

Microorganismos del corredor marino Isla del Coco-Galápagos: diversidad funcional y de especies

Microorganisms of the marine corridor Isla del Coco-Galapagos: functional and species diversity

Keilor Rojas-Jiménez¹

Fecha de recepción: 21 de febrero de 2018
Fecha de aprobación: 15 de mayo de 2018

Rojas-Jiménez, K. Microorganismos del corredor marino Isla del Coco-Galápagos: diversidad funcional y de especies. *Tecnología en Marcha*. Vol. 31-4. Octubre-Diciembre 2018. Pág 157-166.

DOI: <https://doi.org/10.18845/tm.v31i4.3974>



¹ Escuela de Biología, Universidad de Costa Rica, 11501 San José, Costa Rica.
Correo electrónico: keilor.rojas@gmail.com.

Palabras clave

Isla del Coco; Galápagos; microorganismos; biodiversidad.

Resumen

El corredor marino Isla del Coco-Galápagos, en su conjunto, constituye un santuario para la protección de especies marinas en la región oriental del Pacífico Tropical. La alta biodiversidad de macroorganismos ha sido estudiada por décadas, sin embargo, la diversidad de microorganismos del fitoplancton, que constituyen el sustento de la vida en este ecosistema, no se conoce al mismo nivel de detalle. En este trabajo utilizamos datos metagenómicos de libre acceso para examinar la diversidad funcional de proteínas y de especies microbianas presentes en seis puntos de este corredor marino. Se identificó una altísima diversidad de microorganismos que comprende 2132 géneros distribuidos en 68 filos taxonómicos (6 Archaea, 33 Eukaryota, 38 Bacteria y 1 Virus). En particular, los taxones más abundantes fueron *Candidatus Pelagibacter*, *Prochlorococcus*, *Synechococcus*, el Fago T4 y *Pseudomonas*, los cuales coexisten con una gran cantidad de organismos poco representados. Por similitud en su composición, las comunidades pelágicas y litorales pudieron agruparse, sin embargo, estadísticamente no hubo diferencias significativas entre todas las comunidades del corredor. Este trabajo resalta la importancia ecológica del fitoplancton para el ecosistema y representa una justificación adicional protección del corredor en su conjunto.

Keywords

Isla del Coco; Galápagos; microorganisms; biodiversity.

Abstract

The marine corridor Isla del Coco-Galápagos, as a whole, constitutes a sanctuary for the protection of marine species in the eastern region of the Tropical Pacific. The high biodiversity of macroorganisms has been studied for decades, however, the diversity of phytoplankton microorganisms, which constitute the sustenance of life in this ecosystem, is not known at the same level of detail. In this work we use open access metagenomic data to examine the functional diversity of proteins and microbial species present in six points of this marine corridor. A very high diversity of microorganisms was identified, comprising 2132 genera distributed in 68 taxonomic phyla (6 Archaea, 33 Eukaryota, 38 Bacteria and 1 Virus). In particular, the most abundant taxa were *Candidatus Pelagibacter*, *Prochlorococcus*, *Synechococcus*, T4-like virus, and *Pseudomonas*, which coexist with a large number of poorly represented organisms. By similarity in their composition, the pelagic and coastal communities could be grouped, however, statistically there were no significant differences among all the communities of the corridor. This work highlights the ecological importance of microbial phytoplankton for the ecosystem, and represents an additional justification protection for the corridor as a whole.

Introducción

El corredor marino Isla del Coco-Galápagos, que se extiende desde Costa Rica hasta Ecuador, representa un sitio fundamental para la conservación y regeneración de especies marinas de la región oriental del Pacífico Tropical. Asimismo, constituye un laboratorio natural con condiciones geofísicas muy particulares, que han sido objeto de un creciente estudio durante los últimos años.

En términos geológicos esta región forma parte de una cadena de montañas y volcanes submarinos. Los volcanes más antiguos de esta cordillera se originaron hace unos 15 millones de años en el punto de acceso de Galápagos, mientras que la Isla del Coco, ubicada en la parte central, forma parte de una serie de volcanes más jóvenes con edades comprendidas entre 2.2 y 1.5 millones de años [1], [2]. La Isla del Coco es la única porción de tierra de la Placa Tectónica de Cocos emergida sobre el nivel del mar, y por tanto el único lugar donde se pueden realizar estudios precisos sobre el movimiento de esta placa [3].

Esta región es influenciada por la Zona de Convergencia Intertropical (ZCIT). En la Isla del Coco, la colisión de aire cálido y húmedo desde latitudes al norte y al sur del ecuador causa precipitaciones anuales que alcanzan los 7,000 mm. Este factor atmosférico contribuye a la existencia la selva tropical del Pacífico Oriental caracterizada por la presencia de una abundante diversidad florística. En un área de 24 km² la isla presenta siete tipos de vegetación y numerosas plantas endémicas [4], [5].

La interacción de varios factores atmosféricos y oceánicos conduce a que esta región presente una alta biodiversidad marina y altas productividades. La Isla del Coco representa el primer punto de contacto de la Contracorriente Ecuatorial del Norte proveniente del Pacífico Occidental, la cual trae aguas ricas en nutrientes y organismos. Dependiendo de la época del año y la posición de la ZCIT, la isla también es influenciada por el Domo de Costa Rica al norte o por la Corriente de Humboldt al sur [6]–[8]. Estas características, junto con valores relativamente más bajos de salinidad [9], [10], inciden en productividades promedio de 7.8 toneladas de peces por hectárea [11].

El corredor marino Isla del Coco-Galápagos constituye un lugar de encuentro y recuperación para varias poblaciones pelágicas y residencia de numerosas especies costeras, muchas de ellas endémicas. Tanto la Isla del Coco como las Islas Galápagos son consideradas como santuarios naturales que deben ser protegidos y se les ha otorgado el estatus de Patrimonio Natural de la Humanidad [12], [13]. Iniciativas recientes incluso pretenden la consolidación de este corredor como una gran reserva marina internacional.

A modo de ejemplo, se estima que el mar que rodea la Isla del Coco existen al menos 1,668 especies, de las cuales los gasterópodos, peces óseos y crustáceos son las más abundantes. Además, en esta zona se encuentran importantes poblaciones de especies amenazadas como tiburones martillo, tiburones punta blanca, jureles y pargos [11], [14].

La mayoría de investigaciones en esta región se han enfocado en estudiar la biodiversidad de macroorganismos marinos, la cual se conoce relativamente bien. Sin embargo, la diversidad de microorganismos no se ha estudiado con el mismo nivel de detalle. Esto a pesar de que los microorganismos son el sustento de la vida en los mares: forman parte del fitoplancton, intervienen en los procesos biogeoquímicos p.e. descomponiendo materia orgánica y reutilizando nutrientes. Además, son los principales productores primarios y por tanto la base de la transferencia de energía a través de las cadenas tróficas [15]–[18].

En este trabajo se presenta una descripción de la diversidad y estructura de las comunidades de microorganismos en diferentes puntos a lo largo del corredor marino Isla del Coco-Galápagos. Para esto se examinaron bases de datos metagenómicos, debidamente anotadas y de acceso público en la plataforma MG-RAST, las cuales originalmente formaron parte de la Expedición de Muestreo Oceánico realizada por el Instituto J. Craig. Venter [19]–[22]. Se espera que esta información permita una mejor comprensión de la complejidad biológica y ecológica de las comunidades microbianas y de todo el ecosistema del corredor marino. Representa una justificación adicional, desde la perspectiva microbiológica, de la necesidad de conservar este corredor marino en su conjunto.

Materiales y Métodos

La recolecta de muestras de agua marina se realizó entre enero y febrero de 2004 por parte del Instituto J. Craig Venter, en el marco de la Expedición Global de Muestreo de los Océanos. En este estudio se analizaron particularmente seis muestras del Pacífico Tropical Oriental, desde el Golfo de Panamá hasta las Islas Galápagos (Figura 1). Una descripción detallada de la metodología utilizada se puede encontrar en [19]. Brevemente, de cada punto de muestreo se tomaron aproximadamente 200 L de agua superficial utilizando una bomba esterilizada. Los microorganismos en el agua fueron recogidos a través de diferentes tamaños de filtro. Posteriormente se extrajo ADN de los filtros, el cual fue fragmentado aleatoriamente, clonado en vectores y secuenciado con secuenciadores Sanger. El ensamblaje de las secuencias se realizó con programa Celera, las cuales posteriormente fueron cargadas en el repositorio de datos metagenómicos MG-RAST para su anotación [23].

Estación	GS021	GS023	GS025	GS026	GS029	GS036
País	Panamá	Costa Rica	Costa Rica	Ecuador	Ecuador	Ecuador
Profund. muestra (m)	1.6	2	1.1	2	2.1	2.1
Fecha muestreo	19/01/2004	21/01/2004	28/01/2004	01/02/2004	08/02/2004	02/03/2004
Columna de agua (m)	76	1139	30	2386	12	67
Temperatura (°C)	27.6	28.7	28.3	27.8	26.2	25.8
Salinidad (ppm)	3.07	3.26	3.14	3.26	3.45	3.46
Clorofila (µg/kg)	0.5	0.07	0.11	0.22	0.40	0.65
Filtro (µm)	0.1 -0.8	0.1 -0.8	0.8 -3	0.1 -0.8	0.1 -0.8	0.1 -0.8
Habitat	Litoral	Pelágico	Litoral	Pelágico	Litoral	Litoral

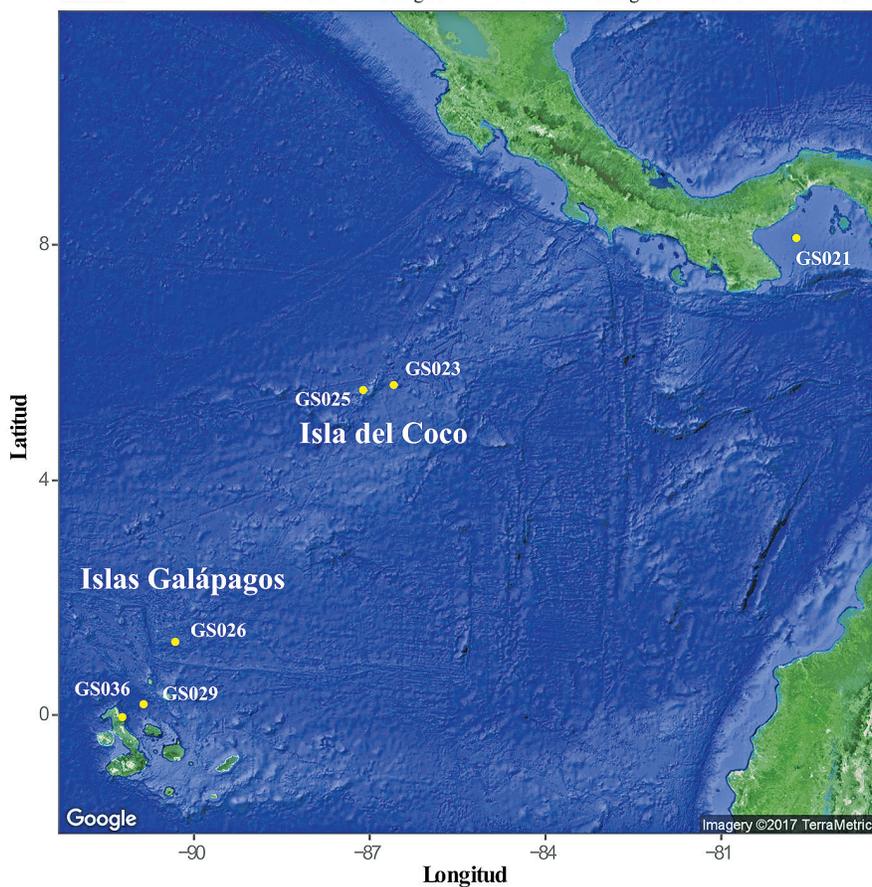


Figura 1. Ubicación geográfica de las estaciones analizadas en este estudio. Asimismo, se muestran las principales variables ambientales de cada punto muestreado. Este mapa se generó utilizando el paquete R ggmap [36] importando imágenes satelitales de Google.

La información de los seis metagenomas analizados en este trabajo es de libre acceso y se encuentra depositada en MG-RAST. Los detalles se muestran en el cuadro 1. Las funciones de las proteínas predichas fueron asignadas a categorías funcionales según las ontologías del sistema de clasificación denominado Subsistemas y utilizando el nivel más alto de jerarquías funcionales [24]. La distribución de las categorías taxonómicas en los diferentes niveles jerárquicos se realizó mediante la comparación de similitudes de las proteínas con la base de datos REFseq [25]. La base de datos incluye 3774 organismos (procariontes, eucariontes y virus) además de registros de 2 879 860 proteínas clasificadas según el sistema de ontología M5NR [26].

Cuadro 1. Principales datos de la secuenciación y anotación de las seis muestras del corredor marino Islas Galápagos-Isla del Coco analizados en este estudio.

Estación	GS021	GS023	GS025	GS026	GS029	GS036
MG-RAST Nr.	mgm4441591.3	mgm4441661.3	mgm4441593.3	mgm4441594.3	mgm4441596.3	mgm4441602.3
Pares de bases	143,454,700	143,626,589	129,781,299	109,049,397	143,822,814	85,757,477
Total secuencias	131,798	133,051	120,671	102,708	131,529	77,538
Longitud media	1,088 ± 70	1,079 ± 76	1,075 ± 95	1,062 ± 72	1,093 ± 81	1,106 ± 80
Porcentaje GC	39 ± 11	36 ± 9	46 ± 12	35 ± 8	36 ± 8	37 ± 9
Proteínas predichas	164,215	171,111	157,264	125,317	151,607	101,648
Prot. identificadas	104,924	123,834	70,021	94,531	109,4	72,254
Cat. funcionales	98,14	114,444	59,149	87,699	102,531	67,041

Fuente: Plataforma MG-RAST <http://metagenomics.anl.gov/>

La visualización de datos y los análisis estadísticos se realizaron en R [27]. Para el análisis de clúster, la matriz géneros-sitios (sin singletons) se transformó en una matriz de distancia Bray-Curtis y el agrupamiento se realizó utilizando el método de enlace promedio. Para el escalamiento multidimensional no métrico (NMDS), y el análisis de varianza con permutaciones (Permanova), la matriz géneros-sitios se normalizó a sus abundancias relativas utilizando la transformación de Helinger. Tanto el análisis de clúster como el NMDS y Permanova se realizaron con el paquete Vegan [28].

Resultados y Discusión

En este estudio se utilizaron varias herramientas bioinformáticas para caracterizar la diversidad funcional y microbiana de seis metagenomas obtenidos a lo largo del corredor marino Isla del Coco-Galápagos. Los perfiles funcionales de las proteínas fueron relativamente similares en todos los sitios (figura 2). La mayor diferencia se observó en el sitio GS025, correspondiente al punto conocido como Dirty Rock en el litoral de la Isla del Coco. En este lugar se determinó un mayor porcentaje de las proteínas relacionadas con fagos y elementos móviles que representó cerca del 14%, mientras que en los otros sitios fue menor al 5%.

Es difícil determinar exactamente las razones de este fenómeno, sin embargo, podría ser que el uso de un tamaño de filtro ligeramente mayor o a condiciones particulares del sitio, por ejemplo,

que en ese momento ocurriera un afloramiento de algas y diatomeas que estuvieran siendo infectadas por virus, como suele ocurrir en ecosistemas marinos [29]–[32].

En todo el corredor marino se identificó una altísima diversidad de microorganismos (figura 3) que comprende 2132 géneros distribuidos en 68 filas taxonómicas (6 Archaea, 33 Eukaryota, 38 Bacteria y 1 Virus). Dentro del dominio Bacteria, el más abundante, los filas con más representantes fueron Proteobacteria, Cyanobacteria, Bacteroidetes, Firmicutes y Actinobacteria. Es importante destacar que las secuencias asociadas a Virus representaron en promedio un 6.63% de las muestras (con variación de 2.93% a 17.4%).

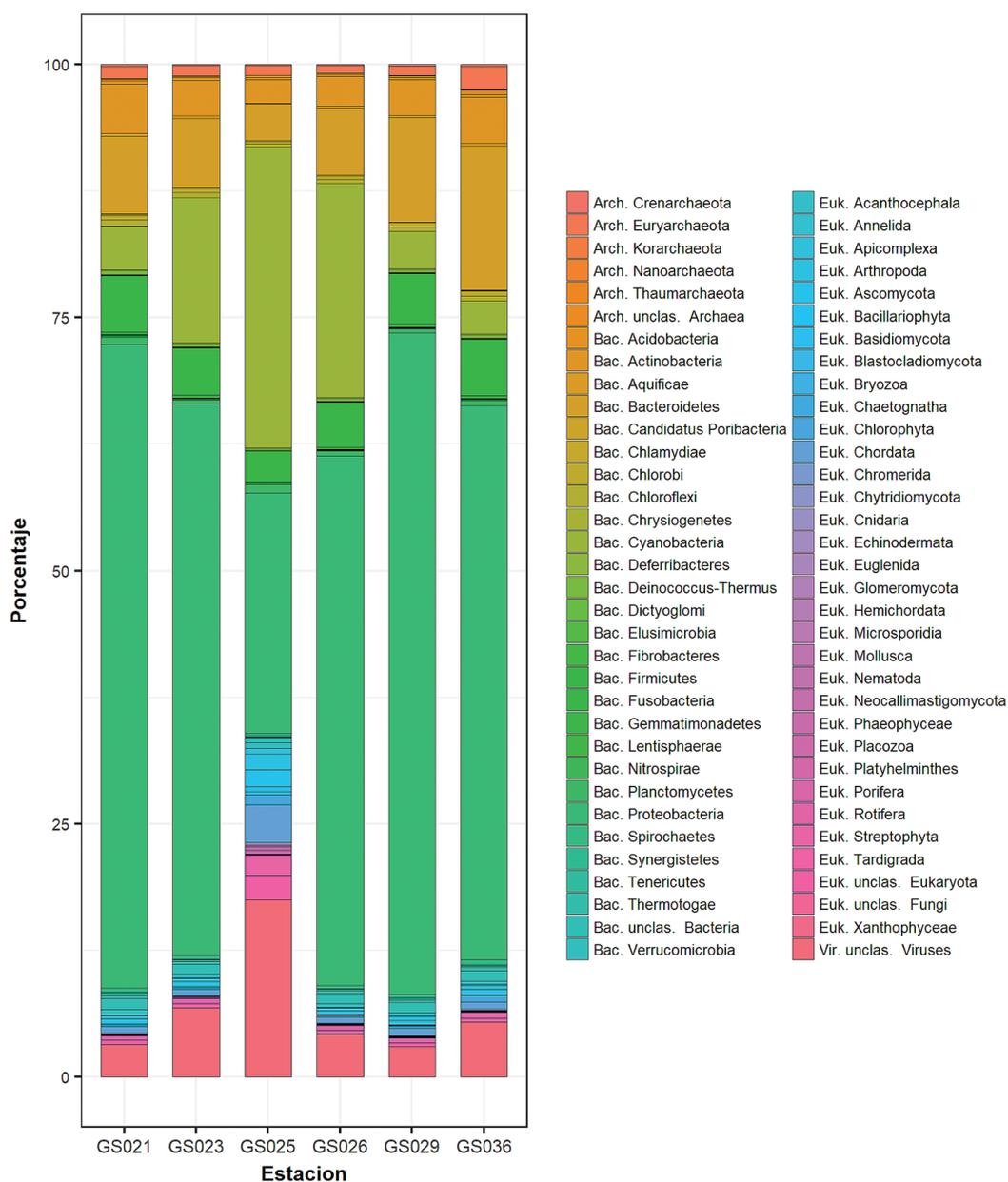


Figura 2. Distribución de las categorías funcionales de las proteínas pronosticadas según la ontología de ortólogos denominada Subsistemas. Se indican los porcentajes relativos de cada categoría en cada uno de los sitios muestreados en el corredor marino Isla del Coco-Galápagos.

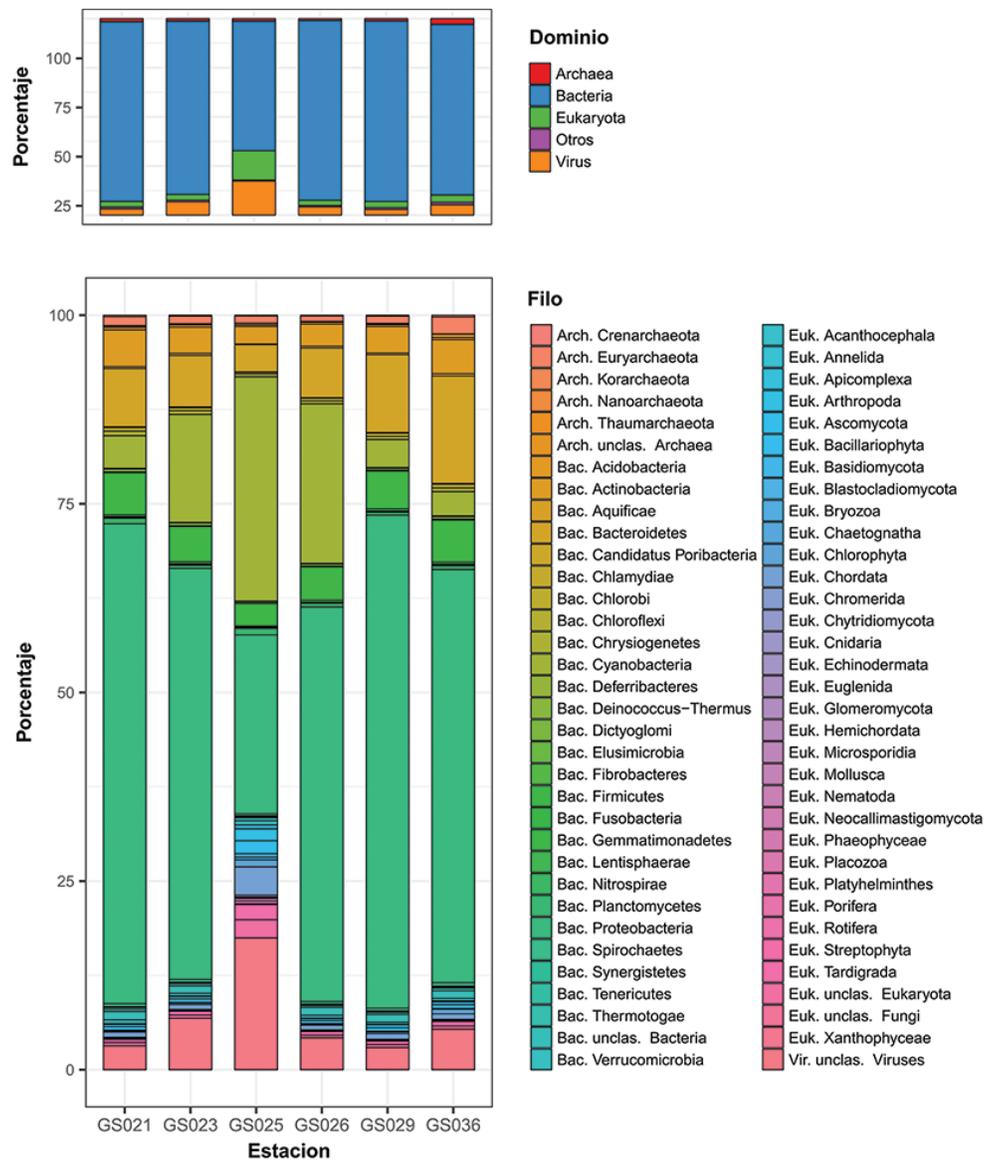


Figura 3. La distribución de los grupos taxonómicos presentes en seis estaciones del corredor marino Isla del Coco-Galápagos. En el panel superior se muestra a nivel de dominio y en el panel inferior a nivel de filo.

En particular, los géneros más abundantes fueron *Candidatus Pelagibacter* (8.82%), *Prochlorococcus* (7.58%), *Synechococcus* (2.83%), Fago T4 (2.71%) y *Pseudomonas* (1.72%). Estos taxones, relativamente más abundantes, coexisten con una gran cantidad de organismos poco representados, aunque también se pudo apreciar pequeñas diferencias entre las estructuras de las comunidades de cada sitio. Por ejemplo, en el sitio G025 hubo una menor proporción de Proteobacteria pero mayor de Cyanobacteria, Euryarchaeota, Chordata, Ascomycota y Virus.

El análisis de clúster y el análisis de escalamiento multidimensional no métrico fueron consistentes en señalar ciertas similitudes en la composición de las comunidades (figura 4), particularmente entre los sitios G023 y G026 (comunidades pelágicas de mar abierto) o entre G021 y G029 (comunidades litorales en Golfo de Panamá y la Isla Galápagos). Sin embargo, la composición de las seis comunidades en conjunto no fue estadísticamente distinta según el Permanova ($p > 0.05$).

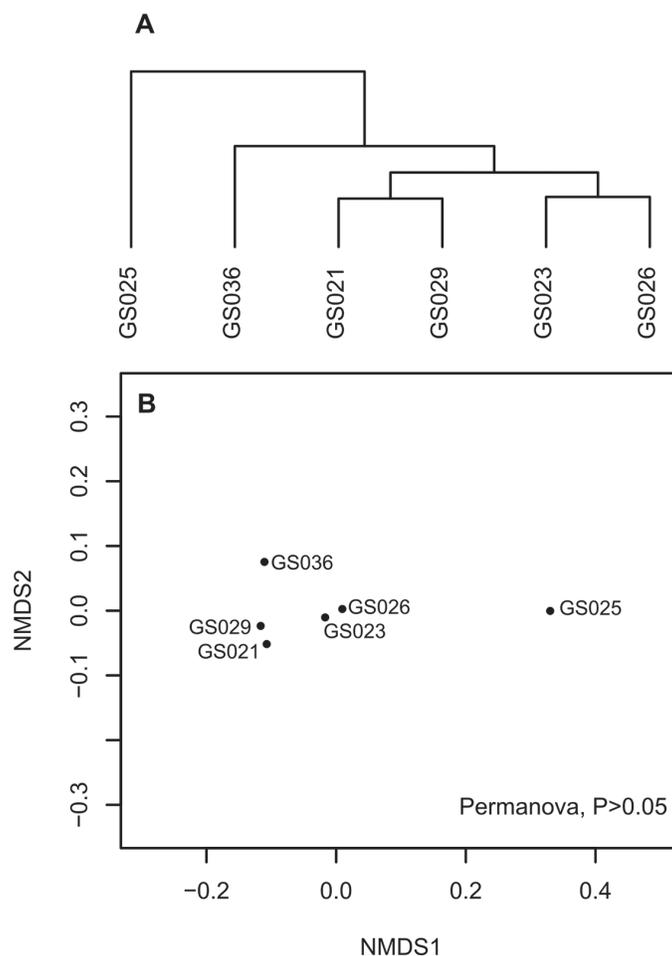


Figura 4. Comparación de la estructura de las comunidades de microorganismos, a nivel de género, provenientes de seis puntos de muestreo a lo largo del corredor marino Isla del Coco-Galápagos. (A) Análisis de cluster. (B) Análisis de escalamiento multidimensional no métrico, indicando el resultado no significativo del Permanova.

La alta abundancia de *Candidatus Pelagibacter* así como otras Alphaproteobacteria es consistente con lo encontrado en estudios realizados en otros océanos y donde se ha señalado la capacidad de estos microorganismos de adaptarse a las condiciones de aguas oligotróficas [15], [18], [19], [33]. Incluso se ha sugerido que *Candidatus Pelagibacter* podría ser una de las bacterias más comunes en los ambientes oceánicos [16]. Por otro lado, la abundancia de Cyanobacteria pertenecientes a los géneros *Prochlorococcus*, *Synechococcus* también ha sido observada en otros ecosistemas, donde se les considera como especies ubicuas del fitoplancton marino debido a su amplia versatilidad para sobrevivir a diferentes profundidades, intensidad de luz y disponibilidad de nutrientes [6], [19], [34], [35].

La gran mayoría de los géneros de microorganismos restantes presentaron abundancias relativas muy bajas. Sin embargo, es importante mencionar que una menor abundancia no representa necesariamente un menor nivel de importancia para el ecosistema, ya que especies poco representadas, o incluso únicas, podrían estar desempeñando funciones importantes para el ecosistema en general.

Al respecto, será importante en futuros estudios considerar una mayor intensidad de muestreo, con una mayor cantidad de puntos a lo largo del corredor, pero también muestreando a

diferentes profundidades y en diferentes épocas del año. Esto para tener una comprensión de la dinámica espacio-temporal del fitoplancton en un corredor marino tan amplio y a la vez complejo.

En este trabajo, el uso del enfoque metagenómico permitió tener un panorama general de la diversidad de grupos de microorganismos presentes en el ecosistema, incluidos los virus, los cuales pueden estar desempeñando funciones esenciales en el control de poblaciones de otros microorganismos, pero que generalmente no se consideran en este tipo de estudios. Sin embargo, una de las limitaciones es que la cuantificación de las especies es influenciada por el tamaño y la anotación de sus respectivos genes. Es por esto que también será importante considerar en futuros estudios la selección de marcadores moleculares específicos.

En conclusión, este estudio permite comprender mejor el componente microbiano de un ecosistema altamente diverso y productivo como lo es el corredor marino Isla del Coco-Galápagos. En esta región factores oceanográficos y climatológicos convergen para proporcionar condiciones adecuadas para el crecimiento de productores primarios y otros microorganismos del fitoplancton, los cuales son alimento para el zooplancton. El zooplancton es, a la vez, consumido por peces planctívoros y estos posteriormente por depredadores grandes como tiburones, jureles y atunes. Este estudio permite comprender mejor la dinámica de las cadenas alimenticias en este ecosistema marino y cómo estas son influenciadas por la actividad microbiana desde el inicio hasta el final.

Agradecimientos

Se agradece al Instituto J. Craig. Venter por hacer de libre acceso las secuencias metagenómicas de la Expedición de Muestreo Oceánico en la plataforma MG-RAST.

Referencias

- [1] O. Lizano, "Batimetría de la plataforma insular alrededor de las Isla del Coco, Costa Rica," *Rev. Biol. Trop.*, vol. 49, pp. 163–170, 2001.
- [2] W. Rojas and G. E. Alvarado, "Marco geológico y tectónico de la Isla del Coco y la región marítima circunvecina, Costa Rica zone off the central Pacific coast of Costa Rica," *Rev. Biol. Trop.*, vol. 60, pp. 15–32, 2012.
- [3] M. Protti, V. González, J. Freymueller, and S. Doelger, "Isla del Coco, on Cocos Plate, converges with Isla de San Andrés, on the Caribbean Plate, at 78mm/yr," *Rev. Biol. Trop.*, vol. 60, pp. 33–41, 2012.
- [4] J. Trusty, H. Kesler, and G. Haug-Delgado, "Vascular Flora of Isla del Coco, Costa Rica," *Proc. Calif. Academy Sci.*, vol. 57, pp. 247–355, 2006.
- [5] J. L. Trusty, A. Tye, T. M. Collins, F. A. Michelangeli, P. Madriz, and J. Francisco-Ortega, "Galápagos and Cocos Islands: Geographically Close, Botanically Distant," *Int. J. Plant Sci.*, vol. 173, no. 1, pp. 36–53, Jan. 2012.
- [6] J. T. Pennington, K. L. Mahoney, V. S. Kuwahara, D. D. Kolber, R. Calienes, and F. P. Chavez, "Primary production in the eastern tropical Pacific: A review," *Prog. Oceanogr.*, vol. 69, no. 2–4, pp. 285–317, 2006.
- [7] W. S. Kessler, "The circulation of the eastern tropical Pacific: A review," *Prog. Oceanogr.*, vol. 69, no. 2–4, pp. 181–217, 2006.
- [8] P. C. Fiedler and L. D. Talley, "Hydrography of the eastern tropical Pacific: A review," *Prog. Oceanogr.*, vol. 69, no. 2–4, pp. 143–180, 2006.
- [9] O. Lizano, "Dinámica de las aguas alrededor de la Isla del Coco, Costa Rica," *Rev. Biol. Trop.*, vol. 56, pp. 31–48, 2008.
- [10] J. Acuña-Gonzales, J. Garcia-Cespedes, E. Gomez-Ramirez, J. Vargas-Zamora, and J. Cortes, "Parametros fisico-quimicos en aguas costeras de la Isla del Coco (2001-2007)," *Rev. Biol. Trop.*, vol. 56, no. 49–56, 2008.
- [11] A. M. Friedlander, B. J. Zgliczynski, E. Ballesteros, O. Aburto-Oropeza, A. Bolaños, and E. Sala, "The shallow-water fish assemblage of Isla del Coco National Park, Costa Rica: structure and patterns in an isolated, predator-dominated ecosystem," *Rev. Biol. Trop.*, vol. 60, no. Suppl. v. 3, pp. 321–338, 2012.



- [12] J. Cortes, "Historia de la investigación marina de la Isla del Coco, Costa Rica," *Rev. Biol. Trop.*, vol. 56, pp. 1–18, 2008.
- [13] A. L. Tobar, *Historia política internacional de la Islas Galápagos*, vol. 2. Editorial Abya Yala, 1997.
- [14] J. Cortés, "Marine biodiversity of an Eastern Tropical Pacific oceanic island, Isla del Coco, Costa Rica," *Rev. Biol. Trop.*, vol. 60, pp. 131–185, 2012.
- [15] J. C. Venter *et al.*, "Environmental genome shotgun sequencing of the Sargasso Sea," *Science (80-)*, vol. 304, no. 5667, pp. 66–74, 2004.
- [16] S. J. Giovannoni *et al.*, "Genome streamlining in a cosmopolitan oceanic bacterium," *Science (80-)*, vol. 309, no. 5738, pp. 1242–1245, 2005.
- [17] E. F. DeLong and D. M. Karl, "Genomic perspectives in microbial oceanography," *Nature*, vol. 437, no. 7057, pp. 336–342, 2005.
- [18] E. F. DeLong *et al.*, "Community genomics among stratified microbial assemblages in the ocean's interior," *Science (80-)*, vol. 311, no. 5760, pp. 496–503, 2006.
- [19] D. B. Rusch *et al.*, "The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific," *Plos Biol.*, vol. 5, no. 3, pp. 398–431, 2007.
- [20] S. Yooseph *et al.*, "The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families," *Plos Biol.*, vol. 5, no. 3, pp. 432–466, 2007.
- [21] S. J. Williamson *et al.*, "The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples," *PLoS One*, vol. 3, no. 1, p. e1456, Jan. 2008.
- [22] T. S. Bibby, Y. Zhang, and M. Chen, "Biogeography of Photosynthetic Light-Harvesting Genes in Marine Phytoplankton," *PLoS One*, vol. 4, no. 2, p. e4601, Feb. 2009.
- [23] F. Meyer *et al.*, "The metagenomics RAST server -- a public resource for the automatic phylogenetic and functional analysis of metagenomes," *BMC Bioinformatics*, vol. 9, no. 1, p. 386, Sep. 2008.
- [24] R. Overbeek *et al.*, "The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes," *Nucleic Acids Res.*, vol. 33, no. 17, pp. 5691–5702, 2005.
- [25] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Res.*, vol. 35, no. suppl_1, pp. D61–D65, Jan. 2007.
- [26] A. Wilke *et al.*, "The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools," *BMC Bioinformatics*, vol. 13, no. 1, p. 141, Jun. 2012.
- [27] R-Core-Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, Vienna, Austria, 2017.
- [28] J. Oksanen *et al.*, "Vegan: Community Ecology Package. R package version 2.4-2." 2017.
- [29] M. B. Sullivan, M. L. Coleman, P. Weigele, F. Rohwer, and S. W. Chisholm, "Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations," *PLoS Biol.*, vol. 3, no. 5, p. e144, Apr. 2005.
- [30] B. L. Hurwitz and M. B. Sullivan, "The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology," *PLoS One*, vol. 8, no. 2, 2013.
- [31] L. Deng, A. Gregory, S. Yilmaz, B. T. Poulos, P. Hugenholtz, and M. B. Sullivan, "Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging," *MBio*, vol. 3, no. 6, pp. e00373-12-e00373-12, Oct. 2012.
- [32] L. Deng *et al.*, "Viral tagging reveals discrete populations in Synechococcus viral genome sequence space," *Nature*, vol. 513, no. 7517, pp. 242–245, Jul. 2014.
- [33] S. G. Tringe *et al.*, "Comparative metagenomics of microbial communities," *Science (80-)*, vol. 308, no. 5721, pp. 554–557, 2005.
- [34] E. R. Zinser *et al.*, "Prochlorococcus ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method," *Appl Env. Microbiol.*, vol. 72, no. 1, pp. 723–732, 2006.
- [35] Z. I. Johnson, E. R. Zinser, A. Coe, N. P. McNulty, E. M. Woodward, and S. W. Chisholm, "Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients," *Science (80-)*, vol. 311, no. 5768, pp. 1737–1740, 2006.
- [36] D. Kahle and H. Wickham, "ggmap: Spatial Visualization with ggplot2," *R J.*, vol. 5, no. 1, pp. 144–161, 2013.