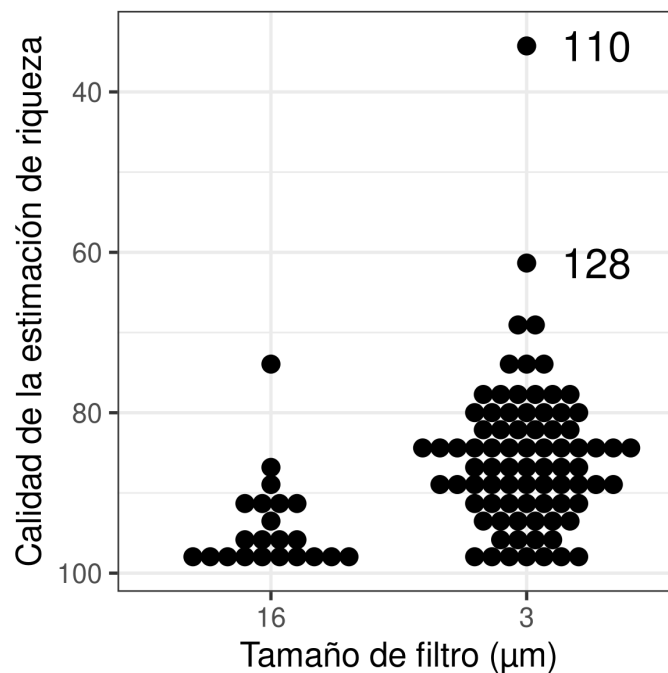


Figura 3.5. Valor del parámetro de ajuste b de la función michaeliana a las curvas de acumulación. *Raw richness quality* se calcula como $100 - 100 * b$.

La biodiversidad también se analizó desde la perspectiva del número de especies efectivas (índice de Hill) por muestra. Dicho valor osciló entre las 300 y 20 especies efectivas, con un outlier que corresponde a la estación 86. Al tratarse de una estación de alta latitud sur es esperable que cuente con un muy bajo número de especies, en especial en el rango de tamaño analizado. Por lo tanto, la estación 86 también fue excluida de análisis futuros.

3.4. Ensamble de una comunidad planctónica

En experimentos controlados diseñados para poner a prueba hipótesis de diversidad y productividad de un ecosistema, las comunidades son ensambladas de forma artificial. En cambio, en ecosistemas naturales, los microorganismos presentes en un momento determinado dependen, como se dijo antes, de los procesos de dispersión, selección, deriva y diversificación. Comprender los patrones de la estructura comunitaria en las muestras de plancton en estudio puede representar un primer paso para la exploración de los mecanismos detrás las relaciones diversidad-productividad. En ese sentido, surge la pregunta acerca de si la composición de una muestra particular es más similar a una muestra cercana geográficamente, o cercana en el contexto ambiental (por ejemplo, con una temperatura similar).

Para responder dicho interrogante calculamos la similitud (índice de Jaccard; dataset mOTUs) entre 34 muestras de superficie con filtro de tamaño entre 0.22 y 3 μm . Luego seleccionamos las dos muestras más similares para cada muestra (siempre y cuando el coeficiente de Jaccard fuera mayor a 0.5) y creamos la Figura 3.6. Se observa que, dada una muestra particular, la muestra más similar en composición no es siempre la muestra geográficamente más cercana. Por ejemplo, la estación 67 (costa oeste de Sudáfrica) es más similar a la estación 93 (costa de Chile) que a las estaciones próximas 68 y 70, también del Atlántico Sur. Otro caso claro es la similitud entre la estación 132 (océano pacífico norte) y la estación 78 (aguas cercanas al sur de Brasil). Más allá de que en el océano las distancias geográficas no se corresponden con las distancias recorridas por el plancton marcadas por el rumbo de las corrientes, estos resultados sugieren que la selección o el filtrado por variables ambientales posee un rol más importante que la dispersión en el ensamblado de las comunidades planctónicas.

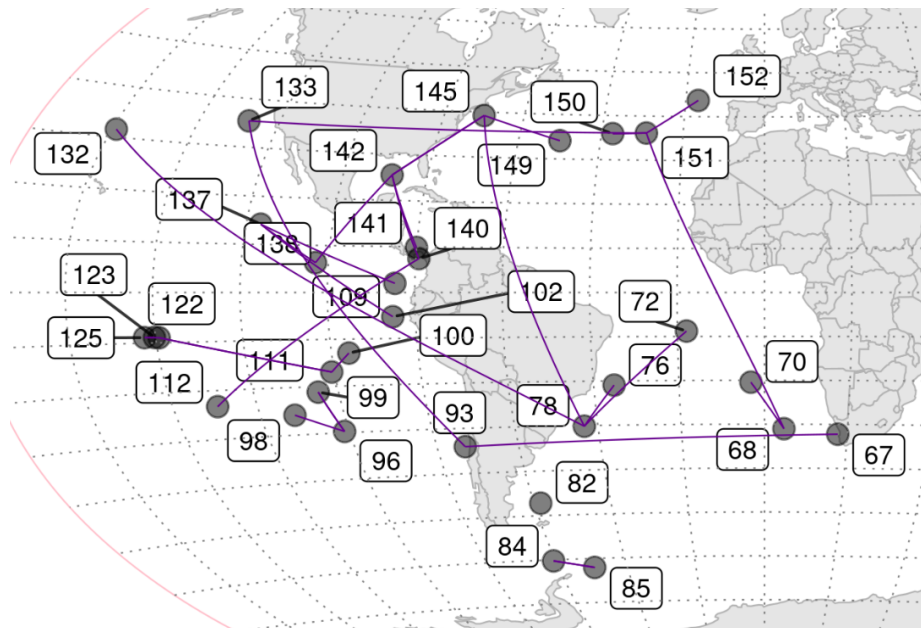


Figura 3.6. Similitud en diversidad para las muestras de superficie con filtro de tamaño entre 0.22 y 3 μm . Cada muestra se encuentra asociada a las dos estaciones más similares (siempre y cuando se cumpla el límite mínimo de similitud dado por Jaccard > 0.5).

Generalmente, los experimentos controlados que evalúan la biodiversidad y el funcionamiento de los ecosistemas muestran relaciones consistentes.^{83–85} Por ejemplo, en el trabajo de O'Connor *et al.* (2017) se analizaron 374 publicaciones y se halló una asociación positiva entre la biomasa y la riqueza de especies. Sin embargo, casi todos los estudios de este tipo se encuentran restringidos a escalas espaciales y/o temporales pequeñas: 1 a 100 m^2 y/o no más de 10 generaciones. Surge entonces la pregunta, ¿hasta que punto las asociaciones entre diversidad y funcionamiento dependen de la escala utilizada? Estudios más “realistas” deberían variar a lo largo de diferentes tipos de ecosistemas y gradientes climáticos. Una ventaja del océano —que se puede ver en este análisis observacional donde no se controla la identidad de las especies— es que la circulación oceánica de superficie homogeniza la distribución de especies. Incluso, el océano tropical se presenta como un buen modelo ecológico para “scale up”.⁸⁶

Como se mencionó previamente, en algunos casos muestras recolectadas en estaciones distantes comparten más taxones que muestras recolectadas en estaciones cercanas. Para analizar si las diferentes composiciones podrían estar

asociadas con la abundancia celular, realizamos, exclusivamente para el grupo de los autótrofos, el análisis de similitud seguido de agrupamiento. Una vez construida la matriz de similitud, la misma fue transformada a distancias y reducida a 4 dimensiones relevantes mediante escalamiento multidimensional métrico (Figura 3.7A). Luego, en base al algoritmo *k-means* agrupamos las muestras cercanas y calculamos la distribución de abundancia y temperatura para los grupos hallados (Figura 3.7B y C). Como resultado, hallamos para el grupo 3 valores de abundancias diferentes a todos los demás grupos. No obstante, esto se puede explicar debido a la temperatura (Figura 3.7C). En cuanto a los grupos 1, 2 y 5, no se hallaron diferencias significativas, lo cual implica que composiciones diferentes de *Synechococcus* y *Prochlorococcus* pueden dar lugar a abundancias celulares similares. Curiosamente, en el grupo 5 la temperatura de las muestras es significativamente menor que en los grupos 1 y 2.

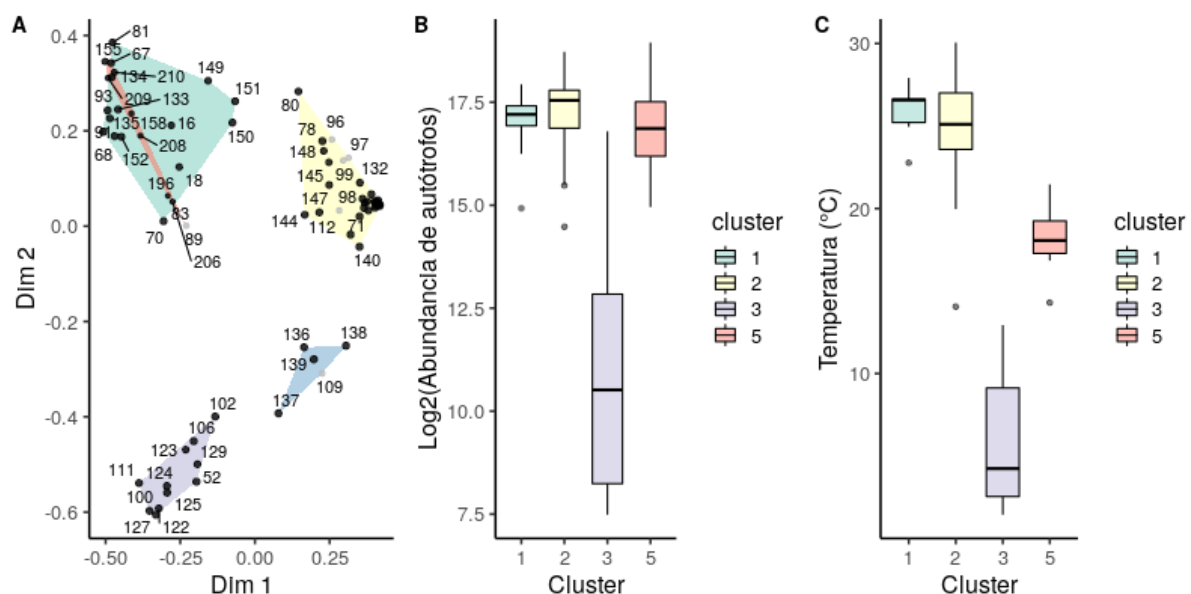


Figura 3.7. A) Proyección de la matriz de distancias en base a la composición taxonómica (solamente mOTUs de autótrofos) usando multidimensional scaling (4 dimensiones) y agrupación en base a *clustering* por *k-means* (5 centros). B) Distribución de abundancia de autótrofos en los diferentes grupos. El grupo 4 fue excluido del análisis por no contar con suficientes datos ($n < 5$). El grupo 3 posee diferencias significativas contra todos los demás grupos. El análisis estadístico se llevó a cabo empleando la ANOVA de Welch seguida del

test post-hoc Dunnett T3 para comparaciones múltiples. C) Distribución de temperaturas en los diferentes grupos.

3.5. Distribución de abundancia de especies

La distribución de abundancia de especies (DAS) siempre revela, tanto para especies micro como macroscópicas, la existencia de pocas especies de mucha abundancia y muchas especies de poca abundancia. Dicho fenómeno se corroboró en este trabajo inspeccionando el dataset de mOTUs. A modo de ejemplo, en la Figura 3.8 se presenta la DAE de la estación 60 de cuatro maneras características. Se observa que existen, en este caso, casi 500 especies con abundancia menor o igual a 3 (Figura 3.8 B), que la especie más abundante tenía 1200 cuentas, y que la segunda más abundante, 600 cuentas (Figura 3.8 A y C). Cuando se aplica la transformación logarítmica, se visualiza un histograma con una forma similar a una distribución normal truncada en el zero (Figura 3.8 B).

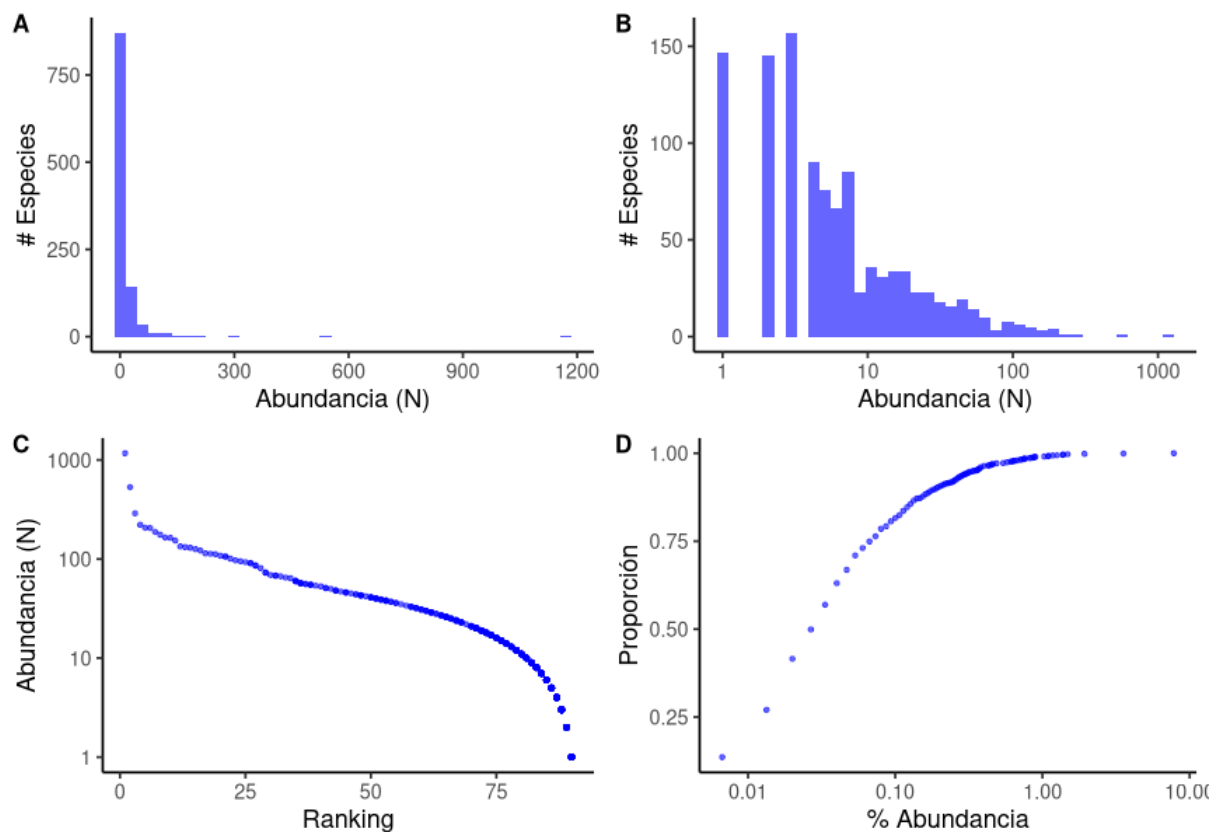


Figura 3.8. Distribución de abundancia de especies para la estación 60 (mOTUs). A) Histograma del número de especies versus abundancia en una escala aritmética. B) Histograma con la abundancia en escala logarítmica de base 10. C) Gráfico de ranking-abundancia en escala logarítmica de base 10. D) Distribución empírica acumulada para la abundancia de especies en escala logarítmica de base 10.

A modo descriptivo, en la Figura 3.9 se observa la distribución empírica acumulada para la abundancia de especies para 15 estaciones elegidas al azar. De la misma se concluye que, (i) en general, todas las muestras poseen una especie con una abundancia relativa mayor al 5-10 %, (ii) y que dada una determinada muestra, más del 50 % de las especies tienen una abundancia relativa menor al 0.1 %, y más del 75 % de las especies una abundancia relativa menor al 1 %. Curiosamente, a pesar de existir especies “dominantes”, dichas especies son siempre distintas de acuerdo a la muestra analizada.

Como ya se mencionó en la introducción, el efecto positivo de la biodiversidad sobre la productividad primaria podría estar dado por un efecto de muestreo en el que a mayor diversidad, mayor la probabilidad de contar con una especie altamente productiva. A su vez, esta especie tiene que ser favorecida por ciertas condiciones ambientales, lo que coincide con un estudio que propone que la productividad primaria marina está guiada por un efecto de selección.⁸⁷ En otras palabras, que la biomasa de la comunidad se asemeja a la biomasa del monocultivo de la especie más productiva. En este sentido, las distribuciones de abundancia asimétricas observadas no contradicen dichos resultados.

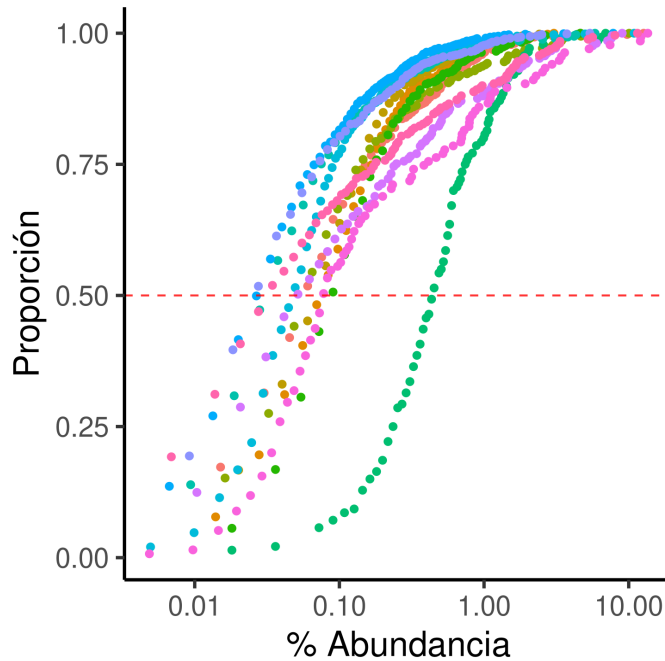


Figura 3.9. Distribución empírica acumulada para la abundancia de especies en escala logarítmica de base 10 para 10 estaciones elegidas al azar.

Se ha observado utilizando muestras de RNA ribosomal 16S de 20.000 localizaciones que la distribución Poisson lognormal posee el mayor éxito en la predicción de los patrones de abundancia de especies (usando la curva ranking-abundancia).⁴⁷ Para confirmar dicha observación y a su vez probar diversos modelos de DAE, ajustamos los datos a tres modelos: Poisson lognormal, METE (series logarítmicas), y distribución de Zipf, también llamada ley de potencias. Basándonos en el criterio de información de akaike (AIC), el mejor modelo fue el Poisson lognormal en 93 ocasiones, el de METE en solamente 4, y en 4 ocasiones el valor de AIC era estadísticamente similar en el modelo Poisson lognormal y METE (78 estaciones con tamaño de filtro 3 μM y 22 con tamaño de filtro 1.6 μm). Es importante notar que el criterio de akaike permite elegir el “mejor” modelo entre varios propuestos, pero no implica que dicho modelo describe efectivamente el proceso generador de los datos. En ese sentido, una inspección visual de los residuales reveló que, para algunas muestras, sería justificable seguir agregando complejidad (Figura S1). A su vez, luego de ajustar el modelo Poisson lognormal es posible construir las curvas predichas de distribución empírica acumulada (Figura

3.8). Si comparamos para todas las muestras la proporción de especies con cierta abundancia relativa predicha versus la observada, se detecta un sesgo en el modelo, ya que el valor predicho es generalmente mayor al observado (Figura S2).

El modelo Poisson lognormal contiene dos parámetros llamados μ y σ que permiten estimar la probabilidad de que una especie tomada al azar esté representada por n individuos. Interesantemente, a pesar de que dichos modelos de distribución son ampliamente utilizados en el campo de la ecología, no hallamos estudios relacionando los valores de dichos parámetros con los índices más comunes de diversidad ("Hill", por ejemplo). Dado que nuestro objetivo es buscar factores que expliquen la productividad de un ecosistema, y sobre todo, que se puedan interpretar, simulamos muestras usando diferentes valores de μ a σ constante y viceversa, y calculamos los índices de Hill (número de especies efectivo) y Pielou (equitatividad). Como resultado, se obtuvo que a σ constante, μ se asocia positivamente con el número efectivo de especies efectivo (índice de Hill) solo si σ es menor a 1, y que μ no tiene un efecto sobre la equitatividad (índice de Pielou) (Figura S3). Por otro lado, a μ constante, σ afecta tanto la diversidad como la equitatividad: para cualquier μ , si σ aumenta, la diversidad y la equitatividad decrecen (Figura S3).

Capítulo 4

Funcionamiento de los ecosistemas microbianos marinos

Este trabajo de Tesis busca comprender los efectos de la biodiversidad del plancton sobre el funcionamiento del ecosistema marino. En particular, profundizamos en los datos recolectados durante la expedición global Tara Oceans para buscar la existencia de relaciones entre biodiversidad y abundancia para dos de los grupos de organismos: bacterias autótrofas y heterótrofas, y evidencias de los tres mecanismos postulados que vinculan la productividad de un sistema con la variedad de especies presentes (Tabla 3.1).

Grupo	Relación con biodiversidad	Evidencia de mecanismo		
		Facilitación	Efecto de muestreo	Complementariedad de nicho
Autótrofos	Si	Si	NA	NA
Heterótrofos	No	NA	No	Si

Tabla 4.1. Relaciones exploradas en este trabajo de Tesis. “Si”, “No” y “NA” implican respectivamente que existe evidencia de asociación, que no, o que dado los datos disponibles, no se halló una manera de probarlo.

4.1 Biodiversidad y productividad

Luego del análisis exploratorio de los datos de productividad y biodiversidad se prosiguió a explorar la relación entre ambas variables. No se analizó la abundancia de picoeucariotas debido al sesgo en el tamaño de filtro utilizado que afecta el muestreo metagenómico y por ende las estimaciones de diversidad (Figura 3.3 A). Usando el dataset de mOTUs o MAGs, las curvas de abundancia de heterótrofos no muestran una clara relación con el índice de biodiversidad de Hill (Figura 4.1 B y C). Por otra parte, usando un subset de los mOTUs correspondientes a bacterias

fotosintéticas dominantes/mayoritarias del tipo *Synechococcus* y *Prochlorococcus*, la abundancia de autótrofos sí muestra una asociación positiva con la biodiversidad (Figura 4.1 A). También se evaluó utilizar a la concentración de clorofila como proxy de productividad, pero en dicho caso no se hallaron relaciones relevantes.

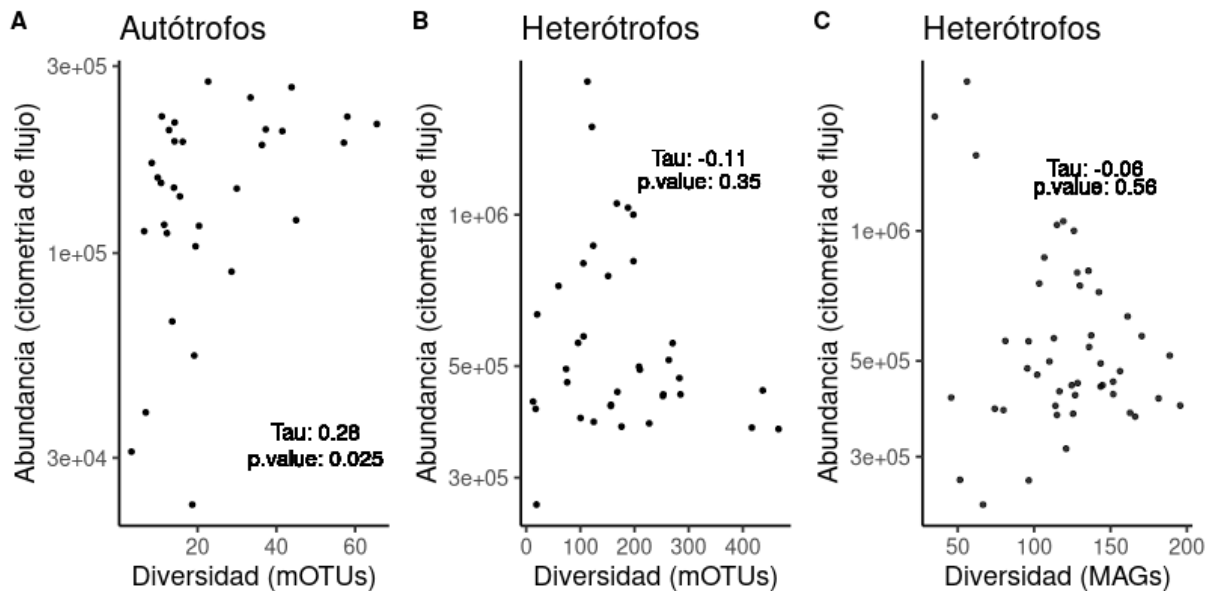


Figura 4.1. Productividad y biodiversidad. Abundancia medida por citometría de flujo (FC, *flow cytometry*) versus número efectivo de especies usando A) los mOTUs de *Synechococcus* y *Prochlorococcus*, B) todos los mOTUs excepto los de *Synechococcus* y *Prochlorococcus*, y C) todos los MAGs excepto los picoeucariotas y del filo Cyanobacteria.

Previamente se demostró que el modelo log normal explica parcialmente la distribución de abundancia de especies (DAE) observadas para 100 estaciones. Surge entonces la pregunta, acerca de si los parámetros ajustados de dicho modelo - media (μ) y desvío estándar (σ) - se relacionan con la productividad. En este aspecto, no se hallaron relaciones significativas usando la correlación de Kendall tau entre la abundancia de heterótrofos/autótrofos, y los valores estimados de μ/σ . Interesantemente, se encontró una asociación (débil) entre la diversidad de autótrofos (mOTUs) y σ (Tau -0.3, p-valor 0.02). Esto último coincide con las simulaciones realizadas en el Capítulo 3 (Figura S3). En trabajos futuros, con mayor disponibilidad de datos, se podrían ajustar las DAE usando subgrupos (por ejemplo

con los autótrofos) y buscar asociaciones entre productividad y μ con σ casi constante, y viceversa.

La productividad del grupo de los autótrofos muestra una asociación débil positiva con la temperatura y con la radiación fotosintética disponible, y una fuerte asociación positiva con la biodiversidad de organismos autótrofos procarióticos (Figura 3.1 y 4.1). Estos resultados son coherentes con los patrones encontrados en ecosistemas terrestres usando experimentos observacionales.²² Es importante resaltar que se trata de un estudio de campo y por ende las correlaciones podrían surgir espuriamente como efectos secundarios de las condiciones ambientales que simultáneamente mejoran diversidad y productividad.²² Los microorganismos heterótrofos, en cambio, no mostraron una asociación particular. En ecosistemas terrestres, los descomponedores, un grupo similar a los heterótrofos marinos aquí analizado, tampoco muestran una clara relación entre la biodiversidad y el funcionamiento de los ecosistemas (estimada como la descomposición de la hojarasca, o basadas en tipos de basura estándar tales como bolsitas de té, celulosa o madera).⁸⁸

4.2. Facilitación

Uno de los mecanismos que vincula un incremento de biodiversidad con un incremento de productividad es la complementariedad por interacción positiva. Un ejemplo clásico en ecosistemas terrestres es la simbiosis entre organismos fijadores de nitrógeno atmosférico y las plantas. En esta interacción, las plantas proveen de los azúcares necesarios a los microorganismos fijadores, y estos a cambio convierten el nitrógeno gaseoso (N_2) a una forma asimilable por las plantas (NO_3^-). Para poner a prueba la posible existencia de este mecanismo con el grupo de las bacterias autótrofas, decidimos explorar la abundancia de genes fijadores de nitrógeno en la comunidad microbiana de cada sitio. En particular nos centramos en el gen *nifH*, que codifica para la enzima nitrogenasa que cataliza la fijación de N_2 .

Primero, se validó el dataset de abundancia de genes mediante la verificación de la función de todos los genes correlacionados con la abundancia de autótrofos. Al hallarse una correlación positiva entre la abundancia de genes involucrados en el metabolismo de la clorofila —un pigmento solamente presente en el grupo de los autótrofos— se continuó con el análisis específico para fijación de nitrógeno (Figura S4). En la Figura 3.10 se presentan las relaciones entre las variables abundancia de autótrofos, *Synechococcus* y *Prochlorococcus* versus la abundancia del gen *nifH* o la temperatura. Esta última variable se limitó al rango de los 12 a 25 °C para desacoplar la relación con la abundancia de autótrofos y temperatura. Encontramos una asociación entre la abundancia de autótrofos y el número de veces que se detectó el gen *nifH* en una muestra (Figura 4.2 A), lo cual sugiere que el mecanismo de facilitación podría ser una explicación para el aumento de la abundancia con la diversidad. A su vez, la abundancia del gen *nifH* depende débilmente de la temperatura (Figura S5) y, con este subconjunto de datos, no hay asociación entre la temperatura y la abundancia de autótrofos.

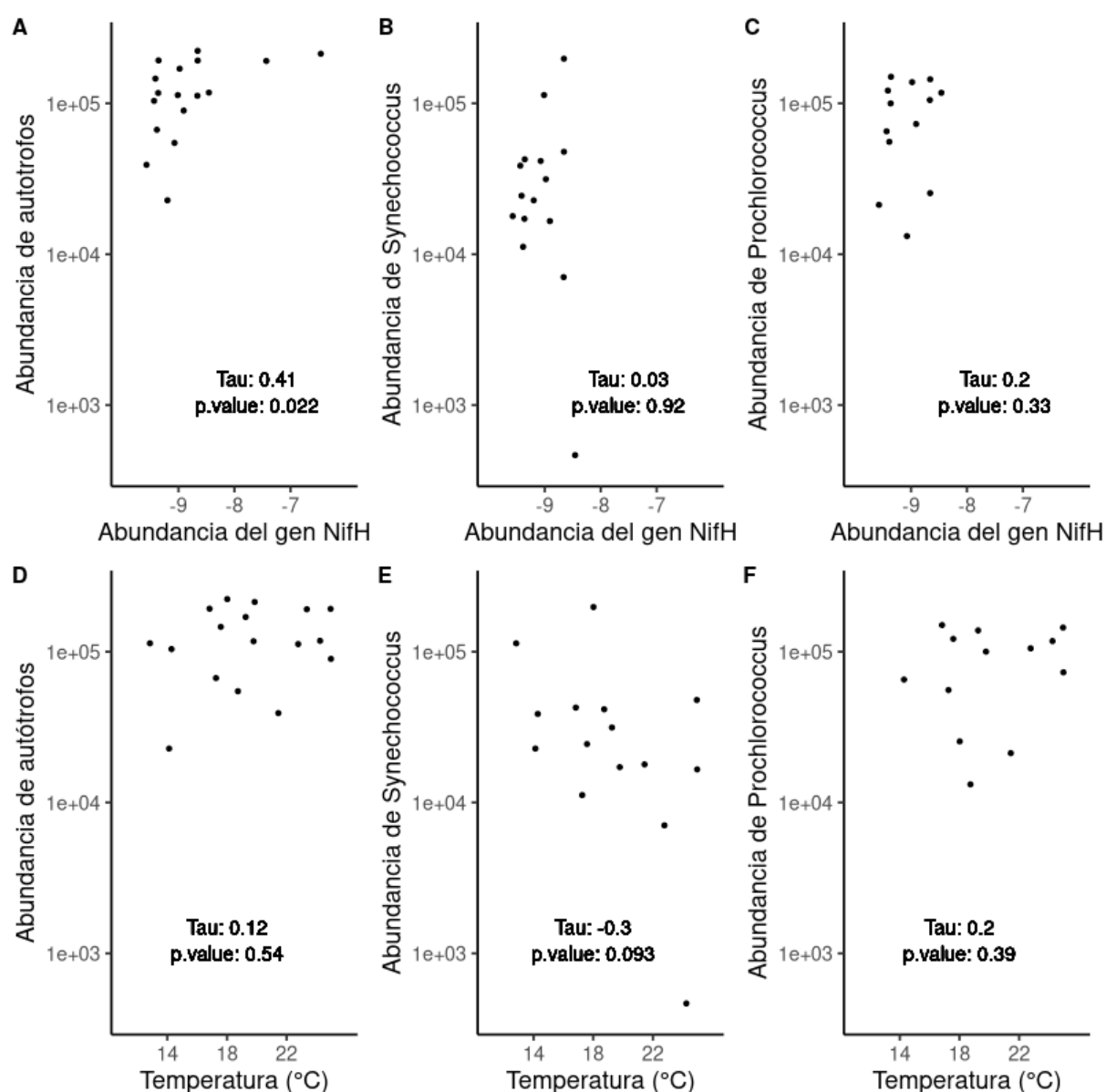


Figura 4.2. Relación entre la abundancia celular y la del gen fijador de N₂ como evidencia de mecanismos de facilitación en el rango de temperatura de 12 a 25 grados. A), B) y C) Abundancia de autótrofos, *Synechococcus* y *Prochlorococcus* versus abundancia del gen *nifH*. Los datos de abundancia de autótrofos provienen de Sunagawa *et. al*, 2015⁴⁹ y los de abundancia de *Synechococcus* y *Prochlorococcus* de Ibarbalz *et. al*, 2019.⁵¹

Se destaca que no parece existir una asociación entre la abundancia del gen *NifH* y cada grupo de autótrofos por separado. Además, el mismo análisis fue repetido para los autótrofos usando temperaturas mayores a 25 grados, no obstante, no se encontró una asociación entre las variables. A nivel global, la abundancia de *Prochlorococcus* y *Synechococcus* mostraron un efecto relativamente pequeño ante

la concentración de nitrato, lo que sería consistente con la falta de respuesta a la facilitación por N aquí explorada para estos dos grupos.^{81,89}

En relación con el mecanismo de facilitación y los heterótrofos, se analizó la relación de la abundancia de los mismos contra la diversidad de los autótrofos. Es decir, se evaluó si a medida que se incrementa la diversidad de autótrofos, también se incrementa la abundancia de heterótrofos. A priori, contrario a lo esperado, podría llegar a existir una asociación negativa (Kendall tau: -0.2, p-valor 0.075) y a mayor diversidad de los autótrofos menor abundancia de heterótrofos (Figura S6).

4.3. Complementariedad de nicho

Otro de los mecanismos que vincula un incremento de biodiversidad con un incremento de productividad es la complementariedad de nicho. En este caso, existe un mayor funcionamiento del sistema debido a que las especies presentes explotan distintos recursos. Para explorar esta hipótesis se utilizó el dataset de MAGs, el cual cuenta con información acerca de la presencia o ausencia de miles de genes para 957 MAGs distintos. Al igual que en los análisis previos, es necesario validar los datos. En ese sentido, no todos los genomas ensamblados a partir de los metagenomas están completos, incluso para algunos se cuenta solo con 10 % o menos del genoma correspondiente.

Explorar la funcionalidad en base a comparar presencia y/o ausencia de genes usando genomas tan incompletos probablemente resulte en conclusiones erróneas. Por lo tanto, exploramos el uso de distintos puntajes de corte para completitud y decidimos proseguir solo con aquellos MAGs con una completitud mayor o igual al 70 % (*anvio completeness*; 693 en total). Esta línea de corte verifica una correspondencia entre la agrupación por similitud de los MAGs en base a la funcionalidad genómica y la filogenia (Figura 4.3). Dicha conclusión se desprende de la tabla de la Figura 4.3., donde se observa que en general los clusters contienen MAGs de un único filo (clusters 1, 2, 5, 8, 9, 10, 11), excepto contadas excepciones, como por ejemplo el cluster 6 que agrupa todas las bacterias del filo chloroflexi,

cuatro de las cinco cianobacterias, algunas actinobacterias, y algunas proteobacterias. Vale recordar que el filo Proteobacteria es altamente diverso y que al menos 3 de sus 5 clases poseen diversidad comparable con otros filos.

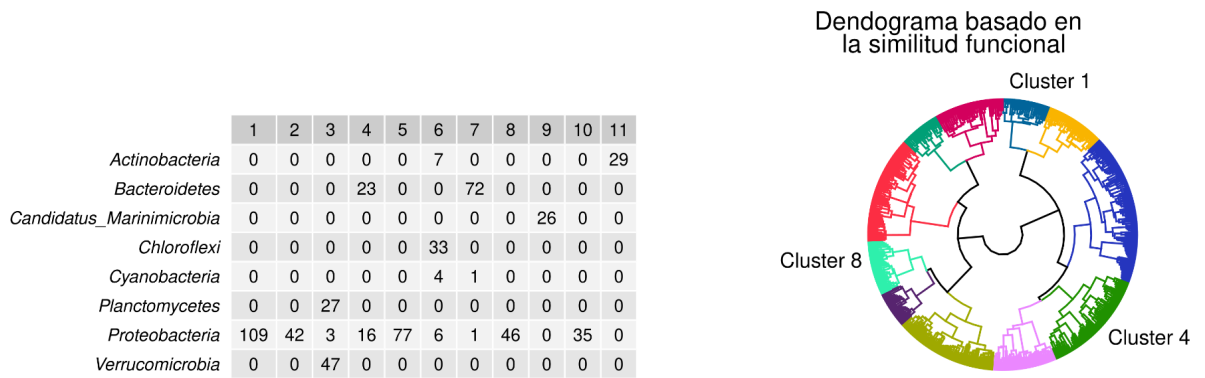


Figura 4.3. Izquierda: Matriz de frecuencia de clusters obtenidos por similitud genómica (numerados desde el 1 a 11) y 8 niveles taxonómicos a nivel de filo (*Phylum*) para el dominio *Bacteria*. Los clusters se consiguieron mediante un algoritmo jerárquico aglomerativo. Ver la sección de Métodos para más detalles.

Ya validada nuestra métrica de similitud funcional entre MAGs y línea de corte para la completitud, proseguimos a explorar la posible evidencia de complementariedad de nicho. Para cada muestra, quitamos los MAGs eucariotas y de cianobacterias, y todos aquellos con baja completitud, y calculamos la similitud funcional como el promedio de la similitud entre todos los pares de MAGs presentes. Mientras menor es el índice de similitud funcional, más alta es la diversidad funcional. Hallamos que la similitud funcional presenta una tendencia a una asociación negativa con la abundancia de heterótrofos aunque con un $p > 0.05$ (Figura 4.4 A). Como era de esperar, en el grupo de los heterótrofos, mientras menor es la similitud genómica, mayor es la diversidad de especies (Figura 4.4 B), no obstante, en concordancia con los resultados hallados con el dataset de mOTUs, la diversidad *per se* no se asocia con la productividad (Figura 4.4 C). Es interesante destacar que si en nuestro procedimiento tenemos en cuenta las bacterias del filo *Cyanobacteria*, la asociación entre diversidad funcional y la abundancia de los heterótrofos es aún mayor y significativa (Figura S7).

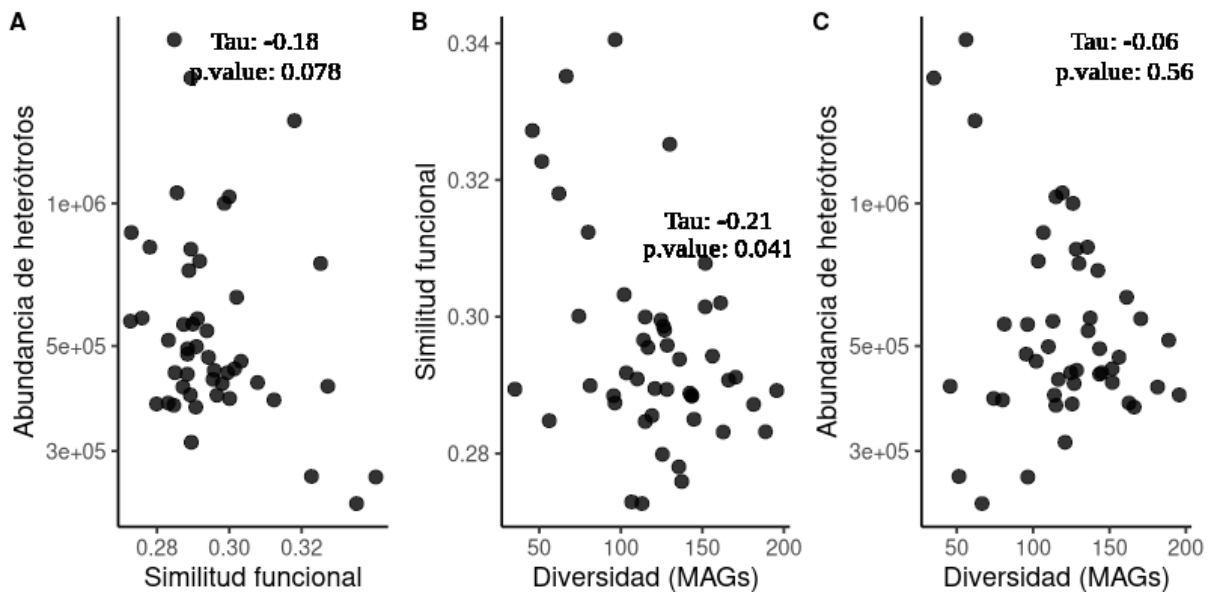


Figura 4.4. A menor similitud funciona (es decir, mayor diversidad) se observa una mayor abundancia de heterótrofos y una mayor diversidad. Los cambios en diversidad no explican los cambios en abundancia. A) Similitud funcional (MAGs) versus abundancia de heterótrofos. B) Diversidad (índice de Hill, MAGs) versus similitud funcional (MAGs). C) Diversidad (índice de Hill, MAGs) versus abundancia de heterótrofos. Para obtener esta figura se tuvieron en cuenta todos los MAGs bacterianos, excepto los pertenecientes al filo *Cyanobacteria*.

Este análisis sugiere que la diversidad estimada como número de especies efectivas no tendría un efecto sobre la abundancia de heterótrofos. En cambio, la diversidad funcional, o la existencia de especies capaces de explotar recursos diferentes, se asocia positivamente con la abundancia. Resultados similares se encontraron analizando especies de zooplancton (malla de 0.33 mm) presentes en estanques.⁹⁰

4.4. Efecto de muestreo

Se ha propuesto que el funcionamiento de un ecosistema podría depender de otro componente de la diversidad, llamado equitatividad (*evenness*). La misma se estima mediante el índice de Pielou, va desde 0 a 1 y pondera la contribución relativa de

cada especie al número total de individuos.⁹¹⁻⁹³ Una mayor equitatividad indica que la abundancia relativa de las especies es similar. A modo de ejemplo, si en una muestra tuviéramos tres especies, todas con el mismo número de individuos, la equitatividad sería igual a uno.

En los paneles A y B de la Figura 4.5 se muestran las relaciones entre el número de especies efectivas y la equitatividad usando todos los mOTUs o una selección que incluye solamente aquellos del género *Prochlorococcus* y *Synechococcus*. En ambos casos, existe una asociación positiva significativa, de lo cual se deduce que un incremento en el número de especies se ve acompañado por una mayor equitatividad, es decir mOTUs con frecuencias relativas similares. Sorprendentemente, a pesar de la relación hallada entre el número efectivo de especies (de los mOTUs de *Prochlorococcus* y *Synechococcus*) y la abundancia de autótrofos (Figura 3.6. D), y la equitatividad, no existe una relación entre estas últimas dos variables. En otras palabras, el número efectivo de especies de *Prochlorococcus* y *Synechococcus* tiene un efecto sobre la abundancia de autótrofos, pero no así las frecuencias relativas de las especies. En el caso de los heterótrofos, la abundancia es también independiente de la equitatividad (Figura 4.5 C). No es posible debido al bajo número de datos evaluar la relación entre productividad y equitatividad usando un número constante de especies (por ejemplo, restringir entre el panel C solamente a aquellos puntos que en el panel A caen en el rango 190-210 del índice de Hill). Además, vale la pena señalar que en el caso de los heterótrofos (Figura 4.5 C), incluso cuando los valores de equitatividad son bajos, sigue habiendo alta abundancia.

La evidencia que encontramos sugiere que el efecto de muestreo no sería un mecanismo tan relevante. Si una especie fuera altamente dominante esperaríamos encontrar que los sitios con menor equitatividad son los de mayor abundancia. El patrón sugiere lo contrario por lo que este mecanismo, si bien no puede ser descartado, no parece ser el más importante para el bacterioplancton. Sin embargo, el efecto de muestreo parece ocurrir en ecosistemas naturales, como en ecosistemas naturales altamente invadidos o dominados por una única especie.²³

En este sentido, las floraciones extensas que se observan a lo largo de la plataforma patagónica podrían ser un ejemplo en el ecosistema marino.⁹⁴

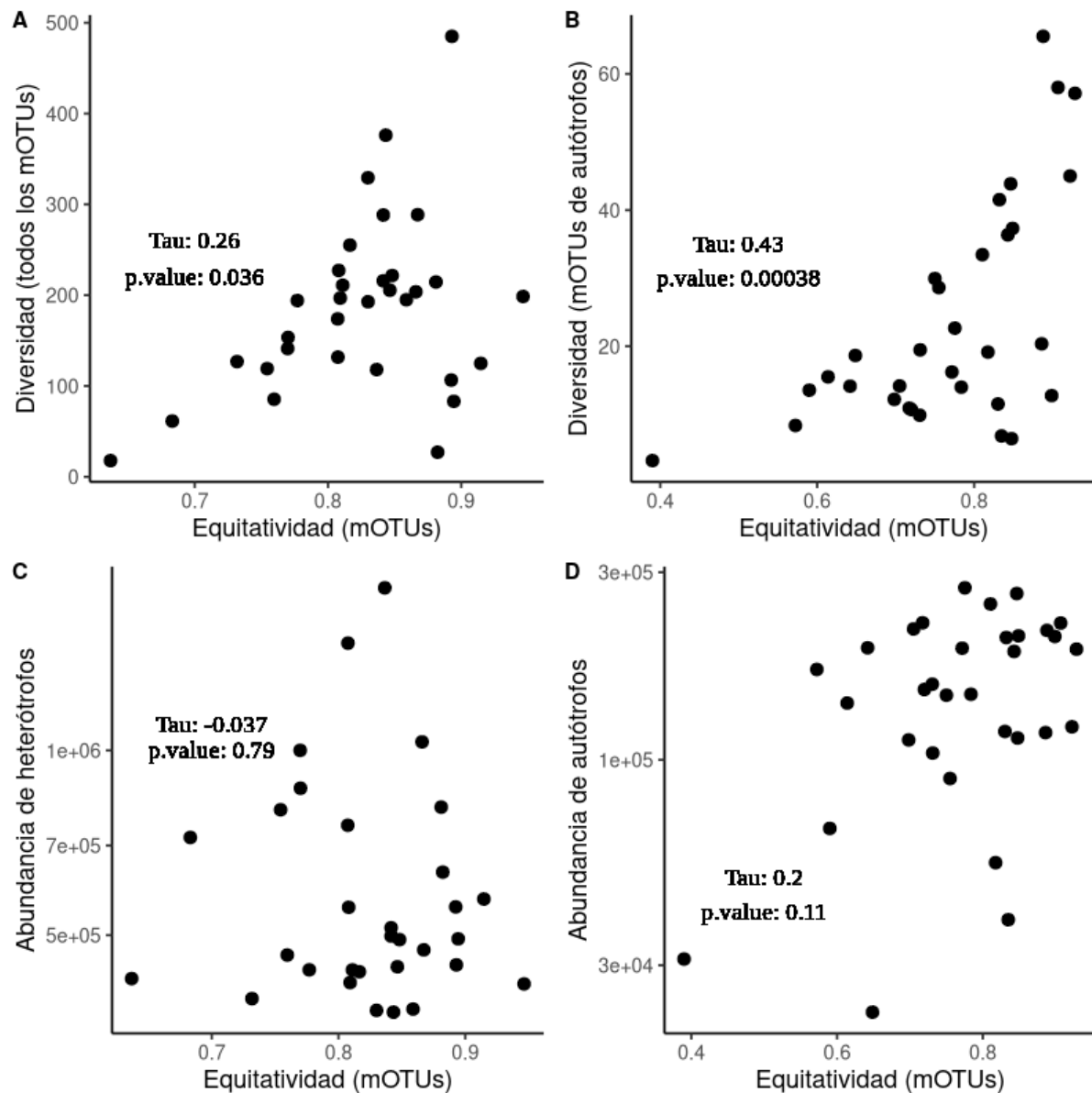


Figura 4.5. Relación entre la equitatividad y el número de especies o abundancia. A) Número efectivo de especies (*Hill index*) versus equitatividad usando todos los mOTUs. Kendall tau de 0.26, *p*-valor 0.04. B) Número efectivo de especies (*Hill index*) versus equitatividad usando solo los mOTUs correspondientes al género *Prochlorococcus* y *Synechococcus*. Kendall tau de 0.43, *p*-valor 0.0004. C) Abundancia de heterótrofos versus equitatividad usando todos los mOTUs. Kendall tau no significativo. D) Abundancia de autótrofos versus equitatividad usando solo los mOTUs correspondientes al género *Prochlorococcus* y *Synechococcus*. Kendall tau no significativo.

En investigaciones previas, Cermeño *et al*, 2016, probaron usando medidas de la comunidad de fitoplancton (sin filtro de tamaño), flujos de nitrato y productividad determinada con carbono radioactivo, que para un ecosistema costero templado la productividad primaria se puede explicar por un efecto de selección.⁸⁷ De hecho, encontraron una asociación positiva entre la dominancia (índice de Berger-Parker en término de la biomasa de Carbono) de una muestra y la biomasa total. Dicho de otro modo, mientras mayor era la biomasa, mayor era la proporción explicada por una sola especie. Nuestros resultados usando el índice de equitatividad de Pielou no corroboran dichas observaciones. Más aún, si analizamos el número de especies necesarias para explicar al menos el 80 % del número total de individuos observados, encontramos en el caso de las bacterias autótrofas que mientras mayor es la biomasa (estimada mediante abundancia total por citometría de flujo), esta se encuentra distribuida entre más especies (Figura 4.6). Estos resultados sugieren que el picofitoplancton bacteriano (picocianobacterias) no está sujeto a las mismas leyes que el nano y microfitoplancton, que sí suelen lograr grandes floraciones.

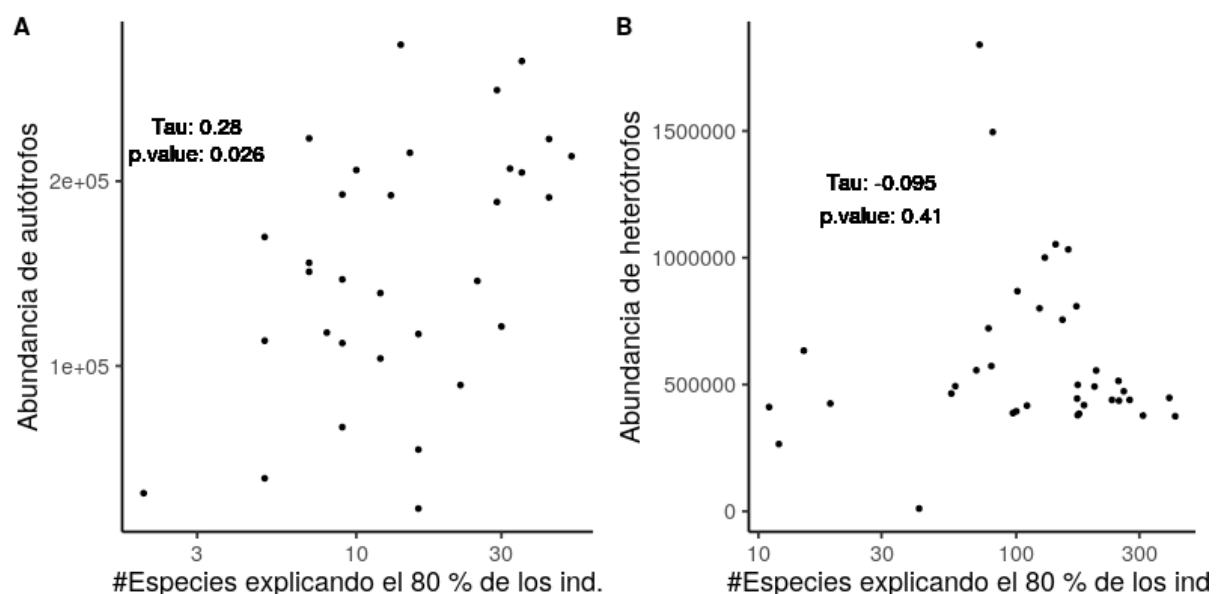


Figura 4.6. Abundancia medida por citometría de flujo versus número de especies distintas cuyo número de individuos combinados representa el 80 % de los individuos observados

(dataset mOTUs). A) Grupo de autótrofos (*Prochlorococcus* y *Synechococcus*). B) Grupo de heterótrofos.

Capítulo 5

Conclusiones

Se prevé que las actividades antropogénicas continúen produciendo cambios ambientales irreversibles y pérdidas de biodiversidad en las siguientes décadas.⁹⁵ Es menester entender los factores que afectan el funcionamiento de un ecosistema para poder anticiparnos e incrementar nuestras posibilidades de adaptación a nuevas condiciones. El plancton marino es esencial para la vida en el planeta y el funcionamiento de los ciclos biogeoquímicos, y sus comunidades no están exentas a los efectos del cambio global. En ecosistemas terrestres, la relación positiva entre diversidad y productividad cuenta con numerosos estudios que la avalan, pero en ecosistemas marinos la evidencia es escasa.

En este trabajo de Tesis, se hizo provecho de los datos a nivel de abundancia de grupos, de datos genómicos y datos físico químicos para evaluar la relación entre la diversidad del plancton y la productividad y los mecanismos asociados: la facilitación, el efecto de muestreo, y la complementariedad de nicho. El hecho de contar con datos moleculares de ADN representa una gran ventaja ya que ofrece alta resolución taxonómica y permite tener información de genes con funciones de impacto ecosistémico. En principio, hallamos una débil (pero significativa) asociación entre la diversidad y abundancia de autótrofos y ninguna asociación en el caso de los heterótrofos. En los siguientes párrafos se discuten las conclusiones de los análisis pertinentes y las posibles causas.

A la hora de investigar las relaciones entre la abundancia y la diversidad se deben considerar varios factores. Por un lado, es importante asegurarse que los valores de diversidad y abundancia provengan de un muestreo homogéneo y que el mismo capture de forma representativa el estado de la comunidad. Debido a este motivo, fue necesario remover del análisis a los picoeucariotas por el sesgo en el tamaño de filtro utilizado, y remover también dos muestras con comunidades sub-muestreadas

(ver Capítulo 3). Además, se debe verificar que no existan covariables asociadas a las variables de estudio. Por ejemplo, se demostró una fuerte relación entre la biodiversidad y la abundancia para el grupo de los autótrofos. No obstante, dicha asociación puede ser explicada en parte por la temperatura. De hecho, se ha propuesto que los factores abióticos son dominantes sobre los procesos ecosistémicos, pero que, de forma secundaria, cambios en la biodiversidad traerán aparejados cambios en la biogeoquímica marina. En el caso de los heterótrofos, los valores de abundancia resultaron ser independientes de la diversidad medida como el número efectivo de especies.

Respecto a los factores que gobiernan el ensamblado de las comunidades marinas, y por lo tanto influyen en la biodiversidad de las mismas, el análisis de similitud entre estaciones sugiere que el océano es un gran reservorio interconectado y de alta dispersión microbiana, y que las condiciones ambientales son el principal factor de selección. Luego, las distribuciones de abundancia revelan la presencia de pocas especies muy abundantes (y muchas especies poco abundantes) y pueden ser parcialmente explicadas por el modelo Poisson log-normal. El hecho de que la especie dominante difiere siempre entre muestras sugiere que existe un proceso de selección transiente y cambiante. Probablemente, de repetirse el muestreo de Tara Oceans en las mismas ubicaciones y momentos del año, nuevamente las especies dominantes sean diferentes a las previamente observadas.

En cuanto a los mecanismos biológicos que explican asociaciones positivas entre la diversidad y la abundancia, hallamos evidencia preliminar de facilitación en el grupo de los autótrofos basándonos en el análisis metagenómico del gen *nifH* (fijación de N_2). Por otra parte, y de forma separada del análisis diversidad-abundancia, en el grupo de los heterótrofos se encontró evidencia potencial para el mecanismo de complementariedad de nicho cuando se analizó la diversidad funcional en diferentes grupos bacterianos. El contenido genético que diferencia a los taxones bacterianos del océano sugiere fuertemente que existen especies que explotan diferentes nichos. Por último, basándonos en nuestro análisis, no hay evidencia suficiente para

postular al efecto de muestreo como un mecanismo que estructura las comunidades de plancton marino.

Las limitaciones del conjunto de datos son varias, ya que no se trata de un muestreo diseñado para poner a prueba esta hipótesis, el número de muestras es reducido, no se cuenta con datos de productividad real (tasa de generación de biomasa por unidad de tiempo y superficie) sino con una aproximación vía abundancias celulares, y ninguna de las variables analizadas fue controlada (estudio observacional). A pesar de las limitaciones mencionadas, este trabajo representa un primer esfuerzo para avanzar en la comprensión del efecto de la diversidad en el funcionamiento de los ecosistemas microbianos marinos a escala global. Futuros experimentos controlados en estaciones de mesocosmos permitirán establecer patrones y mecanismos con mayor claridad.

Información suplementaria

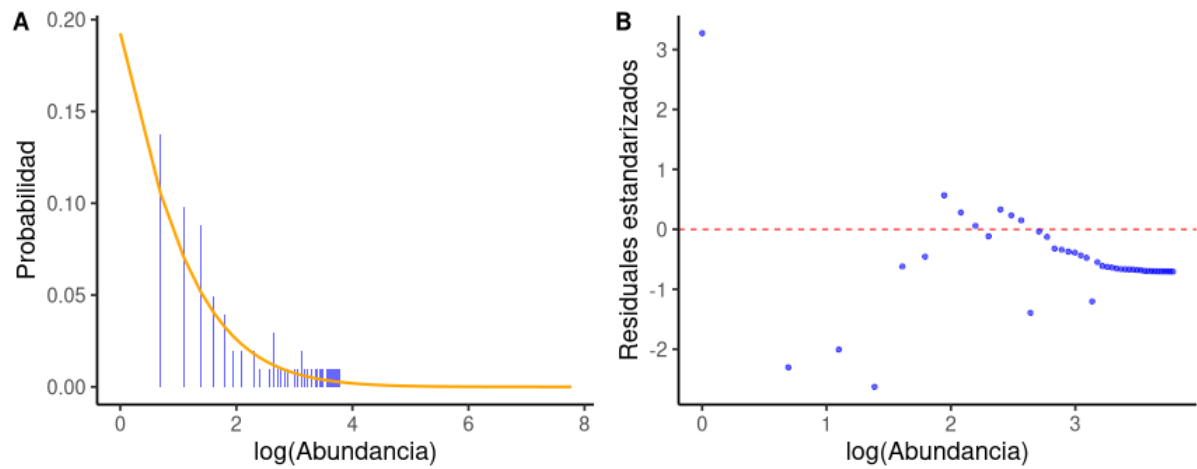


Figura S1. A) Probabilidad observada para la abundancia de especies (mOTUs) versus probabilidad predicha por el ajuste del modelo Poisson lognormal. B) Residuales estandarizados.

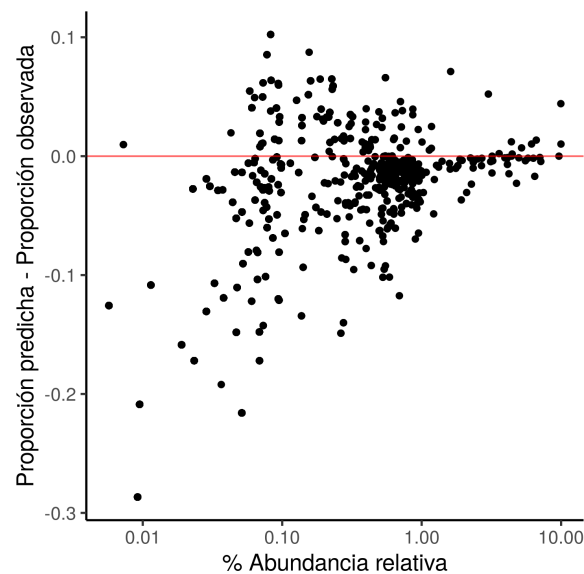


Figura S2. Diferencia entre la proporción acumulada de individuos con cierta abundancia relativa predicha por el modelo Poisson lognormal versus la observada.

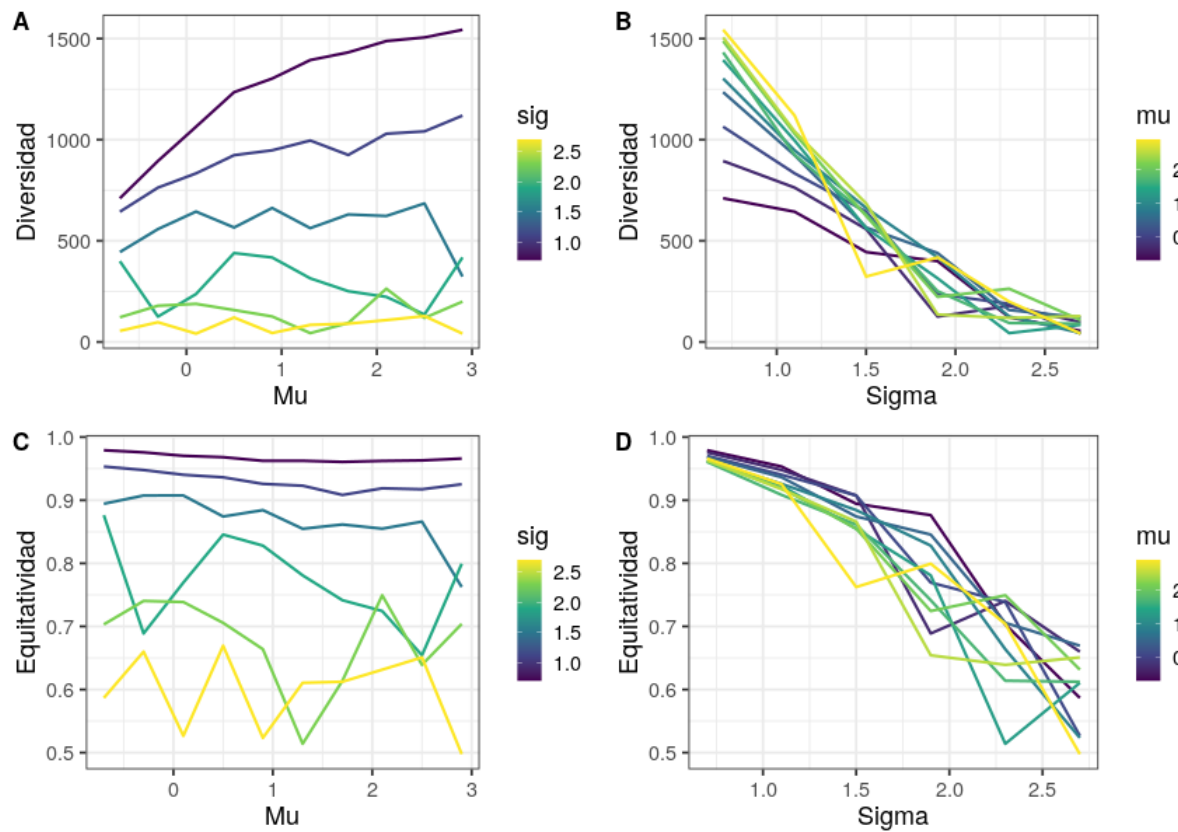


Figura S3. Efecto de los parámetros μ y σ sobre la diversidad y equitatividad de una muestra. Todas las curvas fueron simuladas usando 2200 especies diferentes. A) Número efectivo de especies (índice de Hill) versus μ a distintos valores de σ constante. B) Número efectivo de especies versus σ a distintos valores de μ constante. C) Equitatividad (índice de Pielou) versus μ a distintos valores de σ constante. D) Equitatividad versus σ a distintos valores de μ constante.

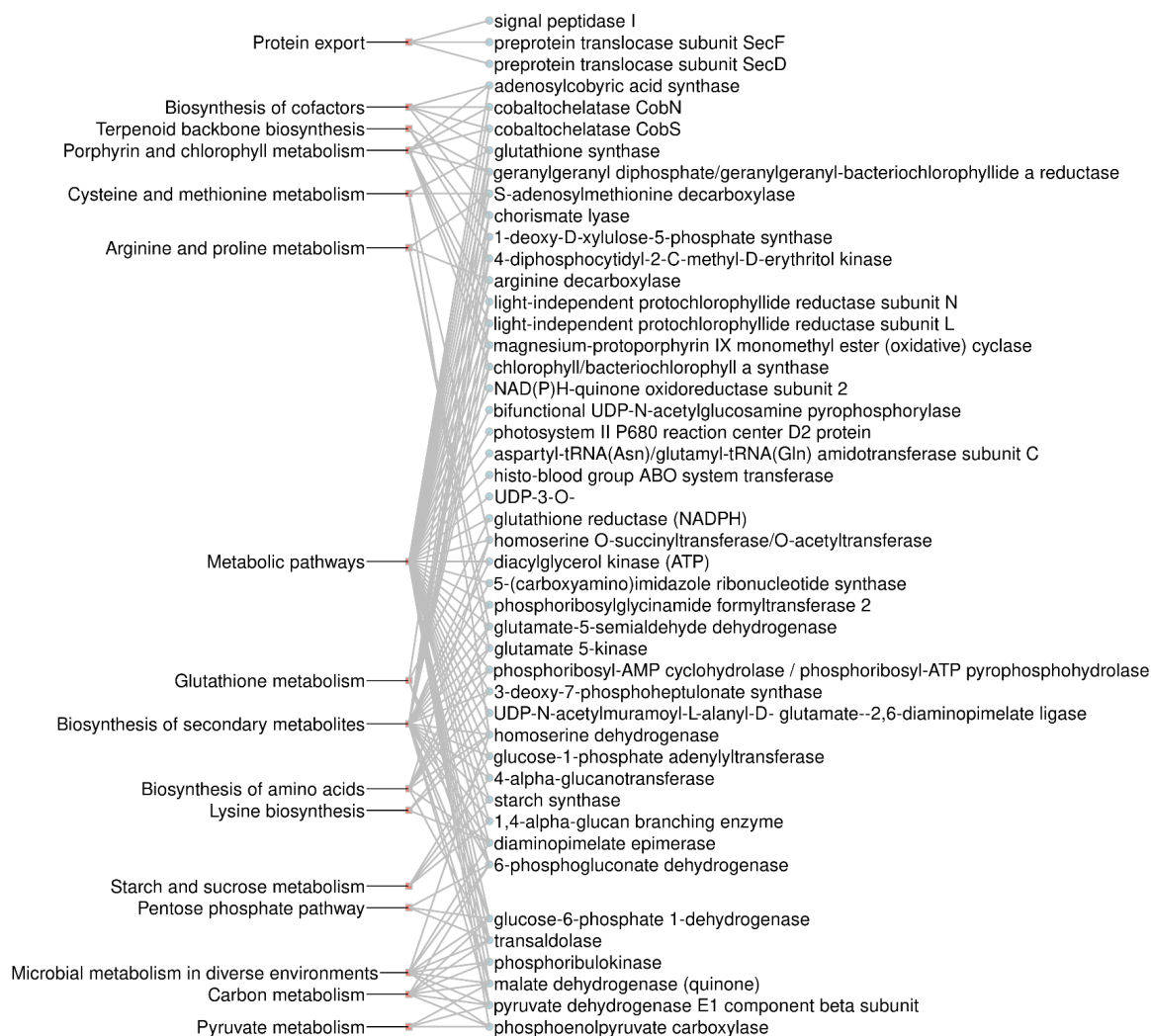


Figura S4. Genes cuya abundancia se hallaba asociada positivamente con la abundancia de autótrofos contados por citometría de flujo. Primero, se calculó para 9000 genes el coeficiente de correlación de Tau y el p -valor bajo la hipótesis nula de que el coeficiente es igual a 0. Luego de corregir por *False Discovery Rate*, se buscó en la base de datos de KEGG, en qué vía metabólica participan todos aquellos genes con p -valor < 0.03.^{96–98} Finalmente, para lograr la figura, se representaron como un grafo bipartito la vía metabólica (izquierda) y los genes (derecha), usando todas las vías metabólicas con al menos 3 genes correlacionados.

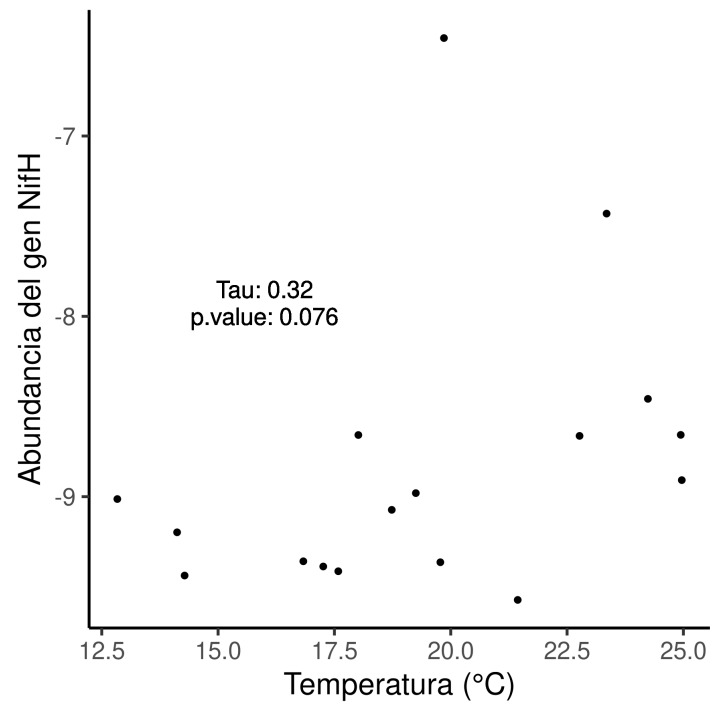


Figura S5. Temperatura versus abundancia del gen nifH

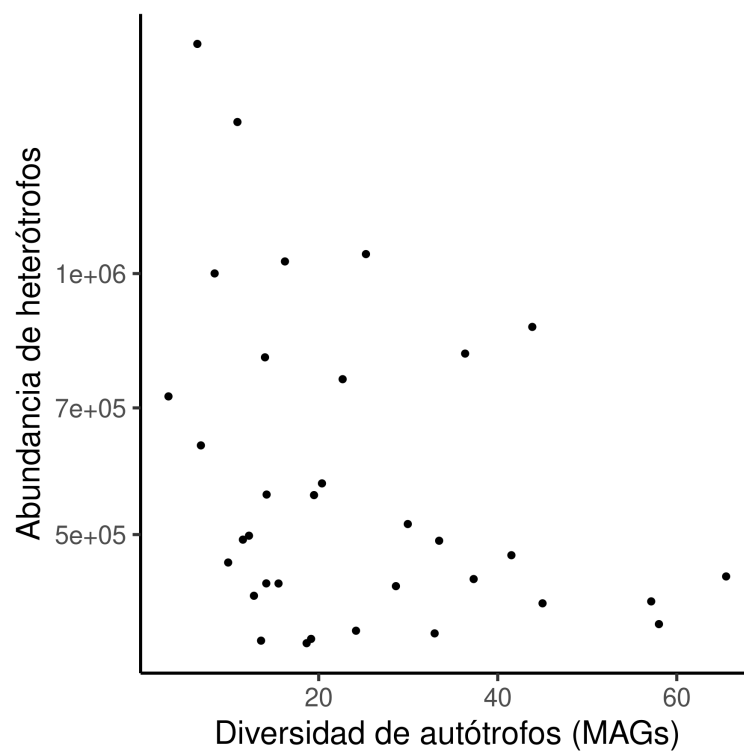


Figura S6. Abundancia de heterótrofos (citometría de flujo) versus diversidad de autótrofos (MAGs). Kendall tau: -0.2, p-valor 0.075.

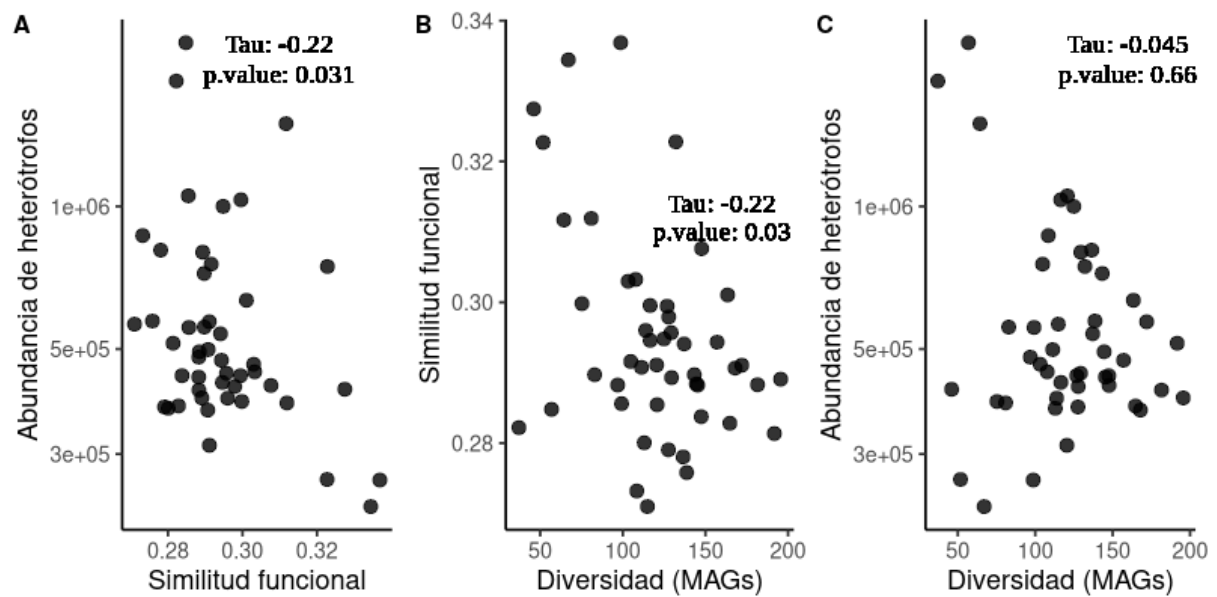


Figura S7. Mayor diversidad genómica implica una mayor abundancia de heterótrofos y una mayor diversidad. Los cambios en diversidad no explican los cambios en abundancia. A) Similitud funcional (MAGs) versus abundancia de heterótrofos. B) Similitud funcional (MAGs) versus diversidad (índice de Hill, MAGs). C) Diversidad (índice de Hill, MAGs) (MAGs) versus abundancia de heterótrofos. Para obtener esta figura se tuvieron en cuenta todos los MAGs bacterianos.

Agradecimientos

Agradezco enormemente a todas las personas que contribuyeron directa, o indirectamente, a la realización de esta Tesis. En especial,

A mis directores Federico M. Ibarbalz y Pedro Flombaum, por su buena onda, por guiarme en el campo de la ecología microbiana marina, por transmitirme su pasión por el tema y por dedicarme su tiempo.

A mis padres Alejandro Burastero y Liliana Giron por brindarme un entorno de trabajo cómodo y positivo durante la cuarentena.

A la Maestría en Explotación de Datos y Descubrimiento del Conocimiento de la Facultad de Ciencias Exactas y Naturales, por la formación brindada.

A Marcelo Soria, previo director de la Maestría en Explotación de Datos y Descubrimiento del Conocimiento, por su buena predisposición y ayuda.

Índice de figuras

Figura 1.1 - Mecanismos hipotéticos que correlacionan positivamente biodiversidad con productividad.

Figura 1.2 - Distintas maneras de graficar DAE. Los datos de abundancia de árboles recolectados por Whittaker en las montañas Siskiyou son graficados en tres formatos distintos.

Figura 1.3 - Esquema de la expedición TARA OCEANS, incluyendo las 210 estaciones de muestreo y los datos adquiridos.

Figura 3.1 - Matriz de correlación de las variables ambientales.

Figura 3.2 - Variables ambientales relacionados entre sí de acuerdo al criterio de información mutua.

Figura 3.3 - Distribución de los valores de citometría de flujo. Relación entre la abundancia de picoeucariotas y la concentración de clorofila. Relación entre la abundancia de autótrofos y la temperatura o la radiación fotosintética disponible.

Figura 3.4 - Curvas de acumulación de nuevos tipos de individuos versus proporción de individuos muestreados (mOTUs).

Figura 3.5 - Valor del parámetro de ajuste b de la función michaeliana a las curvas de acumulación.

Figura 3.6 - Similitud en diversidad para las muestras de superficie con filtro de tamaño entre 0.22 y 3 μm .

Figura 3.7. Composición taxonómica y abundancia celular en el grupo de los autótrofos.

Figura 3.7 - Distribución de abundancia de especies para la estación 60 (mOTUs).

Figura 3.8 - Distribución empírica acumulada para la abundancia de especies en escala logarítmica de base 10 para 10 estaciones elegidas al azar.

Figura 4.1 - Productividad y biodiversidad. Abundancia medida por citometría de flujo versus número efectivo de especies.

Figura 4.2 - Relación entre la abundancia celular y la del gen fijador de N₂ como evidencia de mecanismos de facilitación en el rango de temperatura de 12 a 25 grados.

Figura 4.3 - Clusters obtenidos por similitud funcional.

Figura 4.4 - Relaciones entre similitud funcional, diversidad y abundancia de heterótrofos.

Figura 4.5 - Relación entre la equitatividad y el número de especies o abundancia.

Figura 4.6 - Abundancia medida por citometría de flujo versus número de especies distintas cuyo número de individuos combinados representa el 80 % de los individuos observados.

Índice de tablas

Tabla 2.1 - Descripción de las variables fisicoquímicas asociadas a cada muestra.

Tabla 2.2 - Índices de biodiversidad y fórmulas asociadas.

Tabla 4.1 - Relaciones entre diversidad y mecanismos exploradas en este trabajo de Tesis.

Bibliografía

1. Henson, S. A., Sanders, R. & Madsen, E. Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochemical Cycles* vol. 26 Preprint at <https://doi.org/10.1029/2011gb004099> (2012).
2. Guidi, L. et al. A new look at ocean carbon remineralization for estimating deepwater sequestration. *Global Biogeochemical Cycles* vol. 29 1044–1059 Preprint at <https://doi.org/10.1002/2014gb005063> (2015).
3. Kwon, E. Y., Primeau, F. & Sarmiento, J. L. The impact of remineralization depth on the air–sea carbon balance. *Nature Geoscience* vol. 2 630–635 Preprint at <https://doi.org/10.1038/ngeo612> (2009).
4. Azam, F. et al. The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series* vol. 10 257–263 Preprint at <https://doi.org/10.3354/meps010257> (1983).
5. Schlesinger, W. H. The Global Carbon Cycle. *Biogeochemistry: an Analysis of Global Change* 308–321 Preprint at <https://doi.org/10.1016/b978-0-12-625157-9.50016-2> (1991).
6. Vitousek, P. M. et al. HUMAN ALTERATION OF THE GLOBAL NITROGEN CYCLE: SOURCES AND CONSEQUENCES. *Ecological Applications* vol. 7 737–750 Preprint at [https://doi.org/10.1890/1051-0761\(1997\)007\[0737:haotgn\]2.0.co;2](https://doi.org/10.1890/1051-0761(1997)007[0737:haotgn]2.0.co;2) (1997).
7. Sala, O. E. et al. Biodiversity and ecosystem functioning in grasslands. *Functional roles of biodiversity: a global perspective*. 129–149 (1996).
8. Tilman, D. The Ecological Consequences of Changes in Biodiversity: A Search for General Principles. *Ecology* vol. 80 1455 Preprint at <https://doi.org/10.2307/176540> (1999).
9. Bruno, J. F., Stachowicz, J. J. & Bertness, M. D. Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution* vol. 18 119–125 Preprint at [https://doi.org/10.1016/s0169-5347\(02\)00045-9](https://doi.org/10.1016/s0169-5347(02)00045-9) (2003).
10. Cardinale, B. J., Palmer, M. A. & Collins, S. L. Species diversity enhances ecosystem functioning through interspecific facilitation. *Nature* **415**, 426–429 (2002).
11. Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).

12. Loreau, M. et al. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science* **294**, 804–808 (2001).
13. Tilman, D., Lehman, C. L. & Thomson, K. T. Plant diversity and ecosystem productivity: Theoretical considerations. *Proceedings of the National Academy of Sciences* vol. 94 1857–1861 Preprint at <https://doi.org/10.1073/pnas.94.5.1857> (1997).
14. Aarssen, L. W. High Productivity in Grassland Ecosystems: Effected by Species Diversity or Productive Species? *Oikos* vol. 80 183 Preprint at <https://doi.org/10.2307/3546531> (1997).
15. Hector, A. et al. Plant diversity and productivity experiments in european grasslands. *Science* **286**, 1123–1127 (1999).
16. Loreau, M. & Hector, A. Partitioning selection and complementarity in biodiversity experiments. *Nature* **412**, 72–76 (2001).
17. van Ruijven, J. & Berendse, F. Diversity-productivity relationships: initial effects, long-term patterns, and underlying mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 695–700 (2005).
18. Wright, A. J., Barry, K. E., Lortie, C. J. & Callaway, R. M. Biodiversity and ecosystem functioning: Have our experiments and indices been underestimating the role of facilitation? *Journal of Ecology* vol. 109 1962–1968 Preprint at <https://doi.org/10.1111/1365-2745.13665> (2021).
19. Mulder, C. P., Uliassi, D. D. & Doak, D. F. Physical stress and diversity-productivity relationships: the role of positive interactions. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6704–6708 (2001).
20. Bertness, M. D. & Callaway, R. Positive interactions in communities. *Trends Ecol. Evol.* **9**, 191–193 (1994).
21. Wright, A. J., Mommer, L., Barry, K. & van Ruijven, J. Stress gradients and biodiversity: monoculture vulnerability drives stronger biodiversity effects during drought years. *Ecology* **102**, e03193 (2021).
22. Duffy, J. E., Godwin, C. M. & Cardinale, B. J. Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature* **549**, 261–264 (2017).
23. Flombaum, P., Aragón, R. & Chanton, E. J. A role for the sampling effect in invaded ecosystems. *Oikos* vol. 126 1229–1232 Preprint at <https://doi.org/10.1111/oik.04221> (2017).
24. Currie, D. J. & Paquin, V. Large-scale biogeographical patterns of species richness of trees. *Nature* vol. 329 326–327 Preprint at <https://doi.org/10.1038/329326a0> (1987).

25. Wang, Z., Brown, J. H., Tang, Z. & Fang, J. Temperature dependence, spatial scale, and tree species diversity in eastern Asia and North America. *Proceedings of the National Academy of Sciences* vol. 106 13388–13392 Preprint at <https://doi.org/10.1073/pnas.0905030106> (2009).
26. Hlásny, T. et al. Climatic drivers of forest productivity in Central Europe. *Agricultural and Forest Meteorology* vols 234-235 258–273 Preprint at <https://doi.org/10.1016/j.agrformet.2016.12.024> (2017).
27. Toledo, M. et al. Climate is a stronger driver of tree and forest growth rates than soil and disturbance. *Journal of Ecology* vol. 99 254–264 Preprint at <https://doi.org/10.1111/j.1365-2745.2010.01741.x> (2011).
28. Tilman, D., Reich, P. B. & Isbell, F. Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proceedings of the National Academy of Sciences* vol. 109 10394–10397 Preprint at <https://doi.org/10.1073/pnas.1208240109> (2012).
29. Vellend, M. Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology* vol. 85 183–206 Preprint at <https://doi.org/10.1086/652373> (2010).
30. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M. & Relman, D. A. The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science* vol. 336 1255–1262 Preprint at <https://doi.org/10.1126/science.1224203> (2012).
31. Sogin, M. L. et al. Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proceedings of the National Academy of Sciences* vol. 103 12115–12120 Preprint at <https://doi.org/10.1073/pnas.0605127103> (2006).
32. Lynch, M. D. J. & Neufeld, J. D. Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology* vol. 13 217–229 Preprint at <https://doi.org/10.1038/nrmicro3400> (2015).
33. McGill, B. J. et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* **10**, 995–1015 (2007).
34. Horner-Devine, M. C., Claire Horner-Devine, M., Lage, M., Hughes, J. B. & Bohannan, B. J. M. A taxa–area relationship for bacteria. *Nature* vol. 432 750–753 Preprint at <https://doi.org/10.1038/nature03073> (2004).
35. Hubbell, S. P. The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32). Preprint at <https://doi.org/10.1515/9781400837526> (2011).

36. McGill, B. J. Towards a unification of unified theories of biodiversity. *Ecology Letters* vol. 13 627–642 Preprint at <https://doi.org/10.1111/j.1461-0248.2010.01449.x> (2010).
37. Harte, J. *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. (OUP Oxford, 2011).
38. Putman, R. J. & Putman, R. *Community Ecology*. (Springer Science & Business Media, 1994).
39. Dunbar, J., Barns, S. M., Ticknor, L. O. & Kuske, C. R. Empirical and Theoretical Bacterial Diversity in Four Arizona Soils. *Applied and Environmental Microbiology* vol. 68 3035–3045 Preprint at <https://doi.org/10.1128/aem.68.6.3035-3045.2002> (2002).
40. Curtis, T. P., Sloan, W. T. & Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences* vol. 99 10494–10499 Preprint at <https://doi.org/10.1073/pnas.142680199> (2002).
41. Bohannan, B. J. M. & Hughes, J. New approaches to analyzing microbial biodiversity data. *Current Opinion in Microbiology* vol. 6 282–287 Preprint at [https://doi.org/10.1016/s1369-5274\(03\)00055-9](https://doi.org/10.1016/s1369-5274(03)00055-9) (2003).
42. Schloss, P. D. & Handelsman, J. Status of the Microbial Census. *Microbiol. Mol. Biol. Rev.* (2004) doi:10.1128/MMBR.68.4.686-691.2004.
43. Pedrós-Alió, C. & Manrubia, S. The vast unknown microbial biosphere. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6585–6587 (2016).
44. Jaynes, E. T. *Probability Theory: The Logic of Science*. (Cambridge University Press, 2003).
45. White, E. P., Thibault, K. M. & Xiao, X. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology* vol. 93 1772–1778 Preprint at <https://doi.org/10.1890/11-2177.1> (2012).
46. Baldrige, E., Harris, D. J., Xiao, X. & White, E. P. An extensive comparison of species-abundance distribution models. Preprint at <https://doi.org/10.1101/024802>.
47. Shoemaker, W. R., Locey, K. J. & Lennon, J. T. A macroecological theory of microbial biodiversity. *Nat Ecol Evol* **1**, 107 (2017).
48. Sunagawa, S. et al. Tara Oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* **18**, 428–445 (2020).
49. Sunagawa, S. et al. Ocean plankton. Structure and function of the global ocean microbiome.

- Science* **348**, 1261359 (2015).
50. Salazar, G. *et al.* Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell* **179**, 1068–1083.e21 (2019).
 51. Ibarbalz, F. M. *et al.* Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* **179**, 1084–1097.e21 (2019).
 52. Farrant, G. K. *et al.* Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3365–74 (2016).
 53. Grébert, T. *et al.* Light color acclimation is a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proceedings of the National Academy of Sciences* vol. 115 E2010–E2019 Preprint at <https://doi.org/10.1073/pnas.1717069115> (2018).
 54. Villar, E. *et al.* Ocean plankton. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **348**, 1261447 (2015).
 55. Caputi, N. *et al.* Factors Affecting the Recovery of Invertebrate Stocks From the 2011 Western Australian Extreme Marine Heatwave. *Frontiers in Marine Science* vol. 6 Preprint at <https://doi.org/10.3389/fmars.2019.00484> (2019).
 56. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* **3**, 804–813 (2018).
 57. Milanese, A. *et al.* Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014 (2019).
 58. Eren, A. M. *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology* **6**, 3–6 (2020).
 59. Artale, V., Boffetta, G., Celani, A., Cencini, M. & Vulpiani, A. Dispersion of passive tracers in closed basins: Beyond the diffusion coefficient. *Physics of Fluids* vol. 9 3162–3171 Preprint at <https://doi.org/10.1063/1.869433> (1997).
 60. Okubo, A. Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences. *Deep Sea Research and Oceanographic Abstracts* vol. 17 445–454 Preprint at [https://doi.org/10.1016/0011-7471\(70\)90059-8](https://doi.org/10.1016/0011-7471(70)90059-8) (1970).
 61. Condie, S. & Condie, R. Retention of plankton within ocean eddies. *Global Ecology and Biogeography* vol. 25 1264–1277 Preprint at <https://doi.org/10.1111/geb.12485> (2016).

62. Falkowski, P. G., Barber, R. T. & Smetacek, V., V. Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* **281**, 200–207 (1998).
63. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs* vol. 84 45–67 Preprint at <https://doi.org/10.1890/13-0133.1> (2014).
64. Magurran, A. E. & McGill, B. J. *Biological Diversity: Frontiers in Measurement and Assessment*. (Oxford University Press, 2011).
65. Prado, P. I., Miranda, M. D. & Chalom, A. sads: Maximum Likelihood Models for Species Abundance Distributions. Preprint at <https://CRAN.R-project.org/package=sads> (2018).
66. Dunn, J. C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics* vol. 3 32–57 Preprint at <https://doi.org/10.1080/01969727308546046> (1973).
67. Dunn, J. C. Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics* vol. 4 95–104 Preprint at <https://doi.org/10.1080/01969727408546059> (1974).
68. Wickham, H. et al. Welcome to the tidyverse. *Journal of Open Source Software* vol. 4 1686 Preprint at <https://doi.org/10.21105/joss.01686> (2019).
69. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Preprint at <https://ggplot2.tidyverse.org> (2016).
70. Wei, T. & Simko, V. R package 'corrplot': Visualization of a Correlation Matrix. <https://github.com/taiyun/corrplot> (2017).
71. Dragulescu, A. & Arendt, C. xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files. Preprint at <https://CRAN.R-project.org/package=xlsx> (2020).
72. Kratzer, G. & Furrer, R. varrank: an R package for variable ranking based on mutual information with applications to observed systemic datasets. Preprint at <https://CRAN.R-project.org/package=varrank> (2018).
73. Wickham, H. Reshaping Data with the reshape Package. *Journal of Statistical Software* vol. 21 1–20 Preprint at <http://www.jstatsoft.org/v21/i12/> (2007).
74. Grøtan, V. & Engen, S. poilog: Poisson lognormal and bivariate Poisson lognormal distribution. Preprint at (2008).

75. Pardy, C. mpmi: Mixed-Pair Mutual Information Estimators. Preprint at <https://CRAN.R-project.org/package=mpmi> (2020).
76. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots. Preprint at <https://CRAN.R-project.org/package=ggpubr> (2020).
77. Robinson, D., Hayes, A. & Couch, S. broom: Convert Statistical Objects into Tidy Tibbles. Preprint at <https://CRAN.R-project.org/package=broom> (2022).
78. Website. Elzhov, Timur V., Katharine M. Mullen, and Spiess A. & Bolker Ben. 2022. 'Minpack. Lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, plus Support for Bounds.' <https://CRAN.R-project.org/package=minpack.lm>.
79. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
80. Flombaum, P., Wang, W.-L., Primeau, F. W. & Martiny, A. C. Global picophytoplankton niche partitioning predicts overall positive response to ocean warming. *Nature Geoscience* vol. 13 116–120 Preprint at <https://doi.org/10.1038/s41561-019-0524-2> (2020).
81. Flombaum, P. et al. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9824–9829 (2013).
82. Hughes, J. B., Hellmann, J. J., Ricketts, T. H. & Bohannan, B. J. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67**, 4399–4406 (2001).
83. Balvanera, P. et al. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9**, 1146–1156 (2006).
84. Cardinale, B. J. et al. The functional role of producer diversity in ecosystems. *Am. J. Bot.* **98**, 572–592 (2011).
85. O'Connor et al. A general biodiversity–function relationship is mediated by trophic level. *Oikos*.
86. Gonzalez, A. et al. Scaling-up biodiversity-ecosystem functioning research. *Ecol. Lett.* **23**, 757–776 (2020).
87. Cermeño, P. et al. Marine Primary Productivity Is Driven by a Selection Effect. *Frontiers in Marine Science* vol. 3 Preprint at <https://doi.org/10.3389/fmars.2016.00173> (2016).
88. van der Plas, F. Biodiversity and ecosystem functioning in naturally assembled communities. *Biol.*

- Rev. Camb. Philos. Soc.* **94**, 1220–1245 (2019).
89. Flombaum, P. & Martiny, A. C. Diverse but uncertain responses of picophytoplankton lineages to future climate change. *Limnology and Oceanography* vol. 66 4171–4181 Preprint at <https://doi.org/10.1002/lno.11951> (2021).
 90. Chase, J. M. & Leibold, M. A. Spatial scale dictates the productivity–biodiversity relationship. *Nature* vol. 416 427–430 Preprint at <https://doi.org/10.1038/416427a> (2002).
 91. Wilsey, B. J. & Potvin, C. BIODIVERSITY AND ECOSYSTEM FUNCTIONING: IMPORTANCE OF SPECIES EVENNESS IN AN OLD FIELD. *Ecology* vol. 81 887–892 Preprint at [https://doi.org/10.1890/0012-9658\(2000\)081\[0887:baefio\]2.0.co;2](https://doi.org/10.1890/0012-9658(2000)081[0887:baefio]2.0.co;2) (2000).
 92. Wilsey, B. J. & Wayne Polley, H. Reductions in grassland species evenness increase dicot seedling invasion and spittle bug infestation. *Ecology Letters* vol. 5 676–684 Preprint at <https://doi.org/10.1046/j.1461-0248.2002.00372.x> (2002).
 93. Polley, H. W., Wayne Polley, H., Wilsey, B. J. & Derner, J. D. Do species evenness and plant density influence the magnitude of selection and complementarity effects in annual plant species mixtures? *Ecology Letters* vol. 6 248–256 Preprint at <https://doi.org/10.1046/j.1461-0248.2003.00422.x> (2003).
 94. Garcia, V. M. T. et al. Environmental factors controlling the phytoplankton blooms at the Patagonia shelf-break in spring. *Deep Sea Res. Part 1 Oceanogr. Res. Pap.* **55**, 1150–1166 (2008).
 95. Sala, O. E. et al. Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
 96. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* vol. 28 27–30 Preprint at <https://doi.org/10.1093/nar/28.1.27> (2000).
 97. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
 98. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).