



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS DEL MAR Y LIMNOLOGÍA

**Modelos de distribución de microorganismos con potencial biotecnológico
en lagunas costeras**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS

PRESENTA:

ANDRÉS HIRAM ZAYOLA MEDINA

TUTOR PRINCIPAL:

DRA. MARÍA LETICIA ARENA ORTÍZ
(PARQUE CIENTÍFICO Y TECNOLÓGICO-UNAM)

COMITÉ TUTOR:

DR. CARLOS YAÑEZ ARENAS
(PARQUE CIENTÍFICO Y TECNOLÓGICO-UNAM)

DR. LUIS DANIEL ÁVILA CABADILLA
(ENES MÉRIDA-UNAM)

DRA. CECILIA ELIZABETH ENRÍQUEZ ORTÍZ
(ENES MÉRIDA-UNAM)

DR. EDGAR TORRES IRINEO
(PARQUE CIENTÍFICO Y TECNOLÓGICO-UNAM)

Ciudad Universitaria, Cd. Mx.

abril 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



MODELOS DE DISTRIBUCIÓN DE MICROORGANISMOS CON POTENCIAL
BIOTECNOLÓGICO EN LAGUNAS COSTERAS

TESIS

QUE PARA OPTAR POR EL GRADO ACADÉMICO DE:
**MAESTRO EN CIENCIAS
(BIOLOGÍA MARINA)**

PRESENTA:

ANDRÉS HIRAM ZAYOLA MEDINA

TUTOR PRINCIPAL:

DRA. MARÍA LETICIA ARENA ORTIZ
(PARQUE CIENTÍFICO Y TECNOLÓGICO-UNAM)

COMITÉ TUTOR:

DR. CARLOS YAÑEZ ARENAS
(PARQUE CIENTÍFICO Y TECNOLÓGICO-UNAM)

DR. LUIS DANIEL ÁVILA CABADILLA
(ENES MÉRIDA-UNAM)

DRA. CECILIA ELIZABETH ENRÍQUEZ ORTÍZ
(ENES MÉRIDA-UNAM)

DR. EDGAR TORRES IRINEO
(PARQUE CIENTÍFICO Y TECNOLÓGICO-UNAM)

Índice	
Resumen.	6
Introducción	7
Antecedentes y Fundamentación	9
Ecología Microbiana	9
Distribución espacial de los microorganismos	10
Teoría del nicho ecológico	11
Modelado de nicho ecológico	12
Compuestos orgánicos persistentes.	13
Preguntas de investigación	14
Hipótesis	14
Objetivos	14
Materiales y métodos	15
Área de estudio.	15
Muestreo	15
Muestras de agua	16
Extracción de ADN	16
Cuantificación de compuestos orgánicos por cromatografía de gases	16
Análisis de compuestos orgánicos persistentes	17
Determinación de alcalinidad	17
Determinación de demanda bioquímica de oxígeno	18
Determinación de cloro	18
Determinación de nitrógeno total	18
Determinación de nitrógeno amoniacal	18
Determinación de nitratos	19
Determinación de fósforo	19
Determinación de sulfatos	19
Análisis de coliformes	20
Pretratamiento de muestras para análisis de absorción atómica	20
Interpolación de variables ambientales	20
Compuestos Orgánicos Persistentes.	21
Secuenciación Masiva	22
Análisis de variables ambientales y biológicas	22
Modelos de distribución	23

Validación estadística	25
Resultados	26
Compuestos orgánicos persistentes	26
Análisis de datos de secuenciación masiva	27
Variables ambientales y composición taxonómica.	28
Modelos	31
Bacterias asociadas a bosques de manglar	36
Discusión	38
Conclusiones	40
Referencias	41
Anexos.	55
Mapas de interpolación de variables ambientales.	55
Algoritmos	63
Bioclim (BIO):	63
Domain (DOM).	64
Gaussian Process (GAU).	64
Maxent Default y Simple (MXD/MXS).	65
Random Forest (RDF).	67
Support Vector Machine (SVM).	67
Descripción de especies.	68
Arcobacter lekithochrous.	68
Bacillus cereus.	68
Pelagibacter ubique.	69
Chryseobacterium hominis.	69
Clavibacter michiganensis.	70
Cribrihabitans marinus.	70
Marinobacterium marisflavi.	70
Riemerella anatipestifer.	71
Géneros sintetizadores de PHA's.	73

Índice de Figuras.

Figura 1 Mapa de localidades estudiadas	9
Figura 2 Esquematación del concepto de nicho, diagrama BAM (Soberón & Peterson, 2005).....	12
Figura 3 Mapa de interpolación de concentraciones de Amonio (mg/l)	21
Figura 4 Mapa de contaminantes orgánicos persistentes y concentraciones letales	26
Figura 5 MDS no paramétrico de especies asociadas a COP's.....	27
Figura 6 Comparación del número de especies y lecturas entre lagunas Dulces y Salobres	27
Figura 7 Cluster de similitud de la composición de Filos entre lagunas	28
Figura 8 Shade plot de abundancias y clasificación de similaridad entre sitios	29
Figura 9 Shade plot de abundancias para las 50 especies más importantes	30
Figura 10 MDS del set de especies más importantes para diferenciar entre sitios.....	30
Figura 11 PCA de variables ambientales. PC1 31.7 %, PC2 18.1%	31
Figura 12 Boxplot de desempeño de algoritmos (AUC).....	32
Figura 13 Boxplot de desempeño de algoritmos (pROC).....	32
Figura 14 Mapa de probabilidad de condiciones adecuadas para <i>Arcobacter nitrofigilis</i> (Random Forest).....	33
Figura 15 Mapa binario de probabilidad de condiciones adecuadas de <i>Arcobacter nitrofigilis</i> (Random Forest).....	33
Figura 16 Mapa de probabilidad de condiciones adecuadas para <i>Arcobacter nitrofigilis</i> (Random Forest, restricción por presencias aledañas confirmadas).....	34
Figura 17 Mapa binario de probabilidad de condiciones adecuadas para <i>Arcobacter nitrofigilis</i> (Random Forest, restricción por presencias aledañas confirmadas)	34
Figura 18 Heatmap de porcentajes de contribución de variables por algoritmo para cada especie	35
Figura 19 Heatmap de porcentajes de contribución de variables ambientales por especie.....	36
Figura 20 Mapa de probabilidad de condiciones adecuadas para los géneros bacterianos asociados a bosques de manglar	36
Figura 21 Mapa de probabilidad de condiciones adecuadas para el género <i>Enterobacter</i>	37
Figura 22 Mapa de probabilidad de condiciones adecuadas para el género <i>Flavobacterium</i>	37
Figura 23 Mapa de probabilidad de condiciones adecuadas para el género <i>Pantoea</i>	38

Resumen.

A pesar de que las distribuciones espaciales de los seres vivos han sido ampliamente estudiadas e implementadas en trabajos de conservación, en el caso de la microbiología se siguen caracterizando los insumos necesarios y su interpretación. Esto se debe a que la distribución de las especies es producto de la compleja relación entre procesos locales/regionales con interacciones bióticas. El problema principal es la resolución espacial de los datos colectados en campo, pues la dinámica de la comunidad bacteriana es altamente variable.

El presente trabajo buscó identificar las variables ambientales que rigen la distribución de un grupo representativo de microorganismos de las lagunas costeras de la Península de Yucatán. Con datos de secuenciación masiva y cuantificaciones de diferentes variables, se estudió la composición taxonómica y su relación con el estado de salud ambiental de los sitios, analizados con diferentes técnicas de estadística multivariada. Complementado por modelos de distribución de especies comparando distintos algoritmos para definir cuál funciona mejor dadas las características de los datos.

Los resultados muestran i) una mayor diversidad en las lagunas con aporte de agua dulce, específicamente en los ojos de agua, probablemente por el arrastre de material genético a través del acuífero subterráneo. ii) Diferencias significativas en la composición de especies entre lagunas, iii) diferencias sutiles entre sitios conservados e impactados dentro de cada laguna, iv) el algoritmo Random Forest como el método con menor tasa de omisión, pROC más alto y áreas de predicción de condiciones adecuadas más conservadoras y finalmente v) ejemplos de aplicación de datos de secuenciación masiva para estudiar diferentes procesos.

Los datos de secuenciación masiva provenientes del metagenoma ambiental proveen una “fotografía” de los procesos que se están desarrollando. Tal información, apoyada con estudios de fisiología, predicciones funcionales y modelado de distribuciones pueden encaminar a la ciencia a un monitoreo con alta sensibilidad, costos más rentables y la determinación de aplicaciones potenciales en distintos campos (e.g., farmacéutica, industria, biorremediación, monitoreo ambiental, etc.).

Introducción

México se caracteriza como un país con alta influencia costera, en total cuenta con una línea de costa de 11, 122 km, de los cuales 3,294 km corresponden a la vertiente atlántica, en dicho espacio se desarrollan diversas actividades humanas desde épocas ancestrales (INEGI, 2010). Este trabajo considera la zona costera como el espacio de transición entre los ambientes marino y terrestre que se extiende desde la plataforma continental en el límite oceánico (Rozzi et al., 2007) pasando por zonas inundables adyacentes al mar, hasta el primer cambio importante en el tipo de cobertura vegetal. Las zonas costeras son valiosas en términos ecológicos, antropológicos y socioeconómicos, pues son áreas que por su biodiversidad promueven el desarrollo de actividades económicas, en su mayoría encaminadas al aprovechamiento de recursos naturales.

Las zonas costeras contienen diversos ecosistemas con alta productividad primaria, son zonas de interés prioritario desde el punto de vista de la conservación, se trata de la primera línea de protección para los asentamientos humanos frente a fenómenos hidrometeorológicos. Las actividades que toman lugar en estos sitios se relacionan principalmente con la industria alimentaria, transportista, recreacional, energética y urbanística. No obstante, las consecuencias de dichas acciones modifican los procesos naturales de los ecosistemas, comprometiendo la sustentabilidad de los recursos, mientras degradan la salud de los habitantes producto de la contaminación (Rojas, 2011).

La carencia de un sistema de disposición de residuos eficiente en Yucatán y Quintana Roo, un flujo subterráneo de agua en temporada de lluvias que lixivia y moviliza contaminantes, el cambio en el uso de suelo y las descargas de aguas residuales al manto freático agravan la degradación, modificando el funcionamiento de los ciclos biogeoquímicos con la introducción de metales pesados, pesticidas, material radioactivo, hidrocarburos y sus derivados (Herrera-Silveira & Morales, 2010). Anualmente se descargan en el acuífero de Yucatán 102.3 m³ de aguas residuales, como consecuencia de las actividades que no cuentan con un proceso de tratamiento previo a ser vertidas al acuífero (Herrera-Silveira & Morales, 2010).

La presencia de contaminantes representa un serio problema ambiental, pues su permanencia en el ambiente afecta a toda la red trófica, por ende, se requiere tratamiento de aguas residuales aunado a procesos de remediación. El uso de plaguicidas y derivados de hidrocarburos aumentaron desde inicios del siglo XX, producto del crecimiento industrial del sector petrolero. Su implementación ha causado un incremento desmedido de sus concentraciones en la atmósfera, hidrósfera, suelos y

sedimento, creando problemas de contaminación serios en distintos ambientes (García & Durga, 2012).

Los microorganismos que habitan estos ambientes participan en la gran mayoría de procesos biogeoquímicos, tienen la capacidad de metabolizar y utilizar como fuente de carbono los elementos presentes en el ambiente, incluidos los contaminantes. Sus estrategias adaptativas han impulsado el desarrollo de un campo de investigación enfocado en identificar y comercializar cepas/consorcios bacterianos para atender la problemática de la contaminación ambiental.

Las poblaciones microbianas expuestas a los contaminantes son capaces de sobrevivir con la producción de metabolitos secundarios como enzimas que modifican el estado de oxidación de los compuestos, o pueden utilizarlos como fuente de energía usando alguna de las rutas metabólicas y adaptaciones que poseen.

En el marco del proyecto PAPIIT IT202418 se estudió el efecto de la degradación ambiental producto de la actividad humana en los manglares de los sitios: Celestún, Sisal, Chelem, Dzilam de Bravo, Ría Lagartos, Cancún y Sian Ka'an (Figura 1) analizando el efecto de los cambios en el estado funcional de los manglares analizando la riqueza y diversidad taxonómica de las bacterias, con especial énfasis en organismos relacionados con la presencia de contaminantes. Este trabajo buscó determinar si la comunidad microbiana se modifica ante presiones antropogénicas a pesar de compartir condiciones ambientales similares. Por otro lado, se aporta información sobre cómo influyen los factores ambientales en la composición y distribución microbiana.

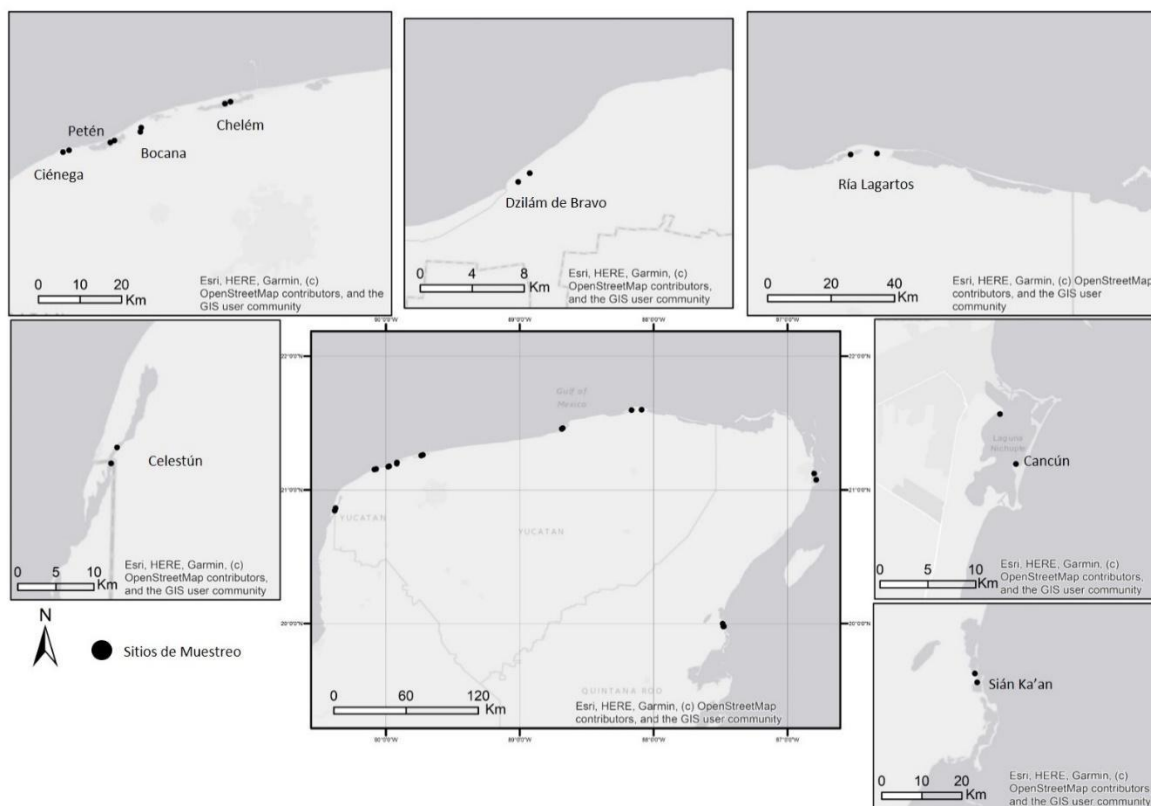


Figura 1 Mapa de localidades estudiadas

Determinar la influencia de los factores que explican la variación de la abundancia y diversidad de las comunidades es una meta central en la ecología (Tuomisto et al., 2003). Aunque la variación espacial de los organismos ha sido considerada como “ruido” en estudios microbiológicos (Ettema & Wardle, 2002), no hay duda en la actualidad de que algunas de las poblaciones de microorganismos de vida libre muestran patrones de distribución claros (Martiny et al., 2006; Ramette & Tiedje, 2007). Por ejemplo, la descripción de relaciones positivas entre disimilaridades genéticas y distancias geográficas (Cho & Tiedje, 2000) y relaciones taxa-área (Horner-Devine et al., 2004; Martiny et al., 2006; Van Der Gast et al., 2005) que son conceptos muy similares para microorganismos con respecto a lo que se describe para organismos eucariontes (Horner-Devine et al., 2004).

Antecedentes y Fundamentación

Ecología Microbiana

Virtualmente todos los procesos ecosistémicos son influenciados por microorganismos y muchos otros son llevados a cabo exclusivamente por ellos, lo que ha llevado a asumir que una descripción más detallada del microbioma debería guiarnos a un mejor entendimiento de los procesos a nivel sistema (Graham et al., 2014). Tal justificación asume que las características del microbioma cuantificables (librerías de genes 16s rRNA, metagenomas, actividades enzimáticas) pueden mejorar nuestra

habilidad para entender mejor y predecir procesos, sin embargo, la información adicional del microbioma no siempre provee un entendimiento más claro de los procesos ecosistémicos más allá de lo que puede predecirse con únicamente factores ambientales (Graham et al., 2016) cuando se analiza en niveles generales como Filo, Clase u Orden.

Distribución espacial de los microorganismos

Los microorganismos están presentes en todo el planeta, sin embargo, poco se sabe sobre sus patrones de distribución espacial y cómo los cambios en las distribuciones de grupos específicos de microbios modifican el ambiente (King et al., 2010). A pesar de la conocida importancia de los microorganismos en el mantenimiento de los ciclos biogeoquímicos del planeta, la relación entre el nicho ecológico de grupos microbianos y su presencia en el ambiente es pobremente entendida (Ettema & Wardle, 2002; Green et al., 2004; Martiny et al., 2006). Esto se debe a la falta de resolución en los datos colectados usando métodos tradicionales, lo que dificulta la identificación de grupos potencialmente importantes en el ambiente, como los organismos no cultivables (Ettema & Wardle, 2002).

Definir cómo la abundancia, distribución y estructura varían en los diversos gradientes ambientales es de vital importancia para el entendimiento de la ecología microbiana (Hongyue Dang et al., 2019). Se ha hipotetizado que los gradientes ambientales aumentan la diversidad de microorganismos y su potencial metabólico, juegan un rol importante en el mantenimiento de la estabilidad y funciones de los ambientes oceánicos y epicontinentales, los cuales han experimentado una multitud de perturbaciones antropogénicas (Damashek & Francis, 2018; H. Dang & Jiao, 2014).

La teoría actual de biogeografía de procariontes y procesos de diversificación propone dos factores para explicar la variación espacial de las comunidades microbianas: (i) La heterogeneidad ambiental (EH, por sus siglas en inglés) asume que los ambientes son intrínsecamente heterogéneos y tanto la abundancia microbiana como la genética responden rápidamente a las condiciones ambientales cambiantes (Travisano & Rainey, 2000). (ii) La distancia espacial (SD, por sus siglas en inglés) puede ser vista como una variable que resume la dinámica de las comunidades y su relación con eventos históricos o perturbaciones (e.g., barreras físicas, actividad antropogénica, historia de dispersión y heterogeneidad pasada) cuyos legados se han mantenido por el aislamiento espacial entre poblaciones (Borcard, D., & Legendre, 1994).

Algunos autores (Ettema & Wardle, 2002) sostienen que, con suficiente información, la variabilidad espacial se vuelve un factor clave más que un obstáculo para entender la estructura y función de la biodiversidad de microorganismos. Estudios previos han demostrado que existen patrones espaciales

en los microorganismos (Green et al., 2004; Horner-Devine et al., 2004; Ramette & Tiedje, 2007; Robertson & Freckman, 1995; Rodrigues et al., 2009) y que los grupos taxonómicos de gruesa escala (Filo o división) muestran una fuerte correlación con parámetros ambientales; por ejemplo Acidobacteria se correlaciona negativamente con el pH del suelo, mientras que Proteobacteria se correlaciona de forma positiva con el pH (Fierer et al., 2009; Fierer & Jackson, 2006).

Diversos autores sostienen que estudiar la biogeoquímica de un sistema cuando se conoce la distribución espacial de grupos taxonómicos a escala gruesa brinda muy poca información adicional. Aunado a ello tenemos conocimiento rudimentario de la escala espacial en la que los microorganismos y los parámetros biogeoquímicos están distribuidos (King et al., 2010). Por lo que este trabajo busca integrar escalas finas para análisis taxonómicos (nivel de especie) y para la cuantificación de variables ambientales.

Teoría del nicho ecológico

Los modelos de distribución de especies relacionan observaciones de campo a variables ambientales predictoras, basados en superficies de respuesta estadística o teóricamente derivadas (Guisan & Zimmermann, 2000).

Hace más de un siglo se propuso que el nicho ecológico de una especie determina su distribución y se definió como las condiciones ecológicas que permiten a una especie mantener su población estable, gracias a la combinación de sus adaptaciones fisiológicas y su comportamiento, que le permiten responder a eventos de cambio o estrés (Grinnell, 1917). Cincuenta años después se define al nicho ecológico como un hipervolumen n-dimensional, en el que cada dimensión es una abstracción de un intervalo de variables limitantes como la temperatura, salinidad, depredación, etc., considerando también que el nicho ecológico es un atributo de la especie y no del ambiente, pues son las especies las que están limitadas fisiológicamente (Hutchinson, 1957).

El nicho realizado según algunos autores (Soberón & Nakamura, 2009; Soberón & Peterson, 2005) tiene una leve diferencia, pues se considera como una parte del nicho potencial, el cual puede ser ocupado por las especies al tomar en cuenta las consecuencias propias de la depredación y la competencia aunado a lo anteriormente mencionado.

En la *Figura 2*, "A" representa la sección del espacio geográfico donde las condiciones abióticas permiten a una especie sobrevivir. Por otro lado "B" representa una región donde las condiciones bióticas son favorables. "M" comprende una región accesible por dispersión o colonización de la especie a través de su historia evolutiva. Al centro "G₀" es el área geográfica ocupada por la especie y "G₁" el área geográfica con potencial de invasión.

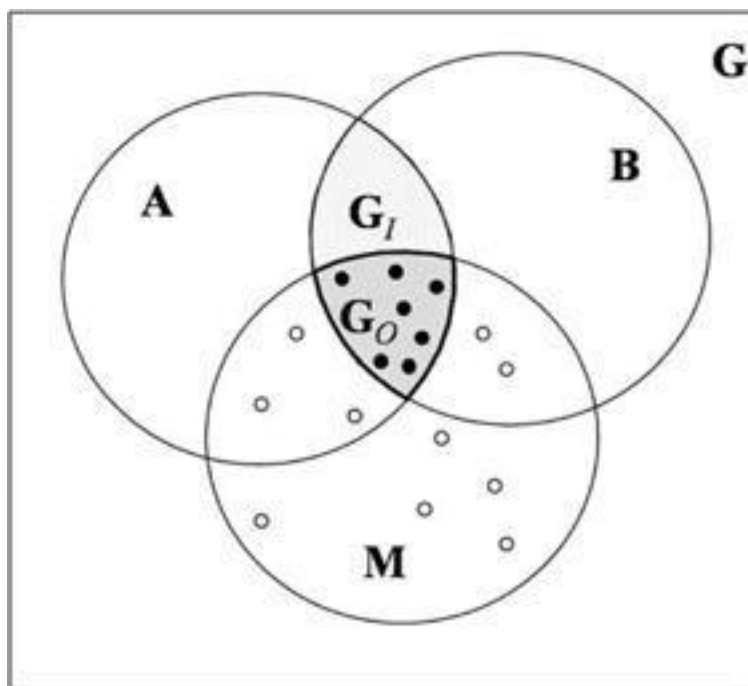


Figura 2 Esquematación del concepto de nicho, diagrama BAM (Soberón & Peterson, 2005)

El nicho fundamental (N_F) se define como el conjunto de condiciones ambientales en las que la especie es capaz de mantener poblaciones sin subsidio por inmigración (Peterson & Anamza, 2015). De esta aseveración se define mediante ecuaciones e inequidades que: Solo un subconjunto de las condiciones asociadas con el nicho ecológico fundamental estarán representadas en el planeta (N_F). El nicho fundamental existente es la reducción del nicho fundamental a las condiciones accesibles por la especie (N_F^*). El nicho realizado es un subconjunto del nicho fundamental existente, idóneo en términos de condiciones bióticas para la especie (N_R). Incluso el N_R será más amplio que el conjunto de ambientes asociados a las ocurrencias conocidas de la especie, siendo este último espacio N_o (el nicho ocupado) (Peterson & Anamza, 2015).

Modelado de nicho ecológico

Los modelos de nicho ecológico permiten estimar la distribución de las especies en un tiempo y espacio definidos. Un modelo de nicho ecológico es una herramienta, generalmente desarrollada con algoritmos, que permite modelar el nicho ecológico de las especies analizando las condiciones ambientales de sus localidades conocidas (J. Elith & Burgman, 2003; Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000). Su objetivo radica en aproximar el nicho fundamental para posteriormente predecir la idoneidad del ambiente para la especie en cuestión, en función de su nicho ecológico (Phillips et al., 2006).

El modelado de nicho ecológico (MNE) es quizás el método más eficiente para estimar la distribución real o potencial de las especies (Guisan & Thuiller, 2005). Su enfoque es cada vez más común e implementado en la toma de decisiones referentes a la conservación, contaminación y restauración ambiental (Ferrier, 2002). No obstante, a pesar de su eficacia en la delimitación de especies terrestres, el uso en los ambientes acuáticos ha sido poco explorada (Wiley et al., 2003). Esto se debe en gran medida a que en los cuerpos epicontinentales la escala geográfica es grande, lo que requeriría datos ambientales finos para crear un modelo, sin embargo, la información disponible de las variables ambientales es comúnmente de mala calidad (McNyset, 2005).

En este trabajo, al no incluir todo el medio biótico ni los factores históricos ni de dispersión se modelaron áreas de distribución, combinando las presencias y las capas ambientales para generar un objeto en el espacio E, posteriormente proyectado en G (dualidad de Hutchinson). Aproximándose al área ocupada (Go) (Soberón et al., 2017)

La precisión de este tipo de modelado puede ser afectada por dos errores comunes, el primero es el de comisión, consiste en mostrar una ausencia como una presencia, pero estrictamente el error no es muy importante, pues es lo que se busca al realizar un modelo de nicho ecológico, es decir, mostrar áreas en las cuales no se conoce la presencia de una especie; y el segundo error es el de omisión, que se basa en mostrar una presencia como una ausencia (Felcísimio et al., 2011).

El ambiente tiene injerencia en la determinación de la distribución de las especies. Jarvis et al. (2005) postula que la distribución de factores como el relieve, el clima y el suelo, indicaría si la especie se encuentra presente o no. La realidad dista de ser tan simple, pues los factores ambientales no son los únicos que determinan la distribución espacial de una especie, pues como se mencionaba anteriormente se deben tomar en cuenta los factores bióticos, así como la información del área geográfica en la cual podría estar distribuida la especie tomando en cuenta su historia biogeográfica.

Compuestos orgánicos persistentes

Internacionalmente conocidos por su acrónimo en inglés POP's (*Persistent Organic Pollutants*), son sustancias químicas que presentan una amenaza para la salud humana y el medio ambiente pues son resistentes a la degradación, son bioacumulables por ser lipofílicos, se biomagnifican, son tóxicos, semi-volátiles y altamente movilizables.

El Convenio de Estocolmo sobre Contaminantes Orgánicos Persistentes del PNUMA, firmado en mayo de 2001 y que entró en vigor el 17 de mayo de 2004, es un instrumento jurídicamente vinculante cuyo objetivo es proteger la salud humana y el medio ambiente, reduciendo o eliminando la producción y el uso de estos contaminantes, así como sus emisiones.

En la actualidad, el Convenio de Estocolmo ha modificado sus anexos, incorporando nuevos Contaminantes Orgánicos Persistentes a los doce iniciales, conocidos como “la docena sucia”, y la Unión Europea ha modificado los anexos del reglamento.

Preguntas de investigación

¿El modelado de distribución de bacterias es capaz de aproximar la distribución espacial real de los organismos?

¿La composición taxonómica de las comunidades de bacterias es diferente en los sitios con afectaciones antropogénicas en comparación con los conservados?

¿La diversidad taxonómica de las comunidades de bacterias varía espacialmente aún en el mismo tipo de ambiente?

Hipótesis

A pesar del amplio espectro de adaptabilidad de las bacterias a las características ambientales, la diversidad de grupos bacterianos en sitios conservados será distinta a aquellos que habitan los sitios perturbados.

Objetivos

- **Objetivo General**

Generar modelos de distribución con restricciones en la predicción/extrapolación de bacterias con interés biotecnológico para aproximar su geografía ocupada en el gradiente ambiental , en lagunas costeras de la Península de Yucatán.

- **Objetivos específicos**

Analizar los datos de secuenciación masiva de nueva generación de ADN de muestras de agua de manglares de Yucatán y Quintana Roo; para conocer la diversidad taxonómica de las comunidades de bacterias entre las localidades estudiadas.

Conocer la correlación entre la presencia de contaminación y la presencia de bacterias reportadas con capacidad de metabolización de contaminantes en las localidades afectadas por causas antropogénicas.

Determinar el efecto de las variables fisicoquímicas y microbiológicas del agua en la composición taxonómica de las comunidades de bacterias metabolizadoras de contaminantes en los ambientes con distinto grado de salud ambiental.

Materiales y métodos

Área de estudio.

La Península de Yucatán se encuentra entre los 19°40'-21°37' N y 87°30'-90°26' O, colinda al Este con el Mar Caribe y al Norte con el Golfo de México. Se han caracterizado tres estaciones climáticas para la región: secas, que van de marzo a mayo; lluvias comprendiendo el periodo de junio a octubre y nortes que va de noviembre a febrero (Schmitter-Soto et al., 2002). Las actividades económicas de mayor importancia para el estado de Yucatán en términos de aportación al Producto Interno Bruto (PIB) son encabezados por los servicios de comercio (21.5%), seguidos por los servicios inmobiliarios de bienes muebles e intangibles (13.1%), la construcción (11.6%) y las actividades referentes a la industria alimentaria (9.9%) (INEGI, 2014). Mientras que en Quintana Roo más del 85% del PIB corresponde a actividades terciarias (INEGI, 2014).

Muestreo

Las localidades de estudio fueron seleccionadas con base en la descripción y resultados de los trabajos realizados durante los últimos 20 años en el grupo de productividad primaria del Dr. Jorge Herrera CINVESTAV-Mérida, considerando el grado de preservación contra aquellos que han presentado deterioro de salud ambiental, para contrastar el efecto antropogénico en la comunidad microbiana. Para este objetivo se tomaron muestras de agua por cada sitio seleccionado (15 litros), una en una zona conservada (o en restauración) y otra donde el impacto antropogénico haya alterado la cobertura vegetal o diversidad, teniendo en total 18 muestras, cada punto fue georreferenciado mediante un GPS marca Garmin modelo eTrex Legend. Se midió temperatura, conductividad, sólidos disueltos totales, salinidad, oxígeno disuelto, concentración de oxígeno, pH y potencial redox *in situ* con una sonda multiparamétrica de marca YSI modelo 556 Handheld. Por otro lado, se determinó la presencia de compuestos orgánicos persistentes, iones mayoritarios, metales pesados, coliformes fecales, nitritos, nitratos, amonio y demanda bioquímica de oxígeno en el Laboratorio de Ingeniería Ambiental de la Facultad de Ingeniería de la Universidad Autónoma de Yucatán (UADY).

Muestras de agua

Las muestras de agua fueron colectadas de la parte superficial de la columna de agua de cada sitio. Las muestras se almacenaron en recipientes de polietileno de baja densidad, se refrigeraron en contenedores y se transportaron al laboratorio de Estudios Ecogenómicos de la UNAM en Yucatán. Se conservaron 12 litros (en 3 recipientes de 4 litros) a 4 °C hasta su procesamiento en un máximo de 24 h, y una porción (3 litros) del agua fue analizada en la UADY para evaluar calidad fisicoquímica y microbiológica.

Extracción de ADN

Las tres muestras de 4 litros de agua por sitio se filtraron en su totalidad (o hasta que el filtro fuera saturado) con una membrana de 0.45 µm. Posteriormente los filtros se mantuvieron a -20°C para realizar la extracción de ADN (Rojas-Herrera et al., 2008). La integridad del ADN extraído se confirmó en un gel de agarosa al 1%. La concentración y pureza se determinó con el Fluorómetro Quantus con el kit QuantiFluor® dsDNA (Promega Corporation, Madison, WI, USA) y el Espectrofotómetro Multiskan Go (Thermo Scientific Waltham, MA, USA), respectivamente de acuerdo con el protocolo del fabricante.

Cuantificación de compuestos orgánicos por cromatografía de gases

La determinación fue realizada con un equipo de cromatografía de gases Trace 1300 GC Ultra (Thermo Scientific, Waltham, MA, U.S.) equipado con un auto-muestreador (TRI PLUS), que incluye un inyector de temperatura programado (PTV) emparejado a un espectrómetro de masas ITQ 900 (GC-MS).

Para la separación de los compuestos se utilizó una columna de capilaridad Elite-17 30 m x 0,25 mm ID x 0.25 µm (Perkin Elmer, Waltham, MA, U.S.). Las condiciones cromatográficas fueron: Helio grado-BIP (99.9999% pureza) a 1 ml/min fue utilizado como gas portador y el puerto de inyección fue dividido a una temperatura de 250°C. La temperatura inicial del horno fue de 60°C por 3 minutos, con incrementos de 10°C/min a 300°C que fueron mantenidos por 5 minutos. El detector de masas fue operado en modo "full scan" (3 escaneos/s) en un rango de 50-500m/z. La ionización fue por impacto electrónico (70 eV) y la temperatura de la fuente de ionización fue de 200°C.

Para el análisis de los compuestos presentes en agua, se utilizó un litro de muestra y la técnica de extracción de fase sólida, en columnas (JT Baker) de 6 ml y 500 mg de C18, usando un gradiente de diclorometano para la elución (1:1).

Análisis de compuestos orgánicos persistentes

La detección de contaminantes orgánicos persistentes en las muestras de agua fue realizada por cromatografía de gases emparejada a un espectrómetro de gases en modo "full scan" (GC-MS/Full Scan) (sensibilidad 1×10^{-3}) con el equipo Thermo Trace GC 1300 Ultra gas emparejado a un detector de masas ITQ 900 (40-550 m/z).

Se importaron los números CAS de las moléculas encontradas en el ambiente al programa ECOSAR™ y se seleccionaron los resultados para el modelo generado a partir de *Lepomis macrochirus*, *Cyprinus carpio*, *Pimephales promelas*, *Poecilia reticulata*, *Oncorhynchus mykiss*, *Oryzias latipes* y *Branchydanio rerio*. Aunque el gradiente de salinidad en las muestras va de agua dulce (1.37 ppm) a agua salobre (34.06), por tratarse de lagunas costeras y dado que los modelos para agua salada en peces, daphnias y algas conservan el orden de los compuestos en agua dulce para peces se optó por tales organismos. Se categorizaron con base en las concentraciones letales medias en un periodo de 96 horas y se generó un mapa (Figura 4).

Utilizando la base de datos NCBI se realizó una búsqueda de microorganismos asociados a tales contaminantes, se seleccionaron 3 géneros (*Halomonas*, *Paracoccus* y *Staphylococcus*) y se proyectaron en un MDS no paramétrico.

Determinación de alcalinidad

La alcalinidad de las muestras fue determinada por el método de titulación (2320 B, Greenberg, 1995). Fenolftaleína o metacresol púrpura fue usado para la evaluación de alcalinidad a un pH de 8.3 o bromocresol verde o una mezcla de bromocresol verde rojo-metil como indicador para pH 4.5. La alcalinidad de la muestra fue determinada a partir de un volumen estándar de ácido requerido para titular una porción a un pH designado. Una curva de titulación fue construida para la estandarización de los reactivos. Para muestras con baja alcalinidad (menos de 20 mg de CaCO_3/l), la técnica de extrapolación fue usada basada en la cercana proporcionalidad de la concentración de iones de hidrógeno con el exceso de titulante más allá de la equivalencia.

Determinación de demanda bioquímica de oxígeno

La demanda bioquímica de oxígeno, definida como la cantidad de un oxidante específico que reacciona con la muestra bajo condiciones controladas, fue realizada utilizando el método de refluo colorimétrico cerrado (5520 D, Greenberg, 1995)). La cantidad de oxidante consumido es expresado en términos de su equivalencia de oxígeno. Cuando una muestra es asimilada, el ion de dicromato oxida el COD en la muestra. Estos resultados se basan en el cambio de cromo del estado hexavalente (VI) al trivalente (III). La absorbancia de lo asimilado, el blanco y las muestras estándar fueron determinadas en un espectrofotómetro DR2800 (Hach, Loveland, CO, U.S.) a una amplitud de onda de 420 nm y 600 nm. A 600 nm, un blanco no asimilado fue utilizado como solución de referencia, y a 420 nm se utilizó agua como solución de referencia.

Determinación de cloro

La determinación de cloro fue realizada con el método argentométrico (4500-Cl-B, Greenberg, 1995). En una solución neutra o alcalina, el dicromato de potasio puede indicar el punto final de la titulación de cloruro de nitrato de plata. El cloruro de plata se precipita cuantitativamente antes de la formación del cromato de plata rojo. 100 ml de muestra ajustados con H_2SO_4 o $NaOH$ a un pH de 7-10 fueron titulados usando 1ml de solución indicadora de K_2CrO_4 y un estándar titulante de $AgNO_3$ hasta el punto final rosa-amarillo. El tritante $AgNO_3$ fue estandarizado y el valor del blanco del reactivo fue establecida por el método de titración descrito anteriormente.

Determinación de nitrógeno total

Con base en el método Macro Kjeldahl (4500-Norg B, Greenberg, 1995). En presencia de ácido sulfúrico (H_2SO_4), sulfato de potasio (K_2SO_4) y sulfato cúprico ($CuSO_4$), el amino nitrógeno de muchos materiales orgánicos es convertido en amonio. El amoniaco libre también se convierte en amonio. Después de la adición de la base, el amoniaco fue destilado de un medio alcalino y absorbido en ácido bórico o sulfúrico. El amonio fue determinado colorimétricamente, por medio de un electrodo selectivo de amonio o por titulación con un estándar de ácido mineral.

Determinación de nitrógeno amoniacal

El nitrógeno amoniacal fue determinado con el método 8038, Nessler (0-2.5 mg/l NH_3-N) con los reactivos del *kit* Hach (Loveland, CO, U.S.), de acuerdo con las instrucciones del fabricante. La muestra fue tratada con una solución de hidróxido de sodio 5 N para ajustarlo a un pH de 7, se añadió un estabilizador mineral y un agente dispersante de alcohol polivinílico, seguido de 1 ml de reactivo

Nessler. Si era necesario, para eliminar la interferencia con la determinación por el método Nessler, la muestra era previamente destilada en una solución de borato a un pH de 9.5 para reducir la hidrólisis de los cianatos y compuestos orgánicos nitrogenados. El amoníaco en el destilado fue tratado como se indicó previamente. El reactivo Nessler y el amoníaco reaccionan para formar un compuesto de color de amarillo a café que absorbe a una amplitud de onda entre 400 y 500 nm, la absorbancia fue determinada en un espectrofotómetro DR2800 (Hach, Loveland, CO, U.S.).

Determinación de nitratos

Los nitratos en las muestras fueron determinados de acuerdo con el método de ácido cromatrópico 10020 que detecta de 0.2 a 30 mg/l de NO₃-N (HR) con el *kit* de tubos de N y los reactivos NitraVerX (Hach, Loveland, CO, U.S.), siguiendo las instrucciones del fabricante. La muestra fue tratada con una solución de hidróxido de sodio 5 N para ajustarlo a un pH de 7. Después, 1 ml de la muestra fue mezclado con el reactivo A NitraVerX en un tubo de prueba N, mezclado por inversión y posteriormente mezclado por inversión con reactivo B NitraVerX, incubado a 5 minutos y leído en un espectrofotómetro DR2800 (Hach, Loveland, CO, U.S.).

Determinación de fósforo

La determinación de fósforo fue realizada de acuerdo con el método de ácido vanadomolibdofosfórico (4500-P C, Greenberg, 1995) . En una solución diluida de ortofosfato, el molibdato de amonio reacciona bajo condiciones ácidas para formar ácido vanadomolibdofosfórico. En presencia de vanadio se forma el ácido amarillo vanadomolibdofosfórico, la intensidad del amarillo es proporcional a la concentración de fosfatos, la intensidad del color se determinó en un espectrofotómetro DR2800 (Hach, Loveland, CO, U.S.) con una longitud de onda de 400 a 490 nm.

Determinación de sulfatos

Los sulfatos fueron determinados siguiendo el método turbidimétrico (4500-SO₄2-E, Greenberg, 1995). El ion sulfato (SO₄²⁻) fue precipitado en medio de ácido acético con cloruro de bario (BaCl₂) para formar cristales de sulfato de bario (BaSO₄) de tamaño uniforme. La absorbancia de la suspensión fue medida con un espectrofotómetro DR2800 (Hach, Loveland, CO, United States), finalmente la concentración de sulfato fue determinada comparando la lectura con una curva estándar.

Análisis de coliformes

La determinación del número más probable de coliformes totales y fecales (termotolerantes) se realizó en concordancia con la norma NMX-AA-42-1987. Para la prueba presuntiva, la muestra fue mezclada y homogenizada, preparando tres diluciones en base 10, con muestras de 0.1 ml, 1 ml y 10 ml en cinco tubos con 10 ml de caldo lactosado en cada dilución. Los tubos fueron incubados a 37°C y examinados en 24 y 48 h, los tubos que mostraron turbidez y producción de gas fueron sembrados para la prueba confirmatoria en un medio selectivo por 48 h a 37°C para detectar coliformes totales y a 44.5°C para la detección de coliformes fecales termotolerantes y *E. coli*. El cálculo del número más probable de coliformes totales, coliformes termotolerantes y *E. coli* presentes en 100 cm³ de muestra fueron determinados por tablas estadísticas de acuerdo con el número de tubos que brindaron resultados positivos en las pruebas confirmatorias.

Pretratamiento de muestras para análisis de absorción atómica

Las muestras de agua fueron filtradas con una membrana con tamaño de poro de 0.45 µm previamente lavado con una solución de ácido nítrico (1%) y enjuagado con agua tipo I antes de ser usado para remover partículas suspendidas, arena y materia orgánica. Subsecuentemente, se ajustó un pH menor a 2 con INSTRA HNO₃ RA, para garantizar la estabilidad de los metales y la acidez de la muestra en el análisis de los iones mayoritarios (nitrógeno amoniacal, nitratos, fósforo, cloro, sulfato, sodio, potasio, calcio y magnesio) y los metales traza (plomo, cadmio, cromo y cobre). El análisis de los iones mayoritarios en las muestras de agua se realizó por absorción atómica de flama de acuerdo con los rangos lineales de las líneas de calibración para esos elementos y llevando a cabo las diluciones necesarias para cada conductividad, ajustando la calibración.

En cuanto a los metales traza, se analizaron mediante la absorción atómica electrotermal en un horno de grafito Varían AA240Z (Agilent, Santa Clara, CA, U.S.) para obtener un límite de detección bajo en la cuantificación de plomo, cadmio, cromo y cobre. La muestra fue directamente analizada por triplicado durante la cuantificación de los diferentes metales traza.

Interpolación de variables ambientales

Se generaron coberturas de las variables ambientales usando las mediciones puntuales georreferenciadas, se construyeron utilizando la unión de los polígonos que definen a las lagunas costeras a analizar, generadas a partir de un archivo *shapefile* de delimitación del país (Escala 1:2,500,000) de CONABIO (CONABIO, 2008). Estas coberturas tipo “raster”, se generaron utilizando la técnica IDW de interpolación (Distancia Inversa Ponderada) (Delaney, J. Van Niel, 2007), en el

Sistema de Información Geográfica ArcMap 10.3® (GIS, ESRI,1999-2019), enmascarando el análisis sólo para la unión de las lagunas costeras de interés. La resolución definitiva (tamaño de píxel) de las coberturas fue de 28 cm (Figura 3).

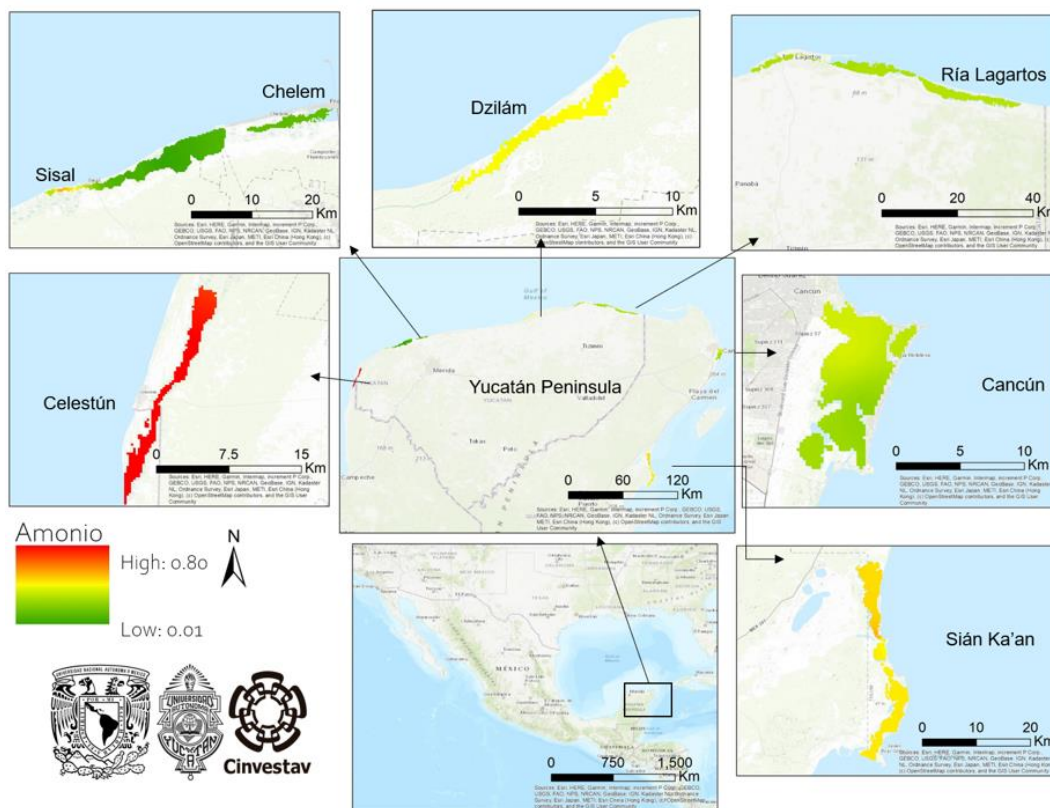


Figura 3 Mapa de interpolación de concentraciones de Amonio (mg/l)

Compuestos Orgánicos Persistentes.

Para determinar el riesgo que cada uno de los compuestos orgánicos persistentes representa en el ambiente se utilizó el software ECOSAR™ (U.S. Environmental Protection Agency) que utiliza QSAR's (Quantitative Structure Activity Relationships), un programa implementado en la predicción de toxicidad de químicos en agua basado en la similitud de la estructura de las moléculas con químicos a los que se han determinado su toxicidad en agua. Los criterios de ECOSAR miden la toxicidad en rangos de pH entre 6 y 8 (replicando condiciones ambientales reales).

Con base en la bibliografía (Gutiérrez, 2018) se analizaron las especies de los géneros *Halomonas*, *Paracoccus*, *Staphylococcus* y *Planobacterium*, reportadas como sintetizadoras de polihidroxicanoatos (PHA's), polímeros lipídicos sintetizados intracelularmente utilizando n-hexadecano, 6 especies fueron seleccionadas (2 por género) y se proyectaron dentro de un MDS no paramétrico.

Secuenciación Masiva

La secuenciación de ADN metagenómico se realizó con un servicio externo en el IBT UNAM. Se utilizó la plataforma de Secuenciación por Síntesis (SBS) con el equipo MiSeq de Illumina, se utilizó el servicio de generación del amplicón 16S y sus respectivas bibliotecas, así como la corrida de la secuenciación en formato 2x300, se emplearon los siguientes primers: 357F 5´CTCCTACGGGAGGCAGCAG 3´ CD (R) 5´CTTGTGCGGGCCCCCGTCAATTC 3´ Diseñados previamente para el uso con la plataforma de Illumina. Anexo del extremo 5´ del primer Delantero: 5' TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG - (secuencia del locus específico). Anexo del extremo 5´ del primer Reverso: 5' GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACAG - (secuencia del locus específico). Los resultados de la secuenciación se analizaron en formato FASTQ.

Se analizaron 15,983 registros (1'124,632 lecturas), las secuencias fueron importadas al software OneCodex, donde se realizó la asignación de secuencias para la identificación de organismos para cada muestra con 5,738 especies en total.

Se comparó el número total de especies para cada tipo de ambiente, el número de especies únicas y el número total de lecturas para cada ambiente.

Análisis de variables ambientales y biológicas

Los datos se analizaron buscando similitud entre sitios con base en su composición taxonómica y características ambientales. El número de lecturas se estandarizó por el total de lecturas por especie y se le aplicó una transformación de raíz cuarta (para conservar las diferencias entre las cuantificaciones pero apreciar mejor el aporte de cada organismo), posteriormente se generó una matriz de similitud con distancias de Bray-Curtis.

En el caso de las variables ambientales, se transformaron de acuerdo con sus características, eliminando variables con correlaciones mayores a 0.95, las variables restantes (16) fueron transformadas según sus atributos, raíz cuadrada para variables no sesgadas y transformaciones severas como logaritmo natural en variables en escalas finas. Finalmente, las variables fueron normalizadas y se generó una matriz de similitud basada en distancias euclidianas.

Los clústeres de similaridad para variables ambientales y composición taxonómica se combinaron en un *shade plot* para ilustrar el cambio de la composición taxonómica entre sitios.

Para identificar el grupo de especies que mejor diferencian a los sitios estudiados se empleó la rutina BEST (bv-step) de PRIMER7, seleccionando un subconjunto de especies que mejor representan el gradiente ambiental analizado. Reduciendo las 5,738 especies a 50 especies basado en el coeficiente

de correlación de spearman y seleccionando algunas especies que después fueron proyectadas en un MDS ilustrando el cambio en su abundancia con respecto a su ubicación.

A partir de la selección de las 50 especies se buscó el menor número de especies que conservaran la diferenciación entre sitios utilizando variables ambientales, es decir seleccionar las especies cuya distribución genera los mismos grupos que si se utilizan variables ambientales únicamente.

El algoritmo aplicado es el *Stepwise*, que utiliza un proceso de selección *Forward* y *Backward*, seleccionando la especie con mayor correlación en un subconjunto limitado de especies, posteriormente se selecciona una segunda especie que incremente la correlación, se agregan especies hasta lograr una correlación de 0.95 y se repite el proceso varias veces con otro subconjunto de especies. A partir de esta subselección de especies se eligieron los organismos con distribuciones más generalistas para tener suficientes datos de ocurrencias para calibrar y evaluar los modelos de distribución, el detalle de las especies se encuentra en los anexos.

Modelos de distribución

Se utilizó la paquetería de R (en desarrollo) “ENMTML” (Andrade et al., 2020) diseñada para la construcción de modelos de nicho ecológico (ENM's por sus siglas en inglés) complejos. El área de calibración fue restringida, para calibrar el modelo sólo alrededor de los sitios donde la especie en cuestión tenía registros de presencia, se eligió la opción de “BUFFER type 1” que establece un área alrededor de los puntos igual a la distancia máxima entre par de ocurrencias para cada especie limitando la autocorrelación espacial debido al bajo número de réplicas por laguna y para evitar un modelo sobre ajustado.

Se aplicó el argumento “colin_var” que permite reducir la colinealidad de las variables con la opción “VIF” (Variance Inflation Factor). Se modificó la opción de generación de pseudoausencias, eligiendo la opción de “GEO_ENV_CONST” donde las pseudoausencias están constreñidas ambientalmente (basado en un modelo simple Bioclim) pero distribuidas geográficamente alejadas de las ocurrencias con base en un *buffer* geográfico, se eligió una distancia de 10 km para aproximar la predicción a la geografía ocupada. La razón entre presencias y ausencias utilizada fue uno a uno.

Para la primera validación del modelo se eligió una partición aleatoria tipo *Bootstrap* con dos réplicas y una proporción de ocurrencias de 75%, es decir se calibra con el 75% de las ocurrencias y se valida con el restante 25%.

Se eligieron nueve algoritmos: Bioclim, Domain, *Gaussian Process*, *Generalized Linear Model*, *Generalized Additive Model*, Maxent (Default y Simple), *Random Forest* y *Support Vector Machine* detallados en anexos.

Para las predicciones binarias se utilizó el umbral “MAX TSS”, donde la suma de la sensibilidad y la especificidad es mayor. Se incluyó la opción de aplicar restricciones espaciales a la proyección del modelo. Dicho método restringe el modelo para tener menor potencial de predicción y acercarlos más a modelos de distribución de especies. Aunque los métodos pueden ser *a Priori* y *a Posteriori* se optó por una restricción *a Posteriori* basada en las ocurrencias (PRES) seleccionando solo áreas donde hubiese datos de ocurrencias confirmados, incluyendo también un mapa binario basado en esta construcción.

Se generaron 252 archivos, cuatro archivos en formato ráster de cada especie para cada algoritmo. El segmento inicial de 63 archivos consistió en mapas de probabilidad de ocurrencia, el segundo segmento de binarizaciones de ocurrencias. El tercer resultado de la aplicación de la restricción de las probabilidades de ocurrencias basado en un umbral de ocurrencias aledañas confirmadas, limitando la predicción y aproximando más los modelos a la geografía ocupada. Finalmente se produjeron las versiones binarizadas de los modelos restringidos.

Se construyeron modelos para géneros seleccionados con base en la bibliografía (Saravanakumar et al., 2016) reportados como productores de enzimas de uso comercial, vinculados a manglares (*Pantoea*, *Flavobacterium* y *Enterobacter*). Los géneros bacterianos endófitos *Pantoea*, y *Endobacter*, fueron seleccionados dado que son comúnmente utilizados de forma comercial por su producción de enzimas (proteasas), implementadas en detergentes, cuero y medicina (Kumar et al., 2004; Zhang et al., 2005). Mientras que el género *Flavobacterium*, fue seleccionado por ser productor de proteasa alcalina, su principal aplicación es como aditivo de limpieza (Greene et al., 1996; Haddar et al., 2009).

Dichos modelos se construyeron con el algoritmo *Random Forest*, pues consistentemente tuvo el AUC y pROC más alto entre especies, la tasa de omisión más baja y los porcentajes de área predichos más conservadores, en comparación con algoritmos que predijeron 100% del área disponible (la tabla con los datos se encuentra en los anexos).

Validación estadística

Los modelos fueron evaluados mediante la técnica ROC (*Receiver Operator Characteristic*) parcial (Peterson et al., 2008) una de las herramientas más utilizadas para validar modelos. La curva describe la tasa de identificación correcta contra los falsos. La ROC parcial es la razón entre la ROC y un modelo construido mediante un clasificador enteramente aleatorio. La idea de la ROC a su vez da origen a la métrica AUC (*Area Under the Curve*), que es el área total bajo la curva, un clasificador aleatorio tiene un AUC de 0.5.

Utilizando NicheToolbox se determinó la ROC parcial, generando quinientos modelos aleatorios para cada especie en cada algoritmo, los valores superiores a uno indican que el modelo es mejor que uno generado aleatoriamente.

Se tomó en cuenta la tasa de omisión (OR) que se refiere a cuántas ocurrencias quedaron fuera de lo que el algoritmo define como probabilidad alta de condiciones adecuadas para la especie. El porcentaje de área predicha también fue tomado en cuenta pues modelos con OR de 0% en muchos casos eran modelos que predecían toda el área con alta probabilidad de tener condiciones adecuadas para la especie analizada.

Resultados

Compuestos orgánicos persistentes

Con base en el análisis realizado con el programa ECOSAR se proyecta en el siguiente mapa (Figura 4) la distribución de los compuestos encontrados en las muestras. Cancún presenta la mayor variación de compuestos, pero la laguna de la Bocana Carbonera, cerca de Sisal presenta el compuesto más peligroso.

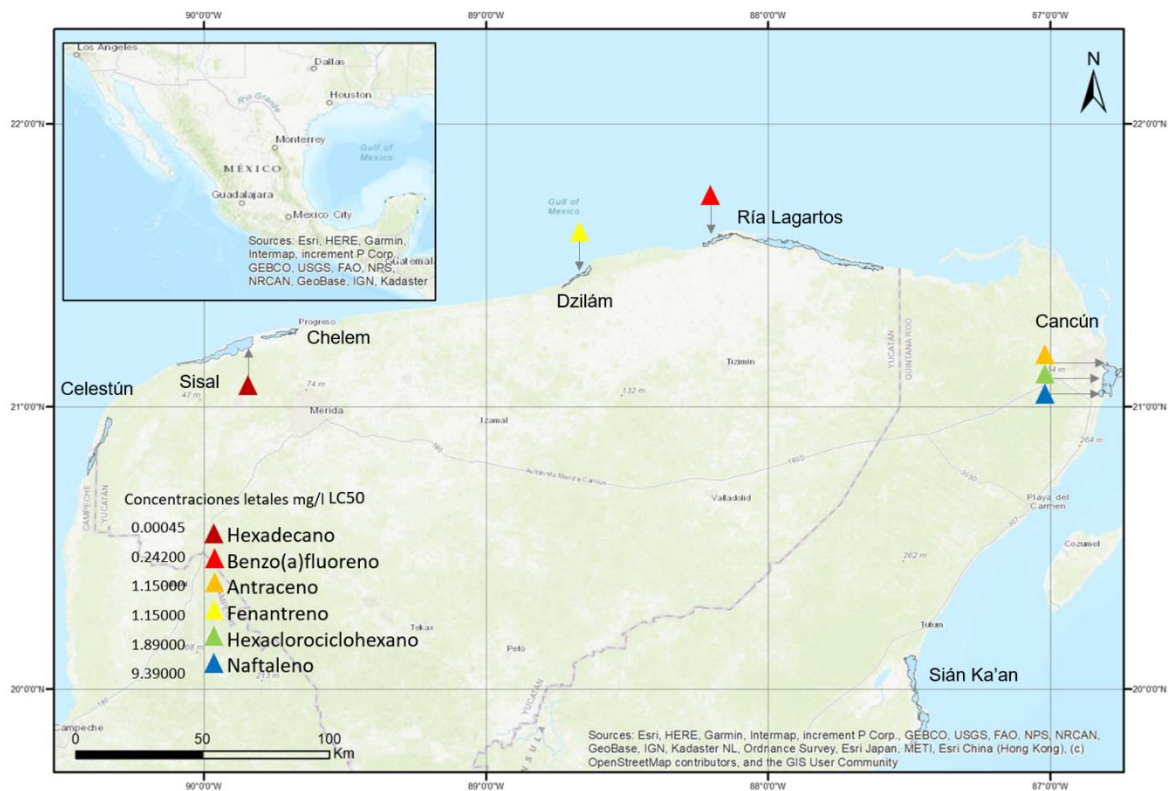


Figura 4 Mapa de contaminantes orgánicos persistentes y concentraciones letales

Se muestra la composición de especies de los géneros *Halomonas*, *Paracoccus* y *Staphylococcus*, reportados como organismos sintetizadores de PHA's (Figura 5). Se aprecia un cambio de la composición en el gradiente ambiental, el más claro presentado en el gradiente de salinidad, en sitios con aporte de agua dulce sólo se encuentra *Halomonas aquamarina*.

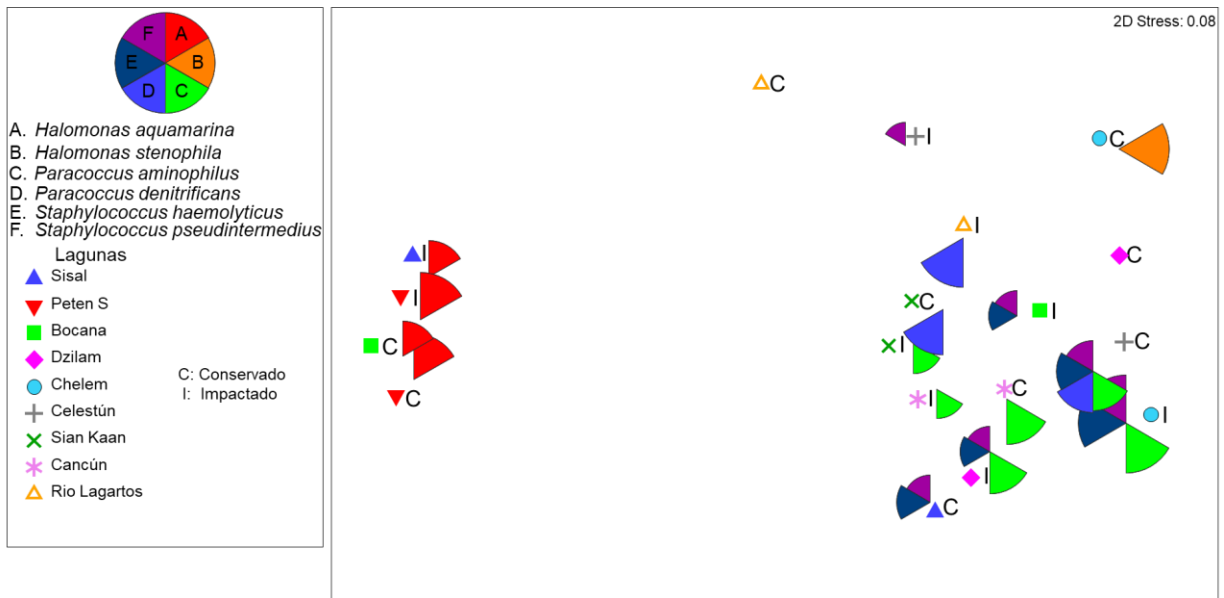


Figura 5 MDS no paramétrico de especies asociadas a COP's

Análisis de datos de secuenciación masiva

Se muestran los resultados al comparar las lagunas con aporte de agua dulce contra las lagunas salobres, el número total de especies es mayor para las lagunas “Dulces” por más de mil especies (Figura 6a), por otro lado, el número total de especies únicas fue mayor en lagunas “Dulces” por más del doble (Figura 6b), finalmente el número total de lecturas fue mayor en las lagunas “Salobres” (Figura 6c).

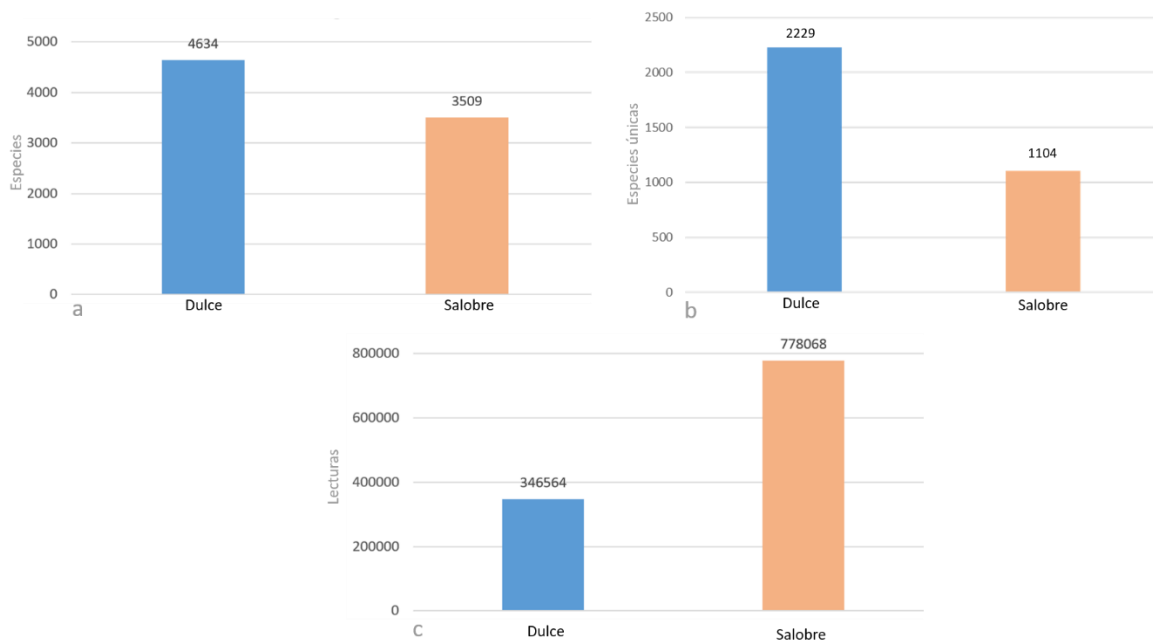


Figura 6 Comparación del número de especies y lecturas entre lagunas Dulces y Salobres

Se muestran los resultados de la aplicación del filtro a nivel Filo y su cluster de similitud correspondiente (Figura 7). Como puede observarse el Filo Proteobacteria es responsable de más del 50% de las lecturas en todos los sitios. En el caso de los sitios con aporte de agua dulce se distingue que, aunque la composición en proporciones se mantiene, la contribución de “Otros” aumenta debido a organismos endémicos de cuerpos dulceacuícolas de Filos variados.

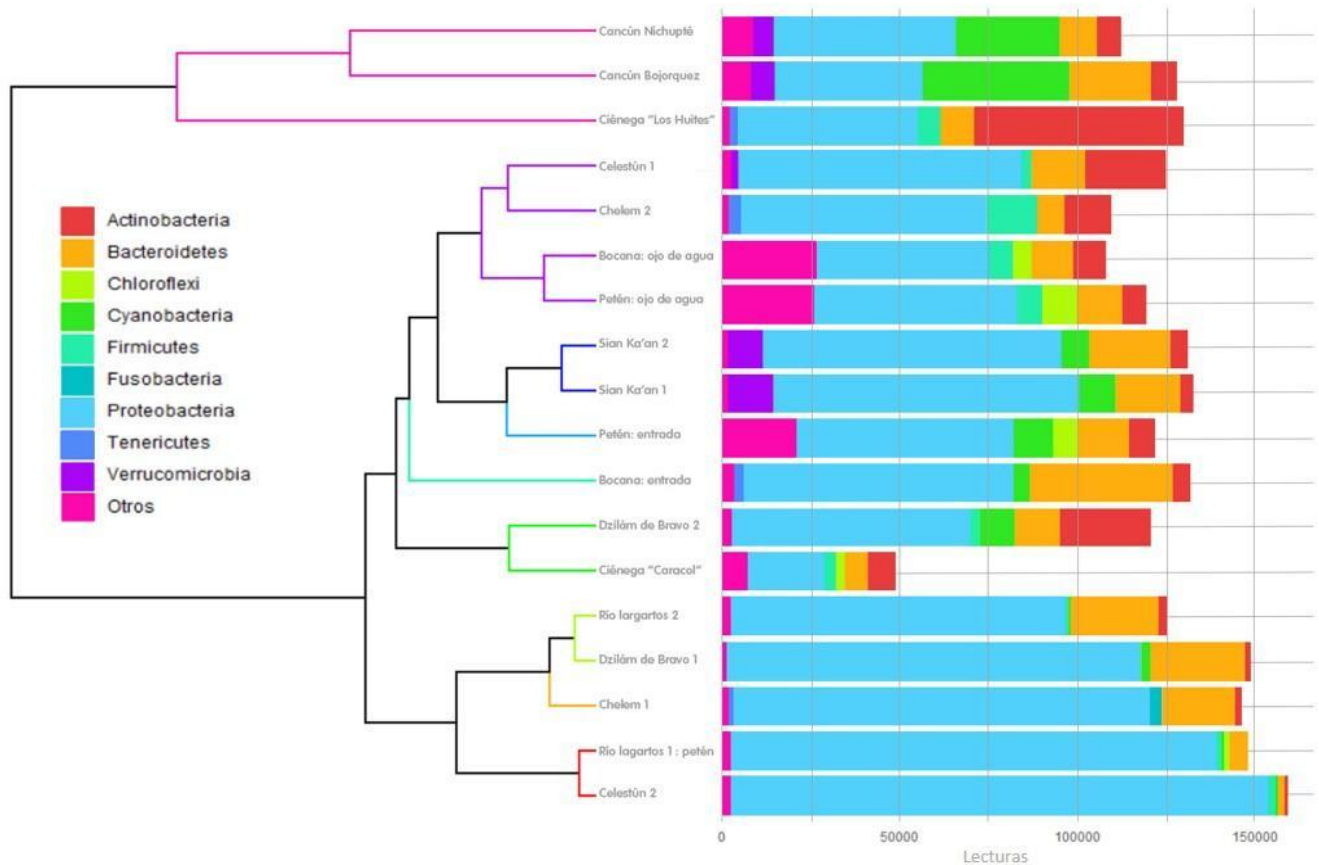


Figura 7 Cluster de similitud de la composición de Filos entre lagunas, los colores en los nodos terminales representan los clústers de agrupación.

Variables ambientales y composición taxonómica.

En el siguiente gráfico se ilustra el cambio de la composición taxonómica entre sitios (Figura 8), hay cambios claros con alta diversidad y abundancia de especies (izquierda) correspondientes a los sitios con aporte de agua dulce, nótese que tanto su composición como sus características ambientales los separan drásticamente de los otros sitios, exceptuando el sitio de Río Lagartos, no es su composición taxonómica sino sus características ambientales las que lo ubican en ese grupo.

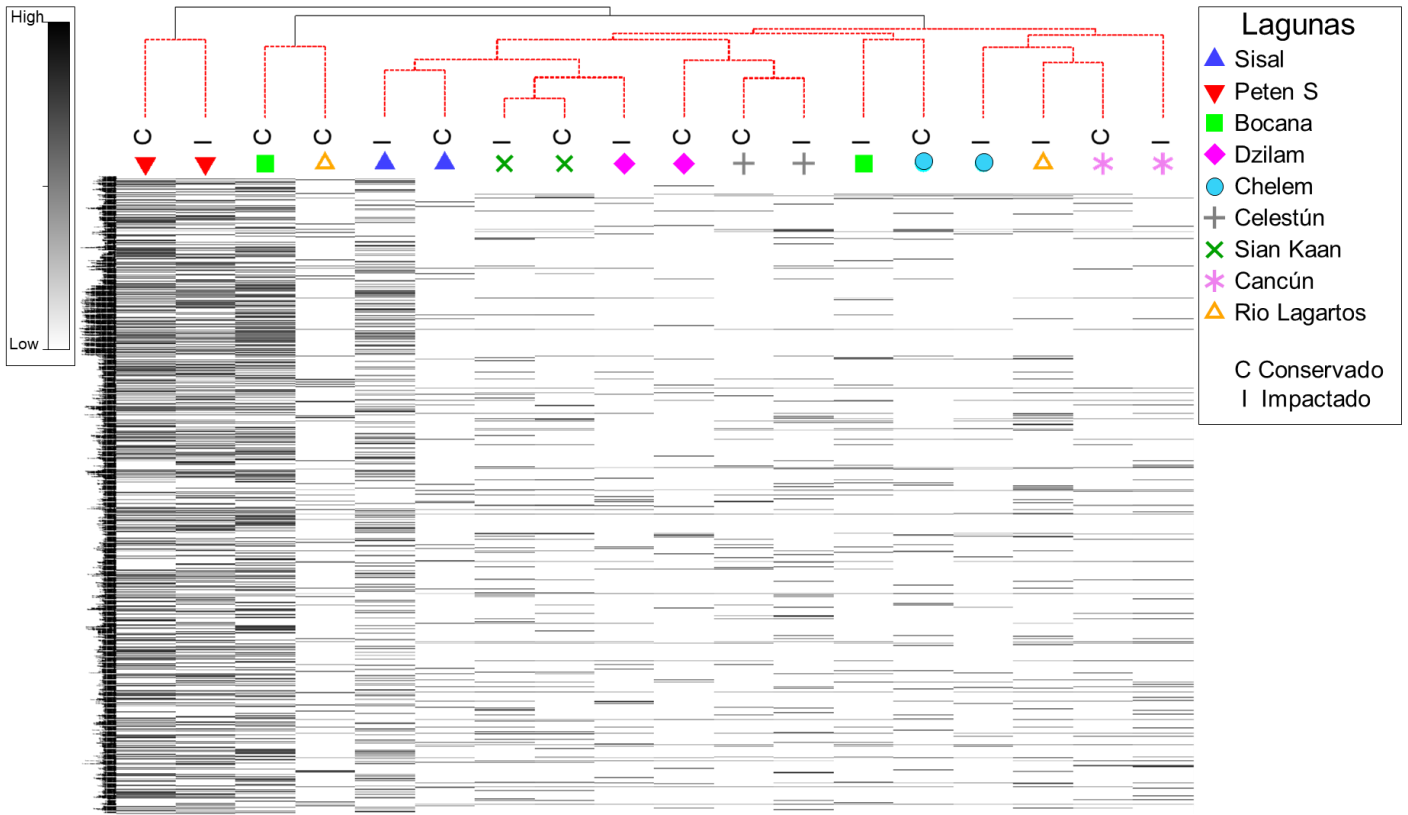


Figura 8 Shade plot de abundancias y clasificación de similitud entre sitios

A partir del análisis BEST (bv-step) se presenta un *shade plot* (Figura 9) integrando el clúster de variables ambientales. Se muestran especies que están en la mayoría de los sitios y su abundancia varía en función de las características ambientales y especies que se encuentran en agrupaciones de lagunas específicas. De nuevo es notorio un patrón de composición que responde al cambio en la salinidad.

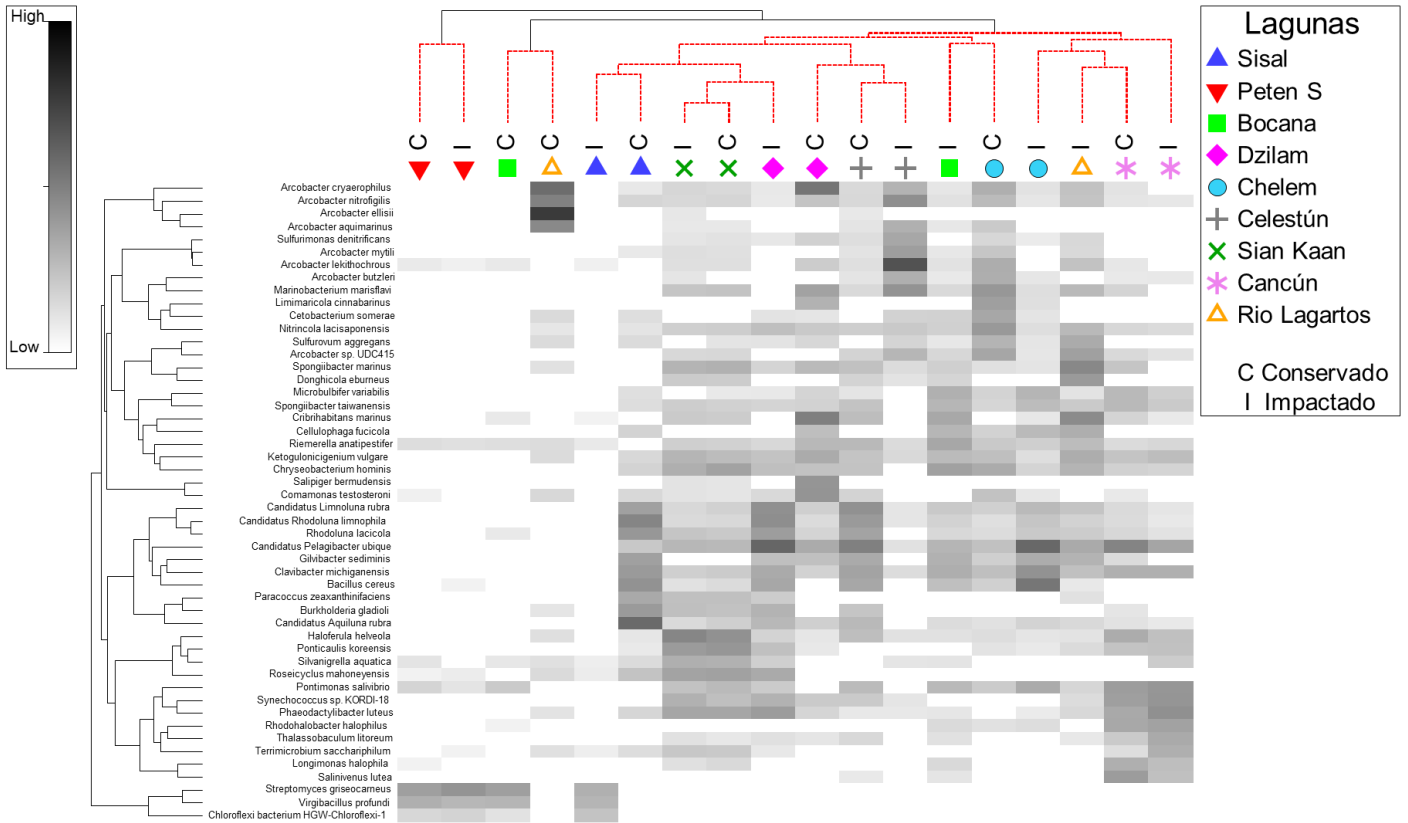


Figura 9 Shade plot de abundancias para las 50 especies más importantes

El análisis BEST (bioenv) el set de organismos resultante fue de 6 con un valor de Rho de 0.95 presentados a continuación (Figura 10). El cambio de las especies ilustra un gradiente de salinidad claro y un gradiente de contaminación multifactorial que se detalla de forma más digerible en el siguiente gráfico.

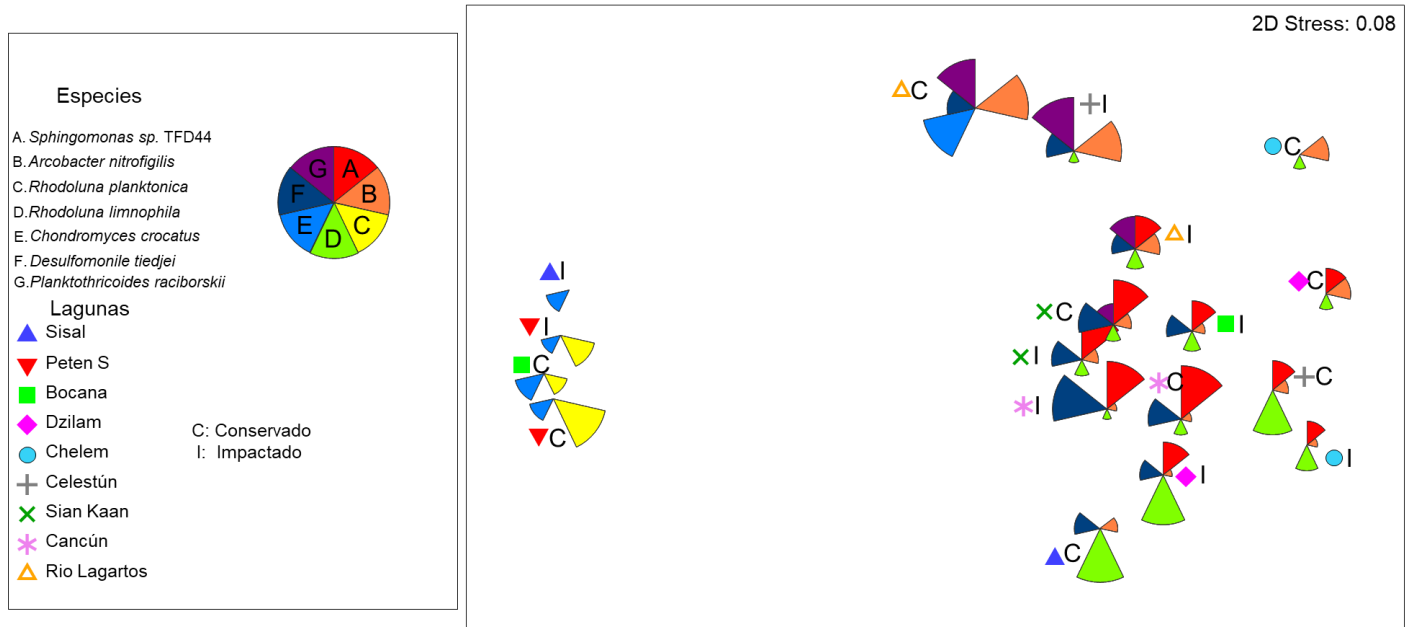


Figura 10 MDS del set de especies más importantes para diferenciar entre sitios

Del PCA (calidad intermedia 50% de la variación explicada con los dos primeros componentes) se observa una diferenciación entre el factor de conservado vs. Impactado (Figura 11). Se observa que consistentemente los sitios “Conservados” se encuentran en la parte más positiva del componente principal 2 y en la región más positiva del componente principal 1. Se debe resaltar que no hay una agrupación fuerte y diferenciada entre sitios “conservados” e “impactados”, aunque hay diferencias internas en cada muestra de las lagunas existen mayores diferencias al comparar entre lagunas.

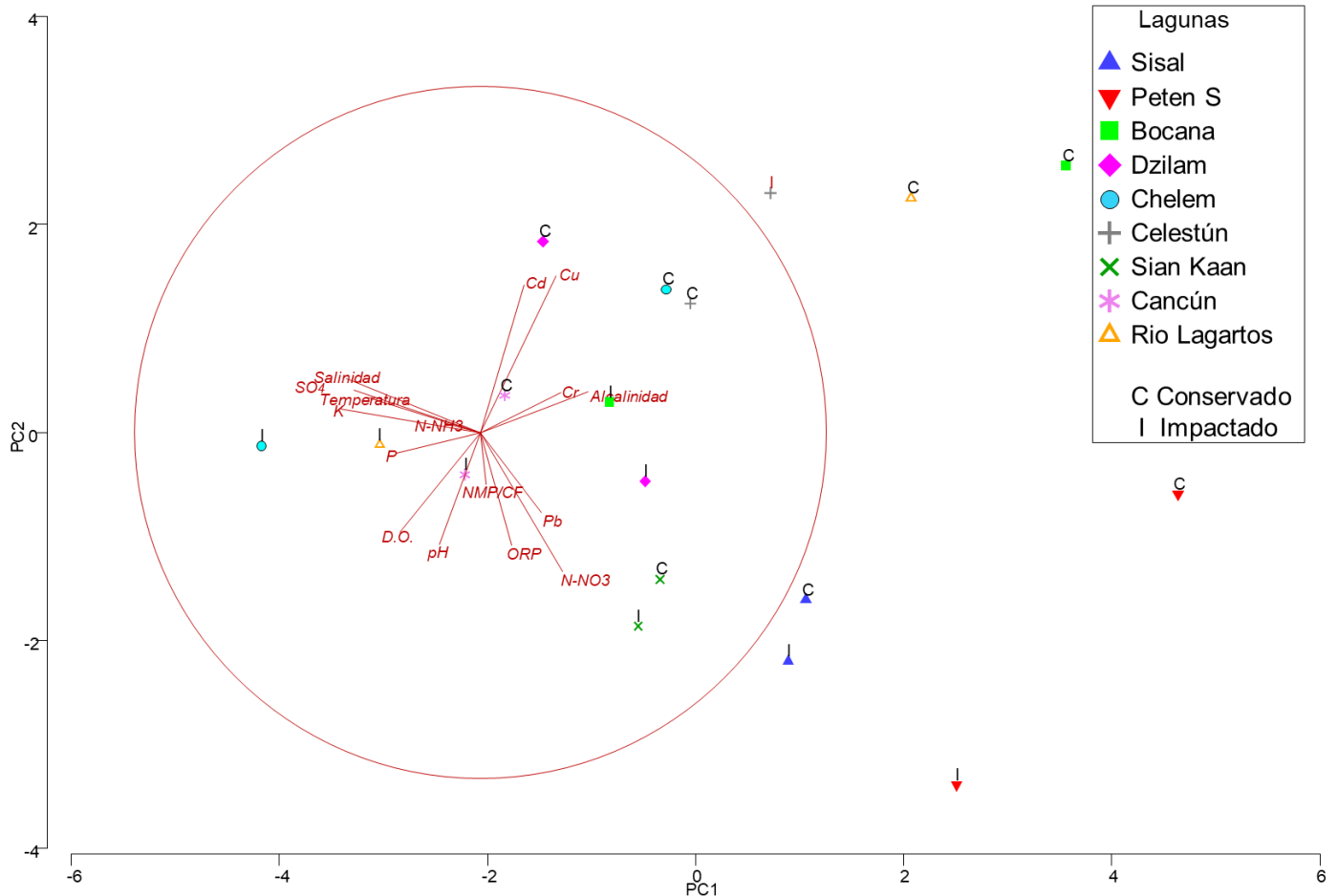


Figura 11 PCA de variables ambientales. PC1 31.7%, PC2 18.1%

Modelos

Se presentan los mapas de la especie *Arcobacter nitrofigilis* realizados con el algoritmo *Random Forest*, pues como se mencionó anteriormente, consistentemente tuvo el AUC (Figura 12) y pROC (Figura 13) más alto entre algoritmos y especies, la tasa de omisión más baja y los porcentajes de área predichos más conservadores, en comparación con algoritmos que predijeron 100% del área disponible. El primer mapa presenta la probabilidad de condiciones adecuadas para la especie (Figura 14), el siguiente mapa es la binarización de tales probabilidades (Figura 15). El tercer mapa es el resultado de la aplicación de la restricción probabilidad de ocurrencias (Figura 16). El cuarto mapa presenta la binarización del modelo con restricciones (Figura 17).

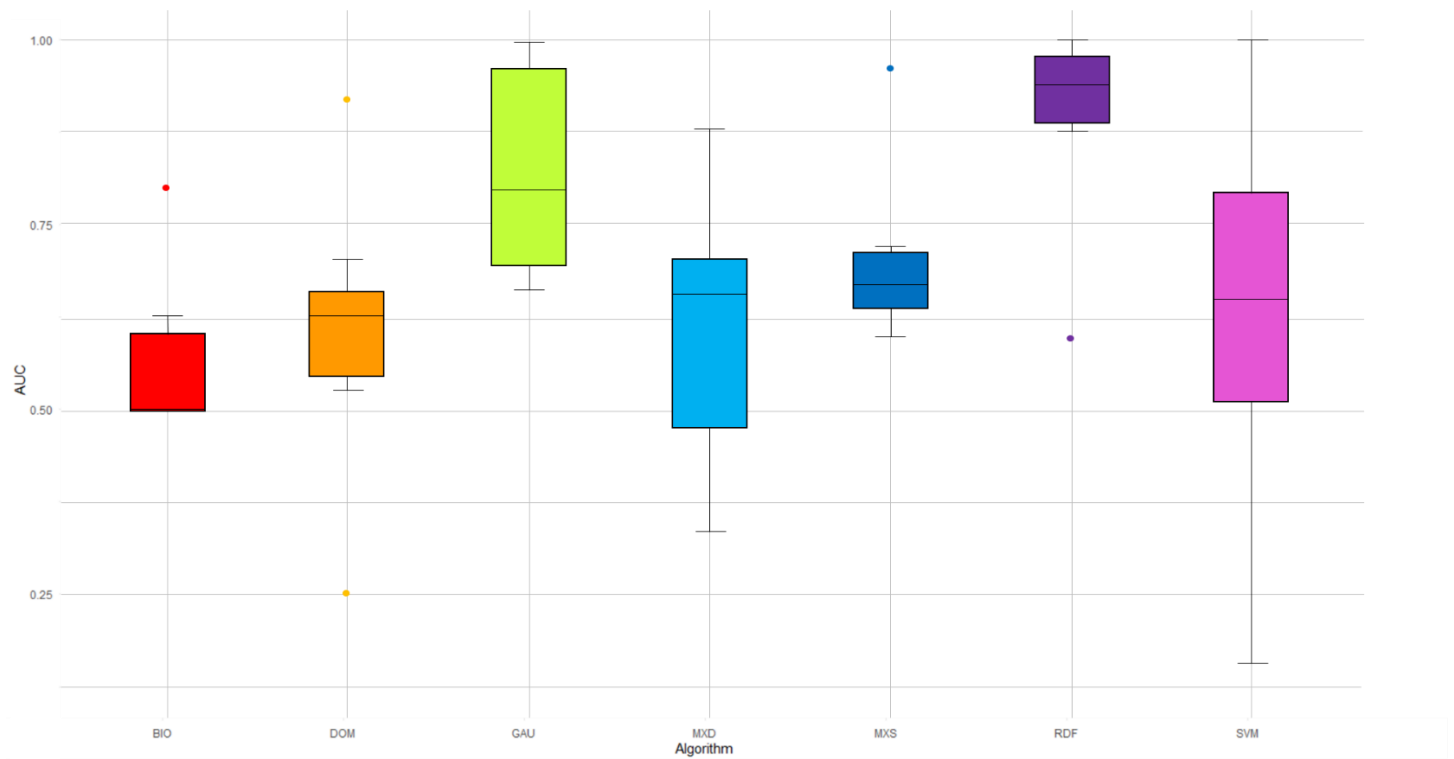


Figura 12 Boxplot de desempeño de algoritmos (AUC)

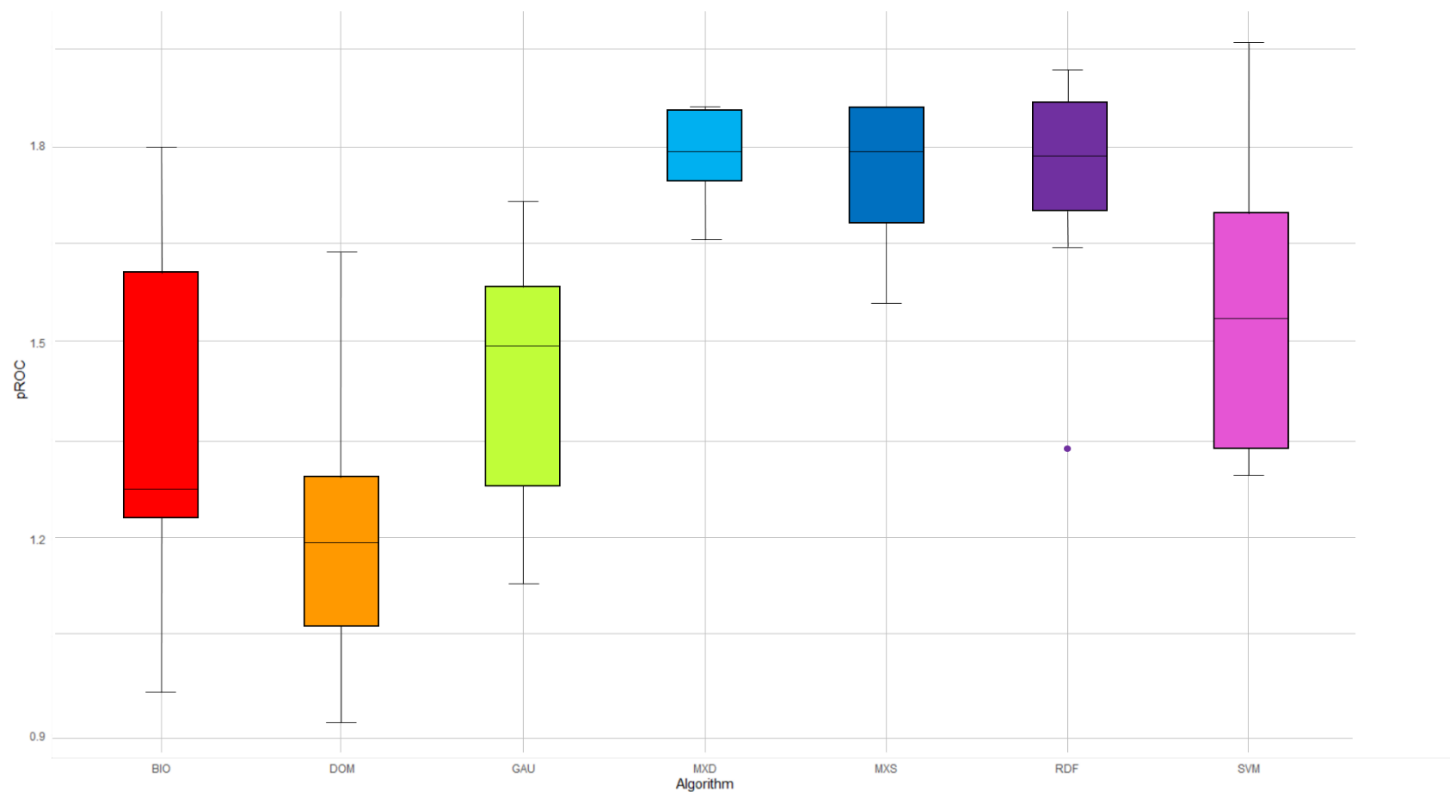


Figura 13 Boxplot de desempeño de algoritmos (pROC)

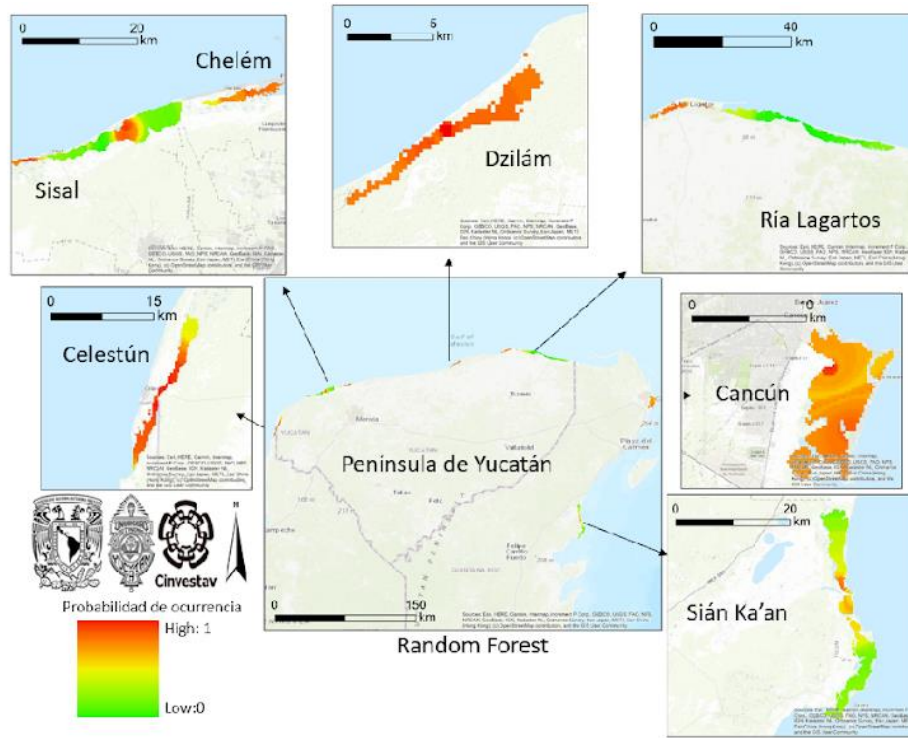


Figura 14 Mapa de probabilidad de condiciones adecuadas para *Arcobacter nitrofigilis* (Random Forest)

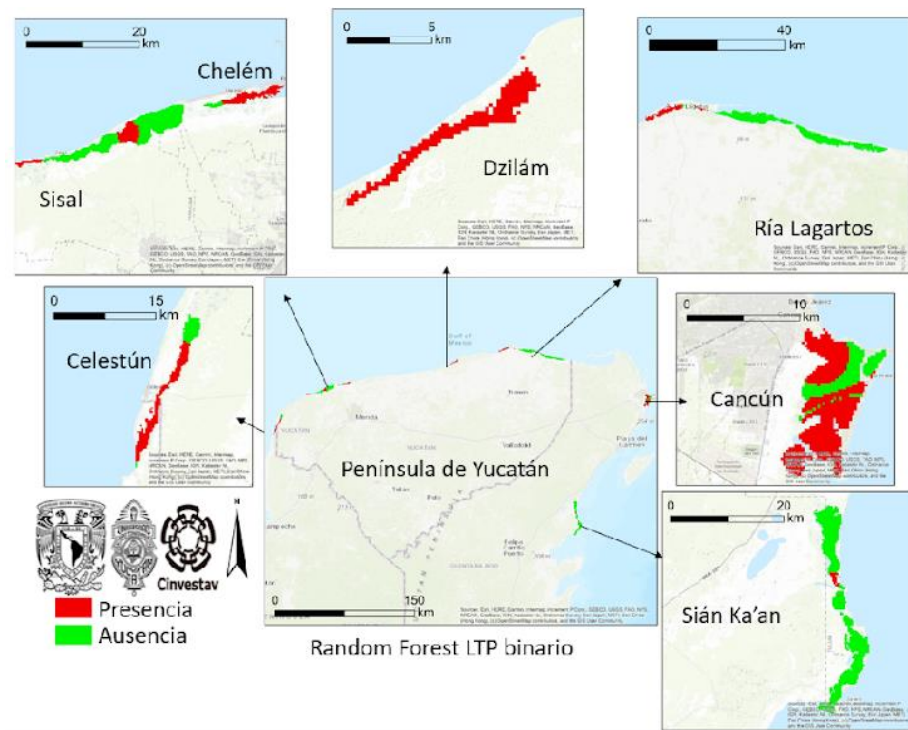


Figura 15 Mapa binario de probabilidad de condiciones adecuadas de *Arcobacter nitrofigilis* (Random Forest)

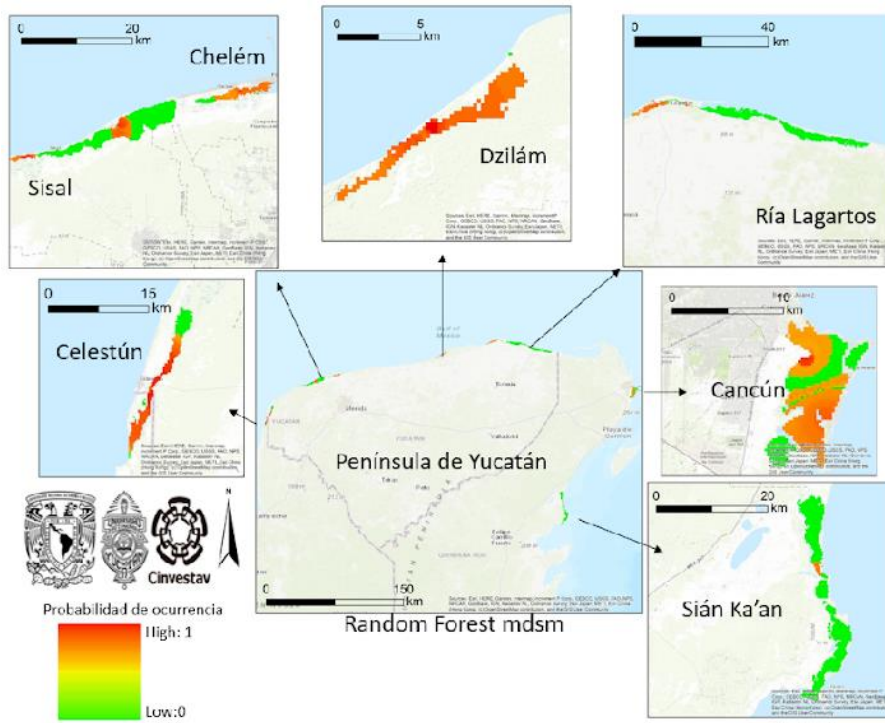


Figura 16 Mapa de probabilidad de condiciones adecuadas para *Arcobacter nitrofigilis* (Random Forest, restricción por presencias aledañas confirmadas)

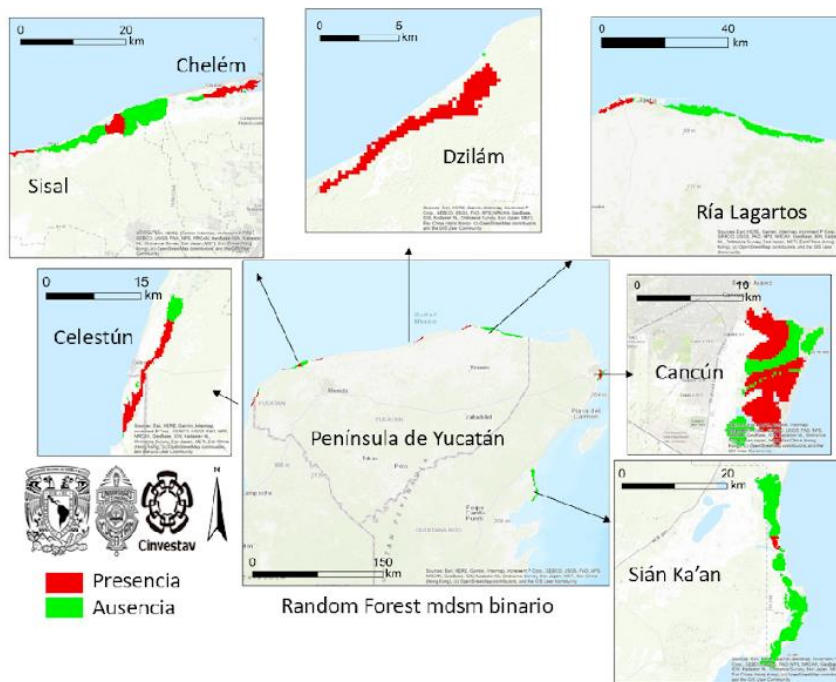


Figura 17 Mapa binario de probabilidad de condiciones adecuadas para *Arcobacter nitrofigilis* (Random Forest, restricción por presencias aledañas confirmadas)

Se analizaron los resultados de aporte de las variables utilizadas en los modelos (Figura 18), es importante recalcar que no todos los algoritmos generan porcentajes de contribución para las variables usadas y que no se encuentran todas las variables disponibles debido a que en la configuración del código se excluyen variables que tengan alta colinealidad, en términos generales, son los carbonatos,

coliformes y pH las variables que contribuyeron más para definir la distribución de las especies analizadas.

De manera más general se promediaron los porcentajes de contribución de cada algoritmo para cada variable, determinando la variable más importante para cada microorganismo, los carbonatos fueron la variable que más aportó para 3 especies y los coliformes fecales para otras dos (Figura 19).

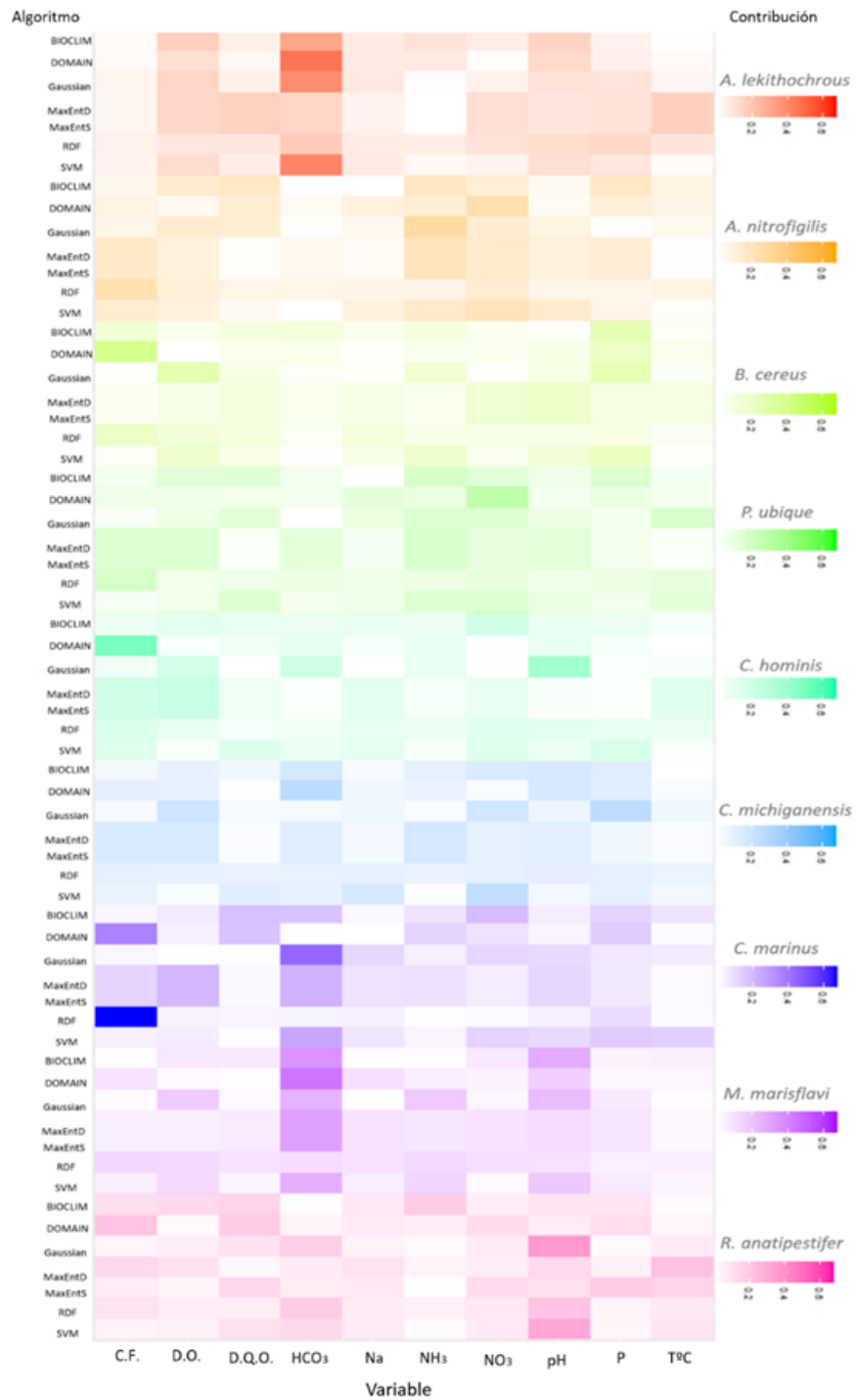


Figura 18 Heatmap de porcentajes de contribución de variables por algoritmo para cada especie

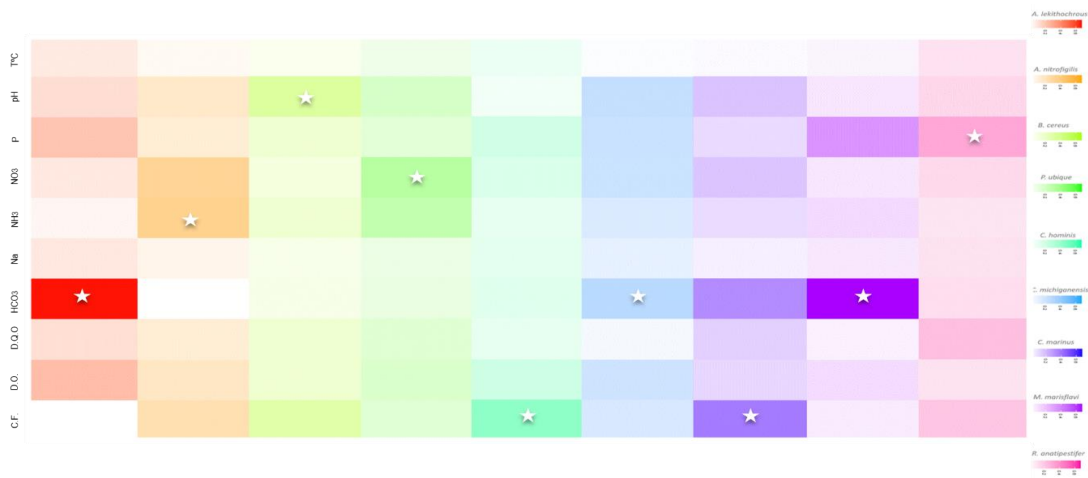


Figura 19 Heatmap de porcentajes de contribución de variables ambientales por especie

Bacterias asociadas a bosques de manglar

Los mapas de probabilidad de condiciones adecuadas para los géneros *Pantoea*, *Enterobacter* y *Flavobacterium*, se muestran a continuación, añadiendo la contribución de cada género en número de lecturas por sitio de muestreo (Figura 20). Aunado a este se presentan los mapas de probabilidad de condiciones adecuadas para cada género (Ilustraciones 21 a 23).

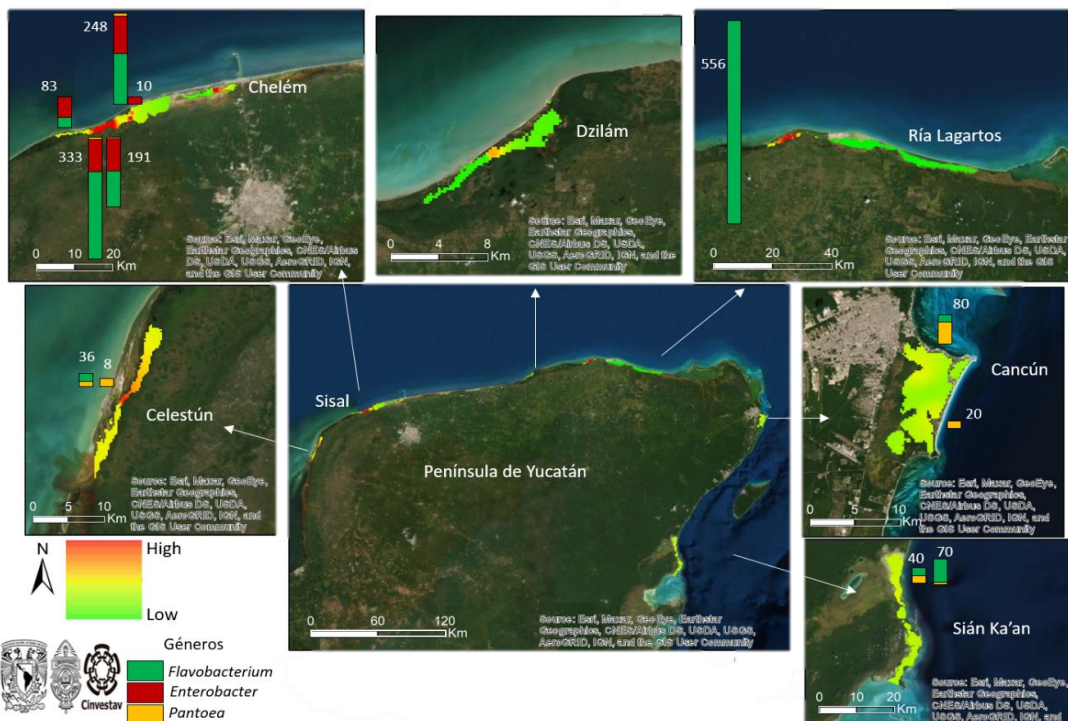


Figura 20 Mapa de probabilidad de condiciones adecuadas para los géneros bacterianos asociados a bosques de manglar

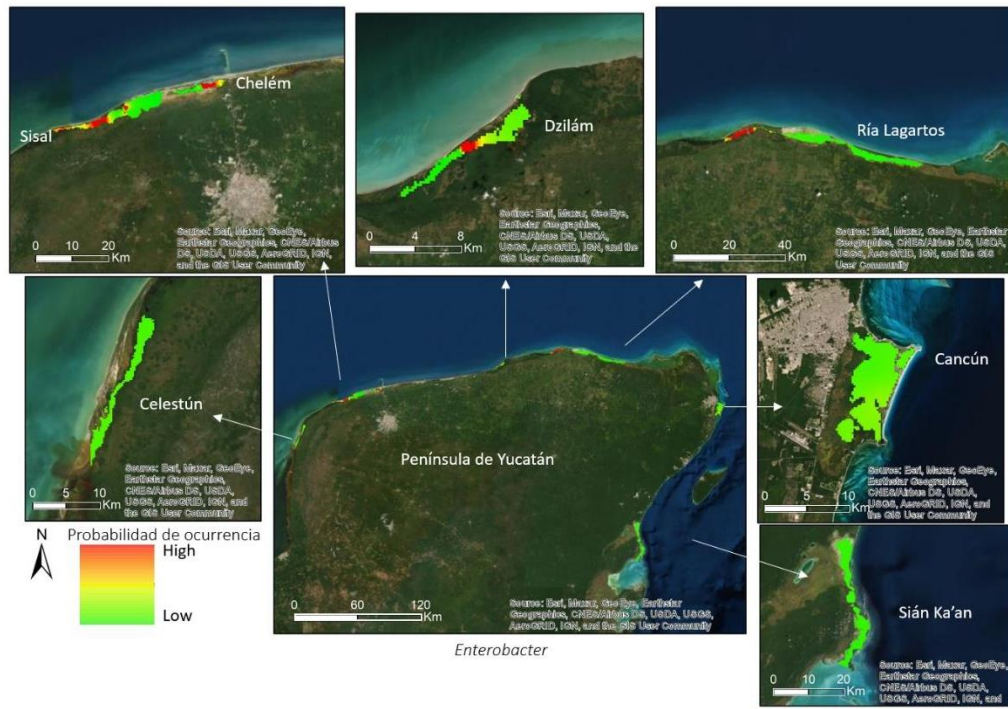


Figura 21 Mapa de probabilidad de condiciones adecuadas para el género *Enterobacter*

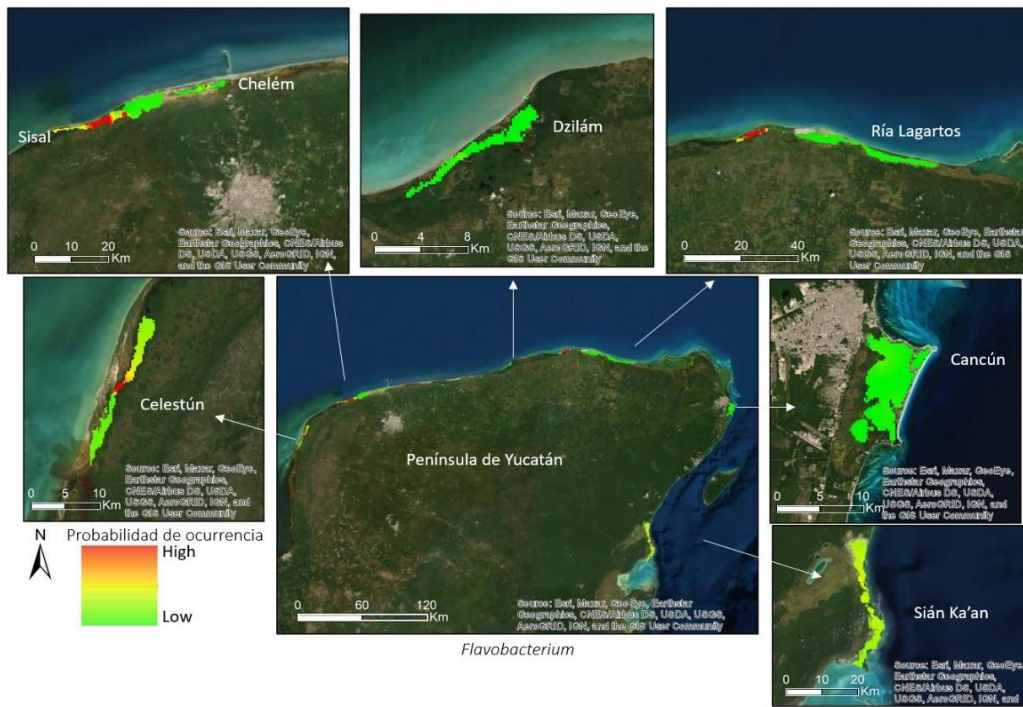


Figura 22 Mapa de probabilidad de condiciones adecuadas para el género *Flavobacterium*

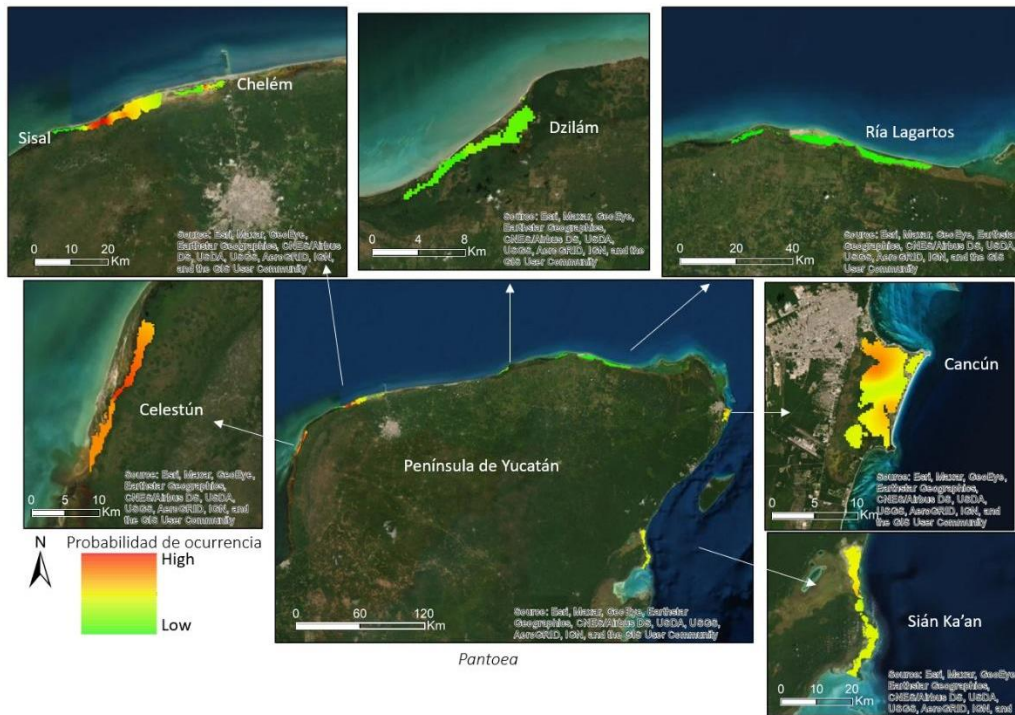


Figura 23 Mapa de probabilidad de condiciones adecuadas para el género *Pantoea*

Discusión

Con base en los análisis estadísticos, los sitios con mayor diversidad de especies son los que tienen aporte de agua dulce, lo que podría explicarse por el arrastre de material genético a través del sistema de cenotes y a la diversidad de microambientes generados en el encuentro con el agua salada. Dentro del grupo de los sitios salobres, fue el sitio de Sisal, Ciénega Caracol el que registró el mayor número de especies (más de 2000) en comparación al promedio de 400 especies por sitio.

Aunque no se encontraron diferencias significativas entre los niveles del factor “Estado” (Conservado vs. Impactado) existen diferencias, evidenciadas en la distribución de los puntos en las proyecciones de los escalamientos multidimensionales. Las diferencias entre lagunas son mayores, esto hace sentido al reflexionar que las condiciones “Impactadas” de un sitio relativamente bien conservado como la “Bocana Carbonera” están mejor conservadas que el sitio “Conservado” en una laguna severamente impactada como “Cancún”

El trabajo presenta el mejor set de especies, variables y al mejor algoritmo para datos de secuenciación masiva en lagunas costeras de la región. Los algoritmos analizados se desempeñaron de formas distintas entre sí y entre especies, en la evaluación de los algoritmos, aquellos que predijeron áreas con alta probabilidad condiciones adecuadas correctamente fueron los modelos realizados con el algoritmo Random Forest, las métricas de evaluación AUC, pROC, OR y proporción de área predicha en conjunto lo destacan como el mejor algoritmo para analizar los datos utilizados. Es importante

recalcar que “el mejor algoritmo” y “el mejor modelo” no serán el mismo para cada especie ni para cada muestreo, comprender lo que ocurre aritméticamente es más importante, así como entender qué representan los mapas de probabilidad y qué determina su apariencia.

Algunos de los sitios mostraron menor diversidad de especies, pero mayor número de lecturas para especies con potencial patógeno (e.g. Cancún, Sian Ka'an), esto podría deberse a fuentes de contaminación continuas que han llevado a que las poblaciones de microorganismos que implementan rutas metabólicas especializadas aumenten su densidad poblacional, desplazando a otros organismos y modificando los nichos disponibles. En el caso de Cancún se debe a la densidad poblacional humana y sus efectos sobre las lagunas analizadas, en el caso de Sian Ka'an se ha reportado transporte de basura mediante corrientes (mismas que transportan el sargazo) provenientes del sur.

En contraste, Sisal y Río Lagartos tienen una gran diversidad de especies que podría ser indicador de un ambiente sano, no obstante, ambos cuerpos de agua están sujetos a diversas fuentes de contaminación, aunque no en las magnitudes de Cancún, es posible vincularlas con microorganismos, especialmente contaminación por aguas residuales, hidrocarburos, basura, actividades cinegéticas y ganadería. Se abordó la contaminación por hidrocarburos en el análisis de géneros sintetizadores de PHA's, destacando el potencial de los datos de secuenciación masiva para encontrar fuentes de contaminación puntuales o difusas y para el monitoreo del estado de salud ambiental.

En cuanto a los modelos construidos con géneros endófitos reportados, es necesario caracterizar más ampliamente la composición del microbioma en distintos estados de salud de los ecosistemas de manglar para elegir con mayor exactitud los géneros y especies para evaluar su potencial como bioindicadores en ecosistemas de manglar, sin embargo la especificidad de los mapas de condiciones adecuadas para géneros con potencial biotecnológico comprueba la sensibilidad de los datos y técnicas implementadas para su monitoreo o estudios de bioprospección.

El trabajo muestra cómo el uso del ADN microbiano proveniente de lagunas, combinado con el monitoreo de diversas variables ambientales y la aplicación de los algoritmos para generar modelos de distribución permiten aproximar la geografía ocupada de microorganismos sin necesidad de aislarlos, con el potencial de integrar variables hidrodinámicas, socioeconómicas y evaluaciones en otros sistemas acuáticos a nivel global.

Conclusiones

En este trabajo se utilizaron distintas técnicas de análisis para variables ambientales, su interpolación, transformación y reducción de colinealidad. Emplea el estudio de secuencias de ADN procesadas, clasificadas, asignadas y georreferenciadas. Los datos se analizaron con diversas técnicas de análisis estadísticos específicos para datos biológicos y ambientales, seguidos de análisis de correlación y selección de las especies/variables más importantes. Se apoya de búsquedas en bases de datos de caracterización de especies y de representaciones visuales para ilustrar distintos procesos. Finalmente incluye la comparación de distintos algoritmos con modificaciones para aproximar la geografía ocupada de microorganismos con potencial biotecnológico determinando con diferentes parámetros el algoritmo con mejor desempeño para los datos analizados.

El trabajo aporta información sobre el comportamiento de los microorganismos a escala regional, mostrando el dinamismo de la comunidad microbiana y el potencial de los datos de secuenciación masiva para llevarnos a un mejor entendimiento de los procesos que se llevan a cabo en el ambiente. El trabajo define a partir de más de 25 variables ambientales, el set de variables que más aportan para diferenciar entre lagunas y que mejor representan las diferencias entre la composición taxonómica.

La diversidad taxonómica de las lagunas costeras está influenciada por la disponibilidad de fuentes de energía (e.g., contaminantes, materia orgánica particulada, materia orgánica disuelta), los sitios con aporte de agua dulce contienen la mayor diversidad de especies: 4,634 en comparación a 3,509 en los sitios salobres. La diversidad de especies “endémicas” de cada tipo de ambiente también es dominada por los cuerpos de agua dulce con 2,229 especies únicas, en contraste los sitios salobres donde solo se cuantificaron 1,104 especies únicas.

El uso de datos de secuenciación masiva, análisis sensibles como la absorción atómica y la búsqueda en bases de datos permitió la selección de especies con capacidad de metabolizar contaminantes. Dichos microorganismos tienen relaciones específicas con tales contaminantes y con los modelos de distribución fue posible encontrarlos en áreas donde los análisis de absorción atómica no detectaron el contaminante en cuestión.

Para los sitios analizados las variables que los diferencian son los nitratos, amonio, sulfatos, potasio y cobre. Para las especies analizadas (organismos con alta plasticidad y distribuciones amplias en la región) los carbonatos, coliformes, pH y nitratos conformaron el set de variables que aportan más para definir su distribución. El análisis especializado que representa la secuenciación masiva puede permitirnos seleccionar especies que funjan como indicadores, organismos asociados a contaminantes específicos o combinaciones de organismos que permitan monitorear el cambio del estado ambiental.

Referencias

- Altermatt, F., Little, C. J., Mächler, E., Wang, S., Zhang, X., & Blackman, R. C. (2020). *Uncovering the complete biodiversity structure in spatial networks: the example of riverine systems*. December 2019, 607–618. <https://doi.org/10.1111/oik.06806>
- Altermatt, F., Seymour, M., & Martinez, N. (2013). *River network properties shape a -diversity and community similarity patterns of aquatic insect communities across major drainage basins*. 2249–2260. <https://doi.org/10.1111/jbi.12178>
- Andrade, A. F. A. de, Velazco, S. J. E., & De Marco Júnior, P. (2020). ENMTML: An R package for a straightforward construction of complex ecological niche models. *Environmental Modelling and Software*, 125(January), 104615. <https://doi.org/10.1016/j.envsoft.2019.104615>
- Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*. <https://doi.org/10.1890/11-1930.1>
- Bae, S. S., Jung, J., Chung, D., & Baek, K. (2018). *Marinobacterium aestuarii* sp. nov., a benzene-degrading marine bacterium isolated from estuary sediment. *International Journal of Systematic and Evolutionary Microbiology*. <https://doi.org/10.1099/ijsem.0.002561>
- Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kausrud, H. (2010). ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. *BMC Microbiology*. <https://doi.org/10.1186/1471-2180-10-189>
- Bennett, D. J., Choimes, A., Collen, B., Day, J., Palma, A. De, Di, S., Edgar, M. J., Feldman, A., Garon, M., Harrison, M. L. K., Alhousseini, T., Echeverria-london, S., Ingram, D. J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Laginha, D., ... Mace, G. M. (2015). *Global effects of land use on local terrestrial biodiversity*. <https://doi.org/10.1038/nature14324>
- Bennett, S. D., Walsh, K. A., & Gould, L. H. (2013). Foodborne disease outbreaks caused by *Bacillus cereus*, *Clostridium perfringens*, and *Staphylococcus aureus* - United States, 1998-2008. *Clinical Infectious Diseases*. <https://doi.org/10.1093/cid/cit244>
- Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). Bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. In *Diversity and Distributions*. <https://doi.org/10.1111/ddi.12144>
- Borcard, D., & Legendre, P. (1994). oribatid mites (Acari, Oribatei). *Environmental and Ecological Statistics*, 1(1), 37–61.

- Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*. <https://doi.org/10.2307/1940179>
- Breiman, L. (2001). Randomforest2001. *Machine Learning*. <https://doi.org/10.1017/CBO9781107415324.004>
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., MacE, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. In *Nature*. <https://doi.org/10.1038/nature11148>
- Carpenter, G., Gillison, A. N., & Winter, J. (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*. <https://doi.org/10.1007/BF00051966>
- Carraro, L., Mächler, E., Wüthrich, R., & Altermatt, F. (2020). Environmental DNA allows upscaling spatial patterns of biodiversity in freshwater ecosystems. *Nature Communications*. <https://doi.org/10.1038/s41467-020-17337-8>
- Chen, Z., Liu, Y., Liu, L. Z., Zhong, Z. P., Liu, Z. P., & Liu, Y. (2014). *Cribrihabitans marinus* gen. nov., sp. nov., isolated from a biological filter in a marine recirculating aquaculture system. *International Journal of Systematic and Evolutionary Microbiology*, 64(PART 4), 1257–1263. <https://doi.org/10.1099/ijs.0.059576-0>
- Cho, J. C., & Tiedje, J. M. (2000). Biogeography and degree of endemism of fluorescent *Pseudomonas* strains in soil. *Applied and Environmental Microbiology*, 66(12), 5448–5456. <https://doi.org/10.1128/AEM.66.12.5448-5456.2000>
- Collado, L., & Figueras, M. J. (2011). Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clinical Microbiology Reviews*. <https://doi.org/10.1128/CMR.00034-10>
- CONABIO. (2008). *División política estatal, escala 1:2,500,000*.
- Damashek, J., & Francis, C. A. (2018). Microbial Nitrogen Cycling in Estuaries: From Genes to Ecosystem Processes. In *Estuaries and Coasts*. <https://doi.org/10.1007/s12237-017-0306-2>
- Dang, H., & Jiao, N. (2014). Perspectives on the microbial carbon pump with special reference to microbial respiration and ecosystem efficiency in large estuarine systems. *Biogeosciences*, 11(14), 3887–3898. <https://doi.org/10.5194/bg-11-3887-2014>
- Dang, Hongyue, Klotz, M. G., Lovell, C. R., & Sievert, S. M. (2019). Editorial: The responses of

- marine microorganisms, communities and ecofunctions to environmental gradients. In *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2019.00115>
- Dang, Hongyue, & Lovell, C. R. (2002). diversity of marine Rhodobacter species during early colonization of submerged surfaces in coastal marine waters as determined by 16S ribosomal DNA sequence. *Appl. Environ. Microbiol.* <https://doi.org/10.1128/AEM.68.2.496>
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*. <https://doi.org/10.1098/rsbl.2014.0562>
- Deiner, K., & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0088786>
- Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J. C., & Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nature Communications*. <https://doi.org/10.1038/ncomms12544>
- Deiner, K., Walser, J. C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*. <https://doi.org/10.1016/j.biocon.2014.11.018>
- Delaney, J. Van Niel, K. P. (2007). *Geographical Information Systems: An Introduction* (K. Van Niel (ed.)). Oxford University Press.
- Diéguez, A. L., Balboa, S., Magnesen, T., & Romalde, J. L. (2017). *Arcobacter lekithochrous* sp. nov., isolated from a molluscan hatchery. *International Journal of Systematic and Evolutionary Microbiology*, 67(5), 1327–1332. <https://doi.org/10.1099/ijsem.0.001809>
- Donachie, S. P., Bowman, J. P., On, S. L. W., & Alam, M. (2005). *Arcobacter halophilus* sp. nov., the first obligate halophile in the genus *Arcobacter*. *International Journal of Systematic and Evolutionary Microbiology*. <https://doi.org/10.1099/ijse.0.63581-0>
- Drake, J. M., Randin, C., & Guisan, A. (2006). Modelling ecological niches with support vector machines. *Journal of Applied Ecology*. <https://doi.org/10.1111/j.1365-2664.2006.01141.x>
- Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. In *Biological Reviews of the Cambridge Philosophical Society*. <https://doi.org/10.1017/S1464793105006950>

- Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0130324>
- Elith, J., & Burgman, M. A. (2003). *Habitat Models for Population Viability Analysis*. https://doi.org/10.1007/978-3-662-09389-4_8
- Elith, Jane, Graham, C. H., Ferrier, S., Guisan, A., Anderson, R. P., Dudi, M., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., ... Zimmermann, N. E. (2009). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*.
- Elith, Jane, Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists: online appendices. *Diversity and Distributions*.
- Ettema, C. H., & Wardle, D. A. (2002). Spatial soil ecology. In *Trends in Ecology and Evolution*. [https://doi.org/10.1016/S0169-5347\(02\)02496-5](https://doi.org/10.1016/S0169-5347(02)02496-5)
- Felícisimo, A. M., Mateo, R. G., Felipe, B., Sánchez, J., Salvatierra, H. C., & Herrera, M. (2011). *Bosques y cambio global. Volumen 1: Modelos de distribución de especies* (Issue January).
- Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, 51(2), 331–363. <https://doi.org/10.1080/10635150252899806>
- Ficetola, G. F., & Miaud, C. (2008). *Species detection using environmental DNA from water samples*. *April*, 423–425. <https://doi.org/10.1098/rsbl.2008.0118>
- Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0507535103>
- Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A., & Cleveland, C. C. (2009). Global patterns in belowground communities. In *Ecology Letters*. <https://doi.org/10.1111/j.1461-0248.2009.01360.x>
- Figueras, M. J., Collado, L., Levican, A., Perez, J., Solsona, M. J., & Yustes, C. (2011). *Arcobacter molluscorum* sp. nov., a new species isolated from shellfish. *Systematic and Applied Microbiology*, 34(2), 105–109. <https://doi.org/10.1016/j.syapm.2010.10.001>
- Franklin, J., & Miller, J. (2014). Mapping Species Distributions: Spatial Inference and Prediction - Janet Franklin - Google Books. *Nature*. <https://doi.org/10.1017/CBO9780511810602>

- García, C., & Durga, G. (2012). PROBLEMÁTICA Y RIESGO AMBIENTAL POR EL USO DE PLAGUICIDAS EN SINALOA. *Ra Ximhai*. <https://doi.org/10.1039/C2BM00138A>
- Garrity, G., & Cole, J. (2007). Taxonomic Outline of the Bacteria and Archaea. In *Taxonomic Outline of the Bacteria and Archaea*. <https://doi.org/10.1601/toba7.7>
- Giovannoni, S. J., Bibbs, L., Cho, J. C., Stapels, M. D., Desiderio, R., Vergin, K. L., Rappé, M. S., Laney, S., Wilhelm, L. J., Tripp, H. J., Mathur, E. J., & Barofsky, D. F. (2005). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature*, *438*(7064), 82–85. <https://doi.org/10.1038/nature04032>
- González, J. M., Mayer, F., Moran, M. A., Hodson, R. E., & Whitman, W. B. (1997). *Microbulbifer hydrolyticus* gen. nov., sp. nov., and *Marinobacterium georgiense* gen. nov., sp. nov., two marine bacteria from a lignin-rich pulp mill waste enrichment community. *International Journal of Systematic Bacteriology*. <https://doi.org/10.1099/00207713-47-2-369>
- Gouda, S., Das, G., Sen, S. K., Shin, H. S., & Patra, J. K. (2016). Endophytes: A treasure house of bioactive compounds of medicinal importance. In *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2016.01538>
- Graham, E. B., Knelman, J. E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., Beman, J. M., Abell, G., Philippot, L., Prosser, J., Foulquier, A., Yuste, J. C., Glanville, H. C., Jones, D. L., Angel, R., Salminen, J., Newton, R. J., Bürgmann, H., Ingram, L. J., ... Nemergut, D. R. (2016). Microbes as engines of ecosystem function: When does community structure enhance predictions of ecosystem processes? *Frontiers in Microbiology*, *7*(FEB), 1–10. <https://doi.org/10.3389/fmicb.2016.00214>
- Graham, E. B., Wieder, W. R., Leff, J. W., Weintraub, S. R., Townsend, A. R., Cleveland, C. C., Philippot, L., & Nemergut, D. R. (2014). Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. *Soil Biology and Biochemistry*. <https://doi.org/10.1016/j.soilbio.2013.08.023>
- Green, J., Holmes, A., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M., & Beattie, A. (2004). Spatial scaling of microbial diversity. *Nature*, *432*(1999), 747–750.
- Greenberg, A. E. (1995). Advances in Standard Methods for the Examination of Water and Wastewater 19th edition. *Proceedings - AWWA Water Quality Technology Conference*, 11–13.
- Greene, R. V., Griffin, H. L., & Cotta, M. A. (1996). Utility of alkaline protease from marine shipworm bacterium in industrial cleansing applications. *Biotechnology Letters*.

<https://doi.org/10.1007/BF00127884>

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M. E., Meng, J., Mulligan, M., ... Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*. <https://doi.org/10.1038/s41586-019-1111-9>

Grinnell, J. (1917). Field Tests of Theories Concerning Distributional Control. *The American Naturalist*. <https://doi.org/10.1086/279591>

Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2014). Maxent is not a presence-absence method: A comment on Thibaud et al. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.12252>

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*. [https://doi.org/10.1016/S0304-3800\(02\)00204-1](https://doi.org/10.1016/S0304-3800(02)00204-1)

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3), 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)

Haddar, A., Agrebi, R., Bougatef, A., Hmidet, N., Sellami-Kamoun, A., & Nasri, M. (2009). Two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: Purification, characterization and potential application as a laundry detergent additive. *Bioresource Technology*, 100(13), 3366–3373. <https://doi.org/10.1016/j.biortech.2009.01.061>

Hänfling, B., Handley, L. L., Read, D. S., Hahn, C., Li, J., Nichols, P., Blackman, R. C., Oliver, A., & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*. <https://doi.org/10.1111/mec.13660>

Herrera-Silveira, J. A., & Morales, S. (2010). Lagunas costeras. *Biodiversidad y Desarrollo Humano En Yucatán*.

Hijmans, R. J., & Elith, J. (2016). Species distribution modeling with R. <http://www.idg.pl/Mirrors/CRAN/Web/Packages/Dismo/Vignettes/Sdm.Pdf>, 87. [https://doi.org/10.1016/S0550-3213\(02\)00216-X](https://doi.org/10.1016/S0550-3213(02)00216-X)

- Ho, H. T. K., Lipman, L. J. A., & Gaastra, W. (2006). Arcobacter, what is known and unknown about a potential foodborne zoonotic agent! *Veterinary Microbiology*, 115(1–3), 1–13.
<https://doi.org/10.1016/j.vetmic.2006.03.004>
- Horner-Devine, M. C., Lage, M., Hughes, J. B., & Bohannon, B. J. M. (2004). A taxa-area relationship for bacteria. *Nature*. <https://doi.org/10.1038/nature03073>
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0), 415–427. <https://doi.org/10.1101/sqb.1957.022.01.039>
- INEGI. (2010). *El sector alimentario en México*.
- INEGI. (2014). *Sistema de cuentas nacionales de México. Producto interno bruto por entidad federativa 2003-2014*.
- IPBES, I. S.-P. P. on B. and E. S. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In *Debating Nature's Value*.
- Jacob, J., Woodward, D., Feuerpfeil, I., & Johnson, W. M. (1998). Isolation of Arcobacter butzleri in raw water and drinking water treatment plants in Germany. *Zentralblatt Für Hygiene Und Umweltmedizin = International Journal of Hygiene and Environmental Medicine*.
- Jetz, W., Mcgeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-karger, F. E., Pereira, H. M., & Regan, E. C. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, 3(April). <https://doi.org/10.1038/s41559-019-0826-1>
- Joseph, B., & Mini Priya, R. (2011). Bioactive compounds from endophytes and their potential in pharmaceutical effect: A review. In *American Journal of Biochemistry and Molecular Biology*. <https://doi.org/10.3923/ajbmb.2011.291.309>
- Kehew, A. (2000). *Applied Chemical Hydrogeology*. Prentice, Hall.
- Kelly, B. R. P., Port, J. A., Yamahara, M., Martone, R. G., Lowell, N., Thomsen, P. F., Mach, M. E., Bennett, M., Prahler, E., Caldwell, M. R., & Crowder, L. B. (2014). *Harnessing DNA to improve environmental management*. 344(6191).
- Kim, H., Choo, Y. J., Song, J., Lee, J. S., Lee, K. C., & Cho, J. C. (2007). Marinobacterium litorale sp. nov. in the order Oceanospirillales. *International Journal of Systematic and Evolutionary Microbiology*, 57(7), 1659–1662. <https://doi.org/10.1099/ijs.0.64892-0>

- Kim, H., Oh, H. M., Yang, S. J., Lee, J. S., Hong, J. S., & Cho, J. C. (2009). *Marinobacterium marisflavi* sp. nov., isolated from a coastal seawater. *Current Microbiology*, *58*(5), 511–515. <https://doi.org/10.1007/s00284-009-9355-5>
- King, A. J., Freeman, K. R., McCormick, K. F., Lynch, R. C., Lozupone, C., Knight, R., & Schmidt, S. K. (2010). Biogeography and habitat modelling of high-alpine bacteria. *Nature Communications*, *1*(5). <https://doi.org/10.1038/ncomms1055>
- Kumar, C. G., Joo, H. S., Koo, Y. M., Paik, S. R., & Chang, C. S. (2004). Thermostable alkaline protease from a novel marine haloalkalophilic *Bacillus clausii* isolate. *World Journal of Microbiology and Biotechnology*. <https://doi.org/10.1023/B:WIBI.00000333057.28828.a7>
- Levican, A., Collado, L., & Figueras, M. J. (2013). *Arcobacter cloacae* sp. nov. and *Arcobacter suis* sp. nov., two new species isolated from food and sewage. *Systematic and Applied Microbiology*. <https://doi.org/10.1016/j.syapm.2012.11.003>
- Loch, T. P., & Faisal, M. (2015). Emerging flavobacterial infections in fish: A review. *Journal of Advanced Research*, *6*(3), 283–300. <https://doi.org/10.1016/j.jare.2014.10.009>
- López, A. C., Minnaard, J., Pérez, P. F., & Alippi, A. M. (2015). A case of intoxication due to a highly cytotoxic *Bacillus cereus* strain isolated from cooked chicken. *Food Microbiology*, *46*, 195–199. <https://doi.org/10.1016/j.fm.2014.08.005>
- Mächler, E., Deiner, K., Steinmann, P., & Altermatt, F. (2014). Utility of Environmental DNA for Monitoring Rare and Indicator Macroinvertebrate Species. *Freshwater Science*. <https://doi.org/10.1086/678128>
- Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Øvreås, L., Reysenbach, A. L., Smith, V. H., & Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. In *Nature Reviews Microbiology*. <https://doi.org/10.1038/nrmicro1341>
- Matu, A., Lum Nde, A., Oosthuizen, L., Hitzeroth, A., Badenhorst, M., Duba, L., Gidaga, M., Klinck, J., Kriek, I.-M., Lekoma, P. J., Nel, L., dos Ramos, S. M., Rossouw, J., Salomane, N., Segone, N., Serobe, S., Tiyani, T., Hugo, C. J., & Newman, J. D. (2019). Draft Genome Sequences of Seven *Chryseobacterium* Type Strains. *Microbiology Resource Announcements*. <https://doi.org/10.1128/mra.01518-18>
- McClung, C. R., Patriquin, D. G., & Davis, R. E. (1983). *Campylobacter nitrofigilis* sp. nov., a

- Nitrogen-Fixing Bacterium Associated with Roots of *Spartina alterniflora* Loisel. *International Journal of Systematic Bacteriology*, 33(3), 605–612. <https://doi.org/10.1099/00207713-33-3-605>
- McNyset, K. M. (2005). Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. In *Ecology of Freshwater Fish*. <https://doi.org/10.1111/j.1600-0633.2005.00101.x>
- Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter - Merow - 2013 - *Ecography* - Wiley Online Library. *Ecography*.
- Moffett, M. L., & Wood, B. A. (1984). Survival of *Corynebacterium michiganense* subsp. *michiganense* within host debris in soil. *Australasian Plant Pathology*. <https://doi.org/10.1071/APP9840001>
- Nordhausen, K. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition by Trevor Hastie, Robert Tibshirani, Jerome Friedman. *International Statistical Review*, 77(3), 482–482. https://doi.org/10.1111/j.1751-5823.2009.00095_18.x
- NT, E., BP, O., CR, T., MA, R., Li, Y., CL, J., AR, M., ME, P., GA, L., & DM, L. (2015). Quantification of mesocosm fish and amphibian species diversity via {eDNA} metabarcoding. *Molecular Ecology Resources*.
- Pala, S., & Radhakrishnan, U. (2014). Genomic diversity of *Riemerella anatipestifer* associated with outbreaks of New Duck disease in India. *Indian Journal of Animal Sciences*.
- Papke, R. T., & Ward, D. M. (2004). The importance of physical isolation to microbial diversification. In *FEMS Microbiology Ecology*. <https://doi.org/10.1016/j.femsec.2004.03.013>
- Pennekamp, F., Pontarp, M., Tabi, A., Altermatt, F., Alther, R., Choffat, Y., Fronhofer, E. A., Ganesanandamoorthy, P., Garnier, A., Griffiths, J. I., Greene, S., Horgan, K., Massie, T. M., Mächler, E., Palamara, G. M., Seymour, M., & Petchey, O. L. (2018). Biodiversity increases and decreases ecosystem stability. *Nature*. <https://doi.org/10.1038/s41586-018-0627-8>
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., ... Wegmann, M. (2013). Essential biodiversity variables. In *Science*. <https://doi.org/10.1126/science.1229931>
- Peterson, A. T., & Anamza, T. (2015). Ecological niches and present and historical geographic distributions of species: A 15-year review of frameworks, results, pitfalls, and promises. *Folia*

Zoologica. <https://doi.org/10.25225/fozo.v64.i3.a3.2015>

- Peterson, A. T., Papeş, M., & Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*. <https://doi.org/10.1016/j.ecolmodel.2007.11.008>
- Pettorelli, N. (2015). Agree on biodiversity metrics to track from space. *Nature*. <https://doi.org/10.1038/523403a>
- PF, T., PR, M., EE, S., SW, K., OA, J., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS One*.
- Pfenninger, M., Grossart, H., Fisheries, I., Taberlet, P., & Vellend, M. (2018). *Environmental DNA time series in ecology*. October. <https://doi.org/10.1016/j.tree.2018.09.003>
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Pigolotti, S., López, C., Hernández-García, E., & Andersen, K. H. (2010). How Gaussian competition leads to lumpy or uniform species distributions. *Theoretical Ecology*, 3(2), 89–96. <https://doi.org/10.1007/s12080-009-0056-2>
- Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H., & Dejean, T. (2018). Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*. <https://doi.org/10.1038/s41598-018-28424-8>
- Ramette, A., & Tiedje, J. M. (2007). Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0610671104>
- Robertson, G. P., & Freckman, D. W. (1995). The spatial distribution of nematode trophic groups across a cultivated ecosystem. *Ecology*, 76(5), 1425–1432. <https://doi.org/10.2307/1938145>
- Rodrigues, D. F., Da C Jesus, E., Ayala-Del-Río, H. L., Pellizari, V. H., Gilichinsky, D., Sepulveda-Torres, L., & Tiedje, J. M. (2009). Biogeography of two cold-adapted genera: Psychrobacter and Exiguobacterium. *ISME Journal*. <https://doi.org/10.1038/ismej.2009.25>
- Rojas-Herrera, R., Narváez-Zapata, J., Zamudio-Maya, M., & Mena-Martínez, M. E. (2008). A simple silica-based method for metagenomic DNA extraction from soil and sediments. *Molecular Biotechnology*. <https://doi.org/10.1007/s12033-008-9061-8>

- Rojas, C. (2011). Notas en torno a la protección de los sitios arqueológicos y paleontológicos en cenotes y cuevas sumergidas en Quintana Roo. *Foro de Investigación Científica, de Desarrollo Tecnológico y Exposición Cultural Del "Sistema Hidrológico de La Cuenca Península de Yucatán*, 32.
- Rozzi, R., Massardo, F., Mansilla, A., Anderson, C. B., Berghöfer, A., Mansilla, M., Gallardo, M. R., Plana, J., Berghöfer, U., Arango, X., Russell, S., Araya, P., & Barros, E. (2007). La Reserva de Biosfera Cabo de Hornos: un desafío para la conservación de la biodiversidad e implementación del desarrollo sustentable en el extremo austral de América. *Anales Instituto Patagonia (Chile)*.
- Ruusunen, M., Salonen, M., Pulkkinen, H., Huuskonen, M., Hellström, S., Revez, J., Hänninen, M. L., Fredriksson-Ahomaa, M., & Lindström, M. (2013). Pathogenic bacteria in finnish bulk tank milk. *Foodborne Pathogens and Disease*. <https://doi.org/10.1089/fpd.2012.1284>
- Ryll, M., Christensen, H., Bisgaard, M., Christensen, J.-P., Hinz, K.-H., & Köhler, B. (2008). Studies on the Prevalence of *Riemerella anatipestifer* in the Upper Respiratory Tract of Clinically Healthy Ducklings and Characterization of Untypable Strains. *Journal of Veterinary Medicine, Series B*. <https://doi.org/10.1111/j.1439-0450.2001.00471.x>
- Saravanakumar, K., Rajendran, N., Kathiresan, K., & Chen, J. (2016). Bioprospects of Microbial Enzymes from Mangrove-Associated Fungi and Bacteria. In *Advances in Food and Nutrition Research* (1st ed., Vol. 79). Elsevier Inc. <https://doi.org/10.1016/bs.afnr.2016.08.003>
- Sasi Jyothsna, T. S., Rahul, K., Ramaprasad, E. V. V., Sasikala, C., & Ramana, C. V. (2013). *Arcobacter anaerophilus* sp. nov., isolated from an estuarine sediment and emended description of the genus *Arcobacter*. *International Journal of Systematic and Evolutionary Microbiology*. <https://doi.org/10.1099/ijs.0.054155-0>
- Schmitter-Soto, J. J., Comín, F. A., Escobar-Briones, E., Herrera-Silveira, J., Alcocer, J., Suárez-Morales, E., Elías-Gutiérrez, M., Díaz-Arce, V., Marín, L. E., & Steinich, B. (2002). Hydrogeochemical and biological characteristics of cenotes in the Yucatan Peninsula (SE Mexico). *Hydrobiologia*. <https://doi.org/10.1023/A:1014923217206>
- Shogren, A. J., Tank, J. L., Andruszkiewicz, E., Olds, B., Mahon, A. R., Jerde, C. L., & Bolster, D. (2017). Controls on eDNA movement in streams: Transport, Retention, and Resuspension /704/158/2464 /704/242 /45/77 article. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-05223-1>
- Soberón, J., & Nakamura, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*.

<https://doi.org/10.1073/pnas.0901637106>

- Soberon, J., & Peterson, A. T. (2005). Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics*. <https://doi.org/10.17161/bi.v2i0.4>
- Spens J, Evans AR, Halfmaerten D, Knudsen SW, Sengupta ME, Mak SST, Sigsgaard EE, & Hellstrom M. (2016). Comparison of capture and storage methods for aqueous microbial {eDNA} using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*.
- Stamova, L., & Sotirova, V. (1987). Reaction of different crops to artificial inoculation with *Corynebacterium michiganense* (E. F. Sm.) H. L. Jensen. *Archives Of Phytopathology And Plant Protection*, 23(3), 211–216. <https://doi.org/10.1080/03235408709438079>
- Stenfors Arnesen, L. P., Fagerlund, A., & Granum, P. E. (2008). From soil to gut: *Bacillus cereus* and its food poisoning toxins. In *FEMS Microbiology Reviews*. <https://doi.org/10.1111/j.1574-6976.2008.00112.x>
- Stingl, U., Tripp, H. J., & Giovannoni, S. J. (2007). Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME Journal*, 1(4), 361–371. <https://doi.org/10.1038/ismej.2007.49>
- Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*. <https://doi.org/10.1111/j.1365-294X.2011.05418.x>
- Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. In *Biological Conservation*. <https://doi.org/10.1016/j.biocon.2014.11.019>
- Travisano, M., & Rainey, P. B. (2000). Studies of adaptive radiation using model microbial systems. *American Naturalist*. <https://doi.org/10.2307/3079225>
- Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of Western Amazonian forests. *Science*. <https://doi.org/10.1126/science.1078037>
- Van Der Gast, C. J., Lilley, A. K., Ager, D., & Thompson, I. P. (2005). Island size and bacterial diversity in an archipelago of engineering machines. *Environmental Microbiology*. <https://doi.org/10.1111/j.1462-2920.2005.00802.x>

- Vaneechoutte, M., Kämpfer, P., De Baere, T., Avesani, V., Janssens, M., & Wauters, G. (2007). *Chryseobacterium hominis* sp. nov., to accommodate clinical isolates biochemically similar to CDC groups II-h and II-c. *International Journal of Systematic and Evolutionary Microbiology*, 57(11), 2623–2628. <https://doi.org/10.1099/ijs.0.65158-0>
- Vilà, B., Fontgibell, A., Badiola, I., Esteve-Garcia, E., Jiménez, G., Castillo, M., & Brufau, J. (2009). Reduction of *Salmonella enterica* var. Enteritidis colonization and invasion by *Bacillus cereus* var. *toyoi* inclusion in poultry feeds. *Poultry Science*, 88(5), 975–979. <https://doi.org/10.3382/ps.2008-00483>
- Voordouw, G., Armstrong, S. M., Reimer, M. F., Fouts, B., Telang, A. J., Shen, Y., & Gevertz, D. (1996). Characterization of 16s rRNA genes from oil field microbial communities indicates the presence of a variety of sulfate-reducing, fermentative, and sulfide-oxidizing bacteria. *Applied and Environmental Microbiology*. <https://doi.org/10.1128/aem.62.5.1623-1629.1996>
- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*. <https://doi.org/10.1038/nature09440>
- Whitaker, R. J., Grogan, D. W., & Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science*. <https://doi.org/10.1126/science.1086909>
- Whiteduck-Léveillé, K., Whiteduck-Léveillé, J., Cloutier, M., Tambong, J. T., Xu, R., Topp, E., Arts, M. T., Chao, J., Adam, Z., Lévesque, C. A., Lapen, D. R., Villemur, R., Talbot, G., & Khan, I. U. H. (2015). *Arcobacter lanthieri* sp. Nov., isolated from pig and dairy cattle manure. *International Journal of Systematic and Evolutionary Microbiology*, 65(8), 2709–2716. <https://doi.org/10.1099/ijs.0.000318>
- Wiley, E. O., McNyset, K. M., Peterson, A. T., Robins, C. R., & Stewart, A. M. (2003). Niche modeling and geographic range predictions in the marine environment using a machine-learning Algorithm. *Oceanography*. <https://doi.org/10.5670/oceanog.2003.42>
- WWF. (2018). Living Planet Report - 2018: Aiming higher. In *Environmental Conservation*. <https://doi.org/10.1080/09528820802312343>
- Yoshimoto, S., & Itaya, K. (2013). Adsorption and assembly of ions and organic molecules at electrochemical interfaces: Nanoscale aspects. *Annual Review of Analytical Chemistry*. <https://doi.org/10.1146/annurev-anchem-062012-092559>
- Zhang, L., An, R., Wang, J., Sun, N., Zhang, S., Hu, J., & Kuai, J. (2005). Exploring novel bioactive

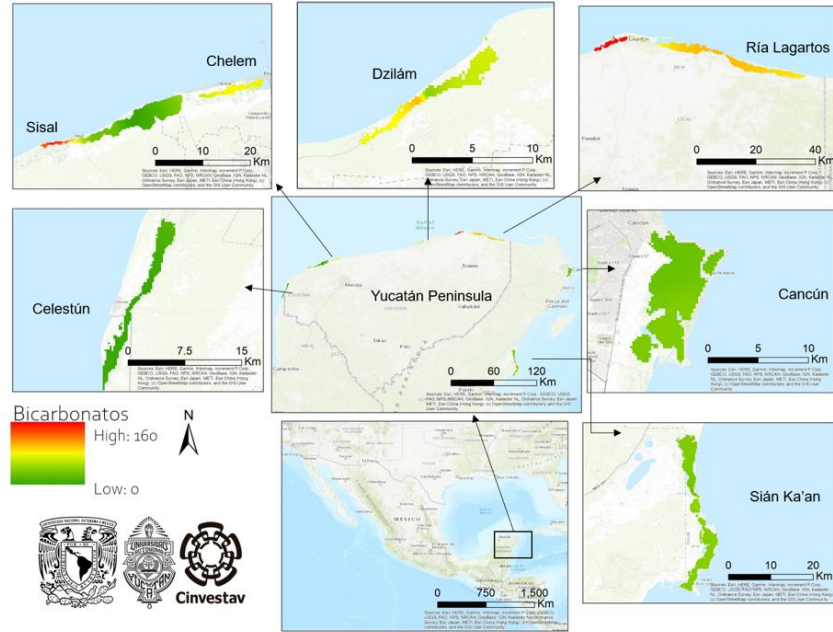
compounds from marine microbes. In *Current Opinion in Microbiology*.

<https://doi.org/10.1016/j.mib.2005.04.008>

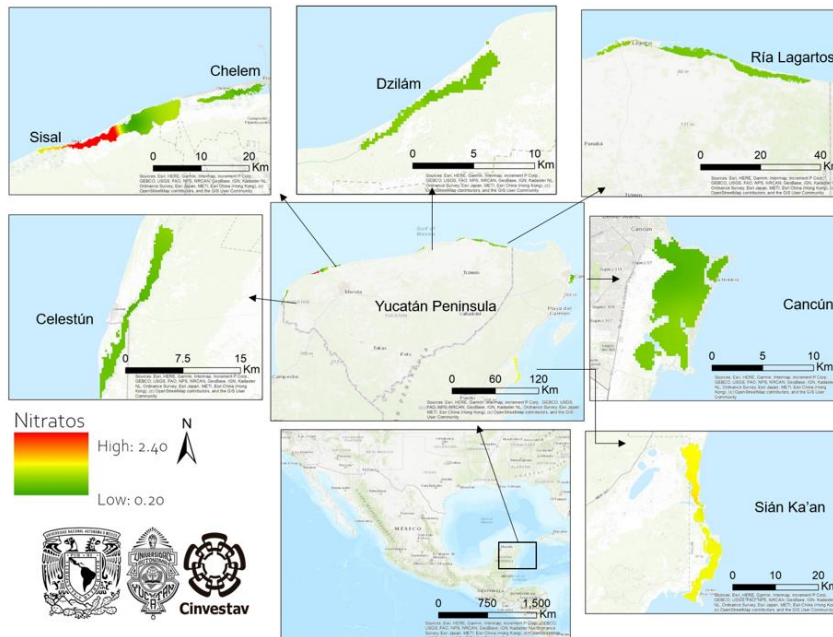
Zimmermann, N. E., Jr, T. C. E., Graham, C. H., Pearman, P. B., & Svenning, J. (2010). *New trends in species distribution modelling*. 985–989. <https://doi.org/10.1111/j.1600-0587.2010.06953.x>

Anexos.

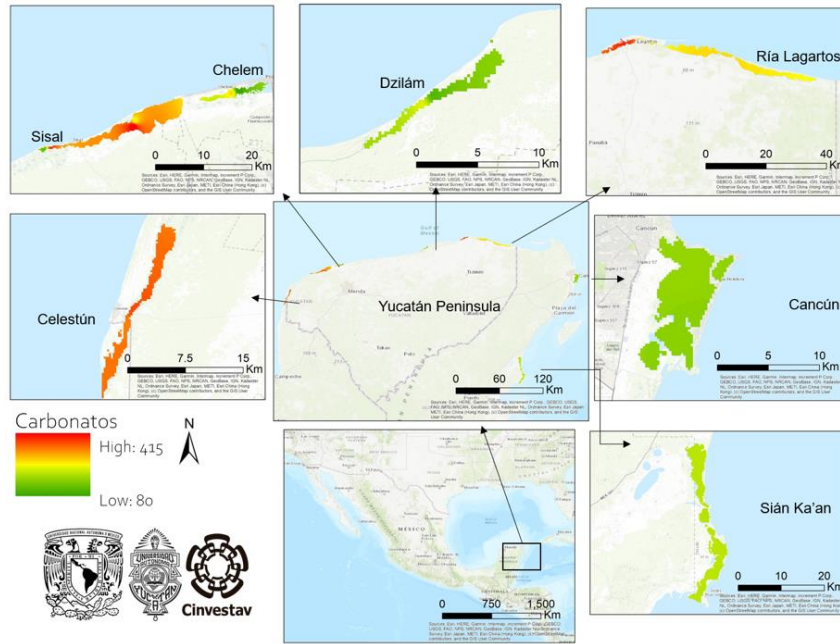
Mapas de interpolación de variables ambientales.



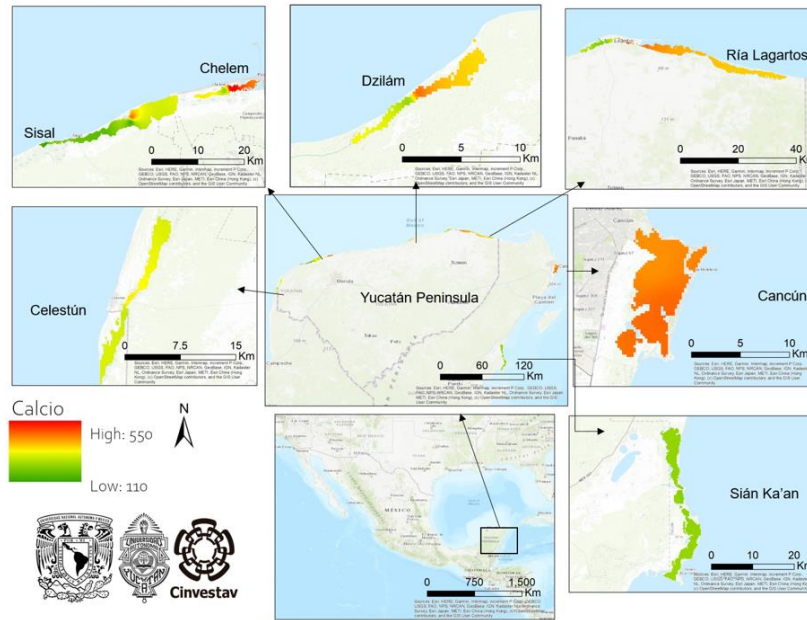
Mapa de interpolación de mediciones de bicarbonatos (mg/l).



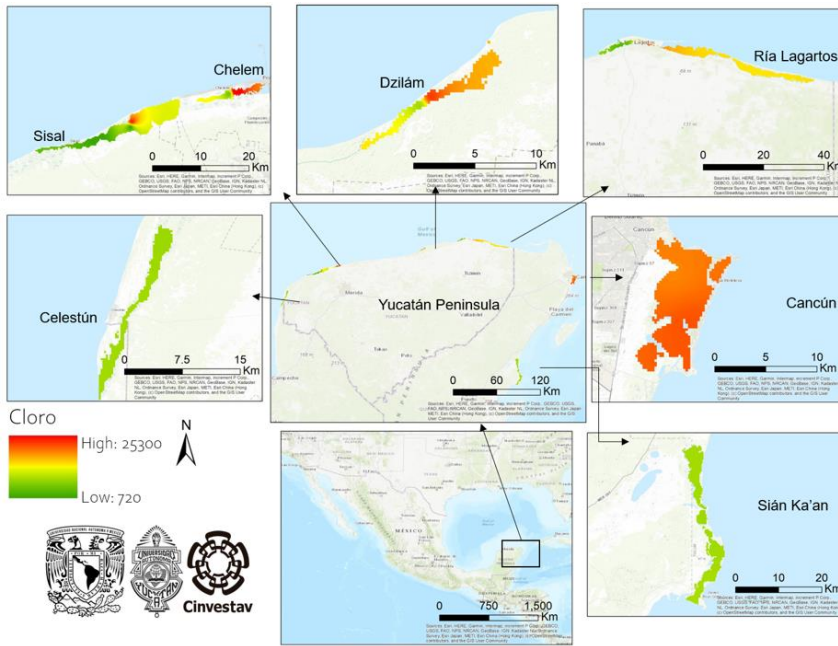
Mapa de interpolación de mediciones de nitratos (mg/l)



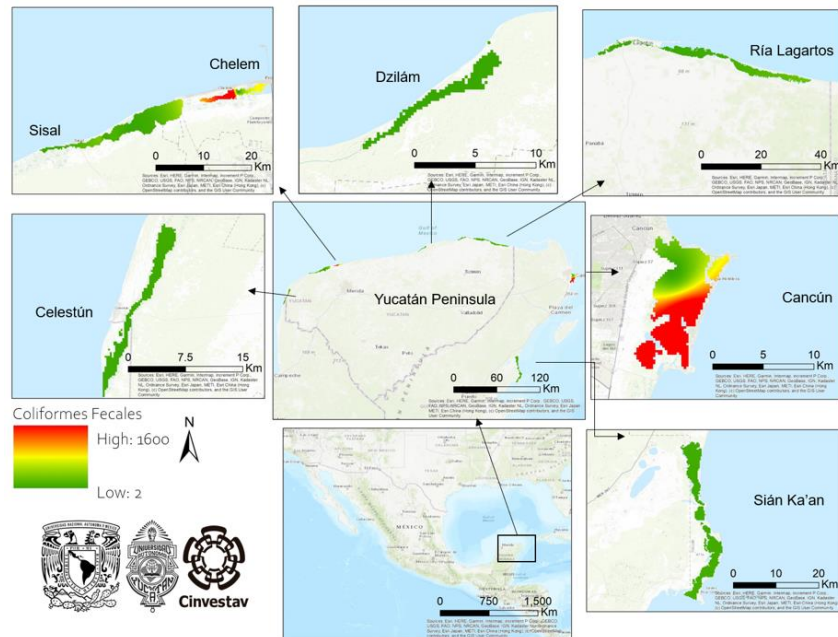
Mapa de interpolación de mediciones de carbonatos (mg/l).



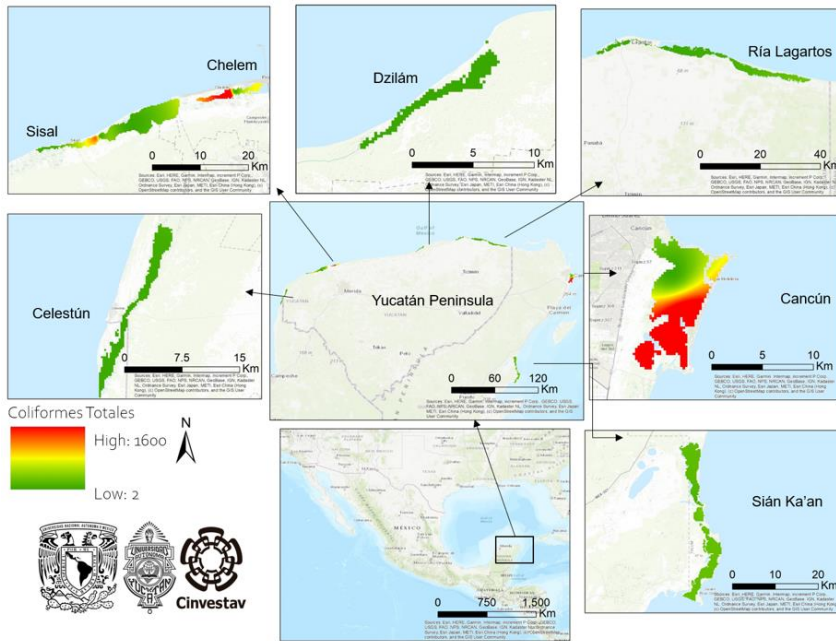
Mapa de interpolación de mediciones de calcio (mg/l).



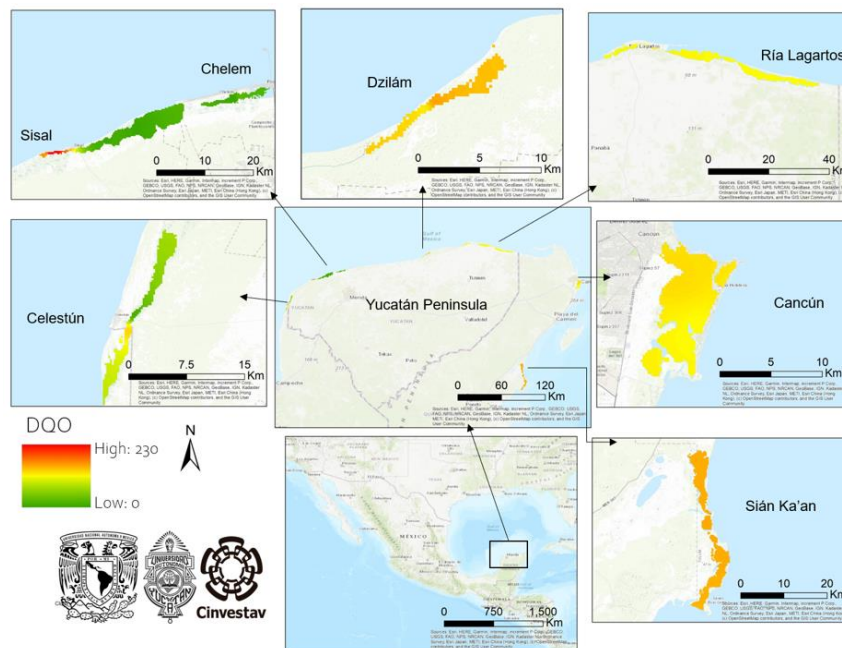
Mapa de interpolación de mediciones de cloro (mg/l).



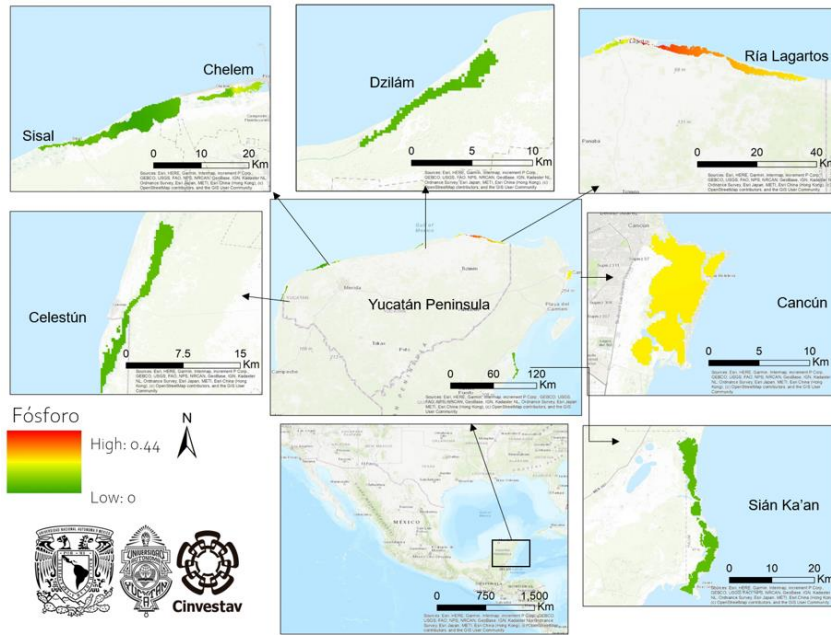
Mapa de interpolación de mediciones de Coliformes Fecales.



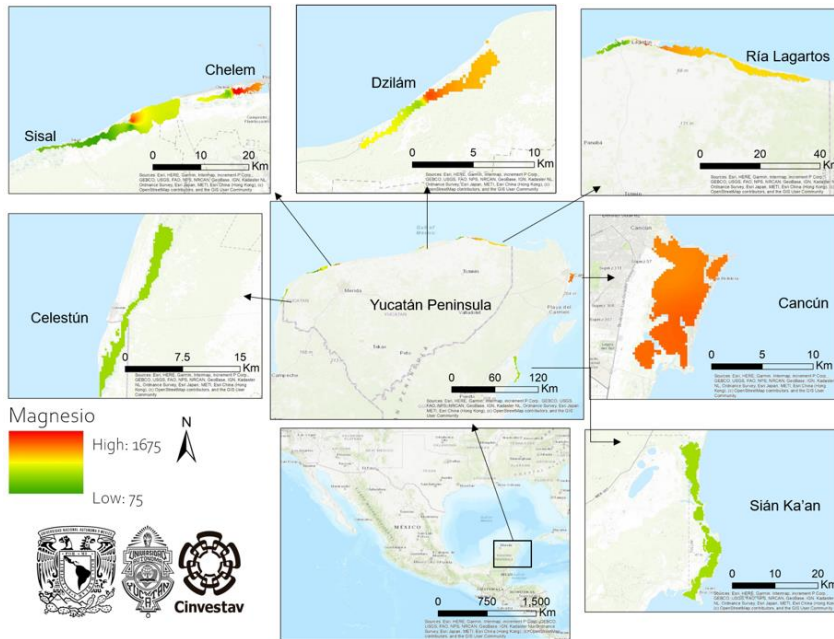
Mapa de interpolación de mediciones de Coliformes Totales.



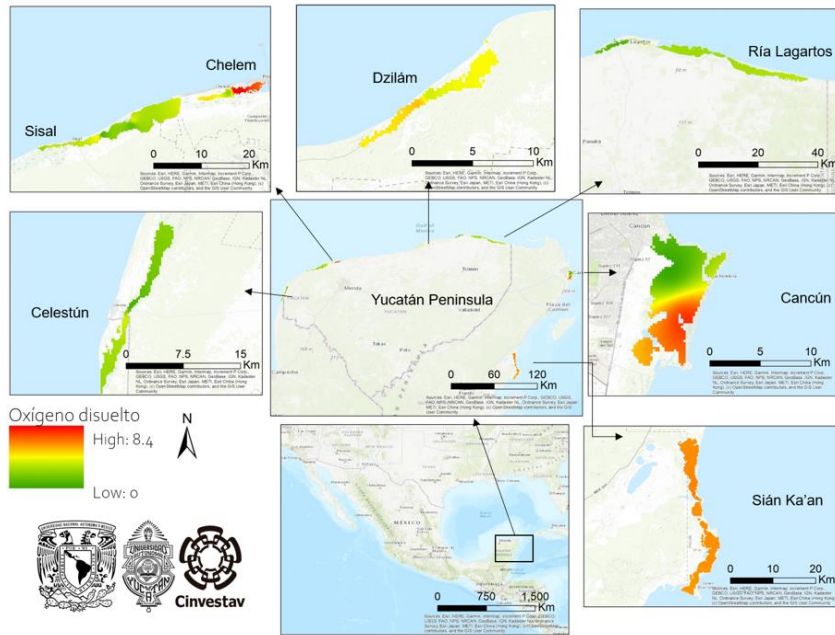
Mapa de interpolación de mediciones de demanda bioquímica de oxígeno (mg/l).



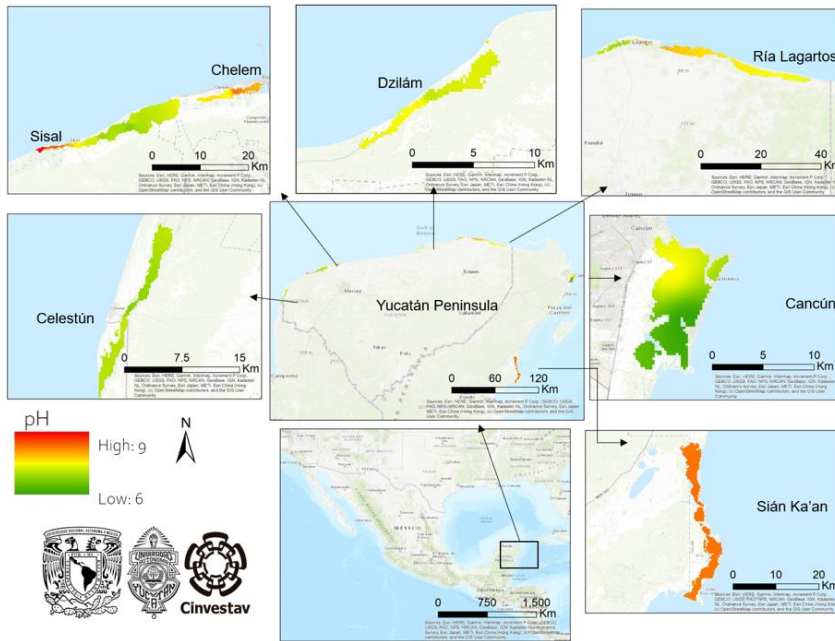
Mapa de interpolación de mediciones de fósforo (mg/l).



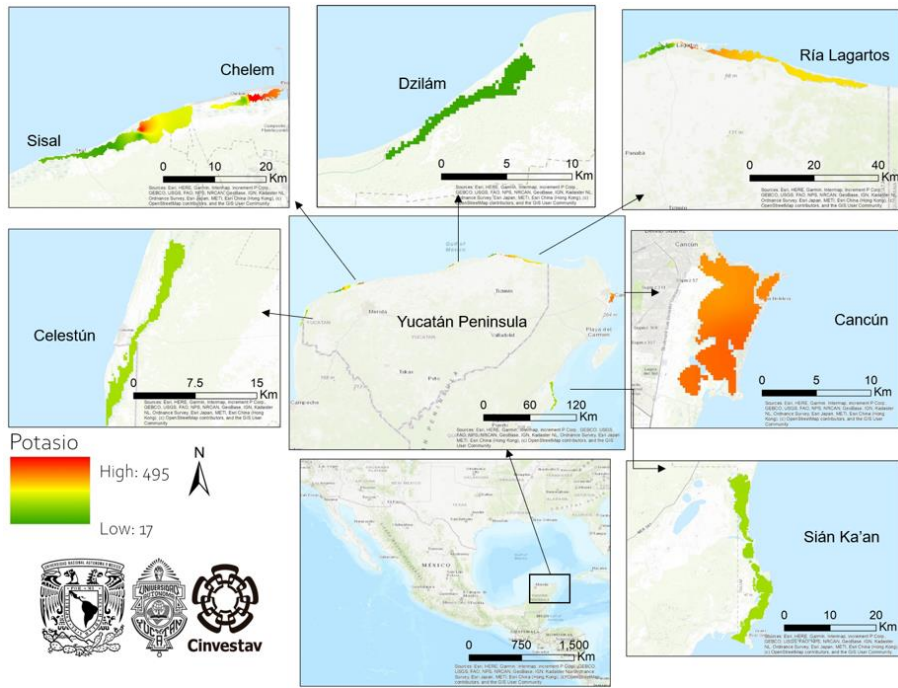
Mapa de interpolación de mediciones de magnesio (mg/l).



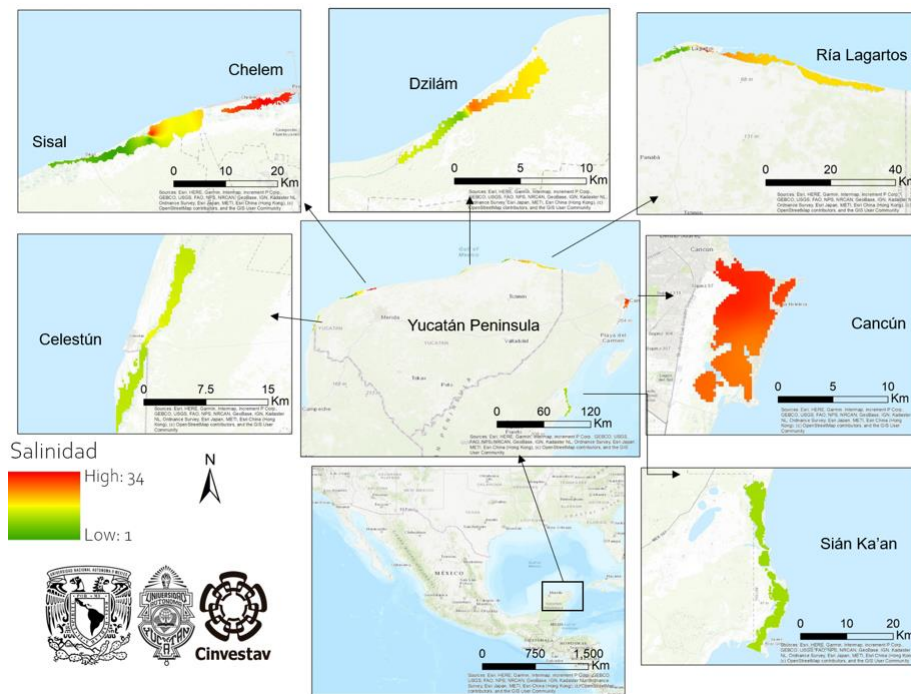
Mapa de interpolación de mediciones de oxígeno disuelto (mg/l).



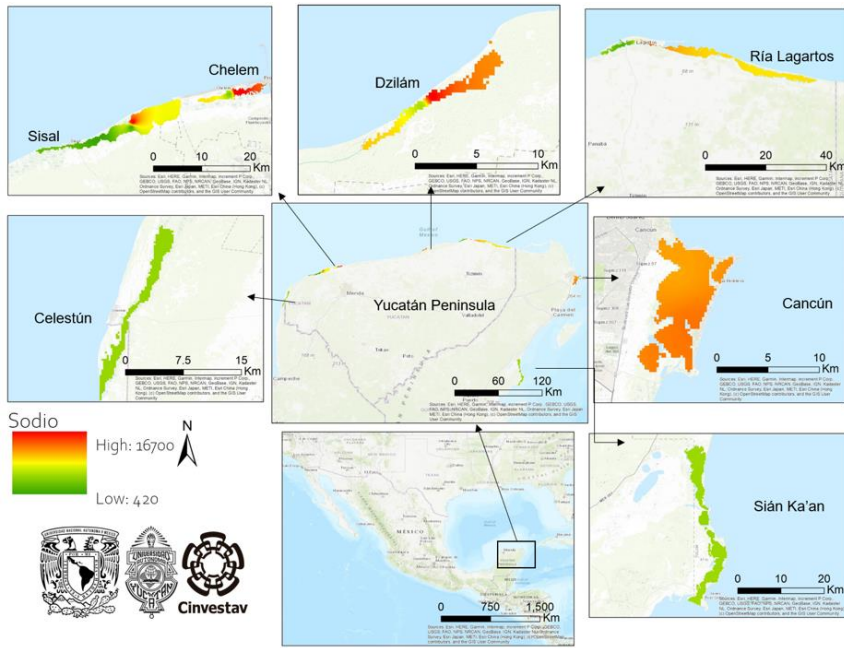
Mapa de interpolación de mediciones de pH.



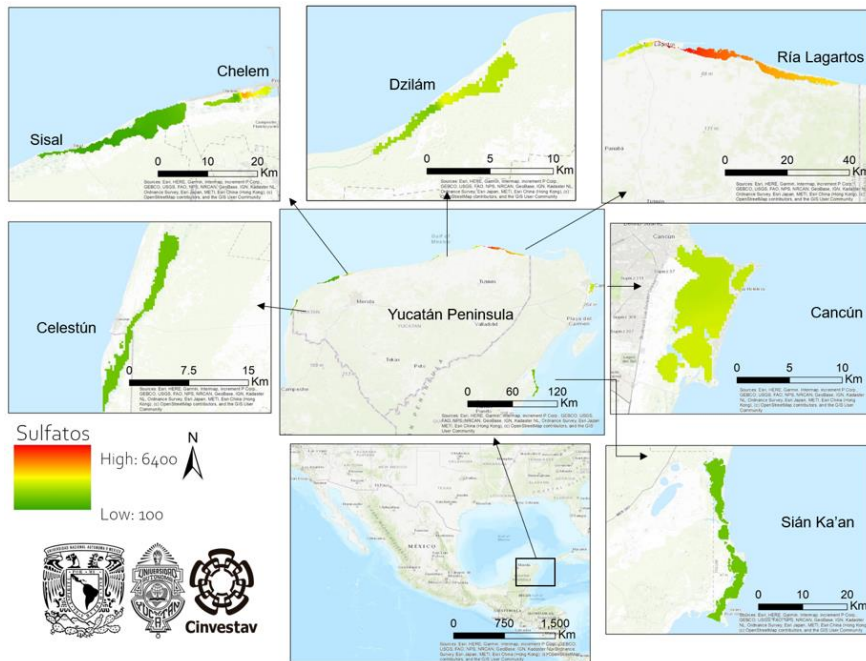
Mapa de interpolación de mediciones de potasio (mg/l).



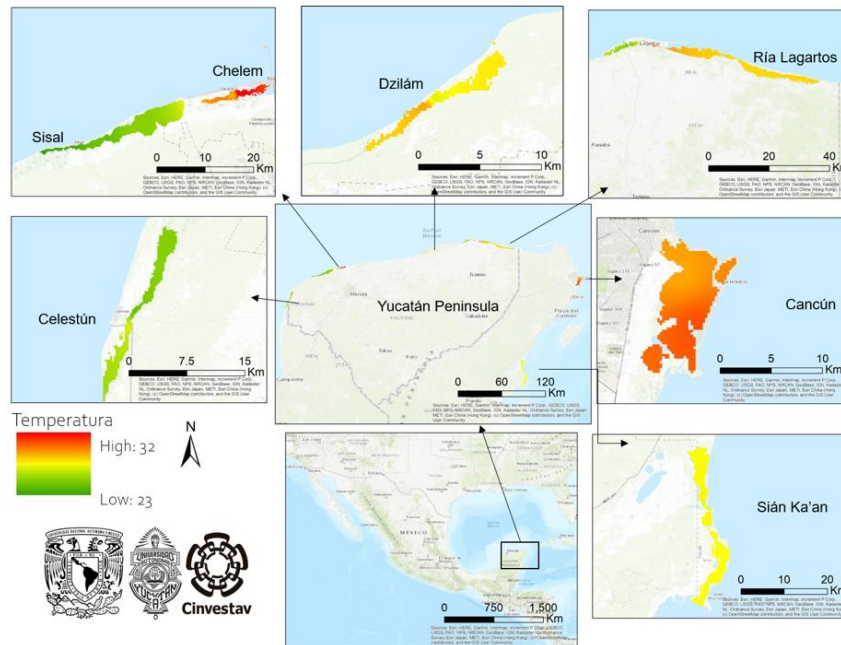
Mapa de interpolación de mediciones de salinidad (ppm).



Mapa de interpolación de mediciones de sodio (mg)



Mapa de interpolación de mediciones de sulfatos (mg/l).



Mapa de interpolación de temperatura °C.

Algoritmos

Bioclim (BIO):

Es un modelo llamado de “envoltura” que utiliza solo datos de ocurrencia para definir un ambiente espacial multidimensional en el que las especies pueden existir. Este espacio ambiental es construido en forma de caja con los valores mínimos y máximos de las variables ambientales de todas las ocurrencias, resultando en una envoltura rectilínea multidimensional (Booth et al., 2014). Para predecir la probabilidad de ocurrencia de las especies en cualquier ubicación, Bioclim compara los valores de las variables ambientales en esa ubicación específica con el percentil de distribución de valores de ubicaciones conocidas. El percentil 50 se refiere a la media, que divide los datos exactamente a la mitad. Entre más cerca esté el valor de la variable ambiental de la ubicación desconocida al percentil 50, más adecuada es la locación para que la especie ocurra, por ende, mayor probabilidad de ocurrencia. Por ello la probabilidad es 1 en el percentil 50, y por ello las colas de la distribución no se distinguen, el percentil 10 es tratado igual que al 90, ambos teniendo el mismo valor de probabilidad. Bioclim combina los puntajes de cada variable ambiental en una probabilidad general de ocurrencia para cada ubicación con pesos iguales para todas las variables ambientales (Araújo & Peterson, 2012).

Bioclim fue el primer paquete de modelado de distribución de especies que vinculaba ocurrencias espacialmente explícitas con mapas de variables ambientales. Se desarrolló en Australia bajo el liderazgo de Henry Nix. El algoritmo es ampliamente utilizado, porque es fácil de entender, pero se

reconoce de forma general que su desempeño no es tan bueno como el de otros métodos (Hijmans & Elith, 2016).

Ventajas: Simple e intuitivo, Solo requiere presencias, Provee rangos de predictores ambientales, útil en enseñanza.

Limitaciones: Susceptible a sobrepredecir, no toma en cuenta las interacciones entre predictores, no puede usar variables categóricas, no genera predicciones cuantitativas o niveles de confianza.

Asunciones: Las ocurrencias están influenciadas por el clima a la escala de las variables climáticas, dichas variables tienen distribución normal.

Domain (DOM).

A diferencia de otras paqueterías de mapeo potencial, DOMAIN puede trabajar eficientemente con pocos puntos de distribución conocida. Estos son correlacionados con los valores de otros valores espacialmente referenciados, variables ambientales tales como la elevación, suelo, vegetación y clima, que se considere influyen la distribución de la especie. Se crea una envoltura definida por los valores mínimos y máximos de las variables ambientales para todas las ocurrencias. Cualquier locación dentro de las condiciones ambientales se incluye en el rango potencial de la especie (Carpenter et al., 1993).

Ventajas: Simple e intuitivo, solo necesita presencias, útil para enseñanza.

Limitaciones: No toma en cuenta interacciones entre predictores, no utiliza variables categóricas, no genera predicciones cuantitativas o niveles de confianza.

Gaussian Process (GAU).

También llamados Gaussian random fields, proveen un método flexible para ajustar modelos estadísticos complejos. Comúnmente ajustados vía inferencia Bayesiana, típicamente requieren el uso de métodos Markov chain Monte Carlo (MCMC).

GP permite al usuario incorporar el conocimiento ecológico en los modelos de distribución manipulando la función media y lengthscale hyperprior. La función media actúa sobre el modelo completo y puede ser usado para incorporar conocimiento específico de la respuesta de la especie a gradientes ambientales. El lengthscale hyperprior determina que tan diferentes son las lengthscales y cuáles pueden ser usadas para informar al modelo, que tan rápido la probabilidad de presencia cambia con diferentes valores de los predictores ambientales. En la ausencia de información prior, se utilizan priors no informativos (Pigolotti et al., 2010).

Maxent Default y Simple (MXD/MXS).

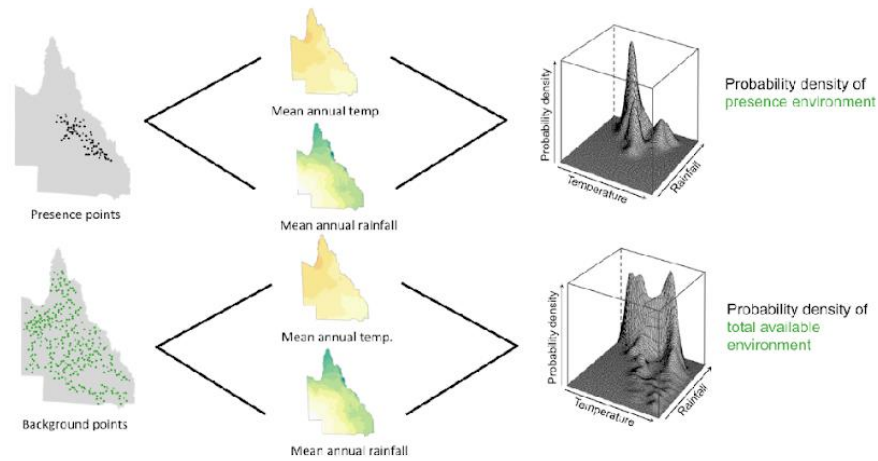
El nombre deriva de *máximum entropy modelling* (modelos de máxima entropía), predice las ocurrencias encontrando la distribución más extensa o cercana a lo uniforme, mientras contempla los límites de las variables ambientales en locaciones conocidas (Phillips et al., 2006). Utiliza solo datos de presencia y el algoritmo compara las ubicaciones donde la especie ha sido encontrada contra todos los ambientes disponibles en la región de estudio. Define estos ambientes muestreando una gran cantidad de puntos en el área de estudio, que son llamados background points. Dado que los background points pueden contener ubicaciones donde se sabe que está la especie, los background points no son los mismo que los puntos de pseudo-ausencia. Los background points definen el ambiente disponible (Merow et al., 2013).

El algoritmo de Maxent desarrollado para el modelado de distribución de especies es un método de machine learning, por lo que construye iterativamente múltiples modelos. Tiene dos componentes principales:

- a) Entropía: El modelo es calibrado para encontrar la distribución más extensa, o más cercana a lo uniforme mediante un estudio de la región.
- b) Constricciones: Las reglas que constriñen la distribución predicha. Estas reglas están basadas en los valores de las variables ambientales (llamadas features) de las locaciones donde las especies han sido observadas.

Maxent considera seis tipos de features, cada una de ellas permite diferentes formas en las curvas de respuesta y tiene diferentes implicaciones en las constricciones. Como default Maxent utiliza todos los tipos de features (Linear, Quadratic, Product, Treshold, Hinge, Categorical).

Para calcular el potencial de distribución de una especie, Maxent primero calcula dos densidades probabilísticas. Para todos los puntos la densidad probabilística describe la probabilidad relativa de todas las variables ambientales en el modelo sobre el rango de esos puntos. Por ejemplo, en la siguiente imagen los valores de temperatura y lluvia bajo el pico de las gráficas en la parte derecha son valores más comunes entre todos los valores de presencia. De forma similar, una densidad de probabilística es calculada entre la región de estudio basada en los background points. Así la densidad probabilística de los background points caracteriza el ambiente disponible dentro de la región de estudio, mientras que la densidad de los puntos de presencia caracteriza el ambiente donde la especie ha sido encontrada. Maxent calcula entonces la razón entre estas dos densidades, lo que determina la idoneidad ambiental para las presencias de una especie en cada punto del área de estudio (Guillera-Arroita et al., 2014).



Representación gráfica del funcionamiento estadístico de Maxent (*Jane Elith et al., 2011*).

Maxent elige la distribución que maximiza la similitud entre las características ambientales de la totalidad de ambientes y aquellas locaciones donde la especie está presente. Esto se conoce como “raw output”. Para facilitar la interpretación de los resultados y para proveer un estimado de la probabilidad que tiene una especie de estar presente en una ubicación, Maxent crea una transformación logística del raw output. El output logístico toma en cuenta la prevalencia de la especie, lo que se refiere a la proporción de locaciones ocupadas. Maxent usa la prevalencia de 0.5, lo que implica que la especie está presente en la mitad de todas las ubicaciones disponibles.

La diferencia entre la versión default radica en usar todas las features, basado en la paquetería MaxNet, mientras que la versión simple utiliza solo features cuadráticas y lineales. Un aspecto importante de Maxent es la regularización, que reduce el sobreajuste del modelo. Se puede llevar a cabo de dos formas:

- a) Relajar las constricciones: En vez de ajustar el modelo usando constricciones exactas (promedio, varianzas, etc.) de las variables ambientales, toma en cuenta intervalos de confianza alrededor de las constricciones. Esto previene el ajuste cercano sobre los datos con los que se crea el modelo.
- b) Penalizar la complejidad: El modelo excluye tipos de features que no añaden mejoras significativas al modelo.

Ventajas: Solo requiere presencias, puede usar variables continuas y categóricas, incluye un protocolo de regularización contra el sobreajuste, muestra buen desempeño predictivo en general.

Limitaciones: Difícil de comparar contra otros algoritmos, pues el resultado es idoneidad ambiental y no probabilidad de ocurrencia. Los resultados logísticos se basan en la asunción de prevalencia, no en la estimación.

Random Forest (RDF).

Son una extensión de los árboles de clasificación simples en donde árboles de decisión múltiples son construidos con subsets aleatorios de la información. Todos los subsets aleatorios tienen el mismo número de puntos y son seleccionados de la matriz completa. Los datos utilizados son colocados de nuevo en la matriz completa y pueden ser seleccionados en árboles subsecuentes (Breiman, 2001). En Random Forest, los subsets aleatorios son seleccionados en un proceso llamado “bagging”, en el que cada punto tiene la misma probabilidad de ser seleccionado para un nuevo subset aleatorio. Cerca de dos tercios de la información es incluida en cada subset aleatorio. El otro tercio no es usado para la construcción de los árboles, esta parte de los datos es llamada “out of the bag” y es utilizada para evaluar el modelo (Franklin & Miller, 2014).

Ventajas: Uno de los algoritmos más precisos disponibles, puede manejar muchas variables predictoras, provee estimados de la importancia de cada variable, mantiene la exactitud aún con cuando grandes proporciones de datos faltantes.

Limitaciones: Puede sobreajustar datos con ruido, para datos con predictores categóricos con diferente número de niveles, random forests da peso a aquellos con más niveles.

Support Vector Machine (SVM).

En machine learning, support-vector machines (support-vector networks) son algoritmos de aprendizaje supervisado con algoritmos de aprendizaje asociados que analizan la información usando una clasificación y un análisis de regresión (Drake et al., 2006). Dado un set de ejemplos de entrenamiento, cada uno se clasifica en una u otra categoría.

En la calibración del modelo se genera un modelo previo que asigna nuevos ejemplos a una u otra categoría, haciéndolo un clasificador binario lineal no-probabilístico. Un modelo SVM es una representación de ejemplo de puntos en el espacio, mapeados de forma que las categorías son divididas por un espacio lo más amplio posible. Los nuevos ejemplos son entonces mapeados en el mismo espacio y se predice a que categoría pertenecen.

En adición para desarrollar una clasificación lineal, SVM's pueden desempeñar eficientemente una clasificación no-lineal usando lo que se llama “kernel trick”, implícitamente mapeando los inputs en features espaciales de alta dimensionalidad.

Ventajas: Con el número suficiente de ocurrencias el modelo genera resultados muy certeros.

Limitaciones: Complejidad, difícil de explicar la contribución de variables, tiene tendencia a sobre ajustar.

Descripción de especies.

Arcobacter lekithochrous.

El género *Arcobacter* pertenece a la familia Campylobacteraceae (McClung et al., 1983). Muestran una distribución global considerando su aislamiento en diferentes regiones geográficas (Levican et al., 2013; Sasi Jyothsna et al., 2013; Whiteduck-Léveillé et al., 2015). De la misma manera el género muestra una gran diversidad de nichos ecológicos: ambientes marinos, aguas residuales, agua para consumo humano, heces fecales, en comunidades de plantas, entre otros (Donachie et al., 2005; Jacob et al., 1998; Voordouw et al., 1996). Algunas especies de *Arcobacter* pueden causar diarrea o bacteremia en humanos y abortos o mastitis en animales, aunque también han sido aislados de animales y seres humanos sanos (Diéguez et al., 2017).

Aunque la ruta zoonótica no es clara, se ha postulado el consumo de comida o agua contaminadas como el medio principal (Collado & Figueras, 2011). Se asume que la prevalencia ha sido subestimada debido a lo complejo que resulta aislar y cultivar estas bacterias (Ho et al., 2006).

Dichas características y supuestos han desencadenado diversos estudios sobre la prevalencia de *Arcobacter* en diferentes ambientes y animales filtradores, tales como moluscos, bajo el supuesto de ser reservorios de estas bacterias debido a sus hábitos (Figueras et al., 2011).

Bacillus cereus.

Compite con organismos como *Salmonella* y *Campylobacter* en el tracto gastrointestinal, su presencia reduce el número de estos microorganismos. En ganado se utilizan algunas cepas inofensivas como probióticos aditivos para reducir la abundancia de *Salmonella* en intestinos, mejorando el crecimiento de los animales (Vilà et al., 2009).

Es la causa común de brotes infecciosos por comida contaminada, en un estudio de 10 años en Estados Unidos, *B. cereus* se ha posicionado entre las cinco bacterias más comunes que causan brotes en alimentos (S. D. Bennett et al., 2013). Los brotes de *B. cereus* comúnmente están relacionados con el consumo de arroz, pasta, carne y lácteos (López et al., 2015; Ruusunen et al., 2013).

Existen dos tipos de afecciones relacionadas a *B. cereus*: intoxicación por cereulida (toxina) preformada o infección seguida de una producción intestinal de enterotoxinas como Nhe (Enterotoxina no hemolítica) y Hbl (hemolisina BL) (Stenfors Arnesen et al., 2008).

Pelagibacter ubique.

Conocida principalmente por ser una de las células de vida libre autoreplicable más pequeñas y simples. Es parte del clado SAR11 pequeñas alfa-proteobacterias que componen cerca del 25% de las células microbianas planctónicas y cerca del 50% durante el verano gracias al aumento de la temperatura, *P. ubique* tiene un ecotipo costero-océano superficial (Giovannoni et al., 2005), adaptado a ambientes oligotróficos. Es el organismo vivo con el genoma más corto (1,308,759 pb), carece de copias de genes, genes virales y ADN basura (Stingl et al., 2007).

Se trata de un organismo eficiente dependiente de materia orgánica, consumiéndola en forma disuelta en el océano. La adición de nutrientes al océano no afecta sus tendencias de alimentación, alimentándose de carbono orgánico disuelto *P. ubique* produce los nutrientes necesarios para que las algas pelágicas se desarrollen.

Chryseobacterium hominis.

El género comprende a más de cien especies de diversos hábitats incluyendo agua dulce, suelos, peces marinos y hospederos humanos (Matu et al., 2019). *Chryseobacterium spp.* han sido aislados de suelos, raíces de plantas, flores y plantas en descomposición. Algunas especies asociadas a plantas son capaces de inhibir hongos patógenos que atacan a las plantas. *Chryseobacterium spp.* han sido aisladas en diversos cuerpos lacustres y su sedimento, sistemas de enfriamiento de agua, lodos de biorreactores, suelos contaminados, sedimento marino y permafrost (Loch & Faisal, 2015). De igual forma diversas especies se han asociado con animales, detectados en mosquitos, cucarachas, heces de milpiés, en copépodos, alas de aves, leche de vaca, carnes crudas y pollo. Común en mucosas de peces aparentemente saludables, aunque se le considera un organismo deteriorador (Loch & Faisal, 2015).

Aunque *C. hominis* es asociado a tortugas, anfibios y peces también se ha encontrado en casos aislados de bacteremias en pacientes inmunocomprometidos (Vanechoutte et al., 2007).

Clavibacter michiganensis.

Se trata de un microorganismo patógeno común en la rizósfera de importancia económica por su patogenicidad, se le considera un microorganismo de riesgo moderado en fitosanidad debido a su distribución global. Aunque el principal hospedero de importancia económica es el jitomate, la infección ha sido reportada en diversas solanáceas (Moffett & Wood, 1984; Stamova & Sotirova, 1987).

Entre los endófitos aislados de manglares se encuentra *Clavibacter*, su implementación está siendo explorada para coadyuvar al crecimiento y desarrollo de plántulas, por ahora con *Avicennia marina* (Gouda et al., 2016).

Los endófitos son un grupo de microorganismos endosimbióticos que colonizan plantas, se sabe que pueden mejorar el crecimiento y la ganancia de nutrientes. Se especula que pueden mejorar la habilidad de la planta para tolerar distintos tipos de estrés (biótico o abiótico) y mejorar la resistencia de las plantas frente a insectos o pestes (Joseph & Mini Priya, 2011).

Cribrihabitans marinus.

Pertenciente a la familia Rhodobacteraceae (Garrity & Cole, 2007), miembros caracterizados por ser metabólica y fisiológicamente diversos, ocupando un amplio rango de nichos en distintos hábitats, la mayoría originados en ambientes marinos, como agua salada, hielo marino, sedimento, tapetes microbianos, algas marinas, tejidos de animales marinos y biopelículas costeras (Hongyue Dang & Lovell, 2002). Se distinguen como colonizadores primarios de superficies bióticas y abióticas de la superficie de los océanos (Chen et al., 2014).

Marinobacterium marisflavi.

El género *Marinobacterium* (González et al., 1997; Kim et al., 2007) es parte del orden Oceanospirillales, el género comprende diecisiete especies. Todos los miembros del género han sido aislados de ambientes marinos, incluyendo aguas litorales, pelágicas, marismas y rizósfera costera (Kim et al., 2009). Entre el potencial biotecnológico del género se encuentra la degradación de benceno, siendo *M. aestuarii* la especie con la mejor actividad observada (Bae et al., 2018). Durante el año 2020 aún se están secuenciando los genomas de las distintas especies y sólo diez han sido publicados en NCBI.

Riemerella anatipestifer.

Es una bacteria causante de septicemia y muerte en patos jóvenes y gansos en todo el mundo. Existen más de veinte setoripos e infecciones distribuidas en las aves. Se ha encontrado también en pavos, gallinas y demás aves de corral (Pala & Radhakrishnan, 2014; Ryll et al., 2008).

Tabla de parámetros de evaluación de algoritmos.

Sp	Algoritmo	% de Área			
		AUC	pROC	OR	predicha
Arcobacter_ lekithochrous	BIO	0.80	1.58	0.00	100.00
Arcobacter_ lekithochrous	DOM	0.92	1.35	0.38	90.08
Arcobacter_ lekithochrous	GAU	0.96	1.38	0.00	85.45
Arcobacter_ lekithochrous	MXD	0.08	1.68	0.00	74.34
Arcobacter_ lekithochrous	MXS	0.60	1.68	0.13	76.68
Arcobacter_ lekithochrous	RDF	1.00	1.65	0.38	78.85
Arcobacter_ lekithochrous	SVM	1.00	1.72	0.63	64.92
Arcobacter_ nitrofigilis	BIO	0.50	1.07	0.83	56.44
Arcobacter_ nitrofigilis	DOM	0.65	0.92	0.67	100.00
Arcobacter_ nitrofigilis	GAU	0.80	1.50	0.67	91.60
Arcobacter_ nitrofigilis	MXD	0.33	1.66	0.83	100.00
Arcobacter_ nitrofigilis	MXS	0.71	1.66	0.50	100.00
Arcobacter_ nitrofigilis	RDF	0.95	1.92	0.50	59.28
Arcobacter_ nitrofigilis	SVM	0.51	1.54	0.00	82.21
Bacillus_ cereus	BIO	0.60	1.61	0.00	100.00
Bacillus_ cereus	DOM	0.70	1.20	0.50	100.00
Bacillus_ cereus	GAU	0.68	1.59	0.83	66.72
Bacillus_ cereus	MXD	0.70	1.85	0.33	100.00
Bacillus_ cereus	MXS	0.66	1.85	0.50	100.00
Bacillus_ cereus	RDF	0.90	1.79	0.00	51.32
Bacillus_ cereus	SVM	0.76	1.62	0.50	60.40
Pelagibacter_ ubique	BIO	0.50	1.24	0.43	80.99

Pelagibacter ubique	DOM	0.63	1.28	0.43	100.00
Pelagibacter ubique	GAU	0.78	1.29	0.00	83.84
Pelagibacter ubique	MXD	0.58	1.86	0.00	99.14
Pelagibacter ubique	MXS	0.64	1.86	0.00	99.11
Pelagibacter ubique	RDF	0.89	1.71	0.00	76.56
Pelagibacter ubique	SVM	0.89	1.30	0.00	80.90
Chryseobacterium_hominis	BIO	0.50	1.28	0.00	100.00
Chryseobacterium_hominis	DOM	0.25	1.19	0.83	89.63
Chryseobacterium_hominis	GAU	0.67	1.58	0.50	89.79
Chryseobacterium_hominis	MXD	0.47	1.75	0.00	98.35
Chryseobacterium_hominis	MXS	0.72	1.75	0.00	96.45
Chryseobacterium_hominis	RDF	0.88	1.87	0.67	67.74
Chryseobacterium_hominis	SVM	0.79	1.51	0.67	68.52
Clavibacter_michiganensis	BIO	0.50	1.23	0.00	100.00
Clavibacter_michiganensis	DOM	0.53	1.07	0.00	100.00
Clavibacter_michiganensis	GAU	0.69	1.27	0.71	76.67
Clavibacter_michiganensis	MXD	0.67	1.86	0.00	97.51
Clavibacter_michiganensis	MXS	0.67	1.86	0.14	95.31
Clavibacter_michiganensis	RDF	0.60	1.70	0.00	89.75
Clavibacter_michiganensis	SVM	0.64	1.34	0.43	76.97
Cribrihabitans_marinus	BIO	0.50	1.72	0.67	90.00
Cribrihabitans_marinus	DOM	0.61	1.30	0.17	98.00
Cribrihabitans_marinus	GAU	0.92	1.65	0.17	72.22
Cribrihabitans_marinus	MXD	0.70	1.86	0.00	100.00
Cribrihabitans_marinus	MXS	0.70	1.86	0.00	100.00
Cribrihabitans_marinus	RDF	0.99	1.87	0.33	62.32
Cribrihabitans_marinus	SVM	0.16	1.96	0.00	81.10
Marinobacterium_marisflavi	BIO	0.60	1.80	0.40	96.87
Marinobacterium_marisflavi	DOM	0.66	1.65	0.40	100.00
Marinobacterium_marisflavi	GAU	0.96	1.72	0.80	70.49
Marinobacterium_marisflavi	MXD	0.88	1.79	0.60	68.48

Marinobacterium_marisflavi	MXS	0.96	1.80	0.20	67.93
Marinobacterium_marisflavi	RDF	0.94	1.86	0.20	51.71
Marinobacterium_marisflavi	SVM	0.50	1.70	0.80	77.49
Riemerella_anatipestifer	BIO	0.63	0.97	0.00	97.00
Riemerella_anatipestifer	DOM	0.55	0.92	0.89	100.00
Riemerella_anatipestifer	GAU	1.00	1.20	0.78	70.49
Riemerella_anatipestifer	MXD	0.66	1.79	0.00	100.00
Riemerella_anatipestifer	MXS	0.64	1.57	0.56	100.00
Riemerella_anatipestifer	RDF	0.98	1.34	0.00	73.67
Riemerella_anatipestifer	SVM	0.65	1.32	0.78	74.06

Géneros sintetizadores de PHA's.

Localidad	Lecturas	Especies totales	Lecturas Halomonas	Lecturas Paracoccus	Lecturas Staphylococcus	Lecturas Planobacterium	Especies Halomonas	Especies Paracoccus	Especies Staphylococcus	Especies Planobacterium
Cienaga, caracol	23	9	15	1	7	0	5	1	3	0
Cienaga, los Huites	2264	5	0	2228	36	0	0	3	2	0
Peten, ojo de agua	45	14	17	4	22	1	4	3	6	2
Peten entrada	45	13	19	4	22	0	5	3	6	0
Bocana entrada	32	3	8	0	24	0	1	0	2	0
Bocana ojo de agua	61	18	9	26	26	0	4	6	8	0
Dzilam de Bravo 2	476	9	0	450	26	0	0	7	2	0
Chelem	66	4	66	0	0	0	4	0	0	0
Chelem 2	332	5	10	68	254	0	1	1	3	0
Celestun	118	6	0	54	64	0	0	3	3	0
Celestun 2	64	3	0	0	64	0	0	0	3	0
Sian Kaan	800	9	0	800	0	0	0	9	0	0
Sian Kaan 2	798	7	0	798	0	0	0	7	0	0
Cancun Bojorquez	108	3	0	108	0	0	0	3	0	0
Cancun Nichupte	132	3	0	132	0	0	0	3	0	0
Rio Lagartos ojo de agua	242	6	0	242	0	0	0	6	0	0

Tabla de géneros relacionados a la producción de PHA's y consumo de n-hexadecano.