



**FACULTAD DE CIENCIAS AGRARIAS
UNIVERSIDAD NACIONAL DE ROSARIO**

**EVALUACIÓN DE MODELOS DE MACHINE LEARNING PARA LA PREDICCIÓN
DE FENOTIPOS MEDIANTE LA COMPOSICIÓN DE LA MICROBIOTA
INTESTINAL: UN ENFOQUE PARA LA IDENTIFICACIÓN DE BIOMARCADORES**

NÉSTOR DENIS PORTELA

**TRABAJO FINAL PARA OPTAR AL TÍTULO DE ESPECIALISTA EN
BIOINFORMÁTICA**

**DIRECTORA: Dra. PESOA, Susana
CODIRECTORA: Dra. ARCE, Débora**

AÑO: 2024

**EVALUACIÓN DE MODELOS DE MACHINE LEARNING PARA LA PREDICCIÓN
DE FENOTIPOS MEDIANTE LA COMPOSICIÓN DE LA MICROBIOTA
INTESTINAL: UN ENFOQUE PARA LA IDENTIFICACIÓN DE BIOMARCADORES**

Néstor Denis Portela

Bioquímico – Universidad Nacional de Córdoba

Este Trabajo Final es presentado como parte de los requisitos para optar al grado académico de Especialista en Bioinformática, de la Universidad Nacional de Rosario y no ha sido previamente presentado para la obtención de otro título en esta u otra Universidad. El mismo contiene los resultados obtenidos en investigaciones llevadas a cabo en LACE Laboratorios, durante el período comprendido entre Enero de 2023 y Junio de 2023, bajo la dirección de Dra. Susana Pesoa.



Bioq. Néstor Denis Portela

Autor



Dra. Susana Pesoa

Directora



Dra. Débora Pamela Arce

Codirectora

Defendida: 12 de marzo de 2024.

Agradecimientos

A mi directora, la Dra. Susana Pesoa, por acompañarme con su experiencia y apoyo incondicional en todo momento, por su predisposición a escucharme y ayudarme no solo en lo académico sino también en lo personal.

A mi co-directora, la Dra. Débora Arce, por su tiempo, dedicación y guiarme en el proceso de elaboración de trabajo final.

A la Dirección del Laboratorio LACE, por darme el espacio y tiempo para continuar con mi formación y apoyarme en este proceso.

A la Facultad de Ciencias Agrarias, Universidad Nacional de Rosario por brindarme un espacio de formación de excelencia, por la calidez humana y la dedicación durante el transcurso de mi formación.

A mis padres, Mariana y Amado, por su apoyo incondicional y darme la oportunidad de estudiar y seguir formándome en lo que me gusta. Y por enseñarme el valor del esfuerzo, el mérito y a no bajar los brazos nunca. Y a mi hermana Marysol por acompañarme siempre y apoyarme en todo.

A Jorge Zimmermman, un amigo que me regaló la especialidad, por todas las tardes y charlas que dedicamos a aprender bioinformática resolviendo problemas y compartiendo anécdotas y risas.

Acrónimos

16S rRNA: ARN ribosomal 16S

ABC: área bajo la curva

ARN: ácido ribonucleico

ASV: (amplicon sequences variant) - variante de secuencias de amplicón

CNN: (Convolutional Neural Networks) - Redes neuronales convolucionales

GBM: (Gradient Boosted Machine Model) - Modelo de máquina potenciada por gradiente

LDA: Linear Discriminant Analysis - Análisis Discriminante Lineal

LMT: (Logistic model tree) - Árbol modelo logístico

Logitboost: (Boosted Logistic Regression) - Regresión logística potenciada

LVQ: (Learning Vector Quantization Model) - Modelo de cuantificación de vectores de aprendizaje

MI: Microbiota Intestinal

ML: (Machine Learning) - Aprendizaje automático

NN: (Neural Network): Red neuronal

PCA: (Principal Component Análisis) - Análisis de componentes principales

Rborist: (Random Forest) - Bosque aleatorio

RFE: (Recursive Feature Elimination) - Eliminación recursiva de características

SA: (Simulated Annealing) - Recocido simulado

SBF: (Selection By Filtering) - Selección por filtrado

SLDA: (Sparse Linear Discriminant Analysis) - Análisis discriminante lineal disperso

stepLDA: Linear Discriminant Analysis with Stepwise Feature Selection - Análisis discriminante lineal con selección de funciones paso a paso

SVM: (Support Vector Machine Model) - Modelo de máquina de vectores de soporte

SVMI: (Support Vector Machine Model with linear kernel) - Modelo de máquina de vectores de soporte con kernel lineal

SVMr: (Support Vector Machine Model with radial kernel) - Modelo de máquina de vectores de soporte con kernel radial

Resumen

En los últimos tiempos, los modelos de aprendizaje automático (ML) han surgido como una herramienta eficaz para comprender la variación interpersonal del microbioma y explorar su posible relación con el desarrollo de enfermedades. A diferencia de los enfoques tradicionales, los modelos de ML consideran la abundancia relativa de cada población bacteriana en el contexto de otras poblaciones, lo que permite una explicación más efectiva. No obstante, es crucial abordar los desafíos relacionados con la estandarización de los modelos, las metodologías y la estructura de datos utilizados, a fin de garantizar la robustez y aplicabilidad clínica de dichos modelos. En este contexto, llevamos a cabo una evaluación exhaustiva de 10 modelos de ML diferentes para la clasificación de fenotipos basados en la composición de la microbiota intestinal. Nuestro enfoque se centró en analizar el impacto de la escala taxonómica y la efectividad de la estrategia de selección de características en el rendimiento de los modelos. Los análisis revelaron diferencias significativas en el rendimiento de los modelos de predicción al evaluarlos en diferentes niveles taxonómicos, obteniéndose mejores resultados en la exactitud y el área bajo la curva en niveles más específicos, como género y especie. Asimismo, encontramos que no todos los métodos de selección de características son igualmente eficaces para mejorar el rendimiento de las predicciones de los modelos. En general, la aplicación de estos métodos tuvo un impacto positivo y significativo en los niveles de familia, género y especie, en la mayoría de los modelos. Mediante el análisis Hold-out, encontramos que el modelo de Random Forest con el método de selección de características SAFS.rf, aplicado a nivel de especie, mantuvo un alto rendimiento consistente en todas las etapas de evaluación, mostrando una menor influencia por la composición y estructura de los conjuntos de datos utilizados. Por último, nuestras estrategias de análisis nos permitieron identificar posibles biomarcadores microbianos asociados al fenotipo delgado/obeso. Estos resultados respaldan la aplicabilidad de las técnicas de ML en el estudio de microbiomas, al demostrar su capacidad para descubrir relaciones no lineales y generar perfiles clasificatorios con mayor precisión. Además, subrayan la importancia de considerar los niveles taxonómicos más específicos, como género y especie, y la implementación de técnicas de selección de características como una estrategia recomendada para mejorar el rendimiento de los

métodos de predicción. Estos hallazgos brindan información relevante sobre la robustez de las técnicas de ML en la predicción de fenotipos relacionados con enfermedades a partir de información metagenómica, y pueden proporcionar nuevos indicios sobre la relevancia de ciertos microorganismos en el desarrollo de diversas patologías.

Palabras claves: Métodos de aprendizaje automático, microbiota intestinal, biomarcadores, técnicas de selección de características.

Abstract:

In recent times, machine learning (ML) models have emerged as an effective tool for understanding the interpersonal variation of the microbiome and exploring its possible relationship with disease development. Unlike traditional approaches, ML models take into account the relative abundance of each bacterial population in the context of other populations, allowing for a more effective explanation. However, it is crucial to address the challenges related to model standardization, methodologies, and data structure to ensure the robustness and clinical applicability of these models. In this context, we conducted a comprehensive evaluation of 10 different ML models for the classification of phenotypes based on the composition of the intestinal microbiota. Our focus was on analyzing the impact of taxonomic scale and the effectiveness of feature selection strategies on model performance. The analyses revealed significant differences in the predictive performance of the models when evaluated at different taxonomic levels, with improved accuracy and area under the curve achieved at more specific levels such as genus and species. Additionally, we found that not all feature selection methods are equally effective in improving the predictive performance of the models. Overall, the application of these methods had a positive and significant impact on the family, genus, and species levels in most models. Through the Hold-out analysis, we found that the Random Forest model with the SAFS.rf feature selection method, applied at the species level, maintained consistently high performance across all evaluation stages, showing less influence from the composition and structure of the datasets used. Lastly, our analysis strategies allowed us to identify potential microbial biomarkers associated with the lean/obese phenotype. This evidence supports the applicability of ML techniques in microbiome studies, demonstrating their ability to uncover non-linear relationships and generate more accurate classification profiles. They also emphasize the importance of considering more specific taxonomic levels such as genus and species, and implementing feature selection techniques as a recommended strategy to improve the performance of prediction methods. These findings provide relevant information on the robustness of ML techniques in predicting disease-related phenotypes from metagenomic information and can provide new insights into the relevance of certain microorganisms in the development of various pathologies.

Keywords: Machine learning methods, gut microbiota, biomarkers, feature selection techniques.

ÍNDICE

ÍNDICE.....	10
1. INTRODUCCIÓN.....	12
2. OBJETIVOS:.....	15
2.1. Objetivo General	15
2.2. Objetivos específicos.....	15
3. MATERIALES Y MÉTODOS:.....	16
3.1. Origen de los datos de composición de microbiota intestinal relacionados al desarrollo de obesidad.....	16
3.2. Selección de modelos de clasificación.....	17
3.2.1. Learning Vector Quantization Model (LVQ).....	18
3.2.2. Gradient Boosted Machine Model (GBM).....	18
3.2.3. Support Vector Machine Model (Radial) (SVMr)	18
3.2.4. Support Vector Machine Model (Linear) (SVMl).....	19
3.2.5. Sparse Linear Discriminant Analysis (SLDA).....	19
3.2.6. Neural Network (NN): La Neural Network (Red neuronal)	19
3.2.7. Random Forest (Rborist): El modelo Random Forest (Bosque Aleatorio)	19
3.2.8. eXtreme Gradient Boosting (xgbLinear).	19
3.2.9. Linear Discriminant Analysis with Stepwise Feature Selection (stepLDA).....	20
3.2.10. Naive Bayes (naive_bayes).....	20
3.3. Generación de dataset de análisis.....	20
3.3. Influencia de la escala taxonómica.....	20
3.4. Entrenamiento, evaluación y validación de los modelos.....	21
3.5. Selección de características relevantes para la clasificación	21
3.5.1. Métodos de envoltura (Wrapper methods)	22
3.5.2. Métodos de filtro (Filters methods)	22
3.6. Entrenamiento de los modelos	23
3.7. Prueba de los modelos	23
3.8. Selección de mejores modelos	23
3.9. Evaluación del rendimiento final de modelos de alta performance	24
3.10. Identificación de Taxas microbianas asociadas al fenotipo Delgado/Obeso	25
3.11. Análisis estadístico de datos y visualización de resultados.....	25
4. RESULTADOS	26
4.1. Preprocesamiento de los datos	26

4.2.	Efecto de la taxonomía sobre la performance de los modelos.....	27
4.3.	Selección de características como estrategia de optimización de modelos	31
4.4.	Evaluación de la Exactitud post selección de características	34
4.5.	Selección de modelos más aptos de clasificación del fenotipo delgado/obeso	37
4.6.	Estrategia gráfica para una evaluación integral de modelos de predicción.....	39
4.7.	Validación de los modelos de mejor performance en el dataset independiente.	41
4.8.	Identificación de Taxas microbianas asociadas al fenotipo Delgado / Obeso:.....	43
4.9.	Evaluación del potencial discriminativo de los biomarcadores bacterianos identificados	45
5.	DISCUSIÓN:.....	48
6.	CONCLUSION	53
7.	BIBLIOGRAFÍA.....	54
8.	MATERIAL SUPLEMENTARIO	62

EVALUACIÓN DE MODELOS DE MACHINE LEARNING PARA LA PREDICCIÓN DE FENOTIPOS MEDIANTE LA COMPOSICIÓN DE LA MICROBIOTA INTESTINAL: UN ENFOQUE PARA LA IDENTIFICACIÓN DE BIOMARCADORES

1. INTRODUCCIÓN

Los avances en las ciencias ómicas y el desarrollo de métodos de secuenciación de alto rendimiento y de análisis bioinformático han proporcionado nuevos conocimientos sobre las estructuras y funciones del microbioma/microbiota (Liu et al., 2021). La microbiota intestinal (MI) es un sistema complejo con interacciones dinámicas entre los microorganismos que la conforman, el huésped y factores ambientales (Villanueva-Millán et al., 2015). Estudios previos han demostrado que la MI desempeña un papel fundamental en el mantenimiento de la salud humana, participando en la regulación del sistema inmunológico, el metabolismo de nutrientes y la protección contra patógenos (Clemente et al., 2012; O'Hara y Shanahan, 2006).

Además de su función en la homeostasis del organismo, se ha descrito que la MI también está implicada en el desarrollo y progresión de diversas patologías y condiciones clínicas, por ejemplo: Obesidad (Cani et al., 2007; Ley et al., 2006; Wu et al., 2022), Cáncer (Khan et al., 2022; Schwabe y Jobin, 2013), Enfermedad Inflamatoria Intestinal (Qiu et al., 2022), Enfermedades neuropsiquiátricas (Hsiao et al., 2013), entre otras. Estas asociaciones entre la MI y la salud humana han despertado un gran interés científico y clínico en comprender mejor los mecanismos subyacentes (Marcos-Zambrano et al., 2021), estratificar a los individuos en función de su configuración microbiana, identificar biomarcadores microbianos con fines diagnósticos/pronósticos y para el desarrollo de estrategias de prevención y tratamiento enfocados en la manipulación de la misma "*Ingeniería de Microbiomas*" (Albright et al., 2022).

Sin embargo, la complejidad de la MI dificulta la interpretación y la conexión de sus diferentes componentes con el desarrollo de dichas patologías cuando se utilizan métodos estadísticos/matemáticos tradicionales. En primer lugar, la MI exhibe una inmensa variabilidad interindividual, influenciada por factores genéticos, ambientales y nutricionales, lo que dificulta la identificación de patrones consistentes (Lloyd-Price et al., 2017). En segundo lugar, los datos obtenidos mediante las técnicas destinadas

a su estudio, como “secuenciación de amplicones” o “secuenciación en escopeta”, presentan desafíos en términos de dimensionalidad y representación, ya que generan tablas de características, en las que cada celda representa la abundancia o presencia de un taxón o función específica de la muestra. (Karlsson et al., 2013). Además, estas tablas tienen un número excesivo de ceros y presentan una alta dimensionalidad. (Hernández Medina et al., 2022; Zhou y Gallins, 2019). Por otro lado, los datos de composición describen a su vez, las relaciones entre sus componentes, por lo que sus partes no son independientes y su suma es arbitraria por lo que no pueden someterse a un análisis convencional directamente, al menos sin una normalización o transformación previa (Quinn et al., 2018). Además, es importante destacar que los métodos estadísticos tradicionales suelen considerar los efectos protectores o de riesgo asociados a cada taxón bacteriano de manera individual. Sin embargo, se ha demostrado que rara vez una única especie bacteriana se asocia con el desarrollo de una patología (Norouzi-Beirami et al., 2020; Segata et al., 2011). En lugar de eso, se ha observado que la interacción compleja y sinérgica entre múltiples especies bacterianas puede desempeñar un papel crucial en el desarrollo de enfermedades relacionadas con la MI. Por lo tanto, es fundamental adoptar enfoques más holísticos y centrados en la comunidad para comprender plenamente el impacto de la microbiota en la salud humana.

Ante estos desafíos, los algoritmos de aprendizaje automático - “*Machine Learning*” (ML) han surgido como una herramienta prometedora para analizar y comprender la complejidad de la MI. Estos han demostrado ser una herramienta eficaz para dilucidar la variación interpersonal de los microbiomas y explicar su posible relación con el desarrollo de enfermedades (Li et al., 2022). Los modelos de ML pueden identificar patrones y relaciones no lineales en los datos, lo que permite descubrir biomarcadores relevantes, estratificar a los individuos en función de su configuración microbiana y predecir fenotipos con precisión (Li et al., 2022). Esto gracias a que, a diferencia de los enfoques tradicionales, los modelos de ML consideran la abundancia relativa en un contexto de población y no como taxas individuales, lo que permite una explicación más efectiva con el fenotipo observado. (Faust y Raes, 2012; Lloyd-Price et al., 2017).

En este contexto, diferentes modelos de ML han sido empleados para predecir enfermedades utilizando los componentes de la MI como variables predictoras con rendimientos destacables (Tabla 1).

Tabla 1: Aplicaciones clínicas del aprendizaje automático para estudios de microbiota intestinal		
Objetivo predictivo	Modelo utilizado	Referencia
Cáncer colorrectal	<i>Bayesian Networks</i> <i>Random Forest</i> <i>Decision trees</i>	(Ai et al., 2017)
Cáncer colorrectal	<i>Random Forest</i>	(Flemer et al., 2018)
Diabetes mellitus tipo II	<i>Random Forest</i> <i>Support vector machine (SVM)</i> <i>Elastic net</i> <i>Lasso</i>	(Karlsson et al., 2013)
Múltiples enfermedades (esclerosis múltiple, artritis idiopática juvenil, encefalomiелitis mialgia / síndrome de fatiga crónica, síndrome de inmunodeficiencia adquirida, accidente cerebrovascular y cáncer colorrectal)	<i>Support vector machine (SVM)</i> <i>Boosted Logistic Regression (logitboost)</i> <i>K nearest neighbor (KNN)</i> <i>Logistic model tree (LMT)</i>	(Bang et al., 2019)
Neoplasia colorrectal	<i>L2-regularized logistic regression,</i> <i>L1- and L2-regularized support vector machines (SVM) with linear and radial kernels,</i> <i>Decision trees,</i> <i>Random forest</i> <i>Gradient boosted trees (xgboost).</i>	(Topçuoğlu et al., 2020)
Diabetes tipo 2 Cirrosis	<i>Convolutional Neural Networks (CNN)</i>	(Sharma et al., 2020)
Enfermedad cardiovascular	<i>Random Forest</i> <i>Support vector machine (SVM)</i> <i>Decision trees</i> <i>Elastic net</i> <i>Neural networks (NN)</i>	(Aryal et al., 2020)

Sin embargo, es fundamental abordar los problemas de estandarización de los modelos, las metodologías y la estructura de datos utilizados, así como la interpretabilidad de los resultados obtenidos, para garantizar la robustez y la aplicabilidad clínica de los modelos de ML aplicados a la MI. A modo de ejemplo, en muchas investigaciones que utilizan técnicas de ML existe la suposición implícita de que los modelos más complejos son mejores, lo cual en ocasiones puede llevar a generar predicciones de difícil interpretabilidad, lo que remarca la importancia de realizar un análisis exhaustivo de los modelos a utilizar bajo un objetivo puntual y frente a un set de datos en particular (Topçuoğlu et al., 2020). Por otro lado, la escala taxonómica utilizada para describir la composición de la MI presenta una estructura jerárquica, donde cada nivel taxonómico está relacionado con los niveles superiores e inferiores. Estas relaciones taxonómicas pueden tener un impacto significativo en el rendimiento de los modelos de ML. Además, ciertos atributos taxonómicos pueden tener una influencia más significativa en la clasificación o predicción de fenotipos que otros (Song y Zhou, 2023). Por lo tanto, es importante considerar estas relaciones al desarrollar modelos de ML para predecir fenotipos basados en la composición de la microbiota.

2. OBJETIVOS:

2.1. Objetivo General

Evaluar y mejorar la performance de 10 modelos de aprendizaje automático en la clasificación de fenotipos basados en la composición de la MI, centrándose en la influencia de la escala taxonómica y la efectividad de la estrategia de selección de características.

2.2. Objetivos específicos.

- Determinar la influencia de la escala taxonómica sobre la performance de los modelos de ML ensayados.
- Evaluar la performance de los modelos incorporando la estrategia de selección de características relevantes.
- Inferir el o los modelos con mejor performance para los fines propuestos.
- Identificar taxones microbianos con potencial predictivo sobre los fenotipos bajo estudio.

A través de esta evaluación, se busca mejorar la comprensión de la performance, capacidad predictiva e interpretabilidad de los modelos en el contexto de la MI. Los resultados de este estudio se espera que tengan un impacto significativo al proporcionar información valiosa sobre la elección adecuada de la escala taxonómica y la estrategia de selección de características. La información obtenida tiene el potencial de contribuir al desarrollo de biomarcadores microbiológicos y mejorar la comprensión de las relaciones entre la MI y el desarrollo de enfermedades. Esto puede tener implicaciones significativas en la prevención y tratamiento de diversas condiciones clínicas, contribuyendo al desarrollo de enfoques más efectivos en el campo de la medicina de precisión.

3. MATERIALES Y MÉTODOS:

3.1. Origen de los datos de composición de microbiota intestinal relacionados al desarrollo de obesidad.

Para la predicción de obesidad basada en la composición de MI a partir de muestras de materia fecal, se utilizaron los datos crudos de secuencia generados previamente en nuestro laboratorio en un proyecto basado en un modelo murino de obesidad inducido por la dieta. Estos datos fueron obtenidos mediante la estrategia de secuenciación de amplicones, utilizando el kit Ion 16S Metagenomics Kit (Thermo Fisher Scientific, Carlsbad, CA, USA), en la plataforma Ion Torrent Personal Genome Machine (PGM) (Thermo Fisher Scientific Carlsbad, CA, EE. UU).

En resumen, se generaron bibliotecas a partir de 20 ng de ADN fecal con el kit mencionado, utilizando una combinación de dos conjuntos de cebadores dirigidos a las regiones hipervariables V2, V4 y V8 del gen 16S rRNA con el conjunto 1 y V3, V6-7 y V9 con el grupo 2. Tras obtener los productos de amplificación para cada conjunto de cebadores, se procedió a la preparación de las muestras para la secuenciación. En este proceso, se tomaron 20 µl de cada producto de amplificación y se combinaron. Posteriormente, se llevó a cabo un proceso de digestión parcial y se ligaron los códigos de barras y adaptadores de secuenciación. Luego, se purificaron mediante perlas Agencourt AMPure XP de acuerdo con el protocolo del fabricante (Beckman Coulter; Pasadena, CA, EE. UU.). La concentración de cada biblioteca se determinó mediante qPCR utilizando el kit de cuantificación Ion Universal Library Quantitation

(Thermo Fisher Scientific Carlsbad, CA, EE. UU.). La preparación del templado se realizó utilizando el kit Ion PGM Hi-Q View OT2 (Thermo Fisher Scientific Carlsbad, CA, EE. UU.) y el sistema Ion OneTouch 2 (Thermo Fisher Scientific Carlsbad, CA, EE. UU.). La biblioteca se diluyó a ~ 10 pM antes de la preparación del templado.

Se secuenció un máximo de 9 muestras con código de barras en un chip Ion 316v2 utilizando el kit de secuenciación Ion PGM Hi-Q View (Thermo Fisher Scientific; Carlsbad, CA, EE. UU.) de acuerdo con las instrucciones del fabricante. Posteriormente, los archivos .fastq resultantes de cada corrida analítica fueron recuperados y destinados al análisis bioinformático. Estos archivos contienen las secuencias obtenidas para cada una de las 7 regiones hipervariables amplificadas en cada muestra.

El control de calidad de las secuencias, la anotación y la asignación taxonómica se realizaron utilizando los paquetes *DADA2 v1.22.0* (Callahan et al., 2016), *phyloseq v1.38.0* (McMurdie y Holmes, 2013) y *microbiome v1.16.0* (Leo Lahti, 2012) en el software R v4.1.2 (R Core Team, 2021). Las secuencias se procesaron siguiendo el flujo estándar de archivos fastq demultiplexados. Para la asignación taxonómica, se utilizó la estrategia de Variante de secuencia de amplicón (Amplicon Sequence Variant, ASV) y la base de datos Silva versión 138.1 con formato *DADA2*, actualizada el 10 de marzo de 2021 (Quast et al., 2013). El recuento de lecturas fue normalizado, mediante la técnica de rarefacción, con la intención de disminuir posibles variaciones asociadas a la calidad/cantidad de la secuenciación de las diferentes muestras, utilizando el paquete *microbiome v1.16.0* (Leo Lahti, 2012).

El producto del análisis de asignación taxonómica se almacenó como un objeto de clase *phyloseq*, conteniendo la tabla de ASVs, la tabla de taxonomía, la tabla de metadatos (que contiene la variable a predecir, en este caso el fenotipo delgado/obeso) y la tabla de información filogenética.

3.2. Selección de modelos de clasificación

En este estudio, se evaluaron 10 modelos de clasificación supervisada previamente utilizados y reportados por otros autores (Tabla 1), los cuales han demostrado desempeño aceptable en la clasificación de fenotipos/enfermedades basados en la composición de la MI. Además, se evaluaron algunos modelos de interés para nuestro

propósito: A continuación, se describen brevemente las características de cada uno de los modelos utilizados:

3.2.1. *Learning Vector Quantization Model (LVQ)*: este modelo es un algoritmo de clasificación supervisada que se basa en el uso de vectores prototipo para representar cada clase o categoría del problema. Durante el entrenamiento, los vectores prototipo se ajustan para adaptarse a los datos de entrada y clasificar correctamente las muestras (Kohonen, 1995).

3.2.2. *Gradient Boosted Machine Model (GBM)*: es un algoritmo de ensamblaje que combina múltiples árboles de decisión débiles para construir un modelo más fuerte. En cada iteración, se ajusta un nuevo árbol de decisión para corregir los errores del modelo anterior, y se combinan los resultados de todos los árboles para realizar una predicción final (Zhou y Gallins, 2019).

3.2.3. *Support Vector Machine Model (Radial) (SVMr)*: El modelo utiliza una técnica de clasificación que se basa en encontrar el hiperplano óptimo que separa las muestras de diferentes clases en un espacio dimensional superior. El kernel radial permite mapear los datos a un espacio no lineal, lo que a su vez permite una clasificación más precisa cuando los datos no son linealmente separables (Zhou y Gallins, 2019).

3.2.4. *Support Vector Machine Model (Linear) (SVMl)*: este modelo es similar al SVMr, pero utiliza un hiperplano lineal para separar las muestras de diferentes clases. Este modelo es adecuado cuando los datos son linealmente separables y puede ser más rápido de entrenar que el SVM con kernel no lineal (Marcos-Zambrano et al., 2021).

3.2.5. *Sparse Linear Discriminant Analysis (SLDA)*: La Sparse Linear Discriminant Analysis (SLDA) es una técnica de reducción de dimensionalidad y clasificación. Busca encontrar una combinación lineal de las características originales que maximice la separación entre las clases. Además, promueve la esparcidad en el espacio de características, lo que significa que solo se utiliza un subconjunto de características relevantes para la clasificación (Clemmensen et al., 2016).

3.2.6. *Neural Network (NN)*: La Neural Network (Red neuronal) es un modelo inspirado en la estructura del cerebro humano, que consiste en múltiples neuronas interconectadas. Estas redes neuronales tienen capas de nodos que procesan y transfieren información. El modelo nnet utiliza algoritmos de aprendizaje para ajustar los pesos y las conexiones de las neuronas con el fin de realizar tareas de clasificación o regresión (Zhou y Gallins, 2019).

3.2.7. *Random Forest (Rborist)*: El modelo Random Forest (Bosque Aleatorio) es un ensamblaje de árboles de decisión donde cada árbol se entrena con una muestra aleatoria del conjunto de datos original. Luego, se combinan las predicciones de los árboles individuales para obtener una predicción final. El Random Forest es conocido por su capacidad para manejar grandes conjuntos de datos y reducir el sobreajuste (Zhou y Gallins, 2019).

3.2.8. *eXtreme Gradient Boosting (xgbLinear)*: El modelo eXtreme Gradient Boosting (XGBoost) es un algoritmo de aprendizaje automático que utiliza el ensamblaje de árboles de decisión y un enfoque optimizado de gradiente. XGBoost mejora el rendimiento del Gradient Boosted Machine al aplicar técnicas de regularización, controlar la complejidad del modelo y optimizar la velocidad y eficiencia del entrenamiento (Alshboul et al., 2022).

3.2.9. *Linear Discriminant Analysis with Stepwise Feature Selection*

(*stepLDA*): El Linear Discriminant Analysis (LDA) es un modelo que busca encontrar una combinación lineal de características que maximice la separación entre las clases. El *stepLDA* agrega una selección de características paso a paso, lo que significa que selecciona iterativamente las características más relevantes para mejorar la precisión del modelo. La diferencia entre *stepLDA* y *SLDA* radica en el enfoque utilizado para seleccionar características relevantes. El *stepLDA* realiza una selección paso a paso, mientras que el *SLDA* impone una penalización de regularización para lograr una representación de características relevantes y significativas (Song et al., 2010).

3.2.10. *Naive Bayes (naive_bayes)*:

es un modelo probabilístico que se basa en el teorema de Bayes y la suposición de independencia condicional entre las características. Este modelo asigna una muestra a la clase más probable en función de la probabilidad condicional de las características dadas de las clases (Marcos-Zambrano et al., 2021).

3.3. **Generación de dataset de análisis**

A partir del objeto “*phyloseq*” generado en el análisis de asignación taxonómica, se creó un dataset primario que contiene la matriz de abundancia de las diferentes ASVs, la asignación taxonómica de cada ASV y el fenotipo correspondiente a cada muestra. El número de lecturas obtenido para cada ASV identificada fue transformado a abundancia relativa mediante la siguiente función del paquete *phyloseq*.

```
transform_sample_counts(Objeto Phyloseq, function(x){x / sum(x)})
```

3.3. **Influencia de la escala taxonómica**

Con el fin de abordar el desafío de las relaciones entre los niveles taxonómicos en MI y su impacto en el rendimiento de los modelos de ML, se llevó a cabo un proceso de aglomeración en diferentes niveles taxonómicos.

Para realizar la aglomeración, se utilizó la función “*tax_glom*” del paquete *phyloseq* v1.38.0 (McMurdie y Holmes, 2013) sobre el dataset primario, esta estrategia permitió agrupar las características microbianas a nivel de Phylum, Clase, Orden, Familia, Género y Especie, generando así 6 nuevos datasets. Cada nivel taxonómico

representó una forma de agrupación de las variables originales, lo que redujo la complejidad y las relaciones entre las características.

3.4. Entrenamiento, evaluación y validación de los modelos

Para evaluar los 10 modelos de ML seleccionados, se realizó una partición estratificada del conjunto de datos original de cada nivel taxonómico. En primer lugar, se seleccionó aleatoriamente el 20% de los datos, los cuales se destinaron exclusivamente como un conjunto Hold-out. Este conjunto se mantuvo completamente separado y se utilizó únicamente al final del proceso para evaluar el rendimiento final de los mejores modelos seleccionados.

A continuación, el 80% restante de los datos originales se dividió aleatoriamente, donde el 80% de los mismos se asignaron a un dataset de entrenamiento y el 20% restante se reservó como dataset de prueba.

Es importante destacar que la partición se realizó de forma estratificada para garantizar la preservación de la distribución de los fenotipos presentes en el dataset original. Esto significa que se mantuvo la proporción relativa de cada clase en los conjuntos de entrenamiento, prueba y Hold-out, lo cual es fundamental para obtener resultados representativos y evitar sesgos en la evaluación de los modelos.

Esta técnica de partición en conjunto Hold-out, entrenamiento y prueba permite evaluar adecuadamente el rendimiento de los modelos de ML. El conjunto de entrenamiento se utilizó para entrenar los modelos, mientras que el conjunto de prueba se utilizó para evaluar su rendimiento y seleccionar el mejor modelo. Por último, el conjunto Hold-out, se utilizó para realizar una evaluación final imparcial de los modelos seleccionados.

3.5. Selección de características relevantes para la clasificación

Con el objetivo de mejorar el rendimiento de los modelos de predicción al reducir el número de variables predictoras y considerar únicamente aquellas de mayor relevancia para la clasificación de los fenotipos, se realizó una selección de características relevantes. Para ello, se utilizaron 4 métodos de selección que pueden clasificarse en dos categorías:

3.5.1. *Métodos de envoltura (Wrapper methods)*: los cuales evalúan múltiples modelos mediante procedimientos que agregan o eliminan predictores para encontrar la combinación óptima que maximiza el rendimiento del modelo (Saeys et al., 2007).

Dentro de estos se utilizó:

- Recocido simulado (Simulated Annealing) (SA): es un método de búsqueda global que realiza pequeños cambios aleatorios (es decir, perturbaciones) en una solución candidata inicial. Si el valor de rendimiento del valor perturbado es mejor que la solución anterior, se acepta la nueva solución. De lo contrario, se determina una probabilidad de aceptación basada en la diferencia entre los dos valores de rendimiento y la iteración actual de la búsqueda. A partir de esto, se puede aceptar una solución subóptima a cambio de que eventualmente produzca una mejor solución en iteraciones posteriores (Kirkpatrick et al., 1983). En este estudio se aplicó SA utilizando dos métodos de control/ajuste. Por un lado, se aplicó SA ajustado por control caret, que es un ajuste determinado por la librería caret, a este método lo denominamos SAFS. Y por otro, se aplicó SA ajustado a un modelo de bosque aleatorio (SAFS.rf) (Johnson y Kuhn, 2019).
- Eliminación recursiva de características (Recursive Feature Elimination) (RFE): este método comienza entrenando un modelo en el conjunto completo de características y luego calcula la importancia de cada característica. A continuación, las características menos importantes se eliminan del conjunto de datos. El proceso se repite iterativamente, ajustando el modelo y eliminando características hasta que se alcance un criterio de parada predefinido, como un número específico de características o una métrica de rendimiento deseada (Johnson y Kuhn, 2019) .

3.5.2. *Métodos de filtro (Filters methods)*: los cuales evalúan la relevancia de los predictores fuera de los modelos predictivos y posteriormente modelan solo los predictores que pasan algún criterio (Saeys et al., 2007). Aplicando el método Selección por Filtrado (Selection By Filtering) (SBF).

Los cuatro métodos de selección mencionados fueron aplicados a los 6 conjuntos de datos generados mediante la aglomeración de niveles taxonómicos. Esta etapa permitió obtener las características/variables más relevantes para la clasificación de los fenotipos en cada nivel taxonómico. Posteriormente, se utilizaron estos resultados para filtrar los sets de datos de cada nivel taxonómico y generar nuevos datasets que sólo incluyen aquellas taxas marcadas como relevantes para cada uno de los métodos de selección. Finalmente, a cada conjunto de datos, se le aplicaron los 10 modelos de clasificación nombrados previamente. Globalmente, la estrategia presentada permitió generar y evaluar 300 modelos diferentes de clasificación, aportando una evaluación global del rendimiento de los mismos.

3.6. Entrenamiento de los modelos

Los diferentes modelos fueron entrenados utilizando un enfoque de validación cruzada repetida (`repeatedcv`) con 10 particiones (`number=10`) y 10 repeticiones (`repeats=10`). La exactitud (`accuracy`) media del total de iteraciones fue la métrica considerada para evaluar el desempeño de los modelos.

3.7. Prueba de los modelos

Para la prueba de los modelos se realizó un muestreo aleatorio con reemplazo, tomando una muestra del 90% de los datos del dataset de prueba en cada iteración, y se repitió el proceso por 30 veces, con el fin de disminuir la influencia de la composición del dataset en la performance de la prueba. En este caso, se consideró la media del valor área bajo la curva (ABC) obtenida de las 30 iteraciones, como métrica para evaluar los modelos.

3.8. Selección de mejores modelos

Se diseñó una estrategia gráfica para determinar los modelos con mejor performance para la clasificación de los fenotipos teniendo en cuenta los parámetros evaluados, Exactitud y ABC.

Para ello se realizó un gráfico de dispersión, donde en el eje X se graficó la media de ABC obtenida del proceso de prueba y en el eje Y la media Exactitud obtenida del proceso de entrenamiento para cada modelo respectivamente. El gráfico se dividió en

diferentes cuadrantes, asignándoles una significancia específica para el propósito deseado.

Los modelos ubicados en el cuadrante representado por valores de Exactitud y ABC menores a 0.75 se consideraron de bajo rendimiento y no aptos para la aplicación en la clasificación de los fenotipos. El mismo criterio se aplicó para aquellos modelos ubicados en los cuadrantes comprendidos por valores de Exactitud > 0.75 y ABC < 0.75 , y Exactitud < 0.75 y ABC > 0.75 , respectivamente.

Por otro lado, los modelos ubicados en los cuadrantes con valores de Exactitud entre 0.75 y 0.85 y ABC entre 0.75 y 0.85 se consideraron de performance intermedia. Estos modelos podrían ser utilizados bajo ciertas condiciones o fines específicos, o sometidos a una evaluación exhaustiva, como la valoración de parámetros de ajuste. Pero no fueron considerados en este proyecto.

Finalmente, los modelos representados en el cuadrante con valores de Exactitud entre 0.85 y 1 y ABC entre 0.85 y 1 se clasificaron como de alto rendimiento y aptos para el propósito deseado.

3.9. Evaluación del rendimiento final de modelos de alta performance

Una vez seleccionados los modelos de alto rendimiento, se procedió a evaluar su desempeño final utilizando el conjunto de datos independiente (Hold-out), con el objetivo de identificar los modelos que mantuvieran un rendimiento consistente independientemente del conjunto de datos utilizado.

En esta evaluación los parámetros de Precisión, Exactitud, Sensibilidad (Recall), tasa de Falso Positivo, Tasa de falso Negativo, F1 score y ABC fueron calculados. Para determinar el modelo de mejor rendimiento, se utilizó nuevamente la estrategia gráfica “ABC vs Exactitud”, descrita anteriormente.

Todos los procedimientos descritos de partición de datasets, entrenamiento, prueba y evaluación Hold-out se realizaron utilizando el software R v4.1.2 (R Core Team, 2021) mediante los paquetes: caret v6.0.94 (Kuhn, 2008), rpart v4.1.19 (Therneau et al., 2022), mlbench v 2.1.3 (Leisch y Dimitriadou, 2021), pROC v 1.18.0 (Robin et al., 2011), randomForest v 4.7.1.1 (Liaw y Wiener, 2002), Hmisc v 5.0.1 (Harrell y Dupont,

2019) y e1071 v 1.7.13 (Meyer et al., 2023). Utilizando en cada función los parámetros de ajuste por defecto.

3.10. Identificación de Taxas microbianas asociadas al fenotipo Delgado/Obeso

Con el fin de identificar taxones microbianos asociados al fenotipo Delgado/Obeso, todos aquellos modelos seleccionados como de alto rendimiento fueron agrupados de acuerdo a su nivel taxonómico, y para cada grupo se determinaron aquellas taxas que estaban presentes en todos los modelos pertenecientes a cada grupo, considerándose estas como posibles biomarcadores relacionados al fenotipo delgado/obeso.

Para demostrar si verdaderamente esta estrategia permite identificar biomarcadores relacionados el fenotipo en estudio, los taxones identificados para cada nivel taxonómico fueron agrupados en un conjunto, el cual fue utilizado para filtrar el objeto phyloseq original, generando uno nuevo que incluía únicamente aquellas ASVs identificadas como posibles biomarcadores de delgadez/obesidad.

Para evaluar el potencial discriminativo de los fenotipos utilizando los biomarcadores bacterianos identificados, se empleó el análisis de componentes principales (PCA). Además, se llevó a cabo el mismo estudio utilizando el conjunto de datos phyloseq completo para realizar una comparación de las clasificaciones. Para evaluar la significancia estadística de los resultados obtenidos en el PCA, se realizó un análisis de permutación estadística (permutest).

Por último, con el fin de comprender mejor la contribución de estos posibles biomarcadores al fenotipo delgado/obeso, se llevó a cabo un análisis de abundancia diferencial de los taxones seleccionados. Estos análisis se realizaron utilizando los paquetes *phyloseq v1.38.0* (McMurdie y Holmes, 2013) y *microbiome v1.16.0* (Leo Lahti, 2012) en el entorno de programación R (R Core Team, 2021).

3.11. Análisis estadístico de datos y visualización de resultados

El análisis estadístico se llevó a cabo utilizando el software R v4.1.2 (R Core Team, 2021). El análisis comparativo y diferencial entre variables se realizó utilizando los tests de Wilcoxon o Kruskal Wallis para variables sin distribución normal, según corresponda. Se realizaron comparaciones pareadas utilizando el test de rangos

con signo de Wilcoxon si el test de Kruskal Wallis arrojaba un resultado significativo. Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni. Se consideró un valor $p < 0.05$ como significativo. La visualización gráfica de resultados se realizó utilizando el paquete ggplot2 v 3.4.2 (Wickham, 2016) y ggpubr v 0.6.0 (Kassambara, 2022) el software R v4.1.2 (R Core Team, 2021).

4. RESULTADOS

4.1. Preprocesamiento de los datos

En el presente proyecto se utilizó información metagenómica de 108 muestras de materia fecal de ratón. Las mismas fueron obtenidas en nuestro laboratorio durante el desarrollo de un protocolo de obesidad inducida por la dieta. El proceso de asignación taxonómica arrojó un total de 1757 ASVs identificadas. El análisis de rarefacción produjo un recuento total de 16663 lecturas por muestra. El diagrama 1 presenta el esquema de trabajo utilizado.

La Tabla 2, resume la distribución de fenotipos y el número de muestras de los diferentes sets de datos generados. Es importante destacar que, como se mencionó anteriormente, la partición se realizó de forma estratificada para garantizar la preservación de la distribución de los fenotipos presentes en el dataset original. Esto asegura que la proporción relativa de cada clase en los conjuntos de entrenamiento, prueba y Hold-out, se mantenga constante, lo cual es fundamental para obtener resultados representativos y evitar sesgos en la evaluación de los modelos.

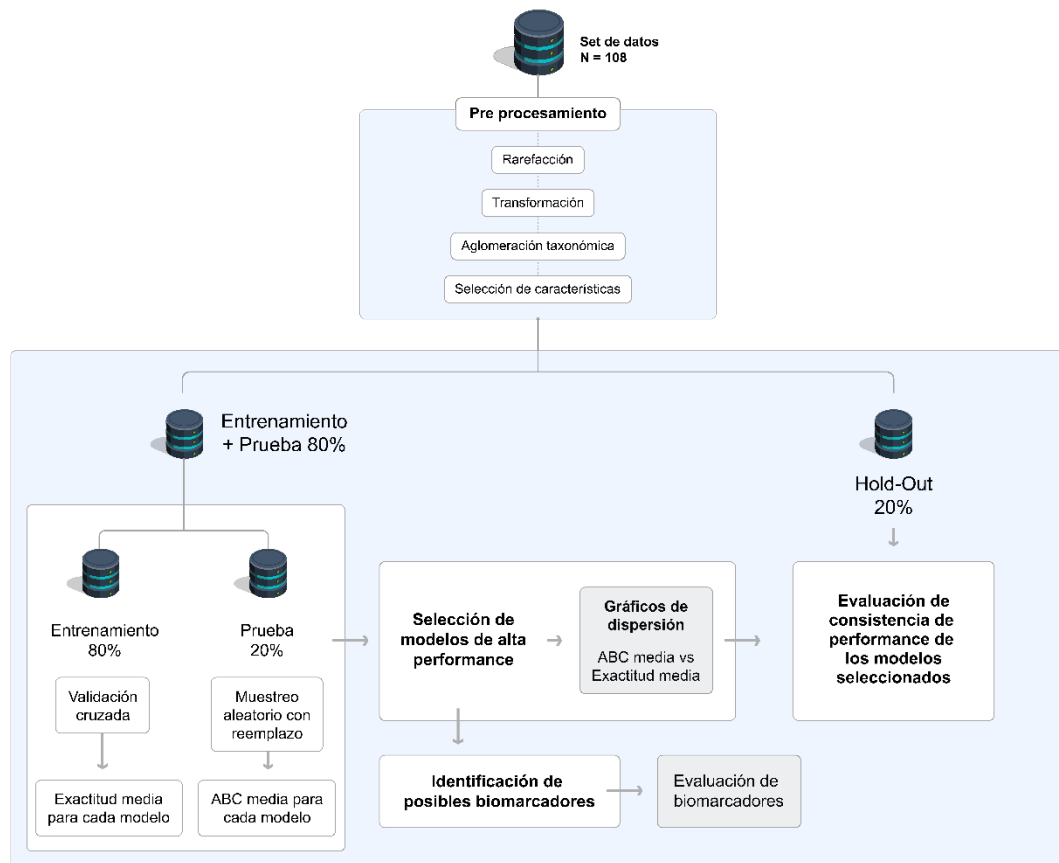


Diagrama 1: Esquema que representa el flujo de trabajo utilizado en el presente proyecto

Tabla 2: Distribución de fenotipos y el número de muestras de los diferentes sets de datos generados

	Número de Muestras	Fenotipo Delgado	Fenotipo Obeso
Dataset completo	108	64 (59.26%)	44 (40.74%)
Dataset de entrenamiento	69	41 (59.42%)	28 (40.58%)
Dataset de testeo	17	10 (58.82%)	7 (41.18%)
Dataset de Holdout	22	13 (59.09%)	9 (40.91%)

4.2. Efecto de la taxonomía sobre la performance de los modelos

Con el fin de evaluar el impacto de los diferentes niveles taxonómicos en la clasificación, se analizó el desempeño de los 10 modelos de clasificación (LVQ,

GBM, SVMI, SVMr, SLDA, NN, Rborist, xgbLinear, stepLDA, naive_bayes) a lo largo de toda la escala taxonómica, considerando los niveles de Phylum, Clase, Orden, Familia, Género y Especie. La Figura 1 muestra el número de características resultantes luego de cada aglomeración.

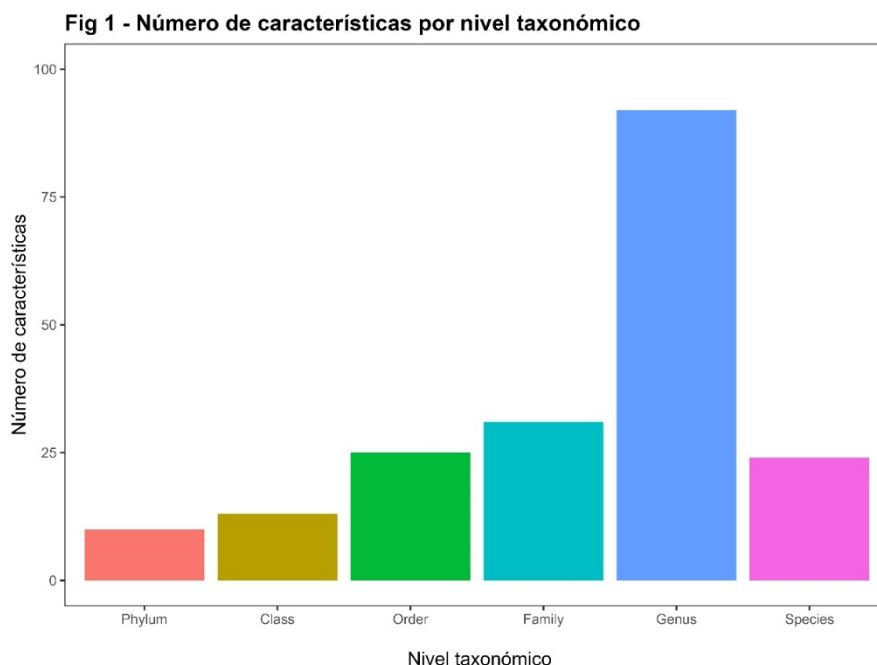


Figura 1: Barplot representando el número de características resultantes para cada nivel taxonómico luego de la aglomeración de ASVs del dataset original.

Phylum (10), Clase (13), Orden (25), Familia (31), Género (92) y Especie (24).

En primer lugar, se valoró la influencia de los niveles taxonómicos sobre cada modelo de manera individual (Fig. 2A). El análisis estadístico reveló que existen diferencias significativas en los valores de exactitud entre la mayoría de los niveles taxonómicos dentro de un mismo modelo de predicción. Es de considerar que los modelos SVMI, SVMr y Rborist presentaron una menor disparidad al comparar los diferentes niveles. Por lo contrario, los modelos LVQ y SLDA fueron los más influenciados por el nivel taxonómico utilizado.

En general, se obtuvieron los mejores valores de exactitud al trabajar con los niveles taxonómicos más específicos, como Género y Especie, mientras que los niveles de Phylum y Clase mostraron los peores resultados en la mayoría de los casos. Sin

embargo, paradójicamente, en el modelo SVMI, la mejor performance estuvo determinada tanto a nivel de Clase como de Género. La Tabla 3 presenta los valores de exactitud para el mejor y el peor nivel taxonómico de cada modelo y en todos los casos se encontraron diferencias estadísticamente significativas entre los mismos. Debido a la dimensionalidad de las comparaciones estadísticas realizadas en este análisis, los valores de exactitud para cada modelo y de significancia entre las comparaciones entre niveles taxonómicos en cada modelo son presentados en la Tabla Suplementaria 1A y 1B, respectivamente

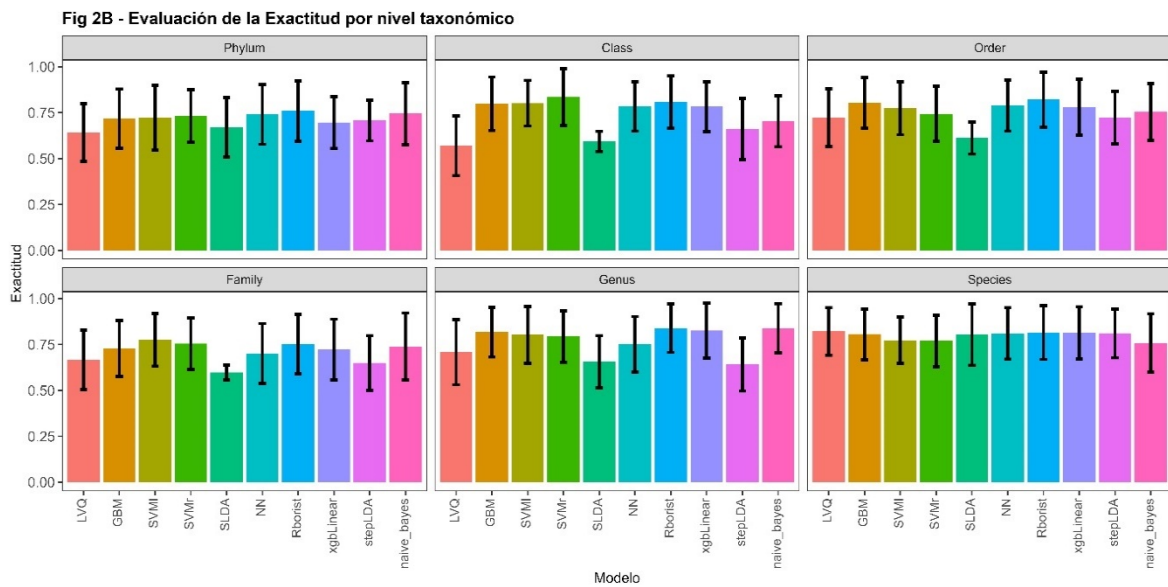
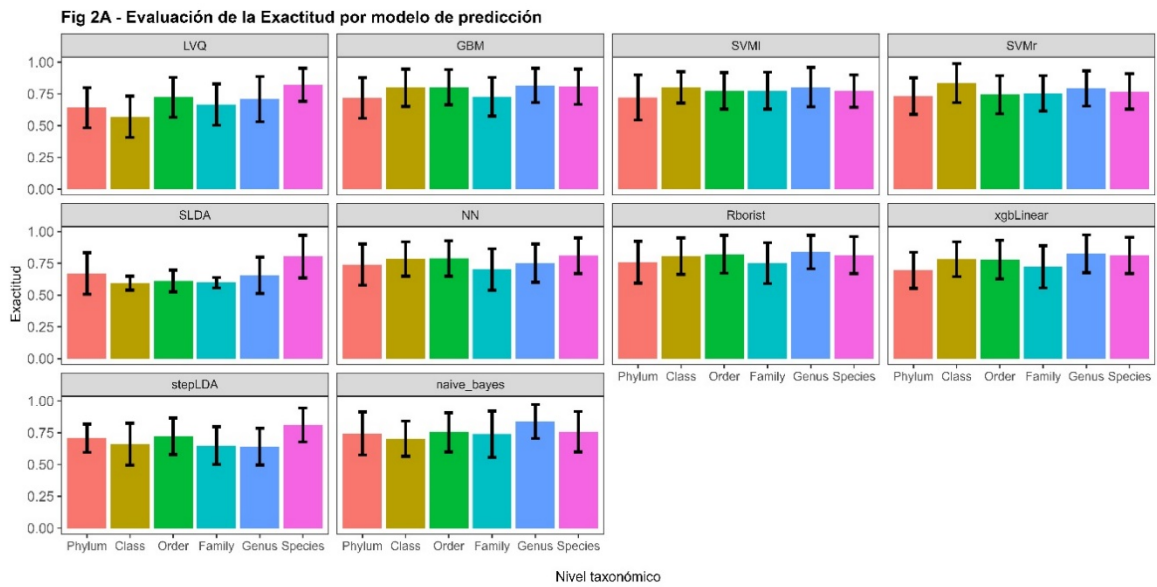


Figura 2: A - Evaluación de la exactitud. Barplot representando la exactitud media del proceso de entrenamiento de los modelos de cada nivel taxonómico en los modelos evaluados. Información expresada como media \pm desviación estándar. Cada color representa un nivel taxonómico diferente. **B** - Evaluación de la exactitud. Barplot representando la exactitud media del proceso de entrenamiento de cada modelo, discriminado por nivel taxonómico. Información expresada como media \pm desviación estándar. Cada color representa un modelo de predicción diferente.

Tabla 3: Influencia de los niveles taxonómicos sobre la performance del modelo. Comparación por modelo de predicción

Modelo	Exactitud (Media + DE)	Nivel Taxonómico
LVQ	0.57 \pm 0.16	Clase
	0.82 \pm 0.13	Especie
GBM	0.72 \pm 0.16	Phylum
	0.82 \pm 0.13	Género
SVMl	0.72 \pm 0.18	Phylum
	0.80 \pm 0.12 / 0.80 \pm 0.15	Clase / Especie
SVMr	0.73 \pm 0.14	Phylum
	0.83 \pm 0.15	Clase
SLDA	0.59 \pm 0.05	Clase
	0.80 \pm 0.17	Especie
NN	0.70 \pm 0.16	Familia
	0.81 \pm 0.14	Especie
Rborist	0.76 \pm 0.16	Familia
	0.84 \pm 0.13	Género
xgbLinear	0.70 \pm 0.14	Phylum
	0.83 \pm 0.15	Género
stepLDA	0.64 \pm 0.14	Género
	0.81 \pm 0.13	Especie
naive_bayes	0.70 \pm 0.14	Clase
	0.84 \pm 0.13	Género

Tabla 3: Describe el peor y el mejor nivel taxonómico de acuerdo al valor de exactitud medio obtenido por cada modelo.

Además, se llevó a cabo un análisis comparativo entre los distintos modelos a cada nivel taxonómico obteniéndose un patrón similar de desempeño en los 6 niveles taxonómicos evaluados (Fig. 2B). En la Tabla 4 se resumen los valores de exactitud para el mejor y peor modelo en cada nivel taxonómico. Resulta interesante destacar que, a nivel de especie, todos los modelos presentaron una performance

comparable, sin mostrar diferencias estadísticas significativas entre ellos. Sin embargo, a nivel de clase se observó una mayor cantidad de diferencias significativas entre los valores de exactitud obtenidos por los modelos. En términos generales, Rborist, mostró valores de exactitud más altos en la mayoría de los niveles taxonómicos (Phylum, Orden, Género y Especie). En cambio, SLDA y LVQ presentaron los peores desempeños en la mayoría de niveles.

Los valores de significancia estadísticas entre las comparaciones entre modelos dentro de un mismo nivel taxonómico son presentados en la Tabla Suplementaria 1C.

Tabla 4: Influencia de los niveles taxonómicos sobre la performance del modelo. Comparación por nivel taxonómico		
Nivel Taxonómico	Exactitud (Media + DE)	Modelo
Phylum	0.64 ± 0.16	LVQ
	0.76 ± 0.16	Rborist
Clase	0.57 ± 0.16	LVQ
	0.83 ± 0.15	SVMr
Orden	0.61 ± 0.09	SLDA
	0.82 ± 0.15	Rborist
Familia	0.60 ± 0.04	SLDA
	0.78 ± 0.14	SVMl
Género	0.64 ± 0.14	stepLDA
	0.84 ± 0.13 / 0.84 ± 0.13	Rborist / naive_bayes
Especie	0.76 ± 0.16	naive_bayes
	0.82 ± 0.13	LVQ

Tabla 4: Describe el peor y mejor modelo de acuerdo al valor de exactitud medio obtenido en cada nivel taxonómico.

4.3. Selección de características como estrategia de optimización de modelos

Con el objetivo de mejorar el rendimiento de los modelos de predicción se implementó una estrategia de selección de características relevantes, utilizando los métodos SAFS, SAFS.rf, RFE y SBF en cada nivel taxonómico. La Figura 3A y la Tabla 5 resumen el número de características finales obtenidas por cada método en cada nivel taxonómico. Para comprender mejor el impacto de su aplicación, la Figura 3B, muestra el porcentaje de disminución del número de características generado en cada nivel taxonómico. Estos datos demuestran que no todos los modelos de selección de características fueron efectivos para reducir el número de variables en

estudio. Específicamente, el método RFE no resultó efectivo a nivel de Phylum y Familia. Además, es importante destacar que los diferentes métodos de selección de características mostraron respuestas distintas incluso dentro del mismo nivel taxonómico. Por ejemplo, en el nivel de Especie, los métodos SAFS y SAFS.rf lograron reducir el número de características a 9, mientras que RFE lo hizo a 11 y SBF a 6, respectivamente. Por otro lado, solo SAFS.rf presentó un rendimiento comparable a lo largo de toda la escala taxonómica. En general, la mayor eliminación de variables se observa para todos los modelos en los niveles taxonómicos más específicos.

Utilizando esta información, generamos nuevos conjuntos de datos en los que nuevamente entrenamos los 10 modelos de predicción utilizados anteriormente.

Fig 3A - Número de características por modelo de selección en cada nivel taxonómico

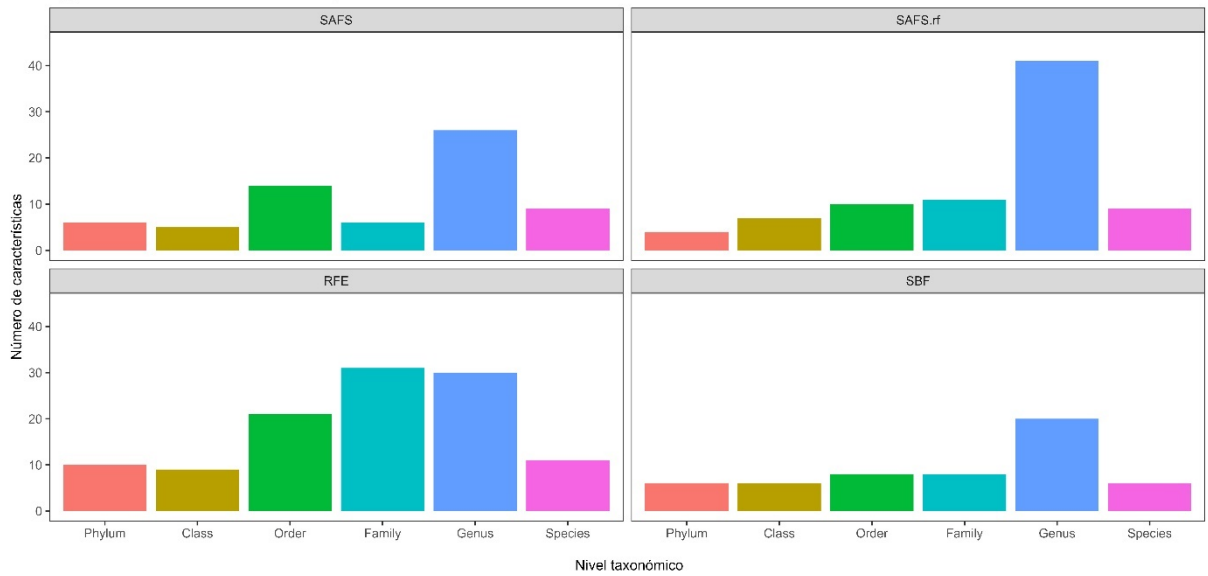


Fig 3B - Porcentaje de disminución de características por modelo de selección en cada nivel taxonómico

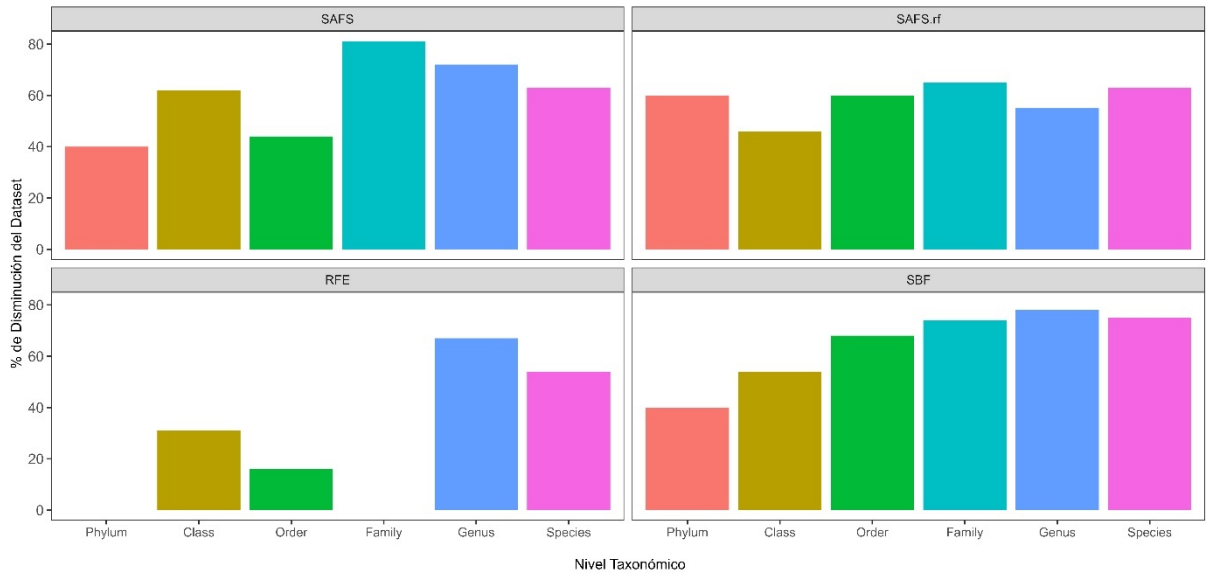


Figura 3: A - Barplot representando el número de características finales de cada dataset luego de la aplicación de métodos de selección discriminado por nivel taxonómico. **B** - Barplot representando el porcentaje de disminución de cada dataset luego de la aplicación de métodos de selección discriminado por nivel taxonómico.

Tabla 5: Número de características relevantes en cada dataset luego de la aplicación Métodos de selección de características.

	Phylum	Clase	Orden	Familia	Género	Especie
Dataset original	10	13	25	31	92	24
Dataset con SAFS	6	5	14	6	26	9
Dataset con SAFS.rf	4	7	10	11	41	9
Dataset con RFE	10	9	21	31	30	11
Dataset con SBF	6	6	8	8	20	6

4.4. Evaluación de la Exactitud post selección de características

La evaluación de la performance de clasificación de los modelos después de aplicar la selección de características reveló que esta estrategia no siempre mejoró la exactitud en comparación con los modelos originales. La Figura 4A muestra los valores de exactitud para cada modelo en diferentes niveles taxonómicos, diferenciados según el método de selección de características utilizado, se incluyen además los resultados de los modelos sin selección de características para comparar de una manera simple los cambios en la performance.

En general, se observaron diferencias tanto en los métodos de selección utilizados como en los niveles taxonómicos donde se aplicaron. Los resultados demuestran que no todos los métodos de selección son igualmente eficaces para mejorar la exactitud de las predicciones de los modelos. Interesantemente, aplicar estos métodos a niveles altos de la escala taxonómica genera una disminución en la exactitud de las predicciones, a excepción del modelo LVQ a nivel de clase. Este modelo, al ser utilizado en combinación con los métodos SAFS, RFE o SBF, presenta una mejora sustancial de su rendimiento por encima del 10% en todos los casos. Otro dato a resaltar es que el método SAFS.rf aplicado a nivel de Familia produjo un aumento del rendimiento de todos los modelos, excepto para el modelo LVQ. Algo similar fue observado para RFE, pero a menor escala y por el contrario SAFS produjo una disminución de la exactitud predictiva en todos los modelos a excepción de stepLDA. Además, a nivel de Orden, la mayoría de los modelos presenta una disminución en el rendimiento, independientemente del método de selección utilizado.

Es importante destacar que la aplicación de los 4 métodos de selección de características tiene un impacto significativo en el rendimiento de los modelos de predicción, especialmente en niveles taxonómicos más específicos, como la Familia, el Género y la Especie. En términos generales, se encontró que los métodos de SAFS.rf y RFE lograron mejorar significativamente la exactitud de los modelos. Estos métodos destacaron al considerar los diez mejores modelos en términos de exactitud de predicciones durante el entrenamiento. Para comprender mejor estos detalles la figura 4B presenta la diferencia porcentual entre el valor de exactitud medio del modelo con la aplicación de un método de selección en relación al mismo modelo sin método de selección. La Tabla 6 presenta el top 10 de modelos con mayor exactitud de predicción durante el entrenamiento considerando nivel taxonómico y método de selección de características. Debido a la alta dimensionalidad de los datos analizados, los valores de exactitud y las comparaciones estadísticas entre los modelos son presentados en las tablas Suplementaria 2A, 2B, 2C y 2D respectivamente.

Tabla 6: Top 10 de modelos de mejor rendimiento durante el entrenamiento			
Modelo	Método de selección	Nivel taxonómico	Exactitud (Media ± DE)
xgbLinear	SAFS.rf	Familia	0.92± 0.10
SVMl	SAFS.rf	Familia	0.92± 0.10
Rborist	SAFS.rf	Familia	0.91± 0.11
GBM	SAFS.rf	Familia	0.91± 0.10
Rborist	RFE	Género	0.90± 0.10
SVMl	RFE	Género	0.88± 0.11
Rborist	RFE	Especie	0.87± 0.12
NN	SAFS.rf	Familia	0.87± 0.11
SVMr	SBF	Género	0.87± 0.12

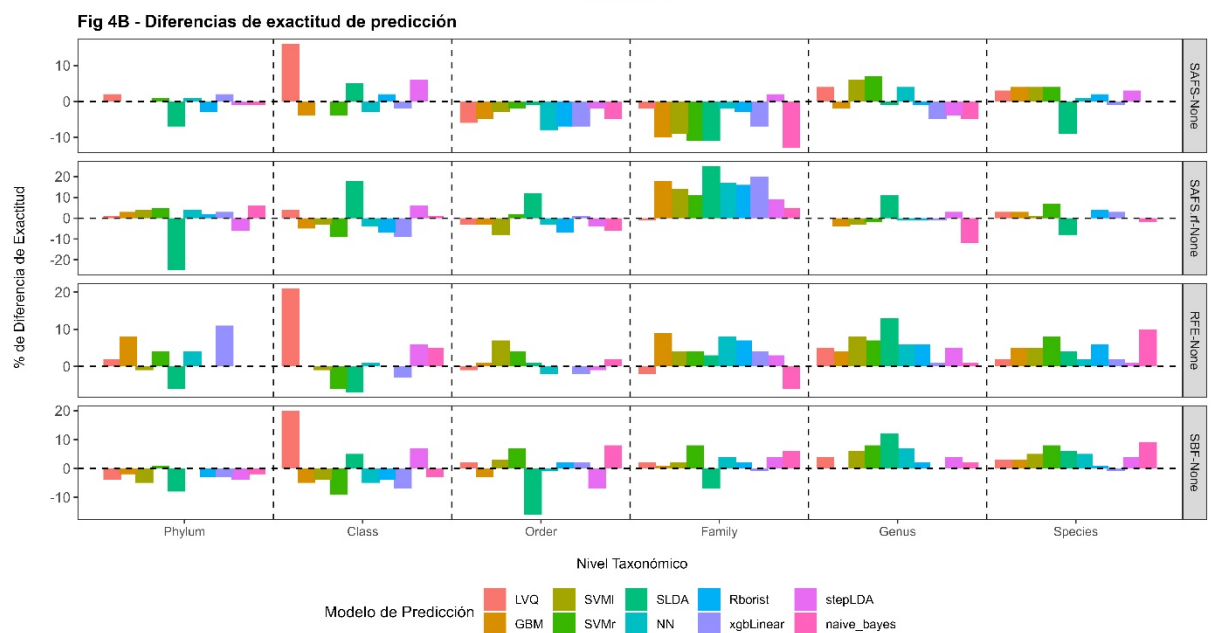
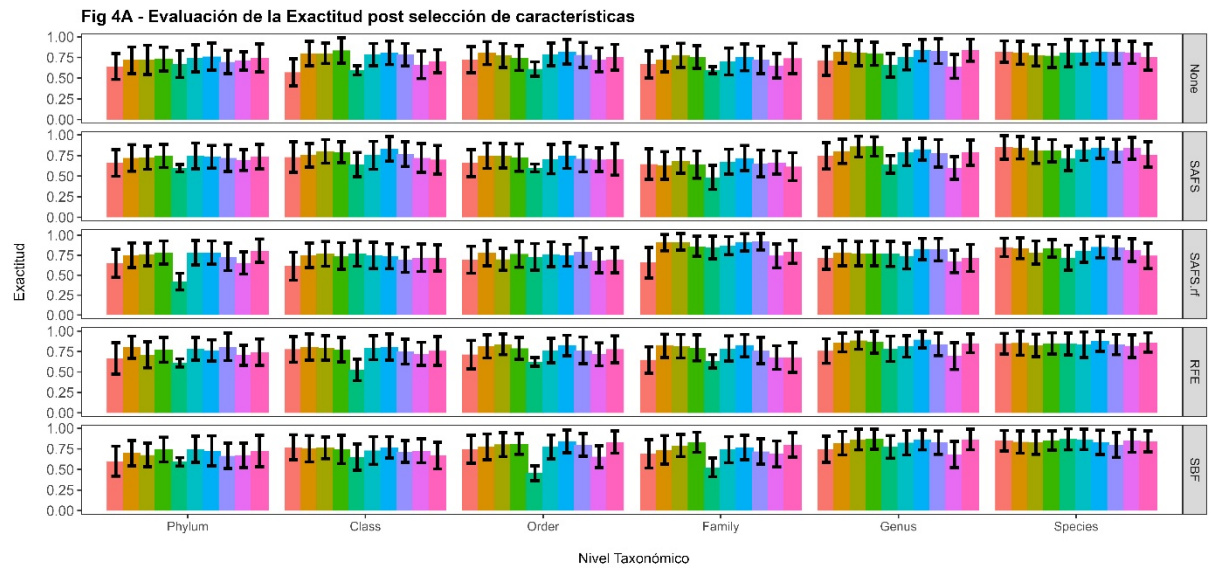


Figura 4: A - Evaluación de la exactitud post selección de características. Gráfico de barras representando la exactitud media luego de proceso de validación cruzada de 10 repeticiones por 10 particiones para cada modelo en los diferentes niveles taxonómicos utilizando diferentes métodos de selección de características. Información mostrada como media \pm desviación estándar. **B** - Barplot representando la diferencia porcentual entre la exactitud media de cada modelo con aplicación de un método de selección en relación al mismo modelo sin selección de características. Cada color representa un método de selección de características diferente.

4.5. Selección de modelos más aptos de clasificación del fenotipo delgado/obeso

Con el objetivo de obtener una evaluación más sólida de los modelos para clasificar el fenotipo delgado/obeso, los modelos previamente entrenados fueron evaluados utilizando un nuevo conjunto de datos de prueba (Dataset de evaluación). Como métrica de evaluación, se utilizó el valor medio del ABC, el cual representa el rendimiento promedio de los modelos en el conjunto de datos empleado.

La Figura 5A, muestra los valores de ABC para cada modelo discriminado por nivel taxonómico y método de selección. Como se puede observar, el rendimiento de los modelos de clasificación varía según el nivel de taxonomía y el método de selección de características utilizado. Los mismos presentan un rango de valores de ABC entre 0.5 y 1, sugiriendo que la mayoría de los modelos tienen cierta capacidad para clasificar las muestras mejor que una clasificación aleatoria. Algunos modelos muestran un rendimiento más consistente y los diferentes métodos de selección de características pueden mejorar el rendimiento de los modelos en ciertos casos.

Dada la dimensionalidad de las comparaciones estadísticas del presente análisis, los valores de ABC y los valores de significancia son presentados en las Tablas Suplementaria 3A, 3B, 3C y 3D respectivamente.

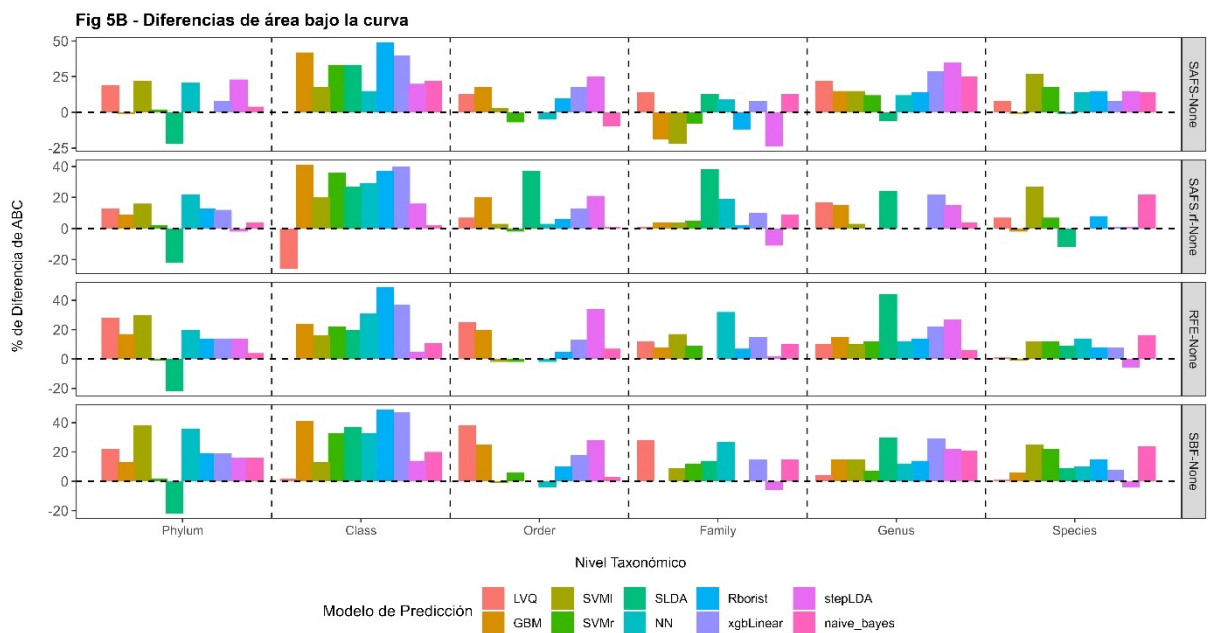
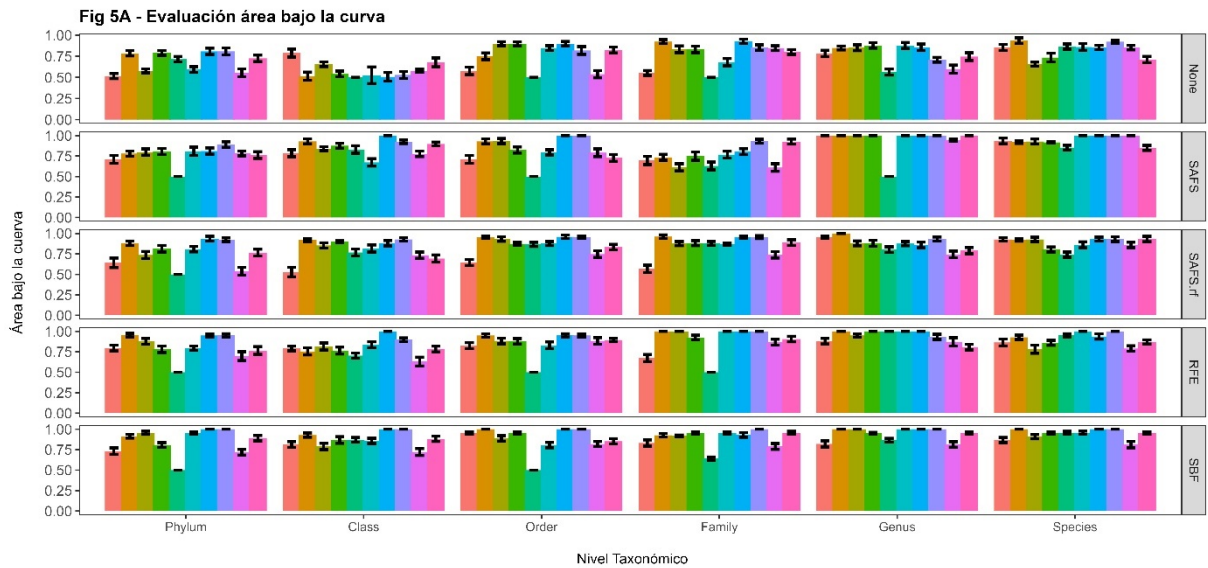


Figura 5: A - Barplot representando la media del área bajo la curva (ABC) para cada modelo en los diferentes niveles taxonómicos utilizando diferentes métodos de selección. Información presentada como media \pm desviación estándar. **B** - Barplot representando la diferencia porcentual entre la ABC media de cada modelo con aplicación de un método de selección en relación al mismo modelo sin selección de características. Cada color representa un modelo de predicción diferente.

4.6. Estrategia gráfica para una evaluación integral de modelos de predicción

El uso de métricas como exactitud o ABC para la selección del mejor modelo de predicción, genera discrepancias en los resultados obtenidos, debido a que cada métrica resalta diferentes aspectos del rendimiento del modelo; esto puede conducir a situaciones en las que un modelo tenga una mayor exactitud, pero menor ABC, o viceversa.

Con el fin de abordar esta variabilidad y obtener una comprensión más completa de la performance de los modelos utilizados, se diseñó una estrategia gráfica que permite discriminar el mejor modelo teniendo en cuenta estas métricas, como se describió en Materiales y Métodos. Los modelos representados en el cuadrante con valores de exactitud y ABC entre 0.85 y 1 se clasificaron como de alta performance y se consideraron aptos para el propósito establecido (Figura 6).

De los 300 modelos evaluados, únicamente 24 fueron clasificados como de alta performance. Es importante remarcar que los mismos fueron exclusivamente modelos aglomerados a nivel de familia, género y especie. La **Tabla 7** resume las métricas obtenidas para estos modelos.

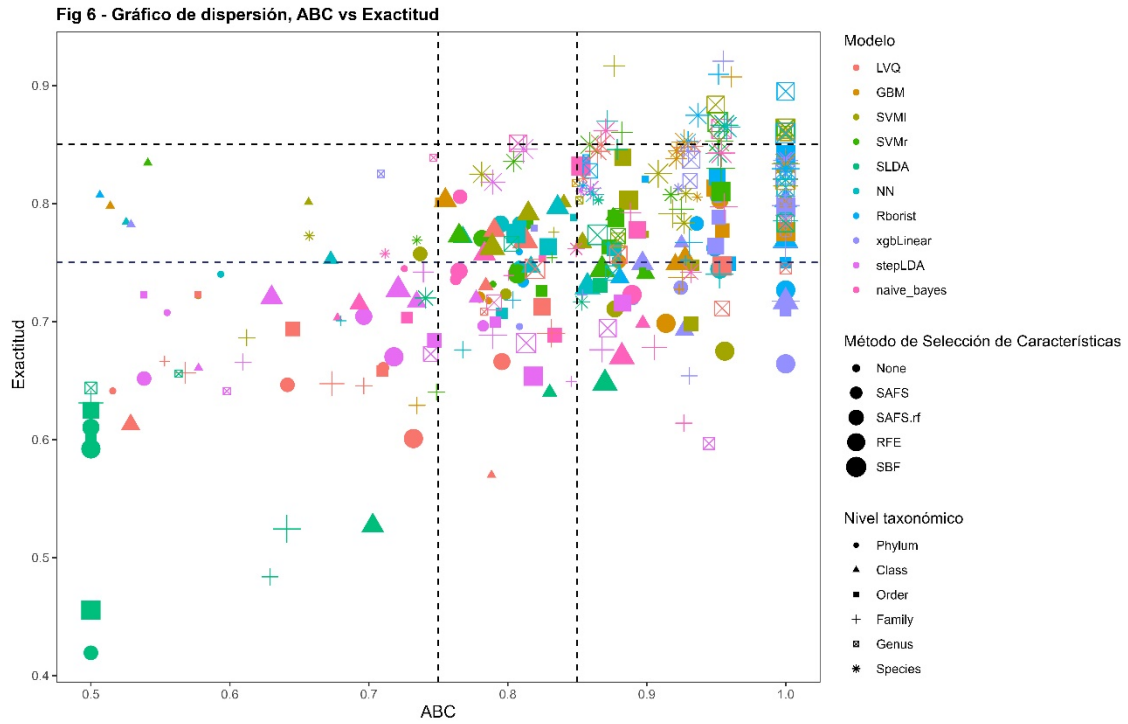


Figura 6: Selección de modelos más aptos para clasificación del fenotipo delgado/obeso. Gráfico de dispersión que evalúa la Exactitud media y la ABC media de cada modelo evaluado, en los diferentes niveles taxonómicos y con o sin aplicación de Métodos de selección de características. Cada punto representa la intersección entre el valor medio de Exactitud y ABC para cada modelo.

Tabla 7: Métricas de modelos de alta performance

Modelo	Nivel Taxonómico	Método de selección	Exactitud	ABC
GBM	Familia	SAFS.rf	0.91	0.96
SVMI	Familia	SAFS.rf	0.92	0.88
SVMr	Familia	SAFS.rf	0.86	0.88
NN	Familia	SAFS.rf	0.87	0.87
Rborist	Familia	SAFS.rf	0.91	0.95
xgbLinear	Familia	SAFS.rf	0.92	0.96
SVMI	Género	SAFS	0.86	1.00
SVMr	Género	SAFS	0.86	1.00
GBM	Género	RFE	0.86	1.00
SVMI	Género	RFE	0.88	0.95
SVMr	Género	RFE	0.86	1.00
Rborist	Género	RFE	0.90	1.00
SVMI	Género	SBF	0.86	1.00
SVMr	Género	SBF	0.87	0.95
Rborist	Género	SBF	0.86	1.00
naive_bayes	Género	SBF	0.86	0.95
Rborist	Especie	SAFS.rf	0.85	0.93
GBM	Especie	RFE	0.85	0.93
Rborist	Especie	RFE	0.87	0.94
naive_bayes	Especie	RFE	0.86	0.87
LVQ	Especie	SBF	0.85	0.87
SVMr	Especie	SBF	0.85	0.95
SLDA	Especie	SBF	0.87	0.96
NN	Especie	SBF	0.86	0.96

4.7. Validación de los modelos de mejor performance en el dataset independiente.

Una vez identificados los modelos de alta performance, se procedió a evaluar su capacidad predictiva en un conjunto de datos independiente (Hold-out) con el objetivo de determinar cuál de ellos mantiene un rendimiento consistente, independientemente del set de datos al que sea aplicado. Durante esta etapa de evaluación se consideraron diversos parámetros, tales como Precisión, Exactitud, Sensibilidad (Recall), Tasa de Falsos Positivos, Tasa de Falsos Negativos, F1 Score y ABC (Tabla 8).

Al analizar estos parámetros en conjunto, se encontró que el modelo "Rborist" con la aplicación de "SAFS.rf" como método de selección de características a nivel de

especie presentó los mejores resultados de rendimiento. Este modelo mantuvo valores de Exactitud y ABC por encima de 0.85 (Figura 7) y demostró bajos niveles de Tasa de Falsos Positivos y Negativos.

En base a este análisis integral, se concluye que el modelo "Rborist" con la aplicación de "SAFS.rf" a nivel de especie es el más adecuado para el estudio de la composición de la MI como predictor del fenotipo delgado/obeso en modelos de obesidad en ratones. Este modelo considera las especies: *Bacteroides sartorii*, *Bacteroides thetaiotaomicron*, *Mobilitalea sibirica*, *Lachnospiraceae UCG-006 bacterium*, *Rikenella microfusum*, *Faecalibaculum rodentium*, *Helicobacter mastomyrinus*, *Bacteroides vulgatus* y *Lachnospiraceae NK4A136 group bacterium* en el análisis.

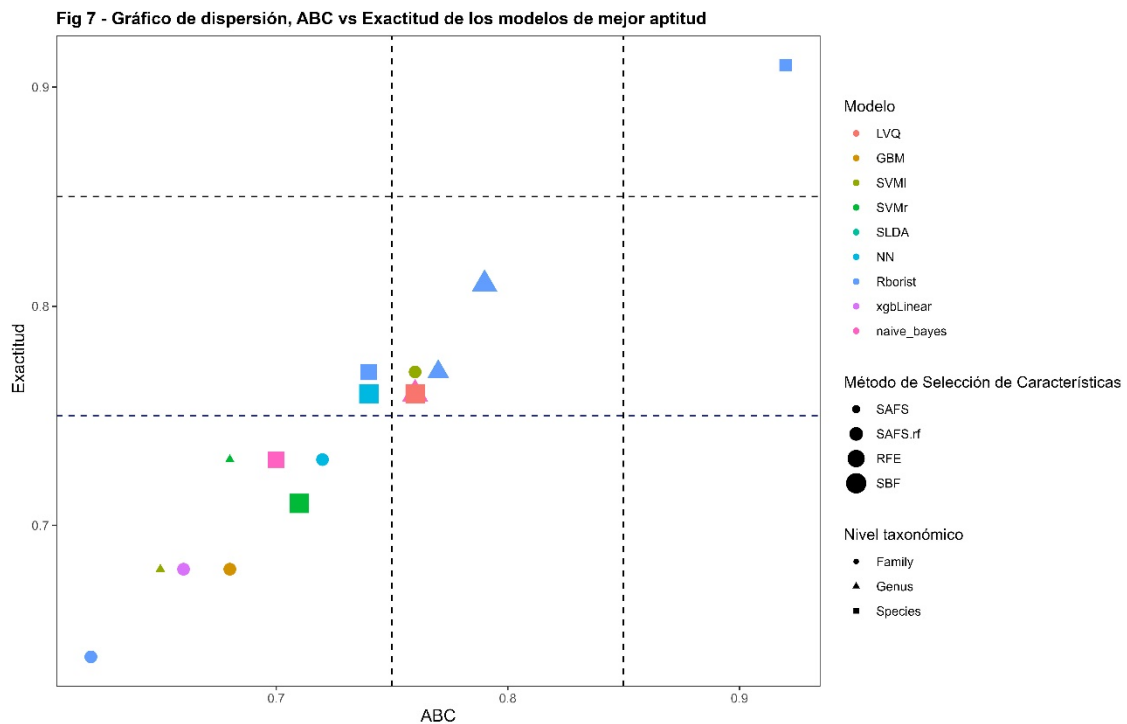


Figura 7: Gráfico de dispersión que evalúa la Exactitud media y la ABC media de los 24 modelos con mejor performance, en un set de datos independiente, en los diferentes niveles taxonómicos con o sin aplicación de Métodos de selección de características. Cada punto representa la intersección entre el valor medio de Exactitud y ABC para cada modelo. El modelo Rborist con aplicación de SAFS.rf a

nivel de especie fue el único que mantuvo sus parámetros de performance al ser evaluado en un set de datos independiente.

Tabla 8: Métricas de performance obtenidas en la etapa de validación

Modelo	Método de selección	Nivel Taxonómico	Precisión	Exactitud	Recall	Tasa de FP	Tasa de FN	F1 Score	ABC
GBM	SAFS.rf	Familia	0.60	0.68	0.31	0.31	0.33	0.41	0.68
SVMr	SAFS.rf	Familia	0.75	0.77	0.15	0.15	0.33	0.26	0.76
SVMI	SAFS.rf	Familia	0.56	0.64	0.31	0.31	0.44	0.40	0.62
NN	SAFS.rf	Familia	0.67	0.73	0.23	0.23	0.33	0.34	0.72
Rborist	SAFS.rf	Familia	0.56	0.64	0.31	0.31	0.44	0.40	0.62
xgbLinear	SAFS.rf	Familia	0.62	0.68	0.23	0.23	0.44	0.34	0.66
SVMr	SAFS	Género	0.67	0.68	0.15	0.15	0.56	0.25	0.65
SVMI	SAFS	Género	0.80	0.73	0.08	0.08	0.56	0.14	0.68
GBM	RFE	Género	0.70	0.77	0.23	0.23	0.22	0.35	0.77
SVMr	RFE	Género	0.70	0.77	0.23	0.23	0.22	0.35	0.77
SVMI	RFE	Género	0.70	0.77	0.23	0.23	0.22	0.35	0.77
Rborist	RFE	Género	0.70	0.77	0.23	0.23	0.22	0.35	0.77
SVMr	SBF	Género	0.86	0.81	0.08	0.08	0.33	0.15	0.79
SVMI	SBF	Género	0.86	0.81	0.08	0.08	0.33	0.15	0.79
Rborist	SBF	Género	0.86	0.81	0.08	0.08	0.33	0.15	0.79
naive_bayes	SBF	Género	0.70	0.76	0.25	0.25	0.22	0.37	0.76
Rborist	SAFS.rf	Especie	0.82	0.91	0.15	0.15	0.00	0.26	0.92
GBM	RFE	Especie	0.71	0.73	0.15	0.15	0.44	0.25	0.70
Rborist	RFE	Especie	0.83	0.77	0.08	0.08	0.44	0.14	0.74
naive_bayes	RFE	Especie	0.71	0.73	0.15	0.15	0.44	0.25	0.70
LVQ	SBF	Especie	0.70	0.76	0.25	0.25	0.22	0.37	0.76
SVMI	SBF	Especie	0.67	0.71	0.25	0.25	0.33	0.36	0.71
SLDA	SBF	Especie	0.83	0.76	0.08	0.08	0.44	0.15	0.74
NN	SBF	Especie	0.83	0.76	0.08	0.08	0.44	0.15	0.74

Tabla 8: Métricas de performance obtenidas por cada modelo de alta performance en el análisis de validación, en rojo se remarca el único método que mantuvo la performance independientemente del dataset utilizado. FP = Falso Positivo, FN = Falso Negativo, ABC = área bajo la curva

4.8. Identificación de Taxas microbianas asociadas al fenotipo Delgado / Obeso:

Con el objetivo de identificar posibles biomarcadores bacterianos relacionados con los fenotipos Delgado/Obeso, se determinaron las taxas de mayor relevancia utilizadas por los 24 modelos seleccionados. Según los resultados previamente descritos, se observó que diferentes métodos de selección de características y solo tres niveles taxonómicos (Familia, Género y Especie) estuvieron involucrados en los modelos de

mejor rendimiento. A continuación, se detallan los modelos seleccionados para cada nivel taxonómico junto con los métodos de selección correspondientes:

A nivel de Familia, los modelos GBM, SVMr, SVMl, NN, Rborist y xgbLinear con la aplicación del método SAFS.rf fueron seleccionados.

A nivel de Género, los modelos SVMl y SVMr con la aplicación del método SAFS; GBM, SVMl, SVMr y Rborist con la aplicación del método RFE, y SVMl, SVMr, Rborist y Naive_bayes con la aplicación del método SBF fueron seleccionados.

A nivel de Especie, los modelos Rborist con la aplicación del método SAFS.rf; GBM, Rborist y Naive_bayes con la aplicación del método RFE, y LVQ, SVMr, SLDA y NN con la aplicación del método SBF fueron seleccionados. Es interesante destacar que SVMr, SVMl, y Rborist son reiteradamente seleccionados a medida que se avanza en la escala taxonómica.

Continuando con el proceso de selección de biomarcadores bacterianos, se realizó un análisis de intersección en cada nivel taxonómico, identificando las variables que fueron consideradas por todos los modelos de predicción utilizados. Este procedimiento, presentó los siguientes resultados:

A nivel de Familia, 11 taxones fueron compartidos por todos los modelos. En cuanto al Género, se identificaron 6 taxones en común y a nivel de Especie, sólo 2 taxones estuvieron presentes en todos los modelos, remarcando su potencial asociación con el fenotipo investigado (Tabla 9).

Tabla 9: Posibles biomarcadores identificados como predictores del fenotipo Delgado/Obeso

Nivel taxonómico		
Familia	Género	Especie
<i>Prevotellaceae</i>	<i>Rikenellaceae RC9 gut group</i>	<i>Bacteroides sartorii</i>
<i>Lachnospiraceae</i>	<i>Candidatus Saccharimonas</i>	<i>Muribaculum intestinale</i>
<i>Saccharimonadaceae</i>	<i>Lachnospiraceae NK4A136 group</i>	
<i>Sutterellaceae</i>	<i>Acetatifactor</i>	
<i>Marinifilaceae</i>	<i>Eubacterium nodatum group</i>	
<i>Tannerellaceae</i>	<i>Shuttleworthia</i>	
<i>Anaerovoracaceae</i>		
<i>Mycoplasmataceae</i>		
<i>Christensenellaceae</i>		
<i>Eubacterium coprostanoligenes group</i>		
<i>Veillonellaceae</i>		

4.9. Evaluación del potencial discriminativo de los biomarcadores bacterianos identificados

A partir de los 19 taxones identificados previamente, se realizó un proceso de filtrado en el objeto phyloseq original, el cual consta de 1757 ASVs. Este procedimiento generó un nuevo objeto que incluye únicamente 240 ASVs asociadas a los taxones seleccionados a partir de los modelos de predicción de mejor rendimiento.

Los resultados del PCA demostraron que la utilización de los biomarcadores identificados presenta un mayor potencial para discriminar los fenotipos en comparación con el uso del total de ASVs (Figura 8). Estos resultados fueron confirmados por el análisis de permutest ($p = 0.001$ vs $p = 0.797$), respaldando la relevancia de los biomarcadores identificados en la clasificación de los fenotipos Delgado/Obeso.

Adicionalmente, para determinar posibles diferencias entre los grupos de estudio se comparó la abundancia relativa de cada uno de los taxones identificados (Figura 9). Este análisis permitió identificar patrones distintivos en la distribución de las taxas seleccionadas entre los grupos, aunque en 7 de las 11 familias estudiadas; *Prevotellaceae*, *Sutterellaceae*, *Marinifilaceae*, *Tannerellaceae*, *Mycoplasmataceae*, *Christensenellaceae* y *Eubacterium coprostanoligenes group*, no se observó abundancia diferencial entre grupos.

Estos hallazgos respaldan la importancia y utilidad de los métodos de ML para detectar relaciones no lineales y no relacionadas a la abundancia de taxas en la discriminación de los fenotipos.

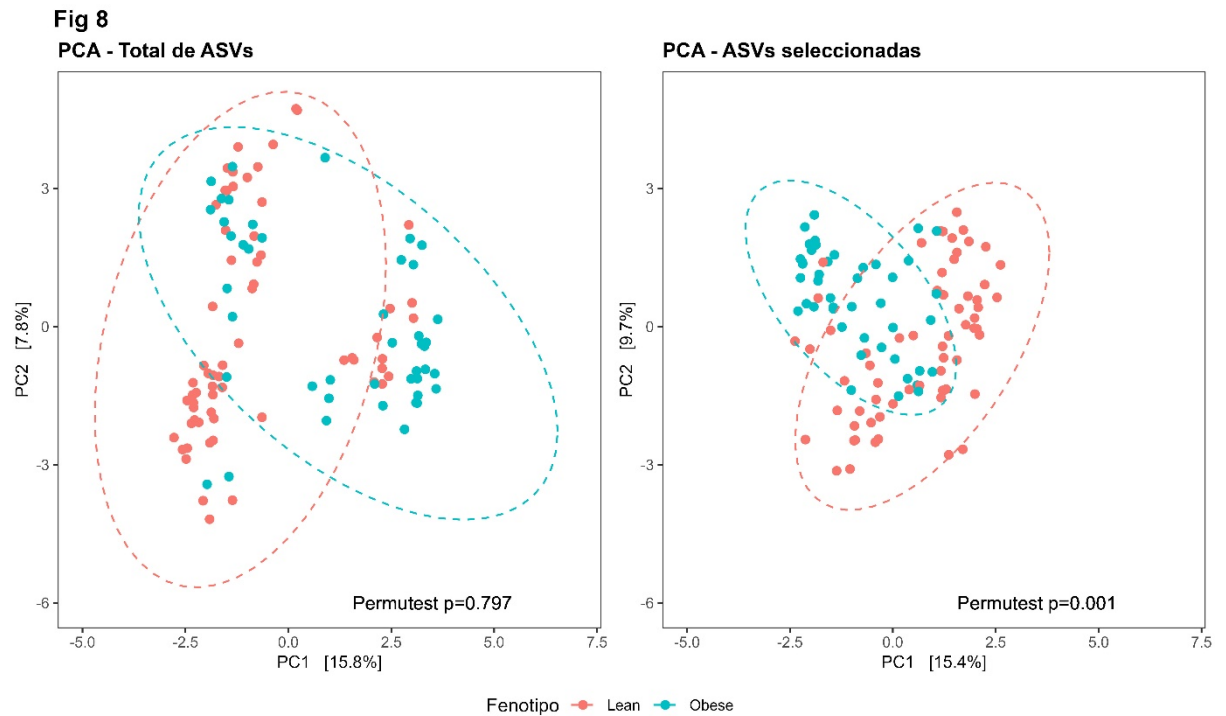


Figura 8: Análisis de Componentes principales (PCA) para el dataset original y el dataset generado a partir de las ASVs seleccionadas como posibles biomarcadores. El análisis de permutación estadística significativo para el dataset de ASVs seleccionadas respalda la relevancia de los biomarcadores identificados como predictores de los fenotipos en estudio.

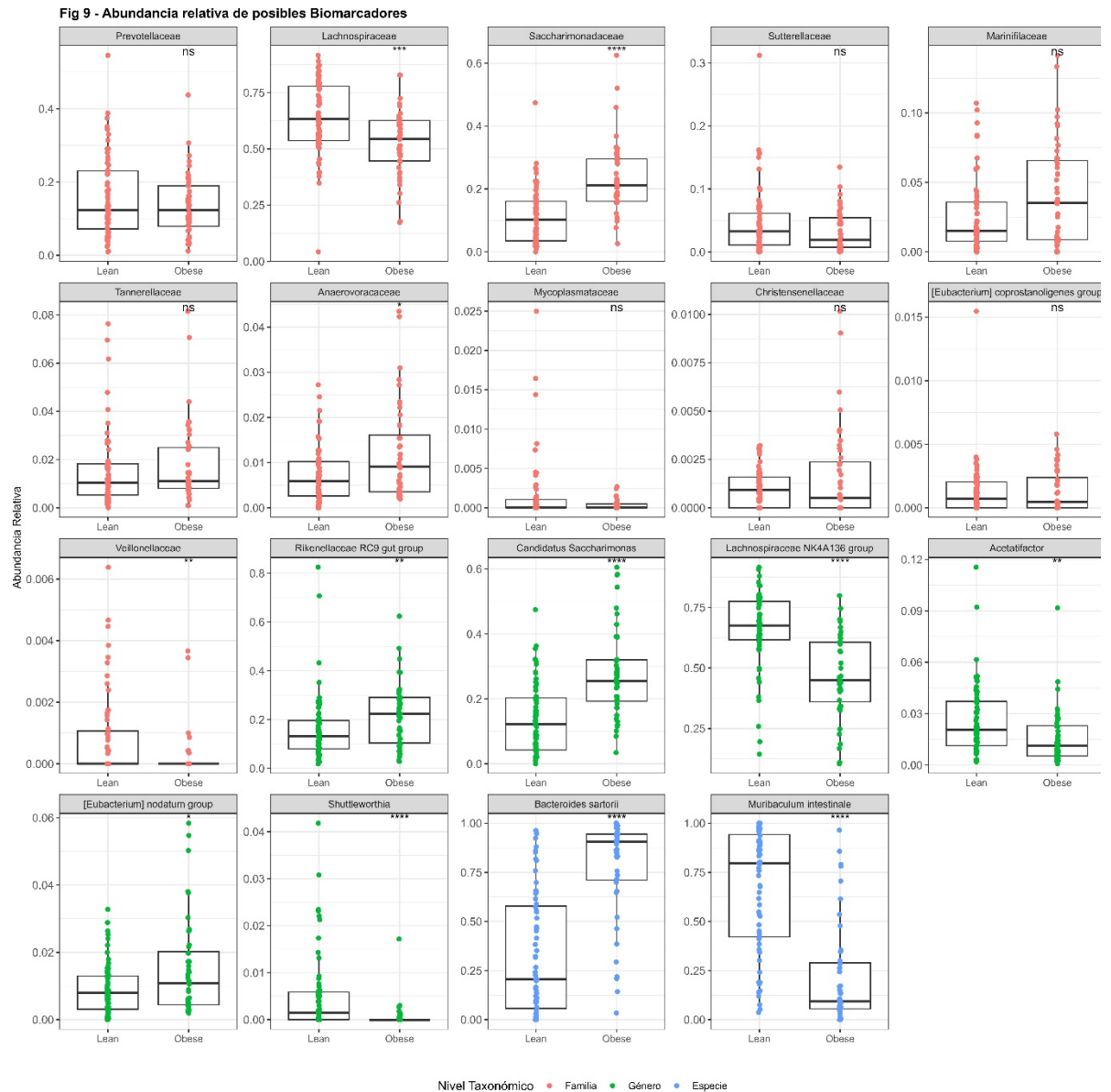


Figura 9: Boxplot representado la abundancia relativa de cada biomarcador potencial observado en cada Fenotipo (Delgado vs Obeso). Las líneas negras continuas indican las medianas, y los límites superior e inferior de la caja representan los cuartiles 25 y 75 %. Los valores atípicos se indican como círculos y representan muestras que se encuentran fuera de los cuartiles 10 y 90 %. Las diferencias significativas se establecieron mediante la prueba de rangos con signo de Wilcoxon (*: $p \leq 0,05$, **: $p \leq 0,01$, ***: $p \leq 0,001$, ****: $p \leq 0,0001$), ns indica que no se encontró significancia estadística. Los diferentes colores denotan el nivel taxonómico correspondiente a cada biomarcador identificado.

5. DISCUSIÓN:

En los últimos años, cada vez más evidencias demuestran el papel fundamental de la MI en la salud humana. Sin embargo, aún se presentan desafíos al momento de definir la conexión entre alteraciones de la MI con el desarrollo de ciertas enfermedades, debido a la complejidad y dimensionalidad de los datos generados por las metodologías utilizadas para su estudio. En este contexto, los modelos de ML han surgido como una herramienta poderosa para analizar y comprender el rol de la estructura de la MI en estados de salud y enfermedad. Aunque estos enfoques han aportado resultados prometedores, es fundamental abordar los problemas de estandarización de los modelos, metodologías y tamaño y estructura de datos utilizados, así como también la interpretabilidad de los resultados obtenidos, para garantizar su robustez, reproducibilidad y aplicación translacional en la práctica clínica. Por estos motivos y con el objetivo de tener una mejor comprensión de la performance de distintos modelos de clasificación aplicados al estudio de microbiota, evaluamos exhaustivamente 10 modelos de predicción considerando tres factores de relevancia que pueden influir la performance de los mismos cuando se los aplica a datos metagenómicos: 1) los niveles taxonómicos, 2) los modelos de predicción propiamente dichos y 3) los métodos de selección de características relevantes.

Nuestros estudios revelaron diferencias significativas en el rendimiento de los modelos de predicción cuando se los evalúa en diferentes niveles taxonómicos. Además, observamos que la exactitud de la predicción mejora al ser aplicados en niveles más específicos, como Género y Especie. Estos hallazgos son consistentes con otros estudios donde se ha observado que la performance de los modelos mejora a medida que se avanza en la escala taxonómica. En particular, el nivel de género mostró los niveles más altos de exactitud, siendo este el nivel más específico estudiado (Bang et al., 2019). Por otro lado, resulta relevante que los modelos de predicción no responden de manera idéntica a un mismo conjunto de datos. Y este patrón se observa a medida que se avanza en la escala taxonómica. Esto sugiere que la aplicabilidad y el rendimiento de los mismos pueden variar según la composición específica de los sets de datos. Song y Zhou, han encontrado un perfil similar al evaluar el rendimiento de los modelos utilizando el Score F1 (Song y Zhou, 2023). Esto abre las puertas a futuras

investigaciones destinadas a evaluar y comprender mejor la aplicabilidad de los modelos en función de la composición de los datos. La exploración de diferentes conjuntos de datos de microbiota permitirá identificar patrones y características específicas que influyen en el rendimiento de los modelos. Es importante destacar también que, en nuestra evaluación, todos los modelos presentaron una performance comparable a nivel de especie, sin mostrar diferencias estadísticas significativas entre ellos. En términos generales, estos resultados se pueden fundamentar desde una perspectiva biológica. La taxonomía permite clasificar los organismos según su parentesco evolutivo, y a niveles taxonómicos más específicos, los microorganismos tienden a tener composición genética y características funcionales más similares (Zhou y Gallins, 2019). Esto implica que comparten vías metabólicas, características fisiológicas y mecanismos de interacción con el huésped. Además, los microorganismos presentes en un determinado nicho pueden interactuar entre sí, con el huésped y con el entorno de diversas formas (Martín et al., 2014). Al utilizar modelos de ML en los niveles taxonómicos más bajos, estos pueden capturar indirectamente patrones estructurales relacionados a estas interacciones, lo que favorece una mejor performance en las predicciones. Por otro lado, la abundancia de los taxones microbianos puede variar en diferentes fenotipos o condiciones de salud (Qin et al., 2010). En niveles más específicos, pueden existir cambios sutiles en la abundancia que los modelos de predicción pueden detectar, lo que favorece un mayor poder discriminativo a estos niveles

Además de considerar las relaciones entre jerarquías en la escala taxonómicas, también aplicamos métodos de selección de características con la intención de mejorar el rendimiento de los modelos. La selección de características es un proceso crucial que busca identificar un subconjunto relevante y discriminativo de características para la predicción de fenotipos, además evitar el sobreajuste y mejorar el rendimiento del modelo (Hernández Medina et al., 2022; Saeys et al., 2007).

Nuestros resultados demuestran que no todos los métodos de selección son igualmente eficaces para mejorar la exactitud de las predicciones de los modelos. En general, aplicar estos métodos a niveles altos de la escala taxonómica genera un aumento no significativo o una disminución en la exactitud de las predicciones, pero tiene un impacto positivo significativo en niveles taxonómicos más específicos, como

Familia, Género y Especie en la mayoría de los modelos. Resultados similares han sido obtenidos por otros investigadores, Bang y col., encontraron que la aplicación de técnicas de *forward selection* (selección hacia adelante) y *backward elimination* (eliminación hacia atrás) mejoran la exactitud predictiva a nivel de Orden, Familia y Género, pero no resultan en cambios significativos a nivel de Phylum y Clase. Al igual que lo presentado en el trabajo, la remoción de características fue más efectiva a medida que se avanza en la escala taxonómica (Bang et al., 2019). Este efecto puede atribuirse a la considerable dimensionalidad inicial de estos niveles, conformados por 31, 92 y 24 taxas, respectivamente y como se explicó anteriormente a la existencia de relaciones entre las variables. Sin duda, la aplicación de los métodos de selección de características, reduce el número de variables predictoras evitando la multicolinealidad en los modelos mejorando así su capacidad predictiva.

Con el objetivo de aumentar la solidez de nuestra evaluación realizamos una evaluación de todos los modelos previamente entrenados en un nuevo dataset. Y en esta ocasión la métrica de evaluación utilizada fue ABC. Los resultados demostraron que al igual que la exactitud de la predicción determinada durante el entrenamiento de los modelos, el ABC varía según el nivel taxonómico utilizado, obteniendo valores más altos a niveles taxonómicos más específicos. Por otro lado, la aplicación de métodos de selección mejora de manera considerable el ABC de casi todos los modelos en todos los niveles taxonómicos, pero con respuestas diferenciales dependiendo de los 3 factores bajo estudio, nivel taxonómico, modelo predictivo y método de selección. Dicho aumento indica que los modelos han mejorado en su capacidad para realizar clasificaciones precisas y discriminar entre las diferentes clases de manera más efectiva. Esto se debe a que los métodos de selección de características permiten a los modelos de predicción enfocarse en las características más relevantes y discriminatorias, lo que a su vez mejora la sensibilidad y la especificidad de las predicciones.

El análisis Hold-out reveló que el modelo Rborist con el método de selección de características SAFS.rf, aplicado a nivel de especie, mantuvo un alto rendimiento consistente en todas las etapas de evaluación. A diferencia de los otros modelos entrenados, este modelo mostró ser menos influenciado por la composición y estructura de los conjuntos de datos utilizados. Esta constatación se encuentra

respaldada por investigaciones previas en diversas enfermedades, donde Random Forest, junto con el empleo de métodos de selección de características, demostró ser el modelo con mejor rendimiento (Aryal et al., 2020; Song y Zhou, 2023; Topçuoğlu et al., 2020)

Por su parte, la variabilidad en la respuesta de los modelos en diferentes datasets destaca una vez más, la necesidad de futuras investigaciones para evaluar y ajustar los modelos según la composición de los datos de microbiota, lo cual permitirá obtener resultados más confiables y precisos. Debemos resaltar también que la estrategia gráfica empleada, es una herramienta útil que permite identificar de manera simple y visual aquellos modelos que tuvieron mejor rendimiento durante el proceso de evaluación. Mediante dicho gráfico pudimos identificar 24 modelos de alta performance sobre los 300 valorados. Esta estrategia permitió evaluar simultáneamente múltiples parámetros, en primer lugar, los 3 factores que pueden influir en la performance de las predicciones (nivel taxonómico, Modelo de predicción y método de selección de variables) y en segundo lugar parámetros de rendimiento, en este caso Exactitud, que presenta la proporción de predicciones correctas realizadas por el modelo con respecto al total de predicciones realizadas y ABC, representando parámetros de sensibilidad y especificidad, vinculadas a la capacidad de un modelo para detectar de manera correcta los casos positivos y los casos negativos, respectivamente.

Finalmente, a pesar de que algunos autores han propuesto métodos más sofisticados para la identificación de biomarcadores microbianos asociados a fenotipos (Li et al., 2022; Su et al., 2022), hemos demostrado satisfactoriamente la eficiencia de la estrategia empleada. Aunque se requieren estudios funcionales más exhaustivos para comprender el rol mecanístico de los taxones identificados en el desarrollo de la obesidad, existen hallazgos previos que respaldan nuestros resultados. Por ejemplo: se ha informado que *Lachnospiraceae NK4A136 group*, un género importante dentro de la MI involucrado en la producción de ácidos grasos de cadena corta, se encuentra disminuido en sujetos obesos en comparación con delgados (Ma et al., 2020). Por el contrario, *Rikenellaceae RC9 gut group* se encuentra enriquecido en ratones que desarrollan obesidad inducida por dietas ricas en grasas (Wang et al., 2020). Además, se ha reportado un aumento de *Bacteroides sartorii* (Yan et al., 2022), y disminución

de *Muribaculum intestinale* (Do et al., 2018) asociados con obesidad inducido por la dieta. Por otro lado, se ha observado que una MI enriquecida en *Ruminococcaceae* y *Lachnospiraceae* se asocia con menor aumento de peso a largo plazo (Menni et al., 2017). Asimismo, la disminución en la abundancia relativa de *Christensenellaceae* se ha asociado con mayor estrés oxidativo e inflamación sistémica de bajo grado, mecanismos claves en el desarrollo de la obesidad y trastornos relacionados (Cani et al., 2008; Coello et al., 2021). Es importante destacar que también se han identificado taxones bacterianos que no presentan diferencias de abundancia entre los fenotipos estudiados, como *Prevotellaceae*, *Sutterellaceae*, *Marinifilaceae*, *Tannerellaceae*, *Mycoplasmataceae*, *Christensenellaceae* y [*Eubacterium*] *coprostanoligenes* *grup.* En este marco, es crucial reconocer que las comunidades microbianas están moldeadas por interacciones complejas que abarcan desde la competencia hasta el mutualismo. La alimentación cruzada (cross-feeding), definida como el intercambio de metabolitos entre diferentes microorganismos, emerge como un actor clave en la formación de comunidades comensales intestinales, con impactos sustanciales en la salud del huésped (Culp y Goodman, 2023). Estas interacciones generan respuestas sinérgicas que a menudo no pueden preverse al estudiar especies individuales por separado. Este fenómeno se ve agravado por el enfoque estático de los análisis tradicionales, que no logran capturar todas las interacciones que configuran la composición de comunidades microbianas complejas, como la microbiota (Saa et al., 2021). Nuestros resultados respaldan la aplicabilidad de las técnicas de ML en el estudio de microbiomas. Estas técnicas no solo demuestran la capacidad para descubrir relaciones no lineales y generar perfiles clasificatorios precisos, independientemente de los niveles de abundancia, sino que también sugieren posibles interacciones entre los componentes microbianos. Incluso aquellos taxones sin diferencias de abundancia, pero que contribuyen a clasificar los fenotipos, podrían desempeñar un papel esencial en la formación de la red de interacciones necesarias para configurar la comunidad microbiana asociada a los diferentes grupos de estudio. Aunque la descripción detallada de estas interacciones no es un objetivo específico del proyecto, nuestros hallazgos proporcionan un punto de partida prometedor que arroja nuevas pistas sobre la relevancia de las interacciones entre ciertos microorganismos en el desarrollo de la obesidad y patologías relacionadas. Este

enfoque global y dinámico abre la puerta a investigaciones más profundas sobre la complejidad de las comunidades microbianas y su papel en las condiciones metabólicas asociadas a diferentes estados fisiológicos.

6. CONCLUSION

Nuestros resultados respaldan la importancia de considerar los niveles taxonómicos más específicos, como género y especie, en los modelos de predicción de fenotipos relacionados con la microbiota. Estos niveles proporcionan información más detallada sobre las características genéticas y funcionales de los microorganismos, capturando interacciones y patrones sutiles en la abundancia de taxones. Además, estos hallazgos tienen implicancias biológicas significativas al revelar la importancia de la composición taxonómica y posibles interacciones entre sus componentes en la variabilidad de los fenotipos estudiados. A su vez, la implementación de técnicas de selección de características es una estrategia recomendada para favorecer un mejor desempeño de los métodos de predicción.

En conclusión, las técnicas de ML son una herramienta poderosa que permiten analizar grandes conjuntos de datos metagenómicos de manera eficiente y efectiva, revelando patrones y asociaciones complejas que pueden ser difíciles de detectar mediante enfoques tradicionales, sin embargo, su rendimiento va de la mano con la implementación de un adecuado protocolo de validación. Por último, su aplicación resulta relevante para generar una mayor comprensión de las relaciones subyacentes entre la MI y el desarrollo de diferentes enfermedades. Esto no solo proporciona una visión más precisa de la contribución de los microorganismos en la salud y la enfermedad, sino que también abre oportunidades para el desarrollo de enfoques de diagnóstico y terapéuticos más precisos y centrados en la medicina de precisión.

7. BIBLIOGRAFÍA

- Ai, L., Tian, H., Chen, Z., Chen, H., Xu, J., & Fang, J.-Y. (2017). Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer. *Oncotarget*, *8*(6), 9546–9556. <https://doi.org/10.18632/oncotarget.14488>
- Alboukadel Kassambara. (2022). *Ggpubr: “ggplot2” Based Publication Ready Plots. R package version 0.5.0*. <https://CRAN.R-project.org/package=ggpubr>
- Albright, M. B. N., Louca, S., Winkler, D. E., Feeser, K. L., Haig, S.-J., Whiteson, K. L., Emerson, J. B., & Dunbar, J. (2022). Solutions in microbiome engineering: Prioritizing barriers to organism establishment. *The ISME Journal*, *16*(2), Article 2. <https://doi.org/10.1038/s41396-021-01088-5>
- Alshboul, O., Shehadeh, A., Almasabha, G., & Almuflih, A. S. (2022). Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction. *Sustainability*, *14*(11), Article 11. <https://doi.org/10.3390/su14116651>
- Aryal, S., Alimadadi, A., Manandhar, I., Joe, B., & Cheng, X. (2020). Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. *Hypertension (Dallas, Tex. : 1979)*, *76*(5), 1555–1562. <https://doi.org/10.1161/HYPERTENSIONAHA.120.15885>
- Bang, S., Yoo, D., Kim, S.-J., Jhang, S., Cho, S., & Kim, H. (2019). Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. *Scientific Reports*, *9*(1), Article 1. <https://doi.org/10.1038/s41598-019-46249-x>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Cani, P. D., Amar, J., Iglesias, M. A., Poggi, M., Knauf, C., Bastelica, D., Neyrinck, A. M., Fava, F., Tuohy, K. M., Chabo, C., Waget, A., Delmée, E., Cousin, B., Sulpice, T., Chamontin, B., Ferrières, J.,

- Tanti, J.-F., Gibson, G. R., Casteilla, L., ... Burcelin, R. (2007). Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes*, *56*(7), 1761–1772. <https://doi.org/10.2337/db06-1491>
- Cani, P. D., Bibiloni, R., Knauf, C., Waaget, A., Neyrinck, A. M., Delzenne, N. M., & Burcelin, R. (2008). Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes*, *57*(6), 1470–1481. <https://doi.org/10.2337/db07-1403>
- Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The impact of the gut microbiota on human health: An integrative view. *Cell*, *148*(6), 1258–1270. <https://doi.org/10.1016/j.cell.2012.01.035>
- Clemmensen, L. (2016). *sparseLDA: Sparse Discriminant Analysis. R package version 0.1-9*.
- Coello, K., Hansen, T. H., Sørensen, N., Ottesen, N. M., Miskowiak, K. W., Pedersen, O., Kessing, L. V., & Vinberg, M. (2021). Affective disorders impact prevalence of Flavonifractor and abundance of Christensenellaceae in gut microbiota. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *110*, 110300. <https://doi.org/10.1016/j.pnpbp.2021.110300>
- Culp, E. J., & Goodman, A. L. (2023). Cross-feeding in the gut microbiome: Ecology and mechanisms. *Cell Host & Microbe*, *31*(4), 485–499. <https://doi.org/10.1016/j.chom.2023.03.016>
- Do, M. H., Lee, E., Oh, M.-J., Kim, Y., & Park, H.-Y. (2018). High-Glucose or -Fructose Diet Cause Changes of the Gut Microbiota and Metabolic Disorders in Mice without Body Weight Change. *Nutrients*, *10*(6), 761. <https://doi.org/10.3390/nu10060761>
- Faust, K., & Raes, J. (2012). Microbial interactions: From networks to models. *Nature Reviews Microbiology*, *10*(8), Article 8. <https://doi.org/10.1038/nrmicro2832>
- Flemer, B., Warren, R. D., Barrett, M. P., Cisek, K., Das, A., Jeffery, I. B., Hurley, E., O'Riordain, M., Shanahan, F., & O'Toole, P. W. (2018). The oral microbiota in colorectal cancer is distinctive and predictive. *Gut*, *67*(8), 1454–1463. <https://doi.org/10.1136/gutjnl-2017-314814>

H. Wickham. (2016). *ggplot2: Elegant Graphics for Data Analysis*.

<https://link.springer.com/book/10.1007/978-3-319-24277-4>

Harrell, & Dupont. (2019). *Hmisc: Harrell Miscellaneous*. <https://CRAN.R-project.org/package=Hmisc>

Hernández Medina, R., Kutuzova, S., Nielsen, K. N., Johansen, J., Hansen, L. H., Nielsen, M., &

Rasmussen, S. (2022). Machine learning and deep learning applications in microbiome

research. *ISME Communications*, 2(1), Article 1. [https://doi.org/10.1038/s43705-022-00182-](https://doi.org/10.1038/s43705-022-00182-9)

9

Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., Codelli, J. A., Chow, J.,

Reisman, S. E., Petrosino, J. F., Patterson, P. H., & Mazmanian, S. K. (2013). Microbiota

modulate behavioral and physiological abnormalities associated with neurodevelopmental

disorders. *Cell*, 155(7), 1451–1463. <https://doi.org/10.1016/j.cell.2013.11.024>

Johnson, M. K., Kjell. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive*

Models. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315108230>

Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C. J., Fagerberg, B., Nielsen, J., &

Bäckhed, F. (2013). Gut metagenome in European women with normal, impaired and

diabetic glucose control. *Nature*, 498(7452), 99–103. <https://doi.org/10.1038/nature12198>

Khan, A. A., Sirsat, A. T., Singh, H., & Cash, P. (2022). Microbiota and cancer: Current understanding

and mechanistic implications. *Clinical & Translational Oncology*, 24(2), 193–202.

<https://doi.org/10.1007/s12094-021-02690-x>

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*,

220(4598), 671–680. <https://doi.org/10.1126/science.220.4598.671>

Kohonen, T. (1995). *Self-Organizing Maps* (Vol. 30). Springer Berlin Heidelberg.

<https://doi.org/10.1007/978-3-642-97610-0>

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical*

Software, 28, 1–26. <https://doi.org/10.18637/jss.v028.i05>

- Leisch, F., & Dimitriadou, E. (2021). *mlbench: Machine Learning Benchmark Problems R package version 2.1-3*.
- Leo Lahti, S. S. (2012). *Microbiome R package*. <http://microbiome.github.io>
- Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature*, *444*(7122), Article 7122. <https://doi.org/10.1038/4441022a>
- Li, P., Luo, H., Ji, B., & Nielsen, J. (2022). Machine learning for data integration in human gut microbiome. *Microbial Cell Factories*, *21*(1), 241. <https://doi.org/10.1186/s12934-022-01973-4>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.
- Liu, Y.-X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., & Bai, Y. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein & Cell*, *12*(5), 315–330. <https://doi.org/10.1007/s13238-020-00724-8>
- Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., & Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, *550*(7674), Article 7674. <https://doi.org/10.1038/nature23889>
- Ma, L., Ni, Y., Wang, Z., Tu, W., Ni, L., Zhuge, F., Zheng, A., Hu, L., Zhao, Y., Zheng, L., & Fu, Z. (2020). Spermidine improves gut barrier integrity and gut microbiota function in diet-induced obese mice. *Gut Microbes*, *12*(1), 1832857. <https://doi.org/10.1080/19490976.2020.1832857>
- Marcos-Zambrano, L. J., Karaduzovic-Hadziabdic, K., Loncar Turukalo, T., Przymus, P., Trajkovic, V., Aasmets, O., Berland, M., Gruca, A., Hasic, J., Hron, K., Klammsteiner, T., Kolev, M., Lahti, L., Lopes, M. B., Moreno, V., Naskinova, I., Org, E., Paciência, I., Papoutsoglou, G., ... Truu, J. (2021). Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. *Frontiers in Microbiology*, *12*. <https://www.frontiersin.org/articles/10.3389/fmicb.2021.634511>

- Martín, R., Miquel, S., Langella, P., & Bermúdez-Humarán, L. G. (2014). The role of metagenomics in understanding the human microbiome in health and disease. *Virulence*, 5(3), 413–423.
<https://doi.org/10.4161/viru.27864>
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8(4), e61217.
<https://doi.org/10.1371/journal.pone.0061217>
- Menni, C., Jackson, M. A., Pallister, T., Steves, C. J., Spector, T. D., & Valdes, A. M. (2017). Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *International Journal of Obesity*, 41(7), Article 7. <https://doi.org/10.1038/ijo.2017.66>
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2023). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group*. <https://CRAN.R-project.org/package=e1071>
- Norouzi-Beirami, M. H., Marashi, S.-A., Banaei-Moghaddam, A. M., & Kavousi, K. (2020). Beyond Taxonomic Analysis of Microbiomes: A Functional Approach for Revisiting Microbiome Changes in Colorectal Cancer. *Frontiers in Microbiology*, 10.
<https://www.frontiersin.org/articles/10.3389/fmicb.2019.03117>
- O'Hara, A. M., & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, 7(7), 688–693. <https://doi.org/10.1038/sj.embor.7400731>
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., ... Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), Article 7285. <https://doi.org/10.1038/nature08821>
- Qiu, P., Ishimoto, T., Fu, L., Zhang, J., Zhang, Z., & Liu, Y. (2022). The Gut Microbiota in Inflammatory Bowel Disease. *Frontiers in Cellular and Infection Microbiology*, 12.
<https://www.frontiersin.org/articles/10.3389/fcimb.2022.733992>

- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596.
<https://doi.org/10.1093/nar/gks1219>
- Quinn, T. P., Erb, I., Richardson, M. F., & Crowley, T. M. (2018). Understanding sequencing data as compositions: An outlook and review. *Bioinformatics*, *34*(16), 2870–2878.
<https://doi.org/10.1093/bioinformatics/bty175>
- R Core Team. (2021). *R Core Team (2022). R: A language and environment for statistical computing.*
<https://www.R-project.org/>.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. <https://doi.org/10.1186/1471-2105-12-77>
- Saa, P., Urrutia, A., Silva-Andrade, C., Martín, A. J., & Garrido, D. (2021). Modeling approaches for probing cross-feeding interactions in the human gut microbiome. *Computational and Structural Biotechnology Journal*, *20*, 79–89. <https://doi.org/10.1016/j.csbj.2021.12.006>
- Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517. <https://doi.org/10.1093/bioinformatics/btm344>
- Schwabe, R. F., & Jobin, C. (2013). The microbiome and cancer. *Nature Reviews. Cancer*, *13*(11), 800–812. <https://doi.org/10.1038/nrc3610>
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), R60.
<https://doi.org/10.1186/gb-2011-12-6-r60>
- Sharma, D., Paterson, A. D., & Xu, W. (2020). TaxoNN: Ensemble of neural networks on stratified microbiome data for disease prediction. *Bioinformatics*, *36*(17), 4544–4550.
<https://doi.org/10.1093/bioinformatics/btaa542>

- Song, F., Mei, D., & Li, H. (2010). Feature Selection Based on Linear Discriminant Analysis. *2010 International Conference on Intelligent System Design and Engineering Application*, 746–749. <https://doi.org/10.1109/ISDEA.2010.311>
- Song, K., & Zhou, Y.-H. (2023). Leveraging Scheme for Cross-Study Microbiome Machine Learning Prediction and Feature Evaluations. *Bioengineering (Basel, Switzerland)*, *10*(2), 231. <https://doi.org/10.3390/bioengineering10020231>
- Su, Q., Liu, Q., Lau, R. I., Zhang, J., Xu, Z., Yeoh, Y. K., Leung, T. W. H., Tang, W., Zhang, L., Liang, J. Q. Y., Yau, Y. K., Zheng, J., Liu, C., Zhang, M., Cheung, C. P., Ching, J. Y. L., Tun, H. M., Yu, J., Chan, F. K. L., & Ng, S. C. (2022). Faecal microbiome-based machine learning for multi-class disease diagnosis. *Nature Communications*, *13*(1), Article 1. <https://doi.org/10.1038/s41467-022-34405-3>
- Therneau, T., Atkinson, B., port, B. R. (producer of the initial R., & maintainer 1999-2017). (2022). *rpart: Recursive Partitioning and Regression Trees* (4.1.19) [Computer software]. <https://cran.r-project.org/web/packages/rpart/index.html>
- Topçuoğlu, B. D., Lesniak, N. A., Ruffin, M. T., Wiens, J., & Schloss, P. D. (2020). A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. *mBio*, *11*(3), e00434-20. <https://doi.org/10.1128/mBio.00434-20>
- Villanueva-Millán, M. J., Pérez-Matute, P., & Oteo, J. A. (2015). Gut microbiota: A key player in health and disease. A review focused on obesity. *Journal of Physiology and Biochemistry*, *71*(3), 509–525. <https://doi.org/10.1007/s13105-015-0390-3>
- Wang, B., Kong, Q., Li, X., Zhao, J., Zhang, H., Chen, W., & Wang, G. (2020). A High-Fat Diet Increases Gut Microbiota Biodiversity and Energy Expenditure Due to Nutrient Difference. *Nutrients*, *12*(10), Article 10. <https://doi.org/10.3390/nu12103197>

Wu, D., Wang, H., Xie, L., & Hu, F. (2022). Cross-Talk Between Gut Microbiota and Adipose Tissues in Obesity and Related Metabolic Diseases. *Frontiers in Endocrinology*, *13*, 908868.

<https://doi.org/10.3389/fendo.2022.908868>

Yan, S., Zhou, J., Zhang, H., Lin, Z., Khambu, B., Liu, G., Ma, M., Chen, X., Chalasani, N., & Yin, X.-M. (2022). Promotion of diet-induced obesity and metabolic syndromes by BID is associated with gut microbiota. *Hepatology Communications*, *6*(12), 3349–3362.

<https://doi.org/10.1002/hep4.2052>

Zhou, Y.-H., & Gallins, P. (2019). A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction. *Frontiers in Genetics*, *10*.

<https://www.frontiersin.org/articles/10.3389/fgene.2019.00579>

8. MATERIAL SUPLEMENTARIO

8.1. Tabla Suplementaria 1A

Presenta la Exactitud obtenidas luego del entramiento de cada Modelo de Predicción en cada nivel taxonómico. El entrenamiento se realizó utilizando un enfoque de validación cruzada repetida (repeatedcv) con 10 particiones (number=10) y 10 repeticiones (repeats=10). En rojo se resalta el valor los más alto obtenido por cada modelo. Valores expresados como Media \pm Desviación estándar (DE).

Tabla Suplementaria 1A		
Modelo	Nivel taxonómico	Media \pm DE
LVQ	Phylum	0.64 \pm 0.16
	Clase	0.57 \pm 0.16
	Orden	0.72 \pm 0.16
	Familia	0.67 \pm 0.16
	Genero	0.71 \pm 0.18
	Especie	0.82 \pm 0.13
GBM	Phylum	0.72 \pm 0.16
	Clase	0.80 \pm 0.15
	Orden	0.80 \pm 0.14
	Familia	0.73 \pm 0.15
	Genero	0.82 \pm 0.13
	Especie	0.81 \pm 0.16
SVMr	Phylum	0.72 \pm 0.18
	Clase	0.80 \pm 0.12
	Orden	0.77 \pm 0.14
	Familia	0.78 \pm 0.14
	Genero	0.80 \pm 0.15
	Especie	0.77 \pm 0.13
SVMl	Phylum	0.73 \pm 0.14
	Clase	0.83 \pm 0.15
	Orden	0.74 \pm 0.15
	Familia	0.75 \pm 0.14
	Genero	0.79 \pm 0.14
	Especie	0.77 \pm 0.14
SLDA	Phylum	0.67 \pm 0.16
	Clase	0.59 \pm 0.05
	Orden	0.61 \pm 0.09
	Familia	0.60 \pm 0.04
	Genero	0.66 \pm 0.14
	Especie	0.80 \pm 0.17

NN	Phylum	0.74 ± 0.16
	Clase	0.78 ± 0.13
	Orden	0.79 ± 0.14
	Familia	0.70 ± 0.16
	Genero	0.75 ± 0.15
	Especie	0.81 ± 0.14
Rborist	Phylum	0.76 ± 0.16
	Clase	0.81 ± 0.14
	Orden	0.82 ± 0.15
	Familia	0.75 ± 0.16
	Genero	0.84 ± 0.13
	Especie	0.81 ± 0.15
xgbLinear	Phylum	0.70 ± 0.14
	Clase	0.78 ± 0.14
	Orden	0.78 ± 0.15
	Familia	0.72 ± 0.17
	Genero	0.83 ± 0.15
	Especie	0.81 ± 0.14
stepLDA	Phylum	0.71 ± 0.11
	Clase	0.66 ± 0.17
	Orden	0.72 ± 0.14
	Familia	0.65 ± 0.15
	Genero	0.64 ± 0.14
	Especie	0.81 ± 0.13
naive_bayes	Phylum	0.74 ± 0.17
	Clase	0.70 ± 0.14
	Orden	0.75 ± 0.15
	Familia	0.74 ± 0.18
	Genero	0.84 ± 0.13
	Especie	0.76 ± 0.16

8.2. Tabla Suplementaria 1B

Presenta los valores de significancia estadística de las comparaciones entre la exactitud de los diferentes modelos entre niveles taxonómicos. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 1B					
Modelo: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	6.62E-05				
Género	2.32E-07	1.03E-01			
Orden	1.84E-09	1.29E-02	4.61E-01		
Phylum	1.78E-03	2.73E-01	7.78E-03	3.67E-04	
Especie	1.28E-19	1.43E-10	8.20E-06	8.90E-05	1.59E-13
Modelo: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	1.32E-03				
Género	5.97E-01	1.29E-04			
Orden	9.27E-01	6.88E-04	7.80E-01		
Phylum	1.32E-03	9.42E-01	1.29E-04	7.13E-04	
Especie	8.73E-01	6.88E-04	8.28E-01	9.56E-01	7.11E-04
Modelo: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	3.28E-04				
Género	6.96E-02	6.96E-02			
Orden	2.35E-04	6.41E-01	4.31E-02		
Phylum	1.16E-05	4.42E-01	6.82E-03	6.41E-01	
Especie	2.03E-03	5.36E-01	2.38E-01	3.60E-01	1.46E-01
Modelo: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	5.69E-01				
Género	6.14E-01	3.06E-01			
Orden	6.14E-01	8.28E-01	3.66E-01		
Phylum	9.75E-03	7.75E-02	9.75E-03	7.75E-02	
Especie	3.30E-01	7.16E-01	1.69E-01	6.14E-01	9.68E-02

Modelo: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	3.22E-01				
Género	1.77E-04	9.61E-04			
Orden	3.56E-02	1.95E-01	3.56E-02		
Phylum	9.24E-05	1.77E-04	4.09E-01	6.52E-03	
Especie	3.65E-23	3.65E-23	1.24E-11	2.68E-19	1.47E-08
Modelo: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	2.13E-03				
Género	1.54E-01	1.26E-01			
Orden	8.16E-01	1.66E-03	1.27E-01		
Phylum	1.54E-01	1.26E-01	9.01E-01	1.26E-01	
Especie	1.31E-01	1.67E-05	9.23E-03	2.34E-01	7.53E-03
Modelo: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	6.64E-02				
Género	1.52E-01	1.44E-03			
Orden	3.94E-01	7.61E-03	6.30E-01		
Phylum	1.56E-01	6.30E-01	6.34E-03	2.51E-02	
Especie	6.30E-01	2.51E-02	3.32E-01	6.30E-01	7.21E-02
Modelo: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	2.44E-02				
Género	3.00E-02	2.89E-05			
Orden	9.74E-01	3.47E-02	3.86E-02		
Phylum	1.31E-04	1.62E-01	3.90E-08	4.10E-04	
Especie	1.21E-01	2.25E-04	5.29E-01	1.37E-01	2.92E-07
Modelo: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	6.98E-01				
Género	3.23E-01	5.33E-01			
Orden	8.16E-03	1.81E-03	2.41E-04		
Phylum	4.51E-02	1.22E-02	6.27E-04	3.23E-01	
Especie	6.76E-11	1.98E-12	2.65E-13	2.84E-05	7.54E-09
Modelo: Naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	8.23E-02				

Género	4.99E-10	5.05E-04			
Orden	1.11E-02	7.52E-01	6.97E-04		
Phylum	6.29E-02	9.96E-01	3.95E-04	7.52E-01	
Especie	1.10E-02	7.52E-01	1.01E-03	9.67E-01	7.52E-01

8.1. Tabla Suplementaria 1C

Presenta los valores de significancia estadística de las comparaciones entre la exactitud de los diferentes modelos en el mismo nivel taxonómico. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 1C									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	2.92E-03								
naive_bayes	4.36E-01	2.11E-04							
NN	4.49E-01	2.11E-04	9.35E-01						
Rborist	1.09E-01	1.90E-05	4.79E-01	4.48E-01					
SLDA	3.99E-02	4.48E-01	4.91E-03	4.91E-03	5.21E-04				
stepLDA	3.21E-01	5.51E-03	6.79E-02	5.99E-02	5.51E-03	1.01E-01			
SVMl	9.96E-01	4.91E-03	4.49E-01	4.79E-01	1.57E-01	5.92E-02	4.79E-01		
SVMr	8.20E-01	4.61E-04	5.25E-01	5.57E-01	1.57E-01	9.11E-03	1.57E-01	7.92E-01	
xgbLinear	3.08E-01	2.76E-02	5.99E-02	5.99E-02	5.51E-03	2.30E-01	8.20E-01	3.93E-01	1.57E-01
Nivel taxonómico: Clase									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	1.55E-16								
ive_bayes	7.13E-06	2.64E-08							
NN	4.13E-01	4.11E-16	1.29E-04						
Rborist	7.13E-01	1.75E-17	2.19E-06	2.66E-01					
SLDA	6.20E-21	1.40E-01	3.13E-12	7.38E-24	2.08E-23				
stepLDA	7.05E-09	1.83E-04	1.03E-01	1.87E-07	2.29E-09	7.13E-06			
SVMl	7.30E-01	3.27E-18	1.73E-06	4.80E-01	5.38E-01	2.59E-26	9.13E-10		
SVMr	6.84E-02	4.54E-19	2.84E-09	6.70E-03	1.50E-01	1.08E-23	3.00E-12	2.24E-02	
xgbLinear	4.09E-01	7.14E-16	8.88E-05	9.35E-01	2.54E-01	3.22E-22	1.05E-07	5.38E-01	5.82E-03
Nivel taxonómico: Orden									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr

LVQ	2.56E-03								
ive_bayes	6.81E-02	3.18E-01							
NN	4.61E-01	2.11E-02	3.00E-01						
Rborist	3.48E-01	7.97E-05	6.05E-03	9.73E-02					
SLDA	2.26E-19	6.41E-08	1.95E-11	1.17E-17	2.16E-20				
stepLDA	4.74E-04	7.84E-01	1.39E-01	7.67E-03	1.13E-05	5.66E-09			
SVMI	3.48E-01	5.01E-02	3.92E-01	8.17E-01	5.40E-02	1.02E-15	1.28E-02		
SVMr	1.28E-02	5.79E-01	5.63E-01	7.41E-02	6.42E-04	1.23E-10	4.17E-01	1.35E-01	
xgbLinear	3.48E-01	5.40E-02	4.14E-01	8.00E-01	6.37E-02	7.91E-15	2.10E-02	9.86E-01	1.54E-01

Nivel taxonómico: Familia

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.07E-02								
ive_bayes	4.26E-01	4.77E-03							
NN	5.26E-01	1.63E-01	1.36E-01						
Rborist	2.63E-01	9.43E-04	8.09E-01	5.51E-02					
SLDA	3.94E-13	3.35E-05	7.70E-12	1.82E-10	3.33E-14				
stepLDA	4.29E-03	5.50E-01	4.50E-04	3.59E-02	3.88E-05	3.69E-04			
SVMI	1.51E-02	5.33E-06	2.89E-01	1.53E-03	4.18E-01	2.02E-20	4.50E-08		
SVMr	2.20E-01	4.50E-04	8.66E-01	5.00E-02	9.15E-01	2.01E-18	1.96E-05	2.64E-01	
xgbLinear	7.83E-01	1.65E-02	4.90E-01	3.55E-01	3.17E-01	3.94E-13	1.53E-03	3.51E-02	3.55E-01

Nivel taxonómico: Género

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	1.94E-05								
ive_bayes	3.57E-01	2.44E-07							
NN	2.76E-03	1.49E-01	5.52E-05						
Rborist	2.57E-01	1.34E-07	8.50E-01	3.48E-05					
SLDA	8.76E-13	1.37E-02	2.17E-15	9.65E-06	1.30E-15				
stepLDA	7.88E-14	4.81E-03	4.63E-16	2.85E-06	2.29E-16	6.78E-01			
SVMI	6.78E-01	3.03E-04	1.93E-01	2.03E-02	1.49E-01	3.58E-10	4.30E-11		
SVMr	3.50E-01	7.36E-04	4.43E-02	4.76E-02	3.15E-02	1.79E-10	3.54E-11	6.19E-01	
xgbLinear	6.19E-01	5.64E-06	7.19E-01	7.36E-04	6.19E-01	8.76E-13	7.70E-14	3.70E-01	1.44E-01

Nivel taxonómico: Especie

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	9.99E-01								
ive_bayes	9.65E-02	6.16E-02							
NN	9.99E-01	9.99E-01	6.16E-02						
Rborist	9.99E-01	9.99E-01	6.16E-02	9.99E-01					
SLDA	9.99E-01	9.99E-01	6.16E-02	9.99E-01	9.99E-01				
stepLDA	9.99E-01	9.99E-01	6.16E-02	9.99E-01	9.99E-01	9.99E-01			
SVMI	1.73E-01	6.16E-02	9.99E-01	6.16E-02	6.16E-02	6.16E-02	6.16E-02		

SVMr	1.28E-01	6.16E-02	9.99E-01	6.16E-02	6.16E-02	6.16E-02	6.16E-02	9.99E-01	
xgbLinear	9.99E-01	9.99E-01	6.16E-02	9.99E-01	9.99E-01	9.99E-01	9.99E-01	6.16E-02	6.16E-02

8.2. Tabla suplementaria 2A

Presenta la Exactitud obtenidas luego del entramiento de cada Modelo de Predicción en cada nivel taxonómico y Método de selección de características. El entrenamiento se realizó utilizando un enfoque de validación cruzada repetida (repeatedcv) con 10 particiones (number=10) y 10 repeticiones (repeats=10). Valores expresados como Media \pm Desviación estándar (DE).

Tabla Suplementaria 2A			
Modelo	Nivel taxonómico	Método de selección de características	Media \pm DE
LVQ	Phylum	SAFS	0.66 \pm 0.16
LVQ	Clase	SAFS	0.73 \pm 0.19
LVQ	Orden	SAFS	0.66 \pm 0.16
LVQ	Familia	SAFS	0.65 \pm 0.18
LVQ	Género	SAFS	0.75 \pm 0.16
LVQ	Especie	SAFS	0.85 \pm 0.14
GBM	Phylum	SAFS	0.72 \pm 0.16
GBM	Clase	SAFS	0.75 \pm 0.14
GBM	Orden	SAFS	0.75 \pm 0.15
GBM	Familia	SAFS	0.63 \pm 0.17
GBM	Género	SAFS	0.8 \pm 0.15
GBM	Especie	SAFS	0.84 \pm 0.14
SVMr	Phylum	SAFS	0.72 \pm 0.14
SVMr	Clase	SAFS	0.8 \pm 0.14
SVMr	Orden	SAFS	0.75 \pm 0.15
SVMr	Familia	SAFS	0.69 \pm 0.15
SVMr	Género	SAFS	0.86 \pm 0.13
SVMr	Especie	SAFS	0.81 \pm 0.15
SVMI	Phylum	SAFS	0.75 \pm 0.14
SVMI	Clase	SAFS	0.79 \pm 0.12
SVMI	Orden	SAFS	0.73 \pm 0.17
SVMI	Familia	SAFS	0.64 \pm 0.17
SVMI	Género	SAFS	0.86 \pm 0.12
SVMI	Especie	SAFS	0.81 \pm 0.14
SLDA	Phylum	SAFS	0.6 \pm 0.05
SLDA	Clase	SAFS	0.64 \pm 0.15
SLDA	Orden	SAFS	0.6 \pm 0.05

SLDA	Familia	SAFS	0.48 ± 0.14
SLDA	Género	SAFS	0.64 ± 0.11
SLDA	Especie	SAFS	0.72 ± 0.15
NN	Phylum	SAFS	0.75 ± 0.15
NN	Clase	SAFS	0.75 ± 0.17
NN	Orden	SAFS	0.71 ± 0.18
NN	Familia	SAFS	0.68 ± 0.15
NN	Género	SAFS	0.79 ± 0.16
NN	Especie	SAFS	0.82 ± 0.13
Rborist	Phylum	SAFS	0.73 ± 0.14
Rborist	Clase	SAFS	0.83 ± 0.15
Rborist	Orden	SAFS	0.75 ± 0.16
Rborist	Familia	SAFS	0.72 ± 0.15
Rborist	Género	SAFS	0.83 ± 0.13
Rborist	Especie	SAFS	0.84 ± 0.12
xgbLinear	Phylum	SAFS	0.72 ± 0.16
xgbLinear	Clase	SAFS	0.77 ± 0.15
xgbLinear	Orden	SAFS	0.71 ± 0.16
xgbLinear	Familia	SAFS	0.65 ± 0.16
xgbLinear	Género	SAFS	0.78 ± 0.17
xgbLinear	Especie	SAFS	0.81 ± 0.14
stepLDA	Phylum	SAFS	0.7 ± 0.13
stepLDA	Clase	SAFS	0.72 ± 0.17
stepLDA	Orden	SAFS	0.7 ± 0.14
stepLDA	Familia	SAFS	0.67 ± 0.14
stepLDA	Género	SAFS	0.6 ± 0.14
stepLDA	Especie	SAFS	0.84 ± 0.13
naive_bayes	Phylum	SAFS	0.74 ± 0.15
naive_bayes	Clase	SAFS	0.7 ± 0.17
naive_bayes	Orden	SAFS	0.7 ± 0.19
naive_bayes	Familia	SAFS	0.61 ± 0.17
naive_bayes	Género	SAFS	0.79 ± 0.15
naive_bayes	Especie	SAFS	0.76 ± 0.15
LVQ	Phylum	SAFS.rf	0.65 ± 0.18
LVQ	Clase	SAFS.rf	0.61 ± 0.17
LVQ	Orden	SAFS.rf	0.69 ± 0.17
LVQ	Familia	SAFS.rf	0.66 ± 0.19
LVQ	Género	SAFS.rf	0.71 ± 0.14
LVQ	Especie	SAFS.rf	0.85 ± 0.12
GBM	Phylum	SAFS.rf	0.75 ± 0.15
GBM	Clase	SAFS.rf	0.75 ± 0.15
GBM	Orden	SAFS.rf	0.78 ± 0.16
GBM	Familia	SAFS.rf	0.91 ± 0.1

GBM	Género	SAFS.rf	0.78 ± 0.16
GBM	Especie	SAFS.rf	0.84 ± 0.13
SVMr	Phylum	SAFS.rf	0.76 ± 0.14
SVMr	Clase	SAFS.rf	0.77 ± 0.15
SVMr	Orden	SAFS.rf	0.7 ± 0.14
SVMr	Familia	SAFS.rf	0.92 ± 0.1
SVMr	Género	SAFS.rf	0.77 ± 0.15
SVMr	Especie	SAFS.rf	0.78 ± 0.14
SVMI	Phylum	SAFS.rf	0.78 ± 0.15
SVMI	Clase	SAFS.rf	0.74 ± 0.17
SVMI	Orden	SAFS.rf	0.76 ± 0.16
SVMI	Familia	SAFS.rf	0.86 ± 0.13
SVMI	Género	SAFS.rf	0.77 ± 0.16
SVMI	Especie	SAFS.rf	0.84 ± 0.11
SLDA	Phylum	SAFS.rf	0.42 ± 0.11
SLDA	Clase	SAFS.rf	0.77 ± 0.16
SLDA	Orden	SAFS.rf	0.73 ± 0.17
SLDA	Familia	SAFS.rf	0.85 ± 0.14
SLDA	Género	SAFS.rf	0.77 ± 0.16
SLDA	Especie	SAFS.rf	0.72 ± 0.16
NN	Phylum	SAFS.rf	0.78 ± 0.15
NN	Clase	SAFS.rf	0.75 ± 0.16
NN	Orden	SAFS.rf	0.76 ± 0.16
NN	Familia	SAFS.rf	0.87 ± 0.11
NN	Género	SAFS.rf	0.74 ± 0.16
NN	Especie	SAFS.rf	0.81 ± 0.15
Rborist	Phylum	SAFS.rf	0.78 ± 0.15
Rborist	Clase	SAFS.rf	0.74 ± 0.16
Rborist	Orden	SAFS.rf	0.75 ± 0.14
Rborist	Familia	SAFS.rf	0.91 ± 0.11
Rborist	Género	SAFS.rf	0.83 ± 0.13
Rborist	Especie	SAFS.rf	0.85 ± 0.13
xgbLinear	Phylum	SAFS.rf	0.73 ± 0.17
xgbLinear	Clase	SAFS.rf	0.69 ± 0.16
xgbLinear	Orden	SAFS.rf	0.79 ± 0.18
xgbLinear	Familia	SAFS.rf	0.92 ± 0.1
xgbLinear	Género	SAFS.rf	0.82 ± 0.14
xgbLinear	Especie	SAFS.rf	0.84 ± 0.14
stepLDA	Phylum	SAFS.rf	0.65 ± 0.14
stepLDA	Clase	SAFS.rf	0.72 ± 0.17
stepLDA	Orden	SAFS.rf	0.68 ± 0.15
stepLDA	Familia	SAFS.rf	0.74 ± 0.15
stepLDA	Género	SAFS.rf	0.67 ± 0.14

stepLDA	Especie	SAFS.rf	0.81 ± 0.14
naive_bayes	Phylum	SAFS.rf	0.81 ± 0.15
naive_bayes	Clase	SAFS.rf	0.72 ± 0.16
naive_bayes	Orden	SAFS.rf	0.69 ± 0.16
naive_bayes	Familia	SAFS.rf	0.79 ± 0.15
naive_bayes	Género	SAFS.rf	0.72 ± 0.17
naive_bayes	Especie	SAFS.rf	0.74 ± 0.16
LVQ	Phylum	RFE	0.67 ± 0.19
LVQ	Clase	RFE	0.78 ± 0.16
LVQ	Orden	RFE	0.71 ± 0.17
LVQ	Familia	RFE	0.65 ± 0.16
LVQ	Género	RFE	0.76 ± 0.15
LVQ	Especie	RFE	0.84 ± 0.13
GBM	Phylum	RFE	0.8 ± 0.13
GBM	Clase	RFE	0.8 ± 0.16
GBM	Orden	RFE	0.81 ± 0.14
GBM	Familia	RFE	0.82 ± 0.14
GBM	Género	RFE	0.86 ± 0.12
GBM	Especie	RFE	0.85 ± 0.15
SVMr	Phylum	RFE	0.71 ± 0.16
SVMr	Clase	RFE	0.79 ± 0.15
SVMr	Orden	RFE	0.84 ± 0.13
SVMr	Familia	RFE	0.81 ± 0.14
SVMr	Género	RFE	0.88 ± 0.11
SVMr	Especie	RFE	0.82 ± 0.14
SVMl	Phylum	RFE	0.77 ± 0.15
SVMl	Clase	RFE	0.77 ± 0.15
SVMl	Orden	RFE	0.79 ± 0.14
SVMl	Familia	RFE	0.8 ± 0.16
SVMl	Género	RFE	0.86 ± 0.13
SVMl	Especie	RFE	0.85 ± 0.13
SLDA	Phylum	RFE	0.61 ± 0.05
SLDA	Clase	RFE	0.53 ± 0.13
SLDA	Orden	RFE	0.62 ± 0.05
SLDA	Familia	RFE	0.63 ± 0.08
SLDA	Género	RFE	0.78 ± 0.16
SLDA	Especie	RFE	0.84 ± 0.13
NN	Phylum	RFE	0.78 ± 0.14
NN	Clase	RFE	0.8 ± 0.15
NN	Orden	RFE	0.76 ± 0.15
NN	Familia	RFE	0.79 ± 0.15
NN	Género	RFE	0.81 ± 0.13
NN	Especie	RFE	0.83 ± 0.16

Rborist	Phylum	RFE	0.76 ± 0.13
Rborist	Clase	RFE	0.8 ± 0.16
Rborist	Orden	RFE	0.82 ± 0.13
Rborist	Familia	RFE	0.82 ± 0.14
Rborist	Género	RFE	0.9 ± 0.1
Rborist	Especie	RFE	0.87 ± 0.12
xgbLinear	Phylum	RFE	0.81 ± 0.17
xgbLinear	Clase	RFE	0.75 ± 0.15
xgbLinear	Orden	RFE	0.76 ± 0.16
xgbLinear	Familia	RFE	0.76 ± 0.16
xgbLinear	Género	RFE	0.84 ± 0.14
xgbLinear	Especie	RFE	0.84 ± 0.12
stepLDA	Phylum	RFE	0.7 ± 0.12
stepLDA	Clase	RFE	0.72 ± 0.14
stepLDA	Orden	RFE	0.72 ± 0.14
stepLDA	Familia	RFE	0.68 ± 0.14
stepLDA	Género	RFE	0.69 ± 0.16
stepLDA	Especie	RFE	0.82 ± 0.14
naive_bayes	Phylum	RFE	0.74 ± 0.16
naive_bayes	Clase	RFE	0.76 ± 0.18
naive_bayes	Orden	RFE	0.78 ± 0.16
naive_bayes	Familia	RFE	0.68 ± 0.18
naive_bayes	Género	RFE	0.85 ± 0.11
naive_bayes	Especie	RFE	0.86 ± 0.12
LVQ	Phylum	SBF	0.6 ± 0.18
LVQ	Clase	SBF	0.77 ± 0.15
LVQ	Orden	SBF	0.75 ± 0.17
LVQ	Familia	SBF	0.69 ± 0.17
LVQ	Género	SBF	0.74 ± 0.16
LVQ	Especie	SBF	0.85 ± 0.12
GBM	Phylum	SBF	0.7 ± 0.15
GBM	Clase	SBF	0.75 ± 0.16
GBM	Orden	SBF	0.78 ± 0.15
GBM	Familia	SBF	0.74 ± 0.17
GBM	Género	SBF	0.82 ± 0.14
GBM	Especie	SBF	0.83 ± 0.14
SVMr	Phylum	SBF	0.67 ± 0.14
SVMr	Clase	SBF	0.76 ± 0.13
SVMr	Orden	SBF	0.8 ± 0.15
SVMr	Familia	SBF	0.79 ± 0.14
SVMr	Género	SBF	0.86 ± 0.13
SVMr	Especie	SBF	0.83 ± 0.14
SVMl	Phylum	SBF	0.74 ± 0.15

SVMI	Clase	SBF	0.74 ± 0.17
SVMI	Orden	SBF	0.81 ± 0.12
SVMI	Familia	SBF	0.83 ± 0.12
SVMI	Género	SBF	0.87 ± 0.12
SVMI	Especie	SBF	0.85 ± 0.12
SLDA	Phylum	SBF	0.59 ± 0.05
SLDA	Clase	SBF	0.65 ± 0.16
SLDA	Orden	SBF	0.46 ± 0.09
SLDA	Familia	SBF	0.52 ± 0.11
SLDA	Género	SBF	0.77 ± 0.14
SLDA	Especie	SBF	0.87 ± 0.13
NN	Phylum	SBF	0.74 ± 0.16
NN	Clase	SBF	0.73 ± 0.17
NN	Orden	SBF	0.77 ± 0.14
NN	Familia	SBF	0.74 ± 0.16
NN	Género	SBF	0.82 ± 0.15
NN	Especie	SBF	0.86 ± 0.12
Rborist	Phylum	SBF	0.73 ± 0.18
Rborist	Clase	SBF	0.77 ± 0.13
Rborist	Orden	SBF	0.84 ± 0.14
Rborist	Familia	SBF	0.77 ± 0.15
Rborist	Género	SBF	0.86 ± 0.12
Rborist	Especie	SBF	0.83 ± 0.15
xgbLinear	Phylum	SBF	0.66 ± 0.15
xgbLinear	Clase	SBF	0.72 ± 0.13
xgbLinear	Orden	SBF	0.8 ± 0.14
xgbLinear	Familia	SBF	0.72 ± 0.15
xgbLinear	Género	SBF	0.82 ± 0.14
xgbLinear	Especie	SBF	0.8 ± 0.15
stepLDA	Phylum	SBF	0.67 ± 0.15
stepLDA	Clase	SBF	0.73 ± 0.14
stepLDA	Orden	SBF	0.65 ± 0.13
stepLDA	Familia	SBF	0.69 ± 0.16
stepLDA	Género	SBF	0.68 ± 0.16
stepLDA	Especie	SBF	0.85 ± 0.14
naive_bayes	Phylum	SBF	0.72 ± 0.19
naive_bayes	Clase	SBF	0.67 ± 0.16
naive_bayes	Orden	SBF	0.83 ± 0.14
naive_bayes	Familia	SBF	0.8 ± 0.15
naive_bayes	Género	SBF	0.86 ± 0.12
naive_bayes	Especie	SBF	0.84 ± 0.13

8.3. Tabla suplementaria 2B

Presenta los valores de significancia estadística de las comparaciones entre la exactitud entre los diferentes modelos a cada nivel taxonómico, luego de la aplicación de cada método de selección de características. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 2B									
Método de selección de características: SAFS									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.94E-02								
naive_bayes	7.28E-01	7.36E-03							
NN	4.93E-01	2.50E-03	7.28E-01						
Rborist	7.28E-01	7.45E-03	9.92E-01	7.28E-01					
SLDA	5.67E-10	2.26E-05	9.53E-14	1.98E-14	1.65E-14				
stepLDA	4.03E-01	3.10E-01	1.35E-01	5.50E-02	1.57E-01	1.92E-12			
SVMI	9.92E-01	3.88E-02	7.24E-01	4.50E-01	7.28E-01	1.33E-12	4.34E-01		
SVMr	4.93E-01	9.70E-04	7.28E-01	9.92E-01	7.24E-01	6.13E-17	2.73E-02	4.19E-01	
xgbLinear	9.96E-01	3.94E-02	7.28E-01	4.93E-01	7.28E-01	5.05E-12	4.34E-01	9.92E-01	4.50E-01
Método de selección de características: SAFS									
Nivel taxonómico: Clase									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	5.06E-01								
naive_bayes	5.92E-02	2.46E-01							
NN	9.11E-01	5.06E-01	6.70E-02						
Rborist	5.97E-04	8.10E-04	8.51E-07	3.39E-03					
SLDA	3.06E-07	8.10E-04	1.39E-02	7.84E-06	1.28E-12				
stepLDA	1.82E-01	6.27E-01	5.78E-01	2.12E-01	3.58E-05	1.80E-03			
SVMI	6.05E-02	2.60E-02	4.08E-04	1.17E-01	1.57E-01	2.43E-10	3.36E-03		
SVMr	1.26E-01	6.70E-02	1.02E-03	2.36E-01	3.57E-02	2.43E-10	8.27E-03	6.06E-01	
xgbLinear	5.06E-01	2.58E-01	1.09E-02	6.06E-01	9.94E-03	3.06E-07	8.23E-02	2.43E-01	5.06E-01
Método de selección de características: SAFS									
Nivel taxonómico: Orden									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	1.15E-04								
naive_bayes	1.89E-01	7.60E-02							

NN	1.92E-01	8.23E-02	9.77E-01						
Rborist	8.25E-01	4.72E-04	2.38E-01	2.36E-01					
SLDA	3.45E-17	7.52E-04	1.22E-08	1.22E-07	1.91E-15				
stepLDA	2.56E-02	1.55E-01	6.25E-01	6.48E-01	6.36E-02	1.22E-08			
SVMI	8.25E-01	4.26E-04	2.53E-01	2.53E-01	9.97E-01	1.91E-15	6.36E-02		
SVMr	4.58E-01	8.90E-03	6.41E-01	6.13E-01	5.78E-01	2.12E-11	2.36E-01	6.25E-01	
xgbLinear	7.60E-02	7.60E-02	8.78E-01	8.89E-01	1.55E-01	1.95E-09	7.63E-01	1.51E-01	4.58E-01

Método de selección de características: SAFS

Nivel taxonómico: Familia

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	6.91E-01								
naive_bayes	4.21E-01	2.81E-01							
NN	8.40E-02	2.69E-01	6.76E-03						
Rborist	9.01E-04	9.95E-03	3.57E-05	1.06E-01					
SLDA	4.42E-09	2.49E-09	1.40E-07	2.57E-14	5.03E-17				
stepLDA	2.69E-01	4.81E-01	3.09E-02	4.75E-01	2.30E-02	5.51E-14			
SVMI	3.06E-02	1.34E-01	1.77E-03	7.05E-01	2.46E-01	2.49E-15	2.70E-01		
SVMr	7.42E-01	8.67E-01	2.69E-01	1.51E-01	2.24E-03	7.49E-10	4.21E-01	7.99E-02	
xgbLinear	4.21E-01	7.09E-01	1.04E-01	3.83E-01	1.60E-02	2.73E-11	7.67E-01	2.42E-01	6.22E-01

Método de selección de características: SAFS

Nivel taxonómico: Género

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.37E-02								
naive_bayes	5.79E-01	1.19E-01							
NN	6.66E-01	1.01E-01	9.28E-01						
Rborist	2.59E-01	7.94E-04	7.67E-02	1.19E-01					
SLDA	1.23E-13	5.03E-07	4.01E-12	5.34E-12	1.12E-17				
stepLDA	1.77E-16	4.28E-10	6.30E-15	1.18E-14	5.17E-20	1.88E-02			
SVMI	7.41E-03	1.15E-06	8.24E-04	2.30E-03	1.01E-01	4.18E-21	9.03E-23		
SVMr	7.41E-03	1.01E-06	8.24E-04	1.87E-03	1.19E-01	9.03E-23	5.43E-24	9.09E-01	
xgbLinear	3.51E-01	2.59E-01	6.89E-01	6.68E-01	4.61E-02	6.00E-10	7.68E-13	7.18E-04	4.90E-04

Método de selección de características: SAFS

Nivel taxonómico: Especie

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	7.19E-01								
naive_bayes	5.35E-04	1.77E-04							
NN	3.17E-01	1.52E-01	1.46E-02						
Rborist	6.66E-01	3.92E-01	2.04E-03	6.05E-01					
SLDA	6.87E-08	4.68E-08	7.06E-02	2.42E-06	9.14E-08				
stepLDA	8.22E-01	5.40E-01	1.06E-03	4.62E-01	8.22E-01	9.14E-08			

SVMI	1.70E-01	9.59E-02	9.59E-02	6.66E-01	3.44E-01	1.77E-04	2.50E-01		
SVMr	1.09E-01	5.14E-02	9.59E-02	5.40E-01	2.32E-01	9.76E-05	1.69E-01	8.98E-01	
xgbLinear	9.59E-02	4.58E-02	1.21E-01	4.87E-01	1.83E-01	1.64E-04	1.50E-01	8.22E-01	9.01E-01
Método de selección de características: SAFS.rf									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	2.23E-05								
naive_bayes	2.52E-02	6.43E-10							
NN	2.17E-01	1.40E-07	3.63E-01						
Rborist	1.72E-01	3.67E-08	3.78E-01	9.34E-01					
SLDA	3.51E-27	3.75E-19	5.18E-29	1.17E-28	2.22E-28				
stepLDA	2.25E-06	9.28E-01	6.63E-12	7.10E-09	7.16E-10	1.27E-22			
SVMI	9.28E-01	6.88E-06	2.77E-02	2.50E-01	1.96E-01	1.17E-28	5.09E-07		
SVMr	1.96E-01	5.04E-08	3.63E-01	9.89E-01	9.28E-01	8.50E-29	1.05E-09	2.27E-01	
xgbLinear	3.59E-01	1.39E-03	1.78E-03	3.21E-02	2.20E-02	4.51E-25	4.54E-04	2.97E-01	2.29E-02
Método de selección de características: SAFS.rf									
Nivel taxonómico: Clase									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	6.54E-07								
naive_bayes	2.74E-01	2.63E-04							
NN	8.99E-01	1.10E-06	2.62E-01						
Rborist	8.88E-01	1.38E-06	3.52E-01	7.79E-01					
SLDA	4.65E-01	6.83E-08	6.33E-02	5.30E-01	3.52E-01				
stepLDA	3.26E-01	2.45E-04	8.99E-01	3.06E-01	4.03E-01	8.13E-02			
SVMI	6.00E-01	6.83E-08	8.31E-02	6.91E-01	4.65E-01	8.89E-01	1.15E-01		
SVMr	8.99E-01	1.38E-06	3.26E-01	8.59E-01	9.46E-01	3.88E-01	3.52E-01	5.09E-01	
xgbLinear	3.92E-02	4.69E-03	4.70E-01	3.92E-02	6.33E-02	4.97E-03	4.18E-01	7.82E-03	6.14E-02
Método de selección de características: SAFS.rf									
Nivel taxonómico: Orden									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.06E-03								
naive_bayes	1.52E-03	9.80E-01							
NN	6.80E-01	8.39E-03	4.83E-03						
Rborist	2.92E-01	2.65E-02	2.11E-02	6.11E-01					
SLDA	9.29E-02	2.54E-01	2.05E-01	2.05E-01	4.56E-01				
stepLDA	6.56E-04	6.80E-01	6.98E-01	1.92E-03	6.15E-03	1.13E-01			
SVMI	1.35E-03	9.80E-01	9.80E-01	3.07E-03	1.67E-02	2.05E-01	6.98E-01		
SVMr	6.11E-01	2.29E-02	1.67E-02	9.08E-01	7.73E-01	2.92E-01	6.03E-03	1.59E-02	
xgbLinear	4.10E-01	3.58E-04	9.45E-05	1.65E-01	3.91E-02	1.45E-02	9.45E-05	9.45E-05	1.97E-01

Método de selección de características: SAFS.rf									
Nivel taxonómico: Familia									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	4.05E-19								
naive_bayes	2.38E-08	1.54E-06							
NN	1.53E-02	3.15E-14	5.70E-04						
Rborist	7.10E-01	2.81E-19	8.47E-09	6.36E-03					
SLDA	6.36E-03	8.81E-12	7.08E-03	6.17E-01	2.49E-03				
stepLDA	9.44E-14	3.22E-03	2.34E-02	1.48E-08	5.34E-14	2.50E-06			
SVMI	4.31E-01	2.14E-20	4.85E-10	1.59E-03	7.10E-01	5.40E-04	2.55E-15		
SVMr	1.09E-02	3.40E-13	1.80E-03	8.26E-01	4.46E-03	7.58E-01	1.54E-07	1.07E-03	
xgbLinear	2.69E-01	2.14E-20	2.89E-10	9.17E-04	4.92E-01	2.42E-04	2.16E-15	7.58E-01	5.41E-04
Método de selección de características: SAFS.rf									
Nivel taxonómico: Género									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	4.50E-03								
naive_bayes	2.68E-02	6.53E-01							
NN	1.40E-01	1.27E-01	3.97E-01						
Rborist	6.69E-02	5.63E-07	2.35E-05	6.51E-04					
SLDA	6.66E-01	9.59E-03	5.77E-02	3.14E-01	1.74E-02				
stepLDA	2.35E-05	1.27E-01	4.92E-02	3.05E-03	1.23E-10	2.94E-05			
SVMI	7.78E-01	8.18E-03	4.75E-02	2.73E-01	2.68E-02	9.33E-01	2.94E-05		
SVMr	7.94E-01	7.78E-03	4.72E-02	2.63E-01	2.99E-02	9.02E-01	2.94E-05	9.53E-01	
xgbLinear	1.27E-01	3.89E-06	8.92E-05	2.85E-03	7.94E-01	4.31E-02	1.16E-09	5.04E-02	5.90E-02
Método de selección de características: SAFS.rf									
Nivel taxonómico: Especie									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	7.88E-01								
naive_bayes	9.08E-05	8.21E-06							
NN	3.20E-01	1.72E-01	6.29E-03						
Rborist	4.38E-01	5.85E-01	4.87E-06	7.67E-02					
SLDA	1.18E-06	1.19E-07	4.44E-01	1.96E-04	1.19E-07				
stepLDA	3.20E-01	1.72E-01	6.13E-03	1.00E+00	7.67E-02	1.83E-04			
SVMI	2.11E-02	6.32E-03	1.23E-01	2.68E-01	1.75E-03	1.03E-02	2.68E-01		
SVMr	6.89E-01	4.66E-01	9.71E-05	4.74E-01	2.16E-01	1.18E-06	4.55E-01	4.36E-02	
xgbLinear	7.13E-01	8.99E-01	1.70E-05	1.72E-01	6.53E-01	2.73E-07	1.72E-01	6.32E-03	4.38E-01
Método de selección de características: RFE									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	1.08E-06								

naive_bayes	8.66E-03	2.09E-02							
NN	3.28E-01	6.90E-05	9.74E-02						
Rborist	3.15E-02	1.36E-03	3.89E-01	3.28E-01					
SLDA	8.68E-22	8.45E-04	1.67E-12	4.04E-19	1.39E-18				
stepLDA	9.10E-07	3.89E-01	8.45E-02	2.01E-04	3.38E-03	1.25E-09			
SVMl	6.90E-05	2.40E-01	2.30E-01	2.78E-03	2.79E-02	3.17E-08	6.80E-01		
SVMr	1.53E-01	3.76E-04	2.12E-01	6.82E-01	5.10E-01	9.84E-18	7.89E-04	8.90E-03	
xgbLinear	5.10E-01	8.22E-07	3.80E-03	1.53E-01	1.07E-02	1.14E-18	8.22E-07	4.47E-05	5.42E-02

Método de selección de características: RFE

Nivel taxonómico: Clase

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	2.72E-01								
naive_bayes	1.31E-01	6.78E-01							
NN	6.22E-01	5.35E-01	2.97E-01						
Rborist	9.53E-01	3.38E-01	1.58E-01	6.78E-01					
SLDA	2.40E-21	1.32E-19	1.91E-16	1.86E-21	2.19E-21				
stepLDA	3.48E-04	2.56E-02	1.27E-01	2.24E-03	6.26E-04	1.20E-15			
SVMl	5.35E-01	6.22E-01	3.82E-01	9.06E-01	6.22E-01	2.19E-21	3.16E-03		
SVMr	2.14E-01	8.34E-01	7.63E-01	4.55E-01	2.42E-01	1.16E-19	4.62E-02	5.01E-01	
xgbLinear	1.70E-02	2.42E-01	5.35E-01	5.66E-02	2.58E-02	1.40E-17	3.94E-01	7.71E-02	3.40E-01

Método de selección de características: RFE

Nivel taxonómico: Orden

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	5.65E-05								
naive_bayes	1.88E-01	9.86E-03							
NN	3.55E-02	4.08E-02	4.75E-01						
Rborist	8.12E-01	8.77E-06	1.02E-01	1.26E-02					
SLDA	2.53E-21	2.47E-04	4.17E-13	8.07E-13	4.76E-23				
stepLDA	8.80E-06	8.66E-01	5.39E-03	2.36E-02	7.52E-07	3.99E-07			
SVMl	2.70E-01	6.01E-07	2.36E-02	1.18E-03	4.75E-01	9.01E-25	1.13E-08		
SVMr	2.19E-01	1.82E-03	8.25E-01	3.24E-01	1.25E-01	2.08E-18	8.11E-04	2.24E-02	
xgbLinear	3.55E-02	4.54E-02	4.90E-01	9.24E-01	1.50E-02	7.67E-12	4.24E-02	1.50E-03	3.11E-01

Método de selección de características: RFE

Nivel taxonómico: Familia

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	1.19E-11								
naive_bayes	4.50E-08	4.09E-01							
NN	1.21E-01	1.13E-07	4.79E-05						
Rborist	8.17E-01	2.06E-11	5.50E-08	1.87E-01					
SLDA	1.54E-18	2.41E-01	5.53E-02	2.71E-12	1.54E-18				

stepLDA	1.76E-10	4.09E-01	9.98E-01	3.12E-06	2.99E-10	9.25E-03			
SVMl	7.67E-01	9.48E-11	1.96E-07	2.34E-01	9.31E-01	3.34E-17	1.93E-09		
SVMr	4.09E-01	1.14E-08	8.50E-06	5.63E-01	5.03E-01	9.70E-14	2.53E-07	5.65E-01	
xgbLinear	1.47E-02	4.11E-06	8.20E-04	4.53E-01	2.83E-02	8.10E-11	5.84E-05	4.10E-02	1.87E-01
Método de selección de características: RFE									
Nivel taxonómico: Género									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	2.83E-06								
naive_bayes	4.26E-01	5.90E-05							
NN	1.73E-02	1.76E-02	9.97E-02						
Rborist	7.78E-02	6.78E-10	1.11E-02	5.44E-05					
SLDA	3.69E-04	3.05E-01	3.77E-03	2.18E-01	3.50E-07				
stepLDA	1.16E-11	1.61E-02	2.57E-10	2.07E-06	1.69E-15	6.34E-04			
SVMl	2.40E-01	1.02E-08	4.53E-02	4.15E-04	5.19E-01	2.83E-06	1.53E-14		
SVMr	6.54E-01	2.29E-06	2.62E-01	1.08E-02	2.27E-01	2.61E-04	1.78E-11	4.99E-01	
xgbLinear	2.98E-01	4.45E-04	7.16E-01	2.28E-01	7.64E-03	1.56E-02	1.02E-08	3.18E-02	2.00E-01
Método de selección de características: RFE									
Nivel taxonómico: Especie									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	6.33E-01								
naive_bayes	9.56E-01	6.33E-01							
NN	6.33E-01	8.60E-01	6.32E-01						
Rborist	7.48E-01	3.65E-01	6.33E-01	3.65E-01					
SLDA	6.33E-01	9.89E-01	6.33E-01	8.60E-01	3.65E-01				
stepLDA	3.65E-01	5.97E-01	3.56E-01	6.33E-01	1.05E-01	5.97E-01			
SVMl	4.15E-01	6.33E-01	3.65E-01	7.89E-01	1.46E-01	6.33E-01	8.17E-01		
SVMr	8.00E-01	8.17E-01	8.00E-01	7.50E-01	5.97E-01	8.17E-01	4.15E-01	5.97E-01	
xgbLinear	6.33E-01	8.60E-01	5.97E-01	9.89E-01	3.65E-01	8.60E-01	6.33E-01	7.65E-01	7.48E-01
Método de selección de características: SBF									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	2.87E-04								
naive_bayes	2.16E-01	7.17E-06							
NN	8.73E-02	5.33E-07	7.59E-01						
Rborist	2.78E-01	1.52E-05	9.37E-01	7.15E-01					
SLDA	6.56E-10	7.17E-01	1.94E-12	2.97E-13	6.39E-10				
stepLDA	1.86E-01	9.20E-03	9.43E-03	3.35E-03	2.15E-02	2.18E-06			
SVMl	2.40E-01	5.72E-03	1.42E-02	4.58E-03	3.03E-02	2.57E-07	8.49E-01		
SVMr	1.34E-01	5.33E-07	8.49E-01	8.49E-01	7.92E-01	7.55E-14	4.51E-03	6.64E-03	
xgbLinear	1.37E-01	1.96E-02	8.24E-03	2.24E-03	1.52E-02	1.08E-05	8.49E-01	7.91E-01	3.38E-03

Método de selección de características: SBF									
Nivel taxonómico: Clase									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	7.49E-01								
naive_bayes	8.47E-04	2.09E-04							
NN	3.45E-01	1.57E-01	6.64E-02						
Rborist	7.49E-01	9.99E-01	6.63E-05	1.15E-01					
SLDA	3.16E-05	3.11E-06	2.86E-01	2.63E-03	8.53E-07				
stepLDA	2.18E-01	1.14E-01	3.76E-02	9.99E-01	7.53E-02	8.47E-04			
SVMI	9.53E-01	7.92E-01	2.09E-04	2.18E-01	7.74E-01	3.11E-06	1.57E-01		
SVMr	8.13E-01	5.68E-01	4.44E-03	5.54E-01	5.54E-01	1.67E-04	4.34E-01	7.49E-01	
xgbLinear	6.64E-02	2.45E-02	1.15E-01	7.12E-01	1.08E-02	3.41E-03	6.23E-01	3.36E-02	1.58E-01
Método de selección de características: SBF									
Nivel taxonómico: Orden									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.90E-01								
naive_bayes	1.23E-02	8.43E-04							
NN	8.45E-01	4.78E-01	4.04E-03						
Rborist	3.67E-03	1.85E-04	6.27E-01	1.15E-03					
SLDA	2.16E-28	7.21E-25	3.29E-31	4.37E-29	1.55E-31				
stepLDA	3.09E-08	1.37E-05	5.63E-15	1.47E-08	1.34E-15	4.73E-22			
SVMI	1.52E-01	2.69E-02	2.54E-01	8.72E-02	1.09E-01	1.77E-29	6.91E-12		
SVMr	1.41E-01	2.12E-02	2.54E-01	8.02E-02	1.09E-01	1.55E-31	6.53E-13	1.00E+00	
xgbLinear	3.70E-01	8.02E-02	8.55E-02	2.49E-01	3.02E-02	4.47E-30	1.34E-11	5.45E-01	5.45E-01
Método de selección de características: SBF									
Nivel taxonómico: Familia									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	4.57E-02								
naive_bayes	3.59E-02	2.21E-05							
NN	9.31E-01	5.38E-02	2.46E-02						
Rborist	2.90E-01	1.35E-03	2.66E-01	2.60E-01					
SLDA	9.61E-17	7.44E-13	8.46E-24	5.19E-18	1.07E-20				
stepLDA	4.10E-02	9.31E-01	1.36E-05	4.93E-02	9.75E-04	7.44E-13			
SVMI	4.95E-02	2.48E-05	7.99E-01	3.66E-02	3.49E-01	4.09E-24	1.52E-05		
SVMr	2.54E-04	6.18E-09	1.59E-01	1.31E-04	7.50E-03	1.76E-27	1.42E-09	6.46E-02	
xgbLinear	3.06E-01	2.71E-01	9.85E-04	3.49E-01	3.30E-02	3.84E-16	2.73E-01	1.35E-03	7.75E-07
Método de selección de características: SBF									
Nivel taxonómico: Género									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr

LVQ	1.24E-03								
naive_bayes	6.75E-02	1.60E-07							
NN	9.13E-01	6.50E-04	8.51E-02						
Rborist	1.05E-01	3.63E-07	8.72E-01	1.25E-01					
SLDA	3.67E-02	1.89E-01	1.57E-05	2.44E-02	3.73E-05				
stepLDA	3.69E-08	1.26E-02	4.42E-13	1.80E-08	7.08E-13	9.61E-05			
SVMI	8.51E-02	3.63E-07	9.39E-01	1.05E-01	9.13E-01	3.71E-05	7.32E-13		
SVMr	4.92E-02	5.03E-08	9.13E-01	6.18E-02	7.47E-01	6.57E-06	1.96E-13	8.72E-01	
xgbLinear	9.62E-01	7.31E-04	6.18E-02	9.35E-01	9.05E-02	3.12E-02	1.77E-08	8.17E-02	4.09E-02
Método de selección de características: SBF									
Nivel taxonómico: Especie									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	6.93E-01								
naive_bayes	8.68E-01	8.68E-01							
NN	5.35E-01	6.64E-01	5.74E-01						
Rborist	8.68E-01	5.95E-01	6.93E-01	3.45E-01					
SLDA	3.87E-01	5.74E-01	5.35E-01	9.24E-01	2.77E-01				
stepLDA	7.73E-01	9.51E-01	9.24E-01	6.20E-01	6.20E-01	5.74E-01			
SVMI	7.84E-01	5.74E-01	6.41E-01	2.77E-01	9.24E-01	2.40E-01	5.74E-01		
SVMr	6.93E-01	9.86E-01	8.68E-01	6.20E-01	5.74E-01	5.74E-01	9.26E-01	5.74E-01	
xgbLinear	3.45E-01	1.67E-01	2.00E-01	2.46E-02	5.46E-01	2.36E-02	1.82E-01	5.74E-01	1.20E-01

8.4. Tabla suplementaria 2C

Presenta los valores de significancia estadística de las comparaciones entre la exactitud de los diferentes niveles taxonómicos en cada modelo de predicción, luego de la aplicación de cada método de selección de características. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 2C					
Método de selección de características: SAFS					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	3.00E-03				
Género	5.99E-01	2.05E-04			
Orden	8.63E-03	5.99E-01	3.26E-04		
Phylum	1.04E-02	5.20E-01	3.30E-04	7.24E-01	

Especie	9.65E-06	2.83E-13	9.65E-06	1.39E-13	7.23E-14
Método de selección de características: SAFS					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	2.67E-07				
Género	1.95E-02	7.91E-11			
Orden	9.46E-01	8.61E-07	2.11E-02		
Phylum	1.60E-01	5.68E-04	6.88E-04	1.80E-01	
Especie	5.61E-06	1.33E-15	4.05E-02	8.98E-06	2.39E-07
Método de selección de características: SAFS					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	1.25E-09				
Género	9.31E-05	9.67E-18			
Orden	1.22E-02	3.61E-04	8.53E-09		
Phylum	3.87E-02	1.32E-05	1.17E-08	5.72E-01	
Especie	2.00E-01	2.28E-11	1.00E-02	5.49E-04	1.60E-03
Método de selección de características: SAFS					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	2.89E-06				
Género	3.84E-03	4.92E-13			
Orden	2.67E-02	1.01E-02	2.37E-07		
Phylum	7.86E-04	1.73E-01	7.90E-10	2.02E-01	
Especie	5.30E-01	7.28E-07	2.67E-02	6.05E-03	1.65E-04
Método de selección de características: SAFS					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	2.00E-11				
Género	5.98E-01	1.33E-14			
Orden	5.49E-02	1.97E-12	1.07E-04		
Phylum	4.34E-02	2.77E-12	9.18E-05	9.70E-01	
Especie	5.44E-04	9.16E-18	1.56E-04	2.00E-11	1.33E-11
Método de selección de características: SAFS					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	3.80E-03				
Género	1.69E-01	9.33E-06			

Orden	1.37E-01	1.77E-01	3.80E-03		
Phylum	6.72E-01	7.72E-03	7.84E-02	2.08E-01	
Especie	4.56E-03	4.07E-10	1.69E-01	9.33E-06	9.04E-04
Método de selección de características: SAFS					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	5.87E-06				
Género	8.47E-01	3.60E-06			
Orden	5.26E-04	3.35E-01	5.16E-04		
Phylum	1.72E-05	7.74E-01	1.11E-05	6.15E-01	
Especie	8.81E-01	9.10E-07	7.81E-01	5.49E-05	1.58E-06
Método de selección de características: SAFS					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	1.68E-05				
Género	6.10E-01	8.55E-06			
Orden	1.71E-02	3.83E-02	7.67E-03		
Phylum	5.64E-02	1.27E-02	2.73E-02	5.87E-01	
Especie	1.05E-01	3.47E-09	2.40E-01	3.21E-05	3.05E-04
Método de selección de características: SAFS					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	2.27E-02				
Género	6.42E-07	3.32E-03			
Orden	3.84E-01	1.13E-01	4.47E-06		
Phylum	3.60E-01	7.91E-02	5.60E-07	9.29E-01	
Especie	7.90E-07	2.10E-13	5.28E-21	3.69E-10	9.04E-12
Método de selección de características: SAFS					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	1.05E-03				
Género	9.61E-04	1.38E-10			
Orden	6.80E-01	3.87E-04	4.23E-03		
Phylum	2.50E-01	1.38E-06	1.82E-02	4.53E-01	
Especie	1.62E-02	2.58E-08	3.25E-01	6.33E-02	2.11E-01
Método de selección de características: SAFS.rf					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum

Familia	8.75E-02				
Género	9.44E-05	1.00E-01			
Orden	2.18E-03	2.81E-01	5.21E-01		
Phylum	2.24E-01	5.21E-01	9.82E-03	6.32E-02	
Especie	1.47E-18	4.10E-12	1.05E-10	9.23E-11	6.70E-15
Método de selección de características: SAFS.rf					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	6.28E-13				
Género	1.95E-01	9.50E-09			
Orden	2.41E-01	6.77E-09	8.98E-01		
Phylum	8.26E-01	3.38E-12	2.65E-01	3.01E-01	
Especie	5.94E-05	3.15E-04	1.71E-02	1.25E-02	1.95E-04
Método de selección de características: SAFS.rf					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	1.06E-06				
Género	2.43E-01	2.04E-04			
Orden	6.09E-01	1.11E-04	6.26E-01		
Phylum	9.35E-02	5.03E-04	6.55E-01	3.45E-01	
Especie	1.11E-04	1.38E-01	1.27E-02	3.43E-03	3.12E-02
Método de selección de características: SAFS.rf					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	3.22E-12				
Género	7.18E-01	3.22E-12			
Orden	3.42E-03	4.99E-22	4.70E-04		
Phylum	7.74E-01	2.54E-14	5.44E-01	4.39E-03	
Especie	5.03E-01	4.37E-11	7.18E-01	6.46E-05	2.99E-01
Método de selección de características: SAFS.rf					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	4.55E-04				
Género	9.80E-01	2.62E-04			
Orden	1.27E-01	7.93E-07	1.10E-01		
Phylum	1.41E-28	7.68E-31	1.92E-27	2.29E-26	
Especie	4.60E-02	3.87E-08	2.99E-02	7.12E-01	1.60E-26
Método de selección de características: SAFS.rf					

Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	7.56E-07				
Género	8.32E-01	2.51E-07			
Orden	5.43E-01	7.96E-06	4.45E-01		
Phylum	2.11E-01	3.19E-04	1.43E-01	4.66E-01	
Especie	1.35E-02	1.80E-02	8.11E-03	5.70E-02	2.85E-01
Método de selección de características: SAFS.rf					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	1.24E-14				
Género	1.06E-04	9.00E-06			
Orden	6.42E-01	1.21E-13	5.33E-04		
Phylum	4.48E-02	4.42E-10	5.71E-02	9.76E-02	
Especie	2.95E-07	2.75E-03	1.29E-01	2.26E-06	7.00E-04
Método de selección de características: SAFS.rf					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	2.29E-20				
Género	6.85E-08	1.79E-07			
Orden	1.80E-05	6.85E-08	5.23E-01		
Phylum	1.26E-01	2.85E-15	1.48E-04	5.52E-03	
Especie	6.67E-11	1.90E-05	1.98E-01	9.21E-02	8.06E-07
Método de selección de características: SAFS.rf					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	2.45E-01				
Género	9.09E-02	2.70E-03			
Orden	1.47E-01	1.32E-02	6.79E-01		
Phylum	2.32E-03	3.10E-05	1.74E-01	1.27E-01	
Especie	1.28E-04	3.15E-03	5.24E-09	1.39E-07	1.12E-11
Método de selección de características: SAFS.rf					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	1.18E-03				
Género	7.27E-01	4.21E-03			
Orden	4.80E-01	4.37E-05	3.06E-01		
Phylum	2.33E-04	5.94E-01	8.72E-04	7.35E-06	
Especie	3.44E-01	3.59E-02	4.80E-01	7.51E-02	7.64E-03

Método de selección de características: RFE					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	8.25E-07				
Género	4.64E-01	1.74E-05			
Orden	9.29E-03	2.52E-02	4.83E-02		
Phylum	2.39E-04	3.02E-01	2.18E-03	2.64E-01	
Especie	2.18E-03	3.22E-14	1.87E-04	1.04E-07	2.11E-10
Método de selección de características: RFE					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	6.19E-01				
Género	4.24E-02	1.41E-01			
Orden	8.20E-01	8.20E-01	5.97E-02		
Phylum	9.11E-01	4.59E-01	3.03E-02	7.97E-01	
Especie	5.97E-02	1.73E-01	9.19E-01	1.02E-01	4.24E-02
Método de selección de características: RFE					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	2.90E-01				
Género	7.01E-05	6.36E-03			
Orden	5.43E-01	5.45E-01	4.11E-04		
Phylum	9.50E-01	2.90E-01	5.41E-05	5.43E-01	
Especie	6.03E-04	5.29E-02	4.90E-01	4.54E-03	4.39E-04
Método de selección de características: RFE					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	2.59E-01				
Género	6.52E-06	1.45E-03			
Orden	2.74E-02	3.97E-01	7.48E-03		
Phylum	1.10E-03	1.41E-05	1.14E-13	5.56E-08	
Especie	9.74E-02	6.77E-01	3.28E-03	6.33E-01	1.10E-06
Método de selección de características: RFE					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	1.09E-09				
Género	6.46E-21	6.26E-15			
Orden	8.28E-10	8.05E-01	4.00E-17		

Phylum	1.69E-08	5.70E-02	2.22E-19	5.70E-02	
Especie	2.18E-26	3.88E-23	6.07E-03	1.48E-25	8.94E-27
Método de selección de características: RFE					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	7.23E-01				
Género	4.78E-01	3.99E-01			
Orden	3.51E-01	4.78E-01	8.60E-02		
Phylum	6.84E-01	8.39E-01	3.44E-01	5.39E-01	
Especie	1.49E-01	8.60E-02	3.99E-01	1.39E-02	7.48E-02
Método de selección de características: RFE					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	5.41E-01				
Género	1.26E-04	2.48E-04			
Orden	5.33E-01	9.40E-01	2.48E-04		
Phylum	6.06E-02	4.36E-03	3.72E-11	3.16E-03	
Especie	2.30E-03	4.36E-03	4.05E-01	4.36E-03	8.42E-09
Método de selección de características: RFE					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	4.06E-01				
Género	1.10E-04	2.53E-03			
Orden	5.16E-01	8.72E-01	2.53E-03		
Phylum	4.06E-03	4.94E-02	4.26E-01	4.77E-02	
Especie	1.10E-04	2.53E-03	8.95E-01	2.53E-03	4.46E-01
Método de selección de características: RFE					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	5.79E-02				
Género	5.06E-01	4.31E-01			
Orden	7.24E-01	1.32E-01	6.94E-01		
Phylum	5.06E-01	2.30E-01	8.75E-01	7.24E-01	
Especie	8.89E-06	1.06E-09	1.02E-06	2.63E-06	8.82E-08
Método de selección de características: RFE					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	2.48E-03				

Género	2.94E-04	5.99E-11			
Orden	4.46E-01	2.16E-04	3.32E-03		
Phylum	4.46E-01	1.13E-02	3.37E-06	1.37E-01	
Especie	2.64E-05	7.38E-12	4.46E-01	4.20E-04	1.93E-07
Método de selección de características: SBF					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	1.74E-03				
Género	4.03E-01	1.48E-02			
Orden	6.31E-01	1.14E-02	7.32E-01		
Phylum	2.73E-09	1.44E-03	8.91E-08	8.91E-08	
Especie	1.04E-04	1.22E-10	1.96E-06	1.99E-05	4.21E-18
Método de selección de características: SBF					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	6.47E-01				
Género	3.87E-03	1.64E-03			
Orden	4.32E-01	2.13E-01	4.41E-02		
Phylum	2.18E-02	8.94E-02	6.86E-07	2.17E-03	
Especie	4.10E-04	1.69E-04	5.63E-01	6.68E-03	1.69E-08
Método de selección de características: SBF					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	4.57E-04				
Género	3.60E-07	2.91E-02			
Orden	8.38E-03	3.34E-01	2.04E-03		
Phylum	7.30E-01	6.77E-05	3.88E-08	1.90E-03	
Especie	5.74E-06	2.03E-01	3.34E-01	2.69E-02	5.16E-07
Método de selección de características: SBF					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	1.04E-01				
Género	6.00E-07	5.08E-04			
Orden	1.57E-02	3.43E-01	9.01E-03		
Phylum	7.28E-05	9.54E-08	3.34E-15	4.80E-09	
Especie	1.73E-03	9.65E-02	9.65E-02	3.43E-01	3.25E-10
Método de selección de características: SBF					
Modelo de predicción: SLDA					

	Clase	Familia	Género	Orden	Phylum
Familia	8.10E-09				
Género	1.03E-08	2.79E-23			
Orden	1.48E-18	1.23E-06	1.09E-29		
Phylum	5.33E-03	3.43E-09	7.85E-22	3.23E-24	
Especie	3.04E-18	1.09E-29	1.85E-06	1.33E-32	1.76E-30
Método de selección de características: SBF					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	6.07E-01				
Género	1.01E-04	5.15E-04			
Orden	6.77E-02	2.22E-01	1.22E-02		
Phylum	5.08E-01	8.63E-01	7.61E-04	2.95E-01	
Especie	9.07E-08	1.75E-07	8.09E-02	1.71E-05	4.38E-07
Método de selección de características: SBF					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	7.70E-01				
Género	5.24E-06	2.89E-05			
Orden	2.29E-04	1.01E-03	4.53E-01		
Phylum	1.71E-01	1.66E-01	7.54E-07	2.04E-05	
Especie	3.27E-03	9.38E-03	1.71E-01	5.74E-01	1.76E-04
Método de selección de características: SBF					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	7.46E-01				
Género	2.21E-07	5.98E-06			
Orden	3.65E-05	4.15E-04	1.81E-01		
Phylum	3.77E-02	3.21E-02	1.76E-10	1.91E-08	
Especie	2.75E-05	3.55E-04	2.88E-01	7.46E-01	1.91E-08
Método de selección de características: SBF					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	1.68E-01				
Género	9.51E-02	7.64E-01			
Orden	4.29E-04	9.51E-02	1.68E-01		
Phylum	1.82E-02	4.13E-01	5.98E-01	3.70E-01	
Especie	2.05E-08	4.39E-11	1.97E-11	3.44E-16	9.90E-14

Método de selección de características: SBF					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	1.67E-07				
Género	3.25E-15	1.52E-03			
Orden	1.38E-11	9.94E-02	1.02E-01		
Phylum	1.21E-02	1.03E-02	1.96E-08	2.00E-05	
Especie	3.44E-13	2.58E-02	2.69E-01	5.47E-01	1.40E-06

8.5. Tabla suplementaria 2D

Presenta los valores de significancia estadística de las comparaciones entre la exactitud de los diferentes métodos de selección de características, en cada modelo de predicción, discriminado por nivel taxonómico. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 2D				
Nivel taxonómico: Phylum				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	2.88E-01			
SAFS	3.79E-01	6.12E-01		
SAFS.rf	8.58E-01	3.79E-01	4.97E-01	
SBF	2.61E-01	5.29E-02	5.29E-02	2.61E-01
Nivel taxonómico: Phylum				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	6.75E-04			
SAFS	9.13E-01	6.75E-04		
SAFS.rf	2.64E-01	4.23E-02	2.51E-01	
SBF	4.41E-01	3.22E-05	4.96E-01	4.40E-02
Nivel taxonómico: Phylum				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	7.59E-01			
SAFS	8.62E-01	7.99E-01		

SAFS.rf	2.08E-01	1.04E-01	1.29E-01	
SBF	1.04E-01	1.81E-01	1.04E-01	1.13E-03
Nivel taxonómico: Phylum				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	1.22E-01			
SAFS	6.80E-01	2.55E-01		
SAFS.rf	7.38E-02	6.80E-01	1.22E-01	
SBF	6.80E-01	2.57E-01	9.62E-01	1.22E-01
Nivel taxonómico: Phylum				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	2.05E-03			
SAFS	1.84E-04	7.22E-02		
SAFS.rf	5.40E-23	3.46E-25	1.19E-24	
SBF	1.07E-04	2.08E-03	1.68E-01	1.20E-24
Nivel taxonómico: Phylum				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	1.85E-01			
SAFS	9.33E-01	1.85E-01		
SAFS.rf	1.85E-01	9.33E-01	1.85E-01	
SBF	9.33E-01	1.85E-01	9.33E-01	1.85E-01
Nivel taxonómico: Phylum				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	7.25E-01			
SAFS	2.85E-01	2.85E-01		
SAFS.rf	5.82E-01	2.85E-01	6.64E-02	
SBF	3.26E-01	3.52E-01	9.54E-01	1.52E-01
Nivel taxonómico: Phylum				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	9.40E-07			
SAFS	3.07E-01	2.45E-04		
SAFS.rf	1.95E-01	1.15E-03	6.74E-01	
SBF	1.95E-01	2.81E-08	3.22E-02	1.41E-02

Nivel taxonómico: Phylum				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	9.61E-01			
SAFS	7.40E-01	7.40E-01		
SAFS.rf	9.54E-03	2.08E-02	2.33E-02	
SBF	1.76E-01	2.21E-01	3.27E-01	3.74E-01
Nivel taxonómico: Phylum				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	8.53E-01			
SAFS	8.53E-01	8.53E-01		
SAFS.rf	2.68E-02	1.38E-02	7.04E-03	
SBF	8.53E-01	8.53E-01	8.53E-01	1.22E-02
Nivel taxonómico: Clase				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	1.01E-13			
SAFS	1.08E-08	1.45E-01		
SAFS.rf	8.51E-02	5.45E-10	2.19E-05	
SBF	1.27E-13	8.08E-01	1.79E-01	1.55E-09
Nivel taxonómico: Clase				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	7.40E-01			
SAFS	4.59E-02	3.78E-02		
SAFS.rf	4.59E-02	3.78E-02	8.27E-01	
SBF	1.26E-01	6.82E-02	7.40E-01	7.40E-01
Nivel taxonómico: Clase				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	9.88E-01			
SAFS	9.88E-01	9.88E-01		
SAFS.rf	3.43E-01	3.76E-01	3.43E-01	
SBF	3.43E-01	3.76E-01	3.43E-01	9.88E-01
Nivel taxonómico: Clase				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf

RFE	7.53E-03			
SAFS	1.23E-02	5.71E-01		
SAFS.rf	1.73E-04	3.40E-01	9.95E-02	
SBF	1.01E-03	5.07E-01	2.46E-01	7.01E-01
Nivel taxonómico: Clase				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	1.18E-06			
SAFS	1.65E-02	2.71E-07		
SAFS.rf	1.35E-15	9.48E-19	1.70E-07	
SBF	4.89E-03	1.70E-07	6.96E-01	1.18E-06
Nivel taxonómico: Clase				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	6.04E-01			
SAFS	4.28E-01	2.46E-01		
SAFS.rf	3.53E-01	1.89E-01	8.47E-01	
SBF	1.11E-01	6.24E-02	4.28E-01	5.11E-01
Nivel taxonómico: Clase				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	9.98E-01			
SAFS	2.83E-01	2.83E-01		
SAFS.rf	9.64E-03	1.18E-02	3.95E-04	
SBF	1.64E-01	1.66E-01	9.64E-03	2.03E-01
Nivel taxonómico: Clase				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	1.15E-01			
SAFS	5.88E-01	3.07E-01		
SAFS.rf	3.14E-04	4.42E-02	2.63E-03	
SBF	2.63E-03	2.50E-01	2.81E-02	3.07E-01
Nivel taxonómico: Clase				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	3.76E-02			
SAFS	4.56E-02	9.85E-01		
SAFS.rf	4.56E-02	9.85E-01	9.85E-01	

SBF	2.27E-02	9.49E-01	9.49E-01	9.49E-01
Nivel taxonómico: Clase				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	3.14E-02			
SAFS	8.18E-01	5.12E-02		
SAFS.rf	6.59E-01	1.15E-01	7.56E-01	
SBF	3.62E-01	3.18E-03	3.02E-01	1.81E-01
Nivel taxonómico: Orden				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	5.23E-01			
SAFS	1.73E-02	1.21E-01		
SAFS.rf	2.57E-01	6.79E-01	1.92E-01	
SBF	3.10E-01	1.21E-01	6.47E-04	4.39E-02
Nivel taxonómico: Orden				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	6.29E-01			
SAFS	5.88E-02	2.77E-02		
SAFS.rf	3.51E-01	2.91E-01	3.33E-01	
SBF	3.33E-01	2.87E-01	3.33E-01	1.00E+00
Nivel taxonómico: Orden				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	5.73E-03			
SAFS	1.13E-01	1.58E-05		
SAFS.rf	1.15E-04	3.62E-11	2.55E-02	
SBF	1.57E-01	2.21E-01	3.34E-03	1.44E-07
Nivel taxonómico: Orden				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	6.95E-02			
SAFS	5.73E-01	2.93E-02		
SAFS.rf	5.02E-01	2.98E-01	2.98E-01	
SBF	5.16E-03	2.98E-01	8.13E-04	6.55E-02
Nivel taxonómico: Orden				

Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	5.86E-02			
SAFS	1.74E-01	2.88E-04		
SAFS.rf	1.65E-08	1.10E-07	5.85E-11	
SBF	5.92E-23	9.89E-26	7.99E-25	6.21E-25
Nivel taxonómico: Orden				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	6.00E-01			
SAFS	1.79E-02	7.27E-02		
SAFS.rf	6.00E-01	9.50E-01	7.27E-02	
SBF	8.05E-01	8.44E-01	5.16E-02	8.44E-01
Nivel taxonómico: Orden				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	8.27E-01			
SAFS	1.12E-03	1.13E-03		
SAFS.rf	1.12E-03	1.12E-03	8.90E-01	
SBF	4.69E-01	2.90E-01	6.92E-05	6.92E-05
Nivel taxonómico: Orden				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	5.11E-01			
SAFS	5.49E-03	5.27E-02		
SAFS.rf	4.39E-01	1.83E-01	7.64E-04	
SBF	4.55E-01	1.77E-01	2.14E-04	7.21E-01
Nivel taxonómico: Orden				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	6.03E-01			
SAFS	3.36E-01	6.03E-01		
SAFS.rf	2.08E-01	3.36E-01	6.03E-01	
SBF	4.38E-03	1.23E-02	7.33E-02	2.13E-01
Nivel taxonómico: Orden				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	3.30E-01			

SAFS	8.22E-02	1.35E-02		
SAFS.rf	6.58E-03	6.10E-04	4.22E-01	
SBF	6.42E-04	2.02E-02	1.49E-06	1.04E-09
Nivel taxonómico: Familia				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	7.42E-01			
SAFS	7.07E-01	8.98E-01		
SAFS.rf	8.98E-01	7.42E-01	7.42E-01	
SBF	7.07E-01	5.34E-01	5.34E-01	7.07E-01
Nivel taxonómico: Familia				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	1.28E-05			
SAFS	1.36E-04	1.42E-13		
SAFS.rf	4.10E-15	2.36E-05	6.29E-24	
SBF	3.25E-01	7.25E-04	1.28E-05	1.87E-12
Nivel taxonómico: Familia				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	6.33E-02			
SAFS	3.53E-05	4.10E-08		
SAFS.rf	1.66E-12	2.49E-07	4.68E-22	
SBF	3.80E-01	2.83E-01	8.23E-07	7.85E-11
Nivel taxonómico: Familia				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	3.53E-02			
SAFS	3.38E-06	2.37E-09		
SAFS.rf	6.12E-07	1.13E-02	3.73E-17	
SBF	7.72E-05	1.91E-01	4.49E-15	1.28E-01
Nivel taxonómico: Familia				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	4.20E-04			
SAFS	7.24E-13	7.75E-15		
SAFS.rf	5.35E-28	1.12E-22	8.50E-28	
SBF	3.58E-10	1.72E-12	8.45E-03	8.50E-28

Nivel taxonómico: Familia				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	7.56E-04			
SAFS	2.47E-01	5.32E-06		
SAFS.rf	1.11E-12	3.78E-04	2.25E-16	
SBF	1.19E-01	9.26E-02	6.50E-03	1.82E-07
Nivel taxonómico: Familia				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	7.59E-03			
SAFS	1.41E-01	1.54E-05		
SAFS.rf	8.78E-12	3.22E-06	1.88E-16	
SBF	4.23E-01	3.90E-02	2.03E-02	3.02E-11
Nivel taxonómico: Familia				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	7.22E-02			
SAFS	3.45E-03	7.78E-06		
SAFS.rf	1.65E-17	6.84E-13	4.34E-23	
SBF	7.63E-01	5.40E-02	8.62E-03	6.16E-18
Nivel taxonómico: Familia				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	4.61E-01			
SAFS	6.63E-01	6.33E-01		
SAFS.rf	3.10E-04	3.40E-03	8.72E-04	
SBF	1.29E-01	4.61E-01	2.72E-01	5.46E-02
Nivel taxonómico: Familia				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	1.68E-02			
SAFS	3.00E-06	1.68E-02		
SAFS.rf	6.16E-02	7.17E-06	1.01E-11	
SBF	3.52E-02	3.41E-06	4.29E-12	7.79E-01
Nivel taxonómico: Género				
Modelo de predicción: LVQ				

	None	RFE	SAFS	SAFS.rf
RFE	1.82E-01			
SAFS	1.82E-01	8.70E-01		
SAFS.rf	1.00E+00	1.82E-01	1.82E-01	
SBF	1.82E-01	8.70E-01	1.00E+00	1.82E-01
Nivel taxonómico: Género				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	5.49E-02			
SAFS	5.30E-01	1.36E-02		
SAFS.rf	1.85E-01	1.59E-03	4.48E-01	
SBF	6.25E-01	1.76E-01	3.92E-01	1.26E-01
Nivel taxonómico: Género				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	4.00E-04			
SAFS	1.85E-02	2.53E-01		
SAFS.rf	1.80E-01	3.84E-07	9.14E-05	
SBF	9.96E-03	4.21E-01	7.21E-01	5.04E-05
Nivel taxonómico: Género				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	6.43E-04			
SAFS	1.62E-03	6.21E-01		
SAFS.rf	4.85E-01	6.09E-05	1.23E-04	
SBF	2.24E-04	8.61E-01	5.55E-01	2.56E-05
Nivel taxonómico: Género				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	2.60E-09			
SAFS	9.52E-01	1.92E-11		
SAFS.rf	7.65E-08	8.07E-01	4.68E-10	
SBF	2.60E-09	9.36E-01	1.92E-11	8.96E-01
Nivel taxonómico: Género				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	6.34E-03			
SAFS	1.02E-01	3.78E-01		

SAFS.rf	9.39E-01	5.99E-03	1.02E-01	
SBF	1.41E-03	4.25E-01	1.03E-01	1.39E-03
Nivel taxonómico: Género				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	8.77E-03			
SAFS	5.09E-01	1.11E-03		
SAFS.rf	5.09E-01	1.11E-03	9.15E-01	
SBF	2.88E-01	1.32E-01	8.68E-02	8.68E-02
Nivel taxonómico: Género				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	7.45E-01			
SAFS	1.03E-01	6.97E-02		
SAFS.rf	7.45E-01	6.77E-01	1.76E-01	
SBF	9.87E-01	7.45E-01	1.03E-01	7.45E-01
Nivel taxonómico: Género				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	2.17E-02			
SAFS	8.92E-02	9.29E-05		
SAFS.rf	9.22E-02	3.77E-01	3.42E-04	
SBF	7.39E-02	6.52E-01	3.42E-04	7.04E-01
Nivel taxonómico: Género				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	6.01E-01			
SAFS	2.05E-02	6.01E-03		
SAFS.rf	7.37E-07	3.52E-08	6.84E-03	
SBF	1.30E-01	2.43E-01	2.35E-04	1.01E-09
Nivel taxonómico: Especie				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	4.84E-01			
SAFS	3.56E-01	7.89E-01		
SAFS.rf	4.49E-01	9.62E-01	7.89E-01	
SBF	3.56E-01	7.89E-01	9.62E-01	8.65E-01

Nivel taxonómico: Especie				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	1.49E-01			
SAFS	2.40E-01	7.06E-01		
SAFS.rf	3.04E-01	4.77E-01	7.77E-01	
SBF	2.88E-01	5.82E-01	8.50E-01	8.83E-01
Nivel taxonómico: Especie				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	1.63E-02			
SAFS	9.62E-02	5.51E-01		
SAFS.rf	5.51E-01	9.62E-02	3.54E-01	
SBF	1.63E-02	8.32E-01	5.42E-01	9.62E-02
Nivel taxonómico: Especie				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	1.71E-04			
SAFS	6.00E-02	5.81E-02		
SAFS.rf	2.97E-03	2.80E-01	2.86E-01	
SBF	4.05E-05	8.36E-01	2.79E-02	1.66E-01
Nivel taxonómico: Especie				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	1.77E-01			
SAFS	1.97E-05	1.67E-08		
SAFS.rf	8.54E-05	1.14E-07	8.37E-01	
SBF	4.77E-03	1.41E-01	1.56E-11	1.24E-10
Nivel taxonómico: Especie				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	4.09E-01			
SAFS	7.19E-01	6.64E-01		
SAFS.rf	9.99E-01	4.09E-01	7.19E-01	
SBF	3.11E-02	3.20E-01	3.98E-02	3.11E-02
Nivel taxonómico: Especie				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf

RFE	2.61E-02			
SAFS	3.97E-01	9.76E-02		
SAFS.rf	1.41E-01	3.97E-01	3.97E-01	
SBF	4.75E-01	9.76E-02	8.74E-01	3.97E-01
Nivel taxonómico: Especie				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	3.58E-01			
SAFS	7.39E-01	2.30E-01		
SAFS.rf	2.30E-01	7.39E-01	1.99E-01	
SBF	7.39E-01	2.30E-01	9.91E-01	1.99E-01
Nivel taxonómico: Especie				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	9.29E-01			
SAFS	3.39E-01	4.33E-01		
SAFS.rf	9.66E-01	9.29E-01	3.39E-01	
SBF	3.03E-01	3.08E-01	7.93E-01	3.03E-01
Nivel taxonómico: Especie				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	6.61E-06			
SAFS	9.15E-01	6.82E-06		
SAFS.rf	4.92E-01	4.19E-07	4.79E-01	
SBF	8.82E-05	5.52E-01	9.44E-05	6.61E-06

8.6. Tabla suplementaria 3A

Presenta el área bajo la curva (ABC) obtenidas para cada Modelo de Predicción en cada nivel taxonómico y Método de selección de características. La evaluación se realizó utilizando un muestreo aleatorio con reemplazo, tomando una muestra del 90% de los datos del dataset de prueba en cada iteración, y se repitió el proceso por 30 veces. Valores expresados como Media \pm Desviación estándar (DE).

Tabla Suplementaria 3A			
Modelo	Nivel taxonómico	Método de selección de características	Media \pm DE
LVQ	Phylum	None	0.52 \pm 0.03
GBM	Phylum	None	0.79 \pm 0.03
SVMI	Phylum	None	0.58 \pm 0.02
SVMr	Phylum	None	0.79 \pm 0.03
SLDA	Phylum	None	0.72 \pm 0.03
NN	Phylum	None	0.59 \pm 0.03
Rborist	Phylum	None	0.81 \pm 0.04
xgbLinear	Phylum	None	0.81 \pm 0.04
stepLDA	Phylum	None	0.55 \pm 0.05
naive_bayes	Phylum	None	0.73 \pm 0.04
LVQ	Clase	None	0.79 \pm 0.05
GBM	Clase	None	0.51 \pm 0.05
SVMI	Clase	None	0.66 \pm 0.03
SVMr	Clase	None	0.54 \pm 0.03
SLDA	Clase	None	0.50 \pm 0.00
NN	Clase	None	0.53 \pm 0.1
Rborist	Clase	None	0.51 \pm 0.05
xgbLinear	Clase	None	0.53 \pm 0.04
stepLDA	Clase	None	0.58 \pm 0.02
naive_bayes	Clase	None	0.68 \pm 0.05
LVQ	Orden	None	0.58 \pm 0.05
GBM	Orden	None	0.75 \pm 0.04
SVMI	Orden	None	0.90 \pm 0.02
SVMr	Orden	None	0.90 \pm 0.02
SLDA	Orden	None	0.50 \pm 0.00
NN	Orden	None	0.85 \pm 0.03
Rborist	Orden	None	0.90 \pm 0.03
xgbLinear	Orden	None	0.82 \pm 0.05
stepLDA	Orden	None	0.54 \pm 0.04
naive_bayes	Orden	None	0.83 \pm 0.04
LVQ	Familia	None	0.55 \pm 0.03

GBM	Familia	None	0.92 ± 0.02
SVMI	Familia	None	0.83 ± 0.04
SVMr	Familia	None	0.83 ± 0.04
SLDA	Familia	None	0.50 ± 0.00
NN	Familia	None	0.68 ± 0.04
Rborist	Familia	None	0.93 ± 0.03
xgbLinear	Familia	None	0.85 ± 0.03
stepLDA	Familia	None	0.85 ± 0.03
naive_bayes	Familia	None	0.80 ± 0.03
LVQ	Género	None	0.78 ± 0.04
GBM	Género	None	0.85 ± 0.02
SVMI	Género	None	0.85 ± 0.04
SVMr	Género	None	0.88 ± 0.03
SLDA	Género	None	0.56 ± 0.04
NN	Género	None	0.88 ± 0.04
Rborist	Género	None	0.86 ± 0.04
xgbLinear	Género	None	0.71 ± 0.03
stepLDA	Género	None	0.60 ± 0.05
naive_bayes	Género	None	0.75 ± 0.05
LVQ	Especie	None	0.85 ± 0.04
GBM	Especie	None	0.94 ± 0.03
SVMI	Especie	None	0.66 ± 0.02
SVMr	Especie	None	0.73 ± 0.05
SLDA	Especie	None	0.87 ± 0.03
NN	Especie	None	0.86 ± 0.04
Rborist	Especie	None	0.85 ± 0.03
xgbLinear	Especie	None	0.92 ± 0.02
stepLDA	Especie	None	0.85 ± 0.03
naive_bayes	Especie	None	0.71 ± 0.04
LVQ	Phylum	SAFS	0.71 ± 0.05
GBM	Phylum	SAFS	0.78 ± 0.03
SVMI	Phylum	SAFS	0.80 ± 0.04
SVMr	Phylum	SAFS	0.81 ± 0.04
SLDA	Phylum	SAFS	0.50 ± 0.00
NN	Phylum	SAFS	0.81 ± 0.05
Rborist	Phylum	SAFS	0.81 ± 0.04
xgbLinear	Phylum	SAFS	0.89 ± 0.03
stepLDA	Phylum	SAFS	0.78 ± 0.03
naive_bayes	Phylum	SAFS	0.76 ± 0.04
LVQ	Clase	SAFS	0.78 ± 0.04
GBM	Clase	SAFS	0.93 ± 0.03
SVMI	Clase	SAFS	0.84 ± 0.02
SVMr	Clase	SAFS	0.88 ± 0.03

SLDA	Clase	SAFS	0.83 ± 0.05
NN	Clase	SAFS	0.67 ± 0.05
Rborist	Clase	SAFS	1.00 ± 0.00
xgbLinear	Clase	SAFS	0.93 ± 0.02
stepLDA	Clase	SAFS	0.78 ± 0.03
naive_bayes	Clase	SAFS	0.90 ± 0.02
LVQ	Orden	SAFS	0.71 ± 0.05
GBM	Orden	SAFS	0.93 ± 0.03
SVMI	Orden	SAFS	0.93 ± 0.03
SVMr	Orden	SAFS	0.82 ± 0.04
SLDA	Orden	SAFS	0.50 ± 0.00
NN	Orden	SAFS	0.80 ± 0.03
Rborist	Orden	SAFS	1.00 ± 0.00
xgbLinear	Orden	SAFS	1.00 ± 0.00
stepLDA	Orden	SAFS	0.79 ± 0.05
naive_bayes	Orden	SAFS	0.73 ± 0.04
LVQ	Familia	SAFS	0.70 ± 0.05
GBM	Familia	SAFS	0.73 ± 0.04
SVMI	Familia	SAFS	0.61 ± 0.04
SVMr	Familia	SAFS	0.75 ± 0.05
SLDA	Familia	SAFS	0.63 ± 0.05
NN	Familia	SAFS	0.77 ± 0.04
Rborist	Familia	SAFS	0.80 ± 0.04
xgbLinear	Familia	SAFS	0.93 ± 0.02
stepLDA	Familia	SAFS	0.61 ± 0.05
naive_bayes	Familia	SAFS	0.93 ± 0.03
LVQ	Género	SAFS	1.00 ± 0.00
GBM	Género	SAFS	1.00 ± 0.00
SVMI	Género	SAFS	1.00 ± 0.00
SVMr	Género	SAFS	1.00 ± 0.00
SLDA	Género	SAFS	0.50 ± 0.00
NN	Género	SAFS	1.00 ± 0.00
Rborist	Género	SAFS	1.00 ± 0.00
xgbLinear	Género	SAFS	1.00 ± 0.00
stepLDA	Género	SAFS	0.94 ± 0.01
naive_bayes	Género	SAFS	1.00 ± 0.00
LVQ	Especie	SAFS	0.93 ± 0.03
GBM	Especie	SAFS	0.92 ± 0.02
SVMI	Especie	SAFS	0.93 ± 0.03
SVMr	Especie	SAFS	0.92 ± 0.01
SLDA	Especie	SAFS	0.85 ± 0.03
NN	Especie	SAFS	1.00 ± 0.00
Rborist	Especie	SAFS	1.00 ± 0.00

xgbLinear	Especie	SAFS	1.00 ± 0.00
stepLDA	Especie	SAFS	1.00 ± 0.00
naive_bayes	Especie	SAFS	0.85 ± 0.03
LVQ	Phylum	SAFS.rf	0.64 ± 0.06
GBM	Phylum	SAFS.rf	0.88 ± 0.03
SVMl	Phylum	SAFS.rf	0.74 ± 0.04
SVMr	Phylum	SAFS.rf	0.81 ± 0.04
SLDA	Phylum	SAFS.rf	0.50 ± 0.00
NN	Phylum	SAFS.rf	0.81 ± 0.03
Rborist	Phylum	SAFS.rf	0.94 ± 0.03
xgbLinear	Phylum	SAFS.rf	0.92 ± 0.02
stepLDA	Phylum	SAFS.rf	0.54 ± 0.05
naive_bayes	Phylum	SAFS.rf	0.77 ± 0.04
LVQ	Clase	SAFS.rf	0.53 ± 0.06
GBM	Clase	SAFS.rf	0.92 ± 0.02
SVMl	Clase	SAFS.rf	0.85 ± 0.03
SVMr	Clase	SAFS.rf	0.90 ± 0.01
SLDA	Clase	SAFS.rf	0.77 ± 0.04
NN	Clase	SAFS.rf	0.82 ± 0.04
Rborist	Clase	SAFS.rf	0.88 ± 0.03
xgbLinear	Clase	SAFS.rf	0.93 ± 0.02
stepLDA	Clase	SAFS.rf	0.73 ± 0.04
naive_bayes	Clase	SAFS.rf	0.69 ± 0.04
LVQ	Orden	SAFS.rf	0.65 ± 0.04
GBM	Orden	SAFS.rf	0.95 ± 0.02
SVMl	Orden	SAFS.rf	0.93 ± 0.03
SVMr	Orden	SAFS.rf	0.87 ± 0.02
SLDA	Orden	SAFS.rf	0.87 ± 0.03
NN	Orden	SAFS.rf	0.88 ± 0.02
Rborist	Orden	SAFS.rf	0.96 ± 0.02
xgbLinear	Orden	SAFS.rf	0.95 ± 0.01
stepLDA	Orden	SAFS.rf	0.75 ± 0.04
naive_bayes	Orden	SAFS.rf	0.83 ± 0.03
LVQ	Familia	SAFS.rf	0.57 ± 0.04
GBM	Familia	SAFS.rf	0.96 ± 0.02
SVMl	Familia	SAFS.rf	0.88 ± 0.03
SVMr	Familia	SAFS.rf	0.88 ± 0.03
SLDA	Familia	SAFS.rf	0.88 ± 0.02
NN	Familia	SAFS.rf	0.87 ± 0.01
Rborist	Familia	SAFS.rf	0.95 ± 0.01
xgbLinear	Familia	SAFS.rf	0.96 ± 0.02
stepLDA	Familia	SAFS.rf	0.74 ± 0.04
naive_bayes	Familia	SAFS.rf	0.89 ± 0.04

LVQ	Género	SAFS.rf	0.95 ± 0.02
GBM	Género	SAFS.rf	1.00 ± 0.00
SVMI	Género	SAFS.rf	0.88 ± 0.03
SVMr	Género	SAFS.rf	0.88 ± 0.04
SLDA	Género	SAFS.rf	0.80 ± 0.04
NN	Género	SAFS.rf	0.88 ± 0.03
Rborist	Género	SAFS.rf	0.86 ± 0.03
xgbLinear	Género	SAFS.rf	0.93 ± 0.02
stepLDA	Género	SAFS.rf	0.74 ± 0.04
naive_bayes	Género	SAFS.rf	0.79 ± 0.04
LVQ	Especie	SAFS.rf	0.93 ± 0.02
GBM	Especie	SAFS.rf	0.92 ± 0.02
SVMI	Especie	SAFS.rf	0.93 ± 0.03
SVMr	Especie	SAFS.rf	0.80 ± 0.03
SLDA	Especie	SAFS.rf	0.74 ± 0.03
NN	Especie	SAFS.rf	0.86 ± 0.04
Rborist	Especie	SAFS.rf	0.93 ± 0.02
xgbLinear	Especie	SAFS.rf	0.93 ± 0.03
stepLDA	Especie	SAFS.rf	0.86 ± 0.03
naive_bayes	Especie	SAFS.rf	0.93 ± 0.03
LVQ	Phylum	RFE	0.80 ± 0.04
GBM	Phylum	RFE	0.95 ± 0.02
SVMI	Phylum	RFE	0.88 ± 0.04
SVMr	Phylum	RFE	0.78 ± 0.04
SLDA	Phylum	RFE	0.50 ± 0.00
NN	Phylum	RFE	0.79 ± 0.03
Rborist	Phylum	RFE	0.95 ± 0.02
xgbLinear	Phylum	RFE	0.95 ± 0.02
stepLDA	Phylum	RFE	0.70 ± 0.05
naive_bayes	Phylum	RFE	0.76 ± 0.05
LVQ	Clase	RFE	0.79 ± 0.03
GBM	Clase	RFE	0.76 ± 0.04
SVMI	Clase	RFE	0.81 ± 0.04
SVMr	Clase	RFE	0.76 ± 0.04
SLDA	Clase	RFE	0.70 ± 0.03
NN	Clase	RFE	0.84 ± 0.04
Rborist	Clase	RFE	1.00 ± 0.00
xgbLinear	Clase	RFE	0.90 ± 0.02
stepLDA	Clase	RFE	0.63 ± 0.05
naive_bayes	Clase	RFE	0.78 ± 0.04
LVQ	Orden	RFE	0.82 ± 0.04
GBM	Orden	RFE	0.95 ± 0.02
SVMI	Orden	RFE	0.88 ± 0.03

SVMr	Orden	RFE	0.88 ± 0.03
SLDA	Orden	RFE	0.50 ± 0.00
NN	Orden	RFE	0.83 ± 0.04
Rborist	Orden	RFE	0.95 ± 0.02
xgbLinear	Orden	RFE	0.95 ± 0.02
stepLDA	Orden	RFE	0.88 ± 0.04
naive_bayes	Orden	RFE	0.89 ± 0.02
LVQ	Familia	RFE	0.67 ± 0.04
GBM	Familia	RFE	1.00 ± 0.00
SVMl	Familia	RFE	1.00 ± 0.00
SVMr	Familia	RFE	0.93 ± 0.03
SLDA	Familia	RFE	0.50 ± 0.00
NN	Familia	RFE	1.00 ± 0.00
Rborist	Familia	RFE	1.00 ± 0.00
xgbLinear	Familia	RFE	1.00 ± 0.00
stepLDA	Familia	RFE	0.87 ± 0.04
naive_bayes	Familia	RFE	0.91 ± 0.03
LVQ	Género	RFE	0.88 ± 0.03
GBM	Género	RFE	1.00 ± 0.00
SVMl	Género	RFE	0.95 ± 0.02
SVMr	Género	RFE	1.00 ± 0.00
SLDA	Género	RFE	1.00 ± 0.00
NN	Género	RFE	1.00 ± 0.00
Rborist	Género	RFE	1.00 ± 0.00
xgbLinear	Género	RFE	0.93 ± 0.03
stepLDA	Género	RFE	0.87 ± 0.05
naive_bayes	Género	RFE	0.81 ± 0.04
LVQ	Especie	RFE	0.86 ± 0.04
GBM	Especie	RFE	0.93 ± 0.03
SVMl	Especie	RFE	0.78 ± 0.05
SVMr	Especie	RFE	0.86 ± 0.03
SLDA	Especie	RFE	0.95 ± 0.02
NN	Especie	RFE	1.00 ± 0.00
Rborist	Especie	RFE	0.94 ± 0.03
xgbLinear	Especie	RFE	1.00 ± 0.00
stepLDA	Especie	RFE	0.79 ± 0.03
naive_bayes	Especie	RFE	0.87 ± 0.03
LVQ	Phylum	SBF	0.73 ± 0.04
GBM	Phylum	SBF	0.91 ± 0.02
SVMl	Phylum	SBF	0.96 ± 0.02
SVMr	Phylum	SBF	0.81 ± 0.03
SLDA	Phylum	SBF	0.50 ± 0.00
NN	Phylum	SBF	0.95 ± 0.01

Rborist	Phylum	SBF	1.00 ± 0.00
xgbLinear	Phylum	SBF	1.00 ± 0.00
stepLDA	Phylum	SBF	0.72 ± 0.04
naive_bayes	Phylum	SBF	0.89 ± 0.03
LVQ	Clase	SBF	0.81 ± 0.04
GBM	Clase	SBF	0.93 ± 0.03
SVMI	Clase	SBF	0.79 ± 0.04
SVMr	Clase	SBF	0.87 ± 0.04
SLDA	Clase	SBF	0.87 ± 0.03
NN	Clase	SBF	0.86 ± 0.03
Rborist	Clase	SBF	1.00 ± 0.00
xgbLinear	Clase	SBF	1.00 ± 0.00
stepLDA	Clase	SBF	0.72 ± 0.04
naive_bayes	Clase	SBF	0.88 ± 0.03
LVQ	Orden	SBF	0.95 ± 0.02
GBM	Orden	SBF	1.00 ± 0.00
SVMI	Orden	SBF	0.89 ± 0.03
SVMr	Orden	SBF	0.95 ± 0.02
SLDA	Orden	SBF	0.50 ± 0.00
NN	Orden	SBF	0.81 ± 0.03
Rborist	Orden	SBF	1.00 ± 0.00
xgbLinear	Orden	SBF	1.00 ± 0.00
stepLDA	Orden	SBF	0.82 ± 0.03
naive_bayes	Orden	SBF	0.85 ± 0.03
LVQ	Familia	SBF	0.83 ± 0.04
GBM	Familia	SBF	0.93 ± 0.02
SVMI	Familia	SBF	0.92 ± 0.01
SVMr	Familia	SBF	0.95 ± 0.02
SLDA	Familia	SBF	0.64 ± 0.02
NN	Familia	SBF	0.95 ± 0.01
Rborist	Familia	SBF	0.93 ± 0.03
xgbLinear	Familia	SBF	1.00 ± 0.00
stepLDA	Familia	SBF	0.79 ± 0.04
naive_bayes	Familia	SBF	0.96 ± 0.02
LVQ	Género	SBF	0.82 ± 0.04
GBM	Género	SBF	1.00 ± 0.00
SVMI	Género	SBF	1.00 ± 0.00
SVMr	Género	SBF	0.95 ± 0.01
SLDA	Género	SBF	0.86 ± 0.02
NN	Género	SBF	1.00 ± 0.00
Rborist	Género	SBF	1.00 ± 0.00
xgbLinear	Género	SBF	1.00 ± 0.00
stepLDA	Género	SBF	0.81 ± 0.03

naive_bayes	Género	SBF	0.95 ± 0.02
LVQ	Especie	SBF	0.87 ± 0.03
GBM	Especie	SBF	1.00 ± 0.00
SVMl	Especie	SBF	0.91 ± 0.03
SVMr	Especie	SBF	0.95 ± 0.01
SLDA	Especie	SBF	0.96 ± 0.02
NN	Especie	SBF	0.96 ± 0.02
Rborist	Especie	SBF	1.00 ± 0.00
xgbLinear	Especie	SBF	1.00 ± 0.00
stepLDA	Especie	SBF	0.81 ± 0.04
naive_bayes	Especie	SBF	0.95 ± 0.02

8.7. Tabla suplementaria 2B

Presenta los valores de significancia estadística de las comparaciones del ABC entre los diferentes modelos a cada nivel taxonómico, luego de la aplicación de cada método de selección de características. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 3B									
Método de selección de características: None									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	4.65E-11								
naive_bayes	1.88E-06	4.65E-11							
NN	4.65E-11	1.32E-09	4.65E-11						
Rborist	3.10E-02	4.65E-11	3.11E-08	4.65E-11					
SLDA	2.40E-07	4.65E-11	7.72E-01	4.65E-11	9.61E-10				
stepLDA	4.65E-11	6.97E-04	4.68E-11	2.23E-03	4.68E-11	4.65E-11			
SVMl	1.00E+00	4.65E-11	3.16E-07	4.65E-11	4.80E-02	2.83E-09	4.65E-11		
SVMr	4.65E-11	2.72E-08	4.65E-11	7.91E-02	4.65E-11	4.65E-11	5.61E-02	4.65E-11	
xgbLinear	2.11E-02	4.65E-11	4.25E-08	4.65E-11	8.70E-01	1.05E-09	4.68E-11	4.24E-02	4.65E-11
Método de selección de características: None									
Nivel taxonómico: Clase									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	6.23E-11								

ive_bayes	6.23E-11	9.01E-09							
NN	9.23E-01	6.23E-11	4.14E-07						
Rborist	7.27E-01	6.23E-11	6.23E-11	8.19E-01					
SLDA	4.00E-01	1.67E-11	1.67E-11	7.03E-01	4.00E-01				
stepLDA	1.09E-08	6.23E-11	6.23E-11	6.91E-02	4.10E-09	2.82E-11			
SVMI	2.37E-02	6.23E-11	6.23E-11	4.35E-01	9.57E-03	2.13E-08	2.22E-05		
SVMr	6.23E-11	6.23E-11	5.36E-01	1.26E-06	6.23E-11	1.67E-11	1.75E-10	1.03E-10	
xgbLinear	2.36E-01	6.23E-11	6.23E-11	7.08E-01	1.08E-01	2.72E-01	3.30E-07	2.56E-01	6.23E-11

Método de selección de características: None

Nivel taxonómico: Orden

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	4.34E-11								
ive_bayes	3.86E-08	4.34E-11							
NN	4.34E-11	4.34E-11	3.67E-04						
Rborist	4.34E-11	4.34E-11	1.22E-08	1.62E-07					
SLDA	6.74E-12	4.34E-11	6.74E-12	6.74E-12	6.74E-12				
stepLDA	4.34E-11	1.34E-03	4.34E-11	4.34E-11	4.34E-11	4.64E-06			
SVMI	4.34E-11	4.34E-11	1.22E-08	8.36E-08	9.56E-01	6.74E-12	4.34E-11		
SVMr	4.34E-11	4.34E-11	6.64E-09	4.07E-08	5.52E-01	6.74E-12	4.34E-11	5.08E-01	
xgbLinear	4.07E-05	4.34E-11	2.12E-02	7.87E-03	4.89E-09	6.74E-12	4.34E-11	4.94E-09	5.65E-09

Método de selección de características: None

Nivel taxonómico: Familia

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	4.20E-11								
ive_bayes	4.20E-11	4.20E-11							
NN	4.20E-11	4.20E-11	4.24E-11						
Rborist	4.94E-01	4.20E-11	4.20E-11	4.20E-11					
SLDA	5.18E-12	5.18E-12	5.18E-12	5.18E-12	5.18E-12				
stepLDA	1.40E-10	4.20E-11	4.34E-07	4.20E-11	6.79E-11	5.18E-12			
SVMI	4.05E-10	4.20E-11	2.67E-03	4.20E-11	3.54E-10	5.18E-12	1.28E-01		
SVMr	2.06E-09	4.24E-11	2.64E-03	4.20E-11	1.60E-09	5.18E-12	2.38E-01	8.85E-01	
xgbLinear	1.17E-09	4.20E-11	2.13E-07	4.20E-11	4.00E-10	5.18E-12	2.72E-01	7.01E-02	1.43E-01

Método de selección de características: None

Nivel taxonómico: Género

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.21E-08								
ive_bayes	7.80E-09	2.16E-05							
NN	3.58E-08	2.37E-10	1.88E-10						

Rborist	7.85E-01	2.75E-07	5.71E-09	3.88E-04					
SLDA	5.36E-11	5.36E-11	5.36E-11	5.36E-11	5.36E-11				
stepLDA	5.36E-11	5.36E-11	7.08E-11	5.36E-11	5.36E-11	2.47E-04			
SVMI	3.21E-08	2.37E-10	1.88E-10	7.85E-01	2.20E-04	5.36E-11	5.36E-11		
SVMr	6.36E-01	1.64E-07	5.71E-09	1.75E-05	6.36E-01	5.36E-11	5.36E-11	1.03E-05	
xgbLinear	5.36E-11	2.10E-09	6.39E-05	5.36E-11	5.36E-11	5.36E-11	1.34E-09	5.36E-11	5.36E-11

Método de selección de características: None

Nivel taxonómico: Especie

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	1.72E-10								
ive_bayes	5.98E-11	5.98E-11							
NN	2.80E-10	1.09E-01	6.60E-11						
Rborist	1.36E-10	8.09E-01	5.98E-11	6.20E-02					
SLDA	3.50E-09	3.10E-01	5.98E-11	1.36E-02	4.22E-01				
stepLDA	1.76E-10	6.39E-01	5.98E-11	2.03E-01	4.54E-01	1.52E-01			
SVMI	5.98E-11	1.80E-10	1.09E-01	7.64E-10	1.36E-10	1.57E-10	2.11E-10		
SVMr	5.98E-11	5.98E-11	1.17E-07	5.98E-11	5.98E-11	5.98E-11	5.98E-11	3.32E-08	
xgbLinear	2.20E-01	2.06E-10	5.98E-11	4.24E-10	1.58E-10	1.02E-08	2.11E-10	5.98E-11	5.98E-11

Método de selección de características: SAFS

Nivel taxonómico: Phylum

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	4.10E-07								
naive_bayes	3.55E-01	8.90E-05							
NN	1.12E-01	2.94E-09	2.91E-02						
Rborist	1.67E-02	2.26E-09	8.35E-03	4.97E-01					
SLDA	5.18E-12	5.18E-12	5.18E-12	5.18E-12	5.18E-12				
stepLDA	8.21E-01	1.93E-07	4.31E-01	2.79E-01	3.00E-02	5.18E-12			
SVMI	2.08E-02	2.94E-09	2.27E-02	7.62E-01	6.16E-01	5.18E-12	2.85E-02		
SVMr	3.39E-01	4.67E-09	1.76E-01	7.62E-01	2.41E-01	5.18E-12	5.85E-01	3.55E-01	
xgbLinear	8.22E-11	8.22E-11	8.22E-11	1.02E-07	5.77E-09	5.18E-12	8.03E-11	8.81E-10	3.37E-09

Método de selección de características: SAFS

Nivel taxonómico: Clase

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	5.01E-11								
ive_bayes	2.44E-05	3.09E-10							
NN	5.01E-11	1.49E-09	5.03E-11						
Rborist	2.66E-10	6.89E-12	6.89E-12	6.96E-12					
SLDA	2.63E-08	1.59E-05	7.58E-07	5.01E-11	6.89E-12				

stepLDA	5.01E-11	4.67E-01	5.01E-11	1.22E-09	6.89E-12	7.23E-08			
SVMI	9.42E-07	3.42E-10	1.73E-05	5.01E-11	6.89E-12	5.25E-05	5.01E-11		
SVMr	7.92E-11	1.33E-07	3.15E-10	5.01E-11	6.89E-12	5.51E-02	1.45E-09	6.07E-09	
xgbLinear	9.17E-01	5.01E-11	3.51E-05	5.01E-11	4.82E-11	3.21E-08	5.01E-11	1.29E-06	6.12E-11
Método de selección de características: SAFS									
Nivel taxonómico: Orden									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.42E-11								
ive_bayes	3.42E-11	6.76E-02							
NN	3.26E-11	2.90E-08	7.11E-07						
Rborist	2.06E-10	2.54E-12	2.54E-12	2.54E-12					
SLDA	2.54E-12	2.54E-12	2.54E-12	2.54E-12	3.71E-13				
stepLDA	4.24E-11	3.24E-06	2.19E-06	3.20E-01	2.54E-12	2.54E-12			
SVMI	3.56E-10	1.03E-09	3.36E-10	7.55E-07	2.54E-12	2.54E-12	1.43E-03		
SVMr	5.00E-01	3.26E-11	3.26E-11	3.06E-11	5.04E-10	2.54E-12	3.26E-11	2.67E-10	
xgbLinear	2.06E-10	2.54E-12	2.54E-12	2.54E-12	-	3.71E-13	2.54E-12	2.54E-12	5.04E-10
Método de selección de características: SAFS									
Nivel taxonómico: Familia									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	4.72E-03								
ive_bayes	5.16E-11	5.16E-11							
NN	2.95E-03	1.66E-05	5.16E-11						
Rborist	4.88E-07	7.14E-09	5.16E-11	6.60E-02					
SLDA	2.95E-09	6.94E-06	5.16E-11	9.12E-11	5.16E-11				
stepLDA	3.55E-10	9.53E-08	5.16E-11	6.88E-11	5.16E-11	2.29E-01			
SVMI	4.03E-01	1.72E-03	5.16E-11	4.05E-02	1.02E-04	4.31E-09	4.21E-10		
SVMr	1.57E-10	1.51E-07	5.16E-11	5.79E-11	5.16E-11	3.93E-01	6.55E-01	2.50E-10	
xgbLinear	5.16E-11	5.16E-11	7.06E-02	5.16E-11	5.16E-11	5.16E-11	5.16E-11	5.16E-11	5.16E-11
Método de selección de características: SAFS									
Nivel taxonómico: Género									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	-								
ive_bayes	-	-							
NN	-	-	-						
Rborist	-	-	-	-					
SLDA	3.58E-14	3.58E-14	3.58E-14	3.58E-14	3.58E-14				
stepLDA	2.74E-12	2.74E-12	2.74E-12	2.74E-12	2.74E-12	1.34E-12			
SVMI	-	-	-	-	-	3.58E-14	2.74E-12		

SVMr	-	-	-	-	-	3.58E-14	2.74E-12	-	
xgbLinear	-	-	-	-	-	3.58E-14	2.74E-12	-	-
Método de selección de características: SAFS									
Nivel taxonómico: Especie									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.11E-01								
ive_bayes	5.56E-10	5.36E-10							
NN	6.43E-12	1.94E-09	3.10E-12						
Rborist	6.43E-12	1.94E-09	3.10E-12	-					
SLDA	4.84E-10	4.09E-10	7.50E-01	3.10E-12	3.10E-12				
stepLDA	6.43E-12	1.94E-09	3.10E-12	-	-	3.10E-12			
SVMI	4.30E-01	8.76E-02	8.42E-10	3.10E-12	3.10E-12	4.84E-10	3.10E-12		
SVMr	8.93E-01	4.30E-01	7.48E-10	2.70E-10	2.70E-10	4.37E-10	2.70E-10	4.24E-01	
xgbLinear	6.43E-12	1.94E-09	3.10E-12	-	-	3.10E-12	-	3.10E-12	2.70E-10
Método de selección de características: SAFS.rf									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.77E-11								
naive_bayes	1.19E-10	1.21E-10							
NN	1.23E-09	3.77E-11	2.43E-04						
Rborist	1.27E-07	3.77E-11	3.77E-11	3.77E-11					
SLDA	5.60E-12	4.93E-11	5.60E-12	5.60E-12	5.60E-12				
stepLDA	3.77E-11	1.94E-08	3.77E-11	3.77E-11	3.77E-11	4.80E-03			
SVMI	4.56E-09	3.77E-11	9.27E-05	5.65E-01	2.74E-10	5.60E-12	3.77E-11		
SVMr	3.77E-11	4.56E-09	4.81E-03	1.29E-07	3.77E-11	5.60E-12	3.77E-11	4.14E-08	
xgbLinear	5.76E-07	3.77E-11	3.77E-11	3.77E-11	3.17E-02	5.60E-12	3.77E-11	3.50E-10	3.77E-11
Método de selección de características: SAFS.rf									
Nivel taxonómico: Clase									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	5.68E-11								
ive_bayes	5.68E-11	7.20E-11							
NN	4.91E-10	5.68E-11	1.38E-10						
Rborist	5.99E-06	5.68E-11	5.68E-11	2.07E-07					
SLDA	5.68E-11	5.68E-11	4.77E-08	6.38E-04	2.92E-10				
stepLDA	5.68E-11	5.68E-11	1.96E-03	4.29E-09	5.73E-11	6.51E-04			
SVMI	4.30E-08	5.68E-11	5.68E-11	4.94E-10	7.52E-06	5.68E-11	5.68E-11		
SVMr	2.90E-09	5.68E-11	5.68E-11	1.37E-03	1.56E-04	1.37E-08	1.66E-10	7.58E-06	
xgbLinear	9.74E-02	5.68E-11	5.68E-11	4.17E-10	4.05E-06	5.68E-11	5.68E-11	8.34E-10	4.02E-10

Método de selección de características: SAFS.rf									
Nivel taxonómico: Orden									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	5.40E-11								
ive_bayes	1.25E-10	5.40E-11							
NN	1.25E-10	5.40E-11	4.64E-05						
Rborist	7.85E-01	5.40E-11	1.17E-10	1.17E-10					
SLDA	6.86E-11	5.40E-11	3.60E-05	1.52E-02	5.40E-11				
stepLDA	5.40E-11	5.40E-11	2.09E-09	5.40E-11	5.40E-11	6.86E-11			
SVMI	5.40E-11	5.40E-11	3.43E-05	8.28E-01	5.40E-11	3.14E-02	5.40E-11		
SVMr	5.94E-07	5.40E-11	3.22E-10	3.71E-08	3.40E-07	3.93E-10	5.40E-11	5.79E-10	
xgbLinear	3.25E-01	5.40E-11	1.25E-10	1.25E-10	1.95E-01	5.40E-11	5.40E-11	5.40E-11	5.94E-07
Método de selección de características: SAFS.rf									
Nivel taxonómico: Familia									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	5.90E-11								
ive_bayes	6.79E-08	5.90E-11							
NN	5.90E-11	5.90E-11	1.18E-01						
Rborist	1.01E-01	5.90E-11	6.89E-07	5.90E-11					
SLDA	1.15E-10	5.90E-11	4.71E-01	3.38E-01	1.48E-10				
stepLDA	5.90E-11	5.90E-11	5.90E-11	5.90E-11	5.90E-11	5.90E-11			
SVMI	2.90E-09	5.90E-11	5.25E-01	4.43E-01	9.66E-09	1.00E+00	6.40E-11		
SVMr	1.35E-10	5.90E-11	1.87E-01	9.45E-01	1.69E-10	4.73E-01	6.31E-11	5.43E-01	
xgbLinear	2.26E-01	5.90E-11	4.36E-07	5.90E-11	7.19E-01	1.46E-10	5.90E-11	7.56E-09	1.66E-10
Método de selección de características: SAFS.rf									
Nivel taxonómico: Género									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	1.04E-10								
ive_bayes	7.14E-12	5.58E-11							
NN	7.14E-12	5.83E-10	1.14E-09						
Rborist	7.14E-12	5.58E-11	6.20E-07	2.40E-03					
SLDA	7.14E-12	5.58E-11	6.76E-02	1.91E-09	3.02E-06				
stepLDA	7.14E-12	5.58E-11	3.26E-04	5.73E-11	1.25E-10	2.07E-06			
SVMI	4.63E-11	1.90E-08	4.89E-10	7.17E-01	4.30E-04	6.33E-10	5.58E-11		
SVMr	7.14E-12	3.85E-09	8.29E-10	8.02E-01	7.74E-04	1.36E-09	5.58E-11	8.60E-01	
xgbLinear	5.58E-11	4.70E-08	4.63E-11	1.90E-08	4.58E-10	4.63E-11	4.63E-11	3.50E-08	4.21E-07
Método de selección de características: SAFS.rf									

Nivel taxonómico: Especie									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	2.06E-01								
ive_bayes	4.11E-01	9.53E-01							
NN	1.46E-08	2.92E-09	1.51E-08						
Rborist	1.28E-01	8.91E-01	8.19E-01	1.99E-09					
SLDA	7.11E-11	7.11E-11	7.27E-11	9.64E-11	7.11E-11				
stepLDA	1.12E-08	2.59E-09	1.08E-08	9.53E-01	1.94E-09	7.69E-11			
SVMl	7.11E-11	7.11E-11	7.11E-11	2.19E-06	7.11E-11	9.78E-09	1.81E-06		
SVMr	7.03E-01	5.69E-01	7.53E-01	1.12E-08	4.11E-01	7.11E-11	9.51E-09	7.11E-11	
xgbLinear	3.99E-01	8.95E-01	9.75E-01	9.51E-09	7.53E-01	7.11E-11	7.27E-09	7.11E-11	7.53E-01
Método de selección de características: RFE									
Nivel taxonómico: Phylum									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	1.08E-06								
naive_bayes	8.66E-03	2.09E-02							
NN	3.28E-01	6.90E-05	9.74E-02						
Rborist	3.15E-02	1.36E-03	3.89E-01	3.28E-01					
SLDA	8.68E-22	8.45E-04	1.67E-12	4.04E-19	1.39E-18				
stepLDA	9.10E-07	3.89E-01	8.45E-02	2.01E-04	3.38E-03	1.25E-09			
SVMl	6.90E-05	2.40E-01	2.30E-01	2.78E-03	2.79E-02	3.17E-08	6.80E-01		
SVMr	1.53E-01	3.76E-04	2.12E-01	6.82E-01	5.10E-01	9.84E-18	7.89E-04	8.90E-03	
xgbLinear	5.10E-01	8.22E-07	3.80E-03	1.53E-01	1.07E-02	1.14E-18	8.22E-07	4.47E-05	5.42E-02
Método de selección de características: RFE									
Nivel taxonómico: Clase									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	2.72E-01								
ive_bayes	1.31E-01	6.78E-01							
NN	6.22E-01	5.35E-01	2.97E-01						
Rborist	9.53E-01	3.38E-01	1.58E-01	6.78E-01					
SLDA	2.40E-21	1.32E-19	1.91E-16	1.86E-21	2.19E-21				
stepLDA	3.48E-04	2.56E-02	1.27E-01	2.24E-03	6.26E-04	1.20E-15			
SVMl	5.35E-01	6.22E-01	3.82E-01	9.06E-01	6.22E-01	2.19E-21	3.16E-03		
SVMr	2.14E-01	8.34E-01	7.63E-01	4.55E-01	2.42E-01	1.16E-19	4.62E-02	5.01E-01	
xgbLinear	1.70E-02	2.42E-01	5.35E-01	5.66E-02	2.58E-02	1.40E-17	3.94E-01	7.71E-02	3.40E-01
Método de selección de características: RFE									
Nivel taxonómico: Orden									
	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr

LVQ	5.65E-05								
ive_bayes	1.88E-01	9.86E-03							
NN	3.55E-02	4.08E-02	4.75E-01						
Rborist	8.12E-01	8.77E-06	1.02E-01	1.26E-02					
SLDA	2.53E-21	2.47E-04	4.17E-13	8.07E-13	4.76E-23				
stepLDA	8.80E-06	8.66E-01	5.39E-03	2.36E-02	7.52E-07	3.99E-07			
SVMl	2.70E-01	6.01E-07	2.36E-02	1.18E-03	4.75E-01	9.01E-25	1.13E-08		
SVMr	2.19E-01	1.82E-03	8.25E-01	3.24E-01	1.25E-01	2.08E-18	8.11E-04	2.24E-02	
xgbLinear	3.55E-02	4.54E-02	4.90E-01	9.24E-01	1.50E-02	7.67E-12	4.24E-02	1.50E-03	3.11E-01

Método de selección de características: RFE

Nivel taxonómico: Familia

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	1.19E-11								
ive_bayes	4.50E-08	4.09E-01							
NN	1.21E-01	1.13E-07	4.79E-05						
Rborist	8.17E-01	2.06E-11	5.50E-08	1.87E-01					
SLDA	1.54E-18	2.41E-01	5.53E-02	2.71E-12	1.54E-18				
stepLDA	1.76E-10	4.09E-01	9.98E-01	3.12E-06	2.99E-10	9.25E-03			
SVMl	7.67E-01	9.48E-11	1.96E-07	2.34E-01	9.31E-01	3.34E-17	1.93E-09		
SVMr	4.09E-01	1.14E-08	8.50E-06	5.63E-01	5.03E-01	9.70E-14	2.53E-07	5.65E-01	
xgbLinear	1.47E-02	4.11E-06	8.20E-04	4.53E-01	2.83E-02	8.10E-11	5.84E-05	4.10E-02	1.87E-01

Método de selección de características: RFE

Nivel taxonómico: Género

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	2.83E-06								
ive_bayes	4.26E-01	5.90E-05							
NN	1.73E-02	1.76E-02	9.97E-02						
Rborist	7.78E-02	6.78E-10	1.11E-02	5.44E-05					
SLDA	3.69E-04	3.05E-01	3.77E-03	2.18E-01	3.50E-07				
stepLDA	1.16E-11	1.61E-02	2.57E-10	2.07E-06	1.69E-15	6.34E-04			
SVMl	2.40E-01	1.02E-08	4.53E-02	4.15E-04	5.19E-01	2.83E-06	1.53E-14		
SVMr	6.54E-01	2.29E-06	2.62E-01	1.08E-02	2.27E-01	2.61E-04	1.78E-11	4.99E-01	
xgbLinear	2.98E-01	4.45E-04	7.16E-01	2.28E-01	7.64E-03	1.56E-02	1.02E-08	3.18E-02	2.00E-01

Método de selección de características: RFE

Nivel taxonómico: Especie

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMl	SVMr
LVQ	6.33E-01								
ive_bayes	9.56E-01	6.33E-01							

NN	6.33E-01	8.60E-01	6.32E-01						
Rborist	7.48E-01	3.65E-01	6.33E-01	3.65E-01					
SLDA	6.33E-01	9.89E-01	6.33E-01	8.60E-01	3.65E-01				
stepLDA	3.65E-01	5.97E-01	3.56E-01	6.33E-01	1.05E-01	5.97E-01			
SVMI	4.15E-01	6.33E-01	3.65E-01	7.89E-01	1.46E-01	6.33E-01	8.17E-01		
SVMr	8.00E-01	8.17E-01	8.00E-01	7.50E-01	5.97E-01	8.17E-01	4.15E-01	5.97E-01	
xgbLinear	6.33E-01	8.60E-01	5.97E-01	9.89E-01	3.65E-01	8.60E-01	6.33E-01	7.65E-01	7.48E-01

Método de selección de características: SBF

Nivel taxonómico: Phylum

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	2.69E-11								
naive_bayes	1.00E-03	2.69E-11							
NN	7.45E-09	2.45E-11	4.49E-08						
Rborist	2.27E-12	2.27E-12	2.27E-12	1.74E-11					
SLDA	2.27E-12	2.27E-12	2.27E-12	2.27E-12	3.71E-13				
stepLDA	2.69E-11	3.42E-01	2.69E-11	2.40E-11	2.27E-12	2.27E-12			
SVMI	2.69E-11	3.30E-09	1.19E-10	2.38E-11	2.27E-12	2.27E-12	2.57E-11		
SVMr	7.20E-08	2.62E-11	1.17E-07	4.10E-01	5.29E-10	2.27E-12	2.57E-11	2.57E-11	
xgbLinear	2.27E-12	2.27E-12	2.27E-12	1.74E-11	-	3.71E-13	2.27E-12	2.27E-12	5.29E-10

Método de selección de características: SBF

Nivel taxonómico: Clase

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.64E-11								
ive_bayes	8.98E-06	1.22E-08							
NN	9.09E-10	1.96E-04	6.17E-06						
Rborist	7.12E-11	4.14E-12	4.14E-12	7.56E-12					
SLDA	1.06E-09	7.94E-06	6.12E-02	1.82E-01	4.14E-12				
stepLDA	3.64E-11	7.25E-11	3.64E-11	3.08E-11	4.14E-12	3.64E-11			
SVMI	3.28E-08	3.67E-05	8.23E-03	2.99E-01	1.05E-11	9.76E-01	3.64E-11		
SVMr	3.74E-11	1.95E-02	3.42E-10	3.83E-09	4.14E-12	1.07E-09	9.02E-08	3.83E-09	
xgbLinear	7.12E-11	4.14E-12	4.14E-12	7.56E-12	-	4.14E-12	4.14E-12	1.05E-11	4.14E-12

Método de selección de características: SBF

Nivel taxonómico: Orden

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	5.13E-11								
ive_bayes	1.98E-12	2.21E-11							
NN	1.98E-12	2.21E-11	1.01E-04						
Rborist	-	5.13E-11	1.98E-12	1.98E-12					

SLDA	2.36E-13	1.98E-12	1.98E-12	1.98E-12	2.36E-13				
stepLDA	1.98E-12	2.94E-11	1.61E-03	8.05E-02	1.98E-12	1.98E-12			
SVMI	3.54E-11	2.79E-01	1.65E-11	1.65E-11	3.54E-11	1.98E-12	2.21E-11		
SVMr	1.98E-12	2.63E-08	9.71E-07	1.11E-09	1.98E-12	1.98E-12	1.11E-09	5.62E-08	
xgbLinear	-	5.13E-11	1.98E-12	1.98E-12	-	2.36E-13	1.98E-12	3.54E-11	1.98E-12

Método de selección de características: SBF

Nivel taxonómico: Familia

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	3.09E-09								
ive_bayes	3.80E-09	1.31E-10							
NN	4.51E-09	1.34E-10	7.72E-01						
Rborist	6.11E-01	3.29E-09	5.05E-07	6.69E-07					
SLDA	3.56E-11	3.56E-11	3.56E-11	3.56E-11	3.56E-11				
stepLDA	3.56E-11	5.06E-05	3.56E-11	3.56E-11	3.70E-11	3.56E-11			
SVMI	3.60E-09	1.31E-10	4.71E-01	6.29E-01	5.09E-07	3.56E-11	3.56E-11		
SVMr	4.71E-01	6.95E-09	3.56E-11	3.56E-11	2.36E-01	3.56E-11	3.70E-11	3.56E-11	
xgbLinear	3.56E-11	9.52E-12	2.11E-10	3.56E-11	2.11E-10	9.52E-12	9.52E-12	5.73E-11	9.52E-12

Método de selección de características: SBF

Nivel taxonómico: Género

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	2.47E-12								
ive_bayes	4.70E-11	2.75E-11							
NN	-	2.47E-12	4.70E-11						
Rborist	-	2.47E-12	4.70E-11						
SLDA	2.47E-12	4.86E-04	2.69E-11	2.47E-12	2.47E-12				
stepLDA	2.47E-12	5.76E-01	2.69E-11	2.47E-12	2.47E-12	1.33E-05			
SVMI	3.89E-12	2.32E-11	7.47E-01	3.89E-12	3.89E-12	2.28E-11	2.28E-11		
SVMr	-	2.47E-12	4.70E-11	-	-	2.47E-12	2.47E-12	3.89E-12	
xgbLinear	-	2.47E-12	4.70E-11	-	-	2.47E-12	2.47E-12	3.89E-12	

Método de selección de características: SBF

Nivel taxonómico: Especie

	GBM	LVQ	naive_bayes	NN	Rborist	SLDA	stepLDA	SVMI	SVMr
LVQ	6.99E-12								
ive_bayes	6.61E-11	6.86E-11							
NN	5.69E-10	6.61E-11	4.81E-01						
Rborist	-	6.99E-12	6.61E-11	5.69E-10					
SLDA	1.45E-10	5.64E-11	5.55E-01	8.64E-01	1.45E-10				
stepLDA	6.99E-12	5.08E-05	4.43E-11	4.43E-11	6.99E-12	4.25E-11			

SVMI	3.10E-11	6.64E-11	8.64E-01	3.74E-01	3.10E-11	4.37E-01	4.43E-11		
SVMr	1.56E-11	4.91E-06	1.97E-08	1.05E-08	1.56E-11	8.61E-09	4.90E-11	2.03E-08	
xgbLinear	-	6.99E-12	6.61E-11	5.69E-10	-	1.45E-10	6.99E-12	3.10E-11	1.56E-11

8.8. Tabla suplementaria 3C

Presenta los valores de significancia estadística de las comparaciones entre el ABC de los diferentes niveles taxonómicos en cada modelo de predicción, luego de la aplicación de cada método de selección de características. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 3C					
Método de selección de características: None					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	4.01E-11				
Género	9.46E-01	4.01E-11			
Orden	4.01E-11	1.13E-01	4.01E-11		
Phylum	4.01E-11	1.73E-06	4.01E-11	1.38E-06	
Especie	4.34E-08	4.01E-11	2.07E-08	4.01E-11	4.01E-11
Método de selección de características: None					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	3.88E-11				
Género	3.88E-11	5.99E-11			
Orden	3.94E-11	3.88E-11	5.87E-11		
Phylum	3.88E-11	3.88E-11	1.85E-09	1.49E-03	
Especie	3.88E-11	8.45E-02	4.79E-11	3.88E-11	3.88E-11
Método de selección de características: None					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	3.27E-11				
Género	3.27E-11	1.39E-01			
Orden	3.27E-11	8.17E-09	4.34E-06		
Phylum	2.33E-10	3.27E-11	3.27E-11	3.27E-11	

Especie	6.26E-01	3.27E-11	3.27E-11	3.27E-11	4.07E-11
Método de selección de características: None					
Modelo de predicción: SVM1					
	Clase	Familia	Género	Orden	Phylum
Familia	4.31E-11				
Género	4.31E-11	6.43E-05			
Orden	4.31E-11	4.93E-08	6.60E-05		
Phylum	4.31E-11	4.26E-08	1.47E-10	4.31E-11	
Especie	4.31E-11	7.50E-09	4.31E-11	4.31E-11	8.00E-06
Método de selección de características: None					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	-				
Género	4.26E-09	4.26E-09			
Orden	-	-	4.26E-09		
Phylum	1.66E-12	1.66E-12	2.12E-11	1.66E-12	
Especie	1.66E-12	1.66E-12	2.12E-11	1.66E-12	2.12E-11
Método de selección de características: None					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	1.94E-08				
Género	4.15E-11	4.15E-11			
Orden	4.15E-11	4.15E-11	6.38E-03		
Phylum	4.69E-02	3.81E-09	4.15E-11	4.15E-11	
Especie	4.15E-11	4.15E-11	5.67E-07	6.72E-01	4.15E-11
Método de selección de características: None					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	5.26E-11				
Género	5.26E-11	4.78E-09			
Orden	5.26E-11	5.45E-04	5.36E-05		
Phylum	5.26E-11	5.26E-11	1.89E-04	5.26E-11	
Especie	5.26E-11	1.08E-10	6.77E-01	9.33E-09	1.19E-04
Método de selección de características: None					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum

Familia	3.66E-11				
Género	3.66E-11	3.66E-11			
Orden	3.66E-11	1.05E-03	3.66E-11		
Phylum	3.66E-11	9.53E-04	3.66E-11	8.44E-01	
Especie	3.66E-11	5.44E-10	3.66E-11	8.50E-10	3.66E-11
Método de selección de características: None					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	3.43E-11				
Género	2.76E-03	3.43E-11			
Orden	2.38E-05	3.43E-11	2.44E-05		
Phylum	6.37E-02	3.43E-11	2.76E-03	1.02E-01	
Especie	3.43E-11	2.21E-01	3.43E-11	3.43E-11	3.43E-11
Método de selección de características: None					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	9.65E-10				
Género	2.36E-06	4.42E-06			
Orden	7.55E-10	1.28E-02	3.71E-08		
Phylum	4.51E-04	1.69E-08	1.70E-01	7.01E-10	
Especie	6.04E-03	7.01E-10	2.72E-03	4.67E-10	2.19E-01
Método de selección de características: SAFS					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	3.21E-07				
Género	4.17E-12	4.17E-12			
Orden	4.68E-06	1.12E-01	4.17E-12		
Phylum	1.26E-06	2.74E-01	4.17E-12	8.52E-01	
Especie	4.62E-11	4.62E-11	2.65E-09	4.62E-11	4.62E-11
Método de selección de características: SAFS					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	2.95E-11				
Género	1.85E-10	6.92E-12			
Orden	1.00E+00	2.95E-11	1.85E-10		
Phylum	2.95E-11	9.60E-06	6.92E-12	2.95E-11	
Especie	6.80E-01	2.95E-11	1.32E-11	6.80E-01	2.95E-11

Método de selección de características: SAFS					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	4.13E-11				
Género	5.04E-12	5.04E-12			
Orden	4.96E-11	4.13E-11	4.49E-10		
Phylum	1.81E-05	4.13E-11	5.04E-12	4.96E-11	
Especie	6.47E-11	4.13E-11	1.73E-10	3.22E-01	6.47E-11
Método de selección de características: SAFS					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	4.92E-11				
Género	3.01E-12	3.01E-12			
Orden	2.41E-06	1.38E-08	3.01E-12		
Phylum	1.26E-09	1.19E-05	3.01E-12	8.35E-04	
Especie	3.67E-06	3.29E-11	3.01E-12	5.54E-10	3.29E-11
Método de selección de características: SAFS					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	2.14E-11				
Género	1.47E-12	1.47E-12			
Orden	1.47E-12	1.47E-12	-		
Phylum	1.47E-12	1.47E-12	-	-	
Especie	2.06E-04	2.55E-11	1.47E-12	1.47E-12	1.47E-12
Método de selección de características: SAFS					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	5.43E-09				
Género	1.90E-12	1.90E-12			
Orden	4.21E-11	2.09E-01	1.90E-12		
Phylum	4.82E-11	6.36E-02	1.90E-12	9.09E-01	
Especie	1.90E-12	1.90E-12	-	1.90E-12	1.90E-12
Método de selección de características: SAFS					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	1.10E-12				

Género	-	1.10E-12			
Orden	-	1.10E-12	-		
Phylum	1.10E-12	2.68E-01	1.10E-12	1.10E-12	
Especie	-	1.10E-12	-	-	1.10E-12
Método de selección de características: SAFS					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	1.49E-01				
Género	1.71E-11	4.16E-11			
Orden	1.71E-11	4.16E-11			
Phylum	1.58E-03	2.64E-04	4.15E-12	4.15E-12	
Especie	1.71E-11	4.16E-11			4.15E-12
Método de selección de características: SAFS					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	3.30E-11				
Género	2.84E-11	2.95E-11			
Orden	4.42E-01	3.30E-11	2.95E-11		
Phylum	5.22E-01	3.28E-11	2.84E-11	6.95E-01	
Especie	3.88E-12	3.88E-12	8.21E-12	3.88E-12	3.88E-12
Método de selección de características: SAFS					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	3.79E-05				
Género	4.00E-12	2.18E-10			
Orden	4.15E-11	4.15E-11	4.00E-12		
Phylum	4.15E-11	4.15E-11	4.00E-12	4.83E-03	
Especie	1.15E-07	1.16E-09	4.00E-12	1.20E-10	5.01E-10
Método de selección de características: SAFS.rf					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	2.37E-02				
Género	4.17E-11	4.17E-11			
Orden	9.79E-10	2.27E-07	4.17E-11		
Phylum	1.35E-08	1.68E-06	4.17E-11	7.83E-01	
Especie	4.17E-11	4.17E-11	5.36E-09	4.17E-11	4.17E-11

Método de selección de características: SAFS.rf					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	4.87E-10				
Género	1.37E-11	6.37E-09			
Orden	5.09E-10	4.73E-01	1.65E-10		
Phylum	9.21E-07	9.73E-10	9.56E-12	1.46E-09	
Especie	9.16E-01	4.87E-10	1.37E-11	4.87E-10	8.80E-07
Método de selección de características: SAFS.rf					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	3.11E-04				
Género	9.69E-05	8.34E-01			
Orden	6.14E-10	1.51E-07	8.79E-07		
Phylum	2.06E-10	8.32E-11	8.32E-11	8.32E-11	
Especie	1.85E-09	3.53E-07	1.50E-06	2.71E-01	8.32E-11
Método de selección de características: SAFS.rf					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	3.37E-04				
Género	8.00E-08	6.99E-01			
Orden	8.00E-08	6.99E-01	7.76E-01		
Phylum	1.51E-09	3.00E-08	1.11E-08	2.30E-08	
Especie	2.05E-10	3.22E-09	1.51E-09	2.46E-09	3.05E-01
Método de selección de características: SAFS.rf					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	1.42E-10				
Género	3.35E-03	3.31E-10			
Orden	1.56E-09	4.85E-03	3.84E-07		
Phylum	2.96E-12	2.96E-12	2.96E-12	2.96E-12	
Especie	1.35E-02	3.94E-11	1.33E-07	4.49E-11	2.96E-12
Método de selección de características: SAFS.rf					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	2.33E-07				
Género	5.44E-07	9.74E-01			

Orden	5.44E-07	9.00E-01	9.00E-01		
Phylum	8.31E-01	2.80E-08	3.93E-08	3.93E-08	
Especie	6.31E-04	1.31E-02	1.67E-02	1.43E-02	3.59E-05
Método de selección de características: SAFS.rf					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	8.54E-07				
Género	1.54E-03	1.17E-10			
Orden	2.56E-07	1.78E-01	1.17E-10		
Phylum	1.84E-06	5.63E-06	1.62E-09	3.03E-06	
Especie	3.03E-06	1.12E-07	2.48E-09	7.56E-08	2.30E-01
Método de selección de características: SAFS.rf					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	1.21E-08				
Género	1.97E-01	7.92E-08			
Orden	1.21E-08	7.80E-01	7.92E-08		
Phylum	3.43E-01	1.21E-08	3.54E-02	1.21E-08	
Especie	7.45E-01	8.54E-07	1.86E-01	9.18E-07	6.47E-01
Método de selección de características: SAFS.rf					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	7.77E-01				
Género	3.76E-01	5.13E-01			
Orden	3.60E-01	5.13E-01	9.21E-01		
Phylum	7.43E-11	7.43E-11	7.43E-11	7.43E-11	
Especie	1.11E-10	1.11E-10	1.11E-10	1.11E-10	7.43E-11
Método de selección de características: SAFS.rf					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	6.54E-11				
Género	1.72E-09	4.93E-10			
Orden	6.54E-11	7.70E-06	9.36E-05		
Phylum	3.57E-08	2.23E-10	6.69E-04	3.78E-08	
Especie	6.54E-11	1.12E-04	6.54E-11	4.78E-10	6.54E-11
Método de selección de características: RFE					

Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	5.17E-10				
Género	2.07E-10	9.11E-11			
Orden	8.91E-08	9.11E-11	4.97E-06		
Phylum	1.19E-01	6.69E-10	9.11E-11	2.56E-03	
Especie	1.21E-08	9.11E-11	2.47E-03	6.44E-04	2.22E-08
Método de selección de características: RFE					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	7.19E-12				
Género	7.19E-12				
Orden	4.87E-11	1.89E-10	1.89E-10		
Phylum	4.87E-11	4.38E-10	4.38E-10	9.35E-02	
Especie	4.87E-11	6.92E-11	6.92E-11	3.84E-08	3.29E-08
Método de selección de características: RFE					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	5.45E-12				
Género	4.75E-11	2.72E-10			
Orden	3.45E-09	5.45E-12	1.98E-07		
Phylum	8.25E-08	1.15E-11	8.00E-09	1.57E-01	
Especie	3.66E-03	5.45E-12	4.75E-11	3.24E-10	3.32E-10
Método de selección de características: RFE					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	5.42E-11				
Género	4.08E-12	9.29E-11			
Orden	1.19E-10	1.13E-05	4.08E-12		
Phylum	4.22E-02	5.42E-11	4.08E-12	1.45E-09	
Especie	7.84E-10	3.83E-09	4.08E-12	4.03E-04	3.54E-08
Método de selección de características: RFE					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	1.15E-12				
Género	1.15E-12	6.74E-14			
Orden	1.15E-12	-	6.74E-14		

Phylum	1.15E-12	-	6.74E-14	-	
Especie	1.43E-11	1.04E-12	9.38E-11	1.04E-12	1.04E-12
Método de selección de características: RFE					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	1.37E-12				
Género	1.37E-12	-			
Orden	3.72E-01	1.37E-12	1.37E-12		
Phylum	8.00E-05	1.37E-12	1.37E-12	1.66E-03	
Especie	1.37E-12	-	-	1.37E-12	1.37E-12
Método de selección de características: RFE					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	-				
Género	-	-			
Orden	2.61E-10	2.61E-10	2.61E-10		
Phylum	1.66E-10	1.66E-10	1.66E-10	9.31E-01	
Especie	1.82E-09	1.82E-09	1.82E-09	2.74E-05	3.16E-05
Método de selección de características: RFE					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	5.51E-12				
Género	2.43E-06	7.51E-10			
Orden	5.87E-09	1.52E-10	4.46E-06		
Phylum	6.10E-09	3.38E-10	3.96E-06	9.57E-01	
Especie	5.51E-12	-	7.51E-10	1.52E-10	3.38E-10
Método de selección de características: RFE					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	5.39E-11				
Género	5.39E-11	8.65E-01			
Orden	5.39E-11	4.52E-03	1.42E-02		
Phylum	4.06E-04	5.39E-11	5.39E-11	5.39E-11	
Especie	7.18E-11	1.30E-09	3.88E-09	5.39E-11	1.43E-07
Método de selección de características: RFE					
Modelo de predicción: naive_bayes					

	Clase	Familia	Género	Orden	Phylum
Familia	6.32E-11				
Género	4.92E-02	6.32E-11			
Orden	6.32E-11	8.96E-02	6.32E-11		
Phylum	1.05E-01	3.63E-10	6.50E-03	4.27E-10	
Especie	6.32E-11	2.32E-07	2.46E-09	4.75E-07	1.31E-09
Método de selección de características: SBF					
Modelo de predicción: LVQ					
	Clase	Familia	Género	Orden	Phylum
Familia	9.08E-04				
Género	6.76E-01	1.48E-01			
Orden	9.95E-11	1.72E-10	9.95E-11		
Phylum	5.44E-09	4.68E-10	5.44E-09	9.95E-11	
Especie	3.68E-05	3.10E-06	8.91E-04	1.22E-10	1.03E-10
Método de selección de características: SBF					
Modelo de predicción: GBM					
	Clase	Familia	Género	Orden	Phylum
Familia	8.14E-01				
Género	4.97E-11	1.54E-11			
Orden	4.97E-11	1.54E-11	-		
Phylum	1.01E-02	5.98E-03	3.92E-12	3.92E-12	
Especie	4.97E-11	1.54E-11	-	-	3.92E-12
Método de selección de características: SBF					
Modelo de predicción: SVMr					
	Clase	Familia	Género	Orden	Phylum
Familia	4.00E-11				
Género	4.55E-12	4.55E-12			
Orden	2.73E-10	2.40E-04	4.55E-12		
Phylum	3.95E-11	3.95E-11	6.49E-10	5.25E-08	
Especie	1.51E-10	6.07E-04	1.25E-11	3.71E-04	4.46E-08
Método de selección de características: SBF					
Modelo de predicción: SVMl					
	Clase	Familia	Género	Orden	Phylum
Familia	4.80E-10				
Género	4.80E-10	5.43E-01			
Orden	4.80E-10	9.47E-01	5.43E-01		
Phylum	2.37E-07	4.71E-11	4.71E-11	4.71E-11	

Especie	4.80E-10	8.47E-01	7.57E-01	8.47E-01	4.71E-11
Método de selección de características: SBF					
Modelo de predicción: SLDA					
	Clase	Familia	Género	Orden	Phylum
Familia	1.87E-11				
Género	6.58E-01	1.87E-11			
Orden	1.71E-12	1.71E-12	1.71E-12		
Phylum	1.71E-12	1.71E-12	1.71E-12	-	
Especie	1.87E-11	1.63E-11	1.85E-11	1.71E-12	1.71E-12
Método de selección de características: SBF					
Modelo de predicción: NN					
	Clase	Familia	Género	Orden	Phylum
Familia	3.27E-10				
Género	1.55E-11	2.86E-11			
Orden	2.28E-05	2.86E-11	9.54E-12		
Phylum	2.86E-10	8.17E-01	2.86E-11	2.86E-11	
Especie	2.86E-10	5.58E-01	5.54E-10	2.86E-11	6.78E-01
Método de selección de características: SBF					
Modelo de predicción: Rborist					
	Clase	Familia	Género	Orden	Phylum
Familia	1.34E-10				
Género	-	1.34E-10			
Orden	-	1.34E-10	-		
Phylum	-	1.34E-10	-	-	
Especie	-	1.34E-10	-	-	-
Método de selección de características: SBF					
Modelo de predicción: xgbLinear					
	Clase	Familia	Género	Orden	Phylum
Familia	-				
Género	-	-			
Orden	-	-	-		
Phylum	-	-	-	-	
Especie	-	-	-	-	-
Método de selección de características: SBF					
Modelo de predicción: stepLDA					
	Clase	Familia	Género	Orden	Phylum

Familia	5.88E-08				
Género	1.38E-10	2.90E-02			
Orden	1.38E-10	7.28E-03	5.79E-01		
Phylum	5.79E-01	3.73E-09	9.59E-11	9.59E-11	
Especie	1.44E-10	7.49E-02	8.25E-01	5.15E-01	9.59E-11
Método de selección de características: SBF					
Modelo de predicción: naive_bayes					
	Clase	Familia	Género	Orden	Phylum
Familia	4.28E-08				
Género	4.28E-08	7.65E-01			
Orden	1.87E-05	7.03E-11	7.03E-11		
Phylum	2.67E-01	9.63E-08	1.17E-07	1.14E-06	
Especie	4.28E-08	7.98E-01	8.94E-01	7.03E-11	1.17E-07

8.9. Tabla suplementaria 3D

Presenta los valores de significancia estadística de las comparaciones entre el ABC de los diferentes métodos de selección de características, en cada modelo de predicción, discriminado por nivel taxonómico. Se utilizó el test de rangos con signo de Wilcoxon ya que para todos los modelos el test de Kruskal Wallis arrojaba un resultado significativo ($p < 0.05$). Los valores de p se ajustaron utilizando el método de corrección de múltiples pruebas de Bonferroni.

Tabla Suplementaria 3D				
Nivel taxonómico: Phylum				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	5.89E-11			
SAFS	5.89E-11	2.87E-08		
SAFS.rf	7.13E-10	7.05E-11	4.05E-06	
SBF	5.89E-11	9.31E-08	9.05E-02	3.44E-08
Nivel taxonómico: Phylum				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	3.39E-11			
SAFS	5.78E-01	3.39E-11		
SAFS.rf	3.39E-11	1.58E-08	3.39E-11	

SBF	3.39E-11	9.98E-06	3.39E-11	1.69E-06
Nivel taxonómico: Phylum				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	3.74E-01			
SAFS	9.16E-02	9.16E-02		
SAFS.rf	4.26E-04	9.16E-02	1.48E-01	
SBF	4.26E-04	1.48E-01	2.26E-01	6.12E-01
Nivel taxonómico: Phylum				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	2.66E-11			
SAFS	2.66E-11	1.12E-08		
SAFS.rf	2.66E-11	3.21E-11	2.86E-05	
SBF	2.66E-11	9.34E-10	2.66E-11	2.66E-11
Nivel taxonómico: Phylum				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	8.31E-13			
SAFS	8.31E-13	-		
SAFS.rf	8.31E-13	-	-	
SBF	8.31E-13	-	-	-
Nivel taxonómico: Phylum				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	3.38E-11			
SAFS	3.38E-11	6.28E-01		
SAFS.rf	3.38E-11	4.45E-02	6.31E-02	
SBF	3.38E-11	3.38E-11	9.01E-11	3.38E-11
Nivel taxonómico: Phylum				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	3.40E-11			
SAFS	7.05E-01	3.40E-11		
SAFS.rf	3.40E-11	4.33E-06	3.40E-11	
SBF	5.26E-12	5.95E-11	5.26E-12	4.87E-10

Nivel taxonómico: Phylum				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	4.01E-11			
SAFS	2.46E-09	4.06E-09		
SAFS.rf	4.01E-11	4.06E-09	1.75E-03	
SBF	5.18E-12	2.41E-10	5.18E-12	3.14E-11
Nivel taxonómico: Phylum				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	5.14E-11			
SAFS	5.14E-11	3.63E-07		
SAFS.rf	6.06E-02	5.14E-11	5.14E-11	
SBF	5.96E-11	5.55E-02	8.11E-09	5.14E-11
Nivel taxonómico: Phylum				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	1.41E-02			
SAFS	7.36E-03	8.86E-01		
SAFS.rf	2.78E-03	1.98E-01	2.31E-01	
SBF	1.11E-10	5.31E-10	1.11E-10	2.25E-10
Nivel taxonómico: Clase				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	7.92E-01			
SAFS	7.92E-01	8.16E-01		
SAFS.rf	6.05E-11	6.05E-11	6.05E-11	
SBF	5.16E-02	1.77E-04	2.06E-02	6.05E-11
Nivel taxonómico: Clase				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	3.33E-11			
SAFS	3.20E-11	3.20E-11		
SAFS.rf	3.20E-11	3.20E-11	4.82E-01	
SBF	3.20E-11	3.20E-11	8.44E-01	3.72E-01

Nivel taxonómico: Clase				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	4.82E-11			
SAFS	4.82E-11	8.56E-11		
SAFS.rf	4.82E-11	4.82E-11	4.44E-05	
SBF	4.82E-11	5.63E-10	5.24E-03	8.84E-04
Nivel taxonómico: Clase				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	3.94E-11			
SAFS	3.94E-11	1.72E-01		
SAFS.rf	3.94E-11	2.54E-02	1.83E-01	
SBF	3.94E-11	2.54E-02	1.61E-07	3.66E-08
Nivel taxonómico: Clase				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	2.41E-12			
SAFS	2.41E-12	3.21E-11		
SAFS.rf	2.41E-12	5.81E-08	2.06E-05	
SBF	2.41E-12	3.46E-11	5.97E-05	8.20E-10
Nivel taxonómico: Clase				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	4.17E-11			
SAFS	1.17E-07	4.17E-11		
SAFS.rf	4.17E-11	5.55E-03	4.17E-11	
SBF	4.17E-11	2.86E-01	4.17E-11	8.47E-04
Nivel taxonómico: Clase				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	1.24E-12			
SAFS	1.24E-12	-		
SAFS.rf	2.13E-11	1.24E-12	1.24E-12	
SBF	1.24E-12	-	-	1.24E-12
Nivel taxonómico: Clase				

Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	2.88E-11			
SAFS	2.76E-11	4.87E-06		
SAFS.rf	2.76E-11	4.64E-06	3.24E-01	
SBF	5.53E-12	5.53E-12	2.19E-11	2.19E-11
Nivel taxonómico: Clase				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	4.60E-05			
SAFS	5.28E-11	1.27E-10		
SAFS.rf	5.28E-11	5.40E-09	2.50E-05	
SBF	5.28E-11	3.50E-08	1.16E-06	4.22E-01
Nivel taxonómico: Clase				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	3.17E-09			
SAFS	3.80E-11	3.80E-11		
SAFS.rf	9.78E-02	1.37E-09	3.80E-11	
SBF	3.80E-11	3.80E-11	2.91E-05	3.80E-11
Nivel taxonómico: Orden				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	3.59E-11			
SAFS	1.07E-10	5.10E-10		
SAFS.rf	1.76E-06	3.59E-11	3.91E-07	
SBF	3.59E-11	3.59E-11	3.59E-11	3.59E-11
Nivel taxonómico: Orden				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	5.31E-11			
SAFS	5.31E-11	4.79E-07		
SAFS.rf	5.31E-11	1.28E-04	5.60E-07	
SBF	1.04E-11	1.94E-10	1.94E-10	8.79E-11
Nivel taxonómico: Orden				
Modelo de predicción: SVMl				

	None	RFE	SAFS	SAFS.rf
RFE	9.91E-05			
SAFS	1.50E-08	2.75E-06		
SAFS.rf	2.36E-08	2.13E-02	4.70E-07	
SBF	1.79E-10	2.06E-09	8.24E-11	8.24E-11
Nivel taxonómico: Orden				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	6.19E-04			
SAFS	6.85E-05	6.85E-05		
SAFS.rf	6.85E-05	6.95E-05	7.92E-01	
SBF	6.19E-04	2.55E-03	6.85E-05	6.85E-05
Nivel taxonómico: Orden				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	-			
SAFS	-	-		
SAFS.rf	9.87E-13	9.87E-13	9.87E-13	
SBF	-	-	-	9.87E-13
Nivel taxonómico: Orden				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	9.61E-04			
SAFS	4.01E-07	7.69E-07		
SAFS.rf	1.03E-02	1.22E-05	1.28E-09	
SBF	2.37E-06	1.60E-04	6.74E-02	2.98E-08
Nivel taxonómico: Orden				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	7.01E-09			
SAFS	4.03E-12	2.35E-10		
SAFS.rf	1.28E-10	2.98E-04	1.65E-09	
SBF	4.03E-12	2.35E-10	-	1.65E-09
Nivel taxonómico: Orden				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf

RFE	2.07E-10			
SAFS	4.12E-12	6.50E-11		
SAFS.rf	1.03E-10	1.04E-02	1.93E-11	
SBF	4.12E-12	6.50E-11	-	1.93E-11
Nivel taxonómico: Orden				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	4.71E-11			
SAFS	4.71E-11	2.62E-09		
SAFS.rf	4.71E-11	4.71E-11	7.70E-04	
SBF	4.71E-11	1.44E-06	1.12E-02	8.46E-08
Nivel taxonómico: Orden				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	2.06E-08			
SAFS	1.01E-09	1.80E-10		
SAFS.rf	1.08E-03	7.70E-10	6.31E-10	
SBF	2.90E-02	2.71E-07	3.07E-10	2.74E-02
Nivel taxonómico: Familia				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	8.79E-11			
SAFS	7.47E-11	7.33E-02		
SAFS.rf	5.18E-01	2.53E-09	6.83E-10	
SBF	5.35E-11	5.35E-11	5.35E-11	5.35E-11
Nivel taxonómico: Familia				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	3.14E-11			
SAFS	3.14E-11	9.23E-12		
SAFS.rf	3.41E-09	5.19E-09	3.14E-11	
SBF	5.66E-01	3.14E-11	3.14E-11	3.11E-09
Nivel taxonómico: Familia				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	5.56E-10			

SAFS	7.71E-09	5.66E-11		
SAFS.rf	1.14E-05	7.64E-06	5.66E-11	
SBF	5.66E-11	5.35E-08	5.66E-11	7.71E-09
Nivel taxonómico: Familia				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	2.59E-12			
SAFS	3.32E-11	2.59E-12		
SAFS.rf	2.60E-06	2.59E-12	3.32E-11	
SBF	2.15E-09	2.59E-12	3.32E-11	4.53E-07
Nivel taxonómico: Familia				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	-			
SAFS	1.65E-12	1.65E-12		
SAFS.rf	1.65E-12	1.65E-12	2.03E-11	
SBF	1.65E-12	1.65E-12	1.41E-01	1.69E-11
Nivel taxonómico: Familia				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	3.47E-12			
SAFS	3.10E-08	3.47E-12		
SAFS.rf	2.18E-11	3.47E-12	2.18E-11	
SBF	2.18E-11	2.18E-11	2.18E-11	2.18E-11
Nivel taxonómico: Familia				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	5.59E-11			
SAFS	3.32E-11	8.64E-12		
SAFS.rf	5.92E-08	3.32E-11	3.32E-11	
SBF	7.20E-01	1.92E-10	3.32E-11	6.06E-07
Nivel taxonómico: Familia				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	3.93E-12			
SAFS	1.23E-10	5.62E-11		

SAFS.rf	5.02E-11	1.51E-10	3.57E-08	
SBF	3.93E-12	-	5.62E-11	1.51E-10
Nivel taxonómico: Familia				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	1.46E-02			
SAFS	6.71E-11	6.71E-11		
SAFS.rf	1.67E-10	1.41E-10	1.67E-10	
SBF	7.42E-08	3.24E-09	6.71E-11	2.04E-06
Nivel taxonómico: Familia				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	1.21E-10			
SAFS	9.19E-11	4.50E-04		
SAFS.rf	3.57E-09	4.50E-04	2.55E-04	
SBF	9.19E-11	3.58E-08	6.03E-07	2.55E-07
Nivel taxonómico: Género				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	1.83E-10			
SAFS	3.52E-12	3.52E-12		
SAFS.rf	4.28E-11	2.25E-09	7.32E-11	
SBF	6.88E-03	1.34E-05	3.52E-12	4.28E-11
Nivel taxonómico: Género				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	5.96E-13			
SAFS	5.96E-13	-		
SAFS.rf	5.96E-13	-	-	
SBF	5.96E-13	-	-	-
Nivel taxonómico: Género				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	3.53E-12			
SAFS	3.53E-12	-		
SAFS.rf	1.05E-03	1.33E-11	1.33E-11	

SBF	4.03E-10	5.00E-12	5.00E-12	1.41E-08
Nivel taxonómico: Género				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	2.62E-11			
SAFS	1.77E-12	1.63E-10		
SAFS.rf	4.62E-05	2.27E-08	1.77E-12	
SBF	1.77E-12	1.63E-10	-	1.77E-12
Nivel taxonómico: Género				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	1.63E-12			
SAFS	4.73E-09	1.69E-13		
SAFS.rf	2.24E-11	1.63E-12	1.63E-12	
SBF	2.24E-11	1.63E-12	1.63E-12	3.88E-07
Nivel taxonómico: Género				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	3.25E-12			
SAFS	3.25E-12	-		
SAFS.rf	9.81E-03	1.92E-12	1.92E-12	
SBF	3.25E-12	-	-	1.92E-12
Nivel taxonómico: Género				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	1.11E-12			
SAFS	1.11E-12	-		
SAFS.rf	5.85E-01	1.11E-12	1.11E-12	
SBF	1.11E-12	-	-	1.11E-12
Nivel taxonómico: Género				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	3.10E-11			
SAFS	4.28E-12	4.83E-10		
SAFS.rf	3.08E-11	1.26E-01	3.10E-11	
SBF	4.28E-12	4.83E-10	-	3.10E-11

Nivel taxonómico: Género				
Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	4.51E-11			
SAFS	4.51E-11	5.89E-09		
SAFS.rf	9.77E-11	1.01E-10	4.51E-11	
SBF	4.51E-11	1.96E-05	4.51E-11	5.67E-08
Nivel taxonómico: Género				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	2.85E-05			
SAFS	3.12E-12	3.12E-12		
SAFS.rf	1.32E-04	4.66E-03	3.12E-12	
SBF	3.09E-11	3.09E-11	5.95E-11	3.09E-11
Nivel taxonómico: Especie				
Modelo de predicción: LVQ				
	None	RFE	SAFS	SAFS.rf
RFE	6.71E-01			
SAFS	7.66E-10	3.28E-08		
SAFS.rf	5.32E-10	1.23E-08	8.38E-01	
SBF	4.94E-01	8.38E-01	1.27E-08	1.23E-08
Nivel taxonómico: Especie				
Modelo de predicción: GBM				
	None	RFE	SAFS	SAFS.rf
RFE	2.29E-01			
SAFS	1.24E-01	8.01E-01		
SAFS.rf	8.41E-02	7.03E-01	8.01E-01	
SBF	3.70E-09	1.15E-10	1.32E-11	1.32E-11
Nivel taxonómico: Especie				
Modelo de predicción: SVMl				
	None	RFE	SAFS	SAFS.rf
RFE	1.24E-10			
SAFS	2.62E-11	2.82E-09		
SAFS.rf	3.51E-08	1.07E-06	2.62E-11	
SBF	2.62E-11	2.62E-11	2.62E-11	2.62E-11

Nivel taxonómico: Especie				
Modelo de predicción: SVMr				
	None	RFE	SAFS	SAFS.rf
RFE	6.20E-11			
SAFS	4.65E-11	4.70E-11		
SAFS.rf	4.65E-11	4.65E-11	8.56E-01	
SBF	4.65E-11	1.27E-10	1.96E-04	1.09E-04
Nivel taxonómico: Especie				
Modelo de predicción: SLDA				
	None	RFE	SAFS	SAFS.rf
RFE	2.48E-11			
SAFS	1.11E-01	2.48E-11		
SAFS.rf	2.48E-11	2.48E-11	4.69E-11	
SBF	2.48E-11	2.07E-04	2.48E-11	2.48E-11
Nivel taxonómico: Especie				
Modelo de predicción: NN				
	None	RFE	SAFS	SAFS.rf
RFE	2.09E-12			
SAFS	2.09E-12	-		
SAFS.rf	6.83E-02	2.09E-12	2.09E-12	
SBF	2.43E-11	4.57E-10	4.57E-10	2.43E-11
Nivel taxonómico: Especie				
Modelo de predicción: Rborist				
	None	RFE	SAFS	SAFS.rf
RFE	6.32E-11			
SAFS	2.01E-12	1.54E-09		
SAFS.rf	5.73E-11	6.53E-01	5.73E-11	
SBF	2.01E-12	1.54E-09	-	5.73E-11
Nivel taxonómico: Especie				
Modelo de predicción: xgbLinear				
	None	RFE	SAFS	SAFS.rf
RFE	6.15E-12			
SAFS	6.15E-12	-		
SAFS.rf	7.88E-01	1.54E-10	1.54E-10	
SBF	6.15E-12	-	-	1.54E-10
Nivel taxonómico: Especie				

Modelo de predicción: stepLDA				
	None	RFE	SAFS	SAFS.rf
RFE	3.72E-09			
SAFS	2.44E-12	2.44E-12		
SAFS.rf	5.02E-01	3.62E-09	2.44E-12	
SBF	1.20E-03	1.28E-01	2.44E-12	3.61E-04
Nivel taxonómico: Especie				
Modelo de predicción: naive_bayes				
	None	RFE	SAFS	SAFS.rf
RFE	4.15E-11			
SAFS	4.15E-11	1.50E-05		
SAFS.rf	4.15E-11	3.97E-08	5.46E-10	
SBF	4.15E-11	5.86E-11	4.15E-11	4.45E-06