

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

**Predicción de las funciones del microbioma del suelo utilizando
herramientas bioinformáticas**

Pablo Gabriel Erazo García

Ingeniería en Biotecnología

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en Biotecnología

Quito, 17 de noviembre de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias Biológicas y Ambientales

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Predicción de las funciones del microbioma del suelo utilizando
herramientas bioinformáticas**

Pablo Gabriel Erazo García

Nombre del profesor, Título académico

Antonio León, PhD

Quito, 17 de diciembre de 2021

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Pablo Gabriel Erazo García

Código: 00200719

Cédula de identidad: 0931066674

Lugar y fecha: Quito, 17 de diciembre de 2021

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

El reciente auge del secuenciamiento de siguiente generación ha facilitado su aplicación en una gran cantidad de estudios del microbioma asociado al suelo y plantas. Sin embargo, varios de estos estudios se limitan a describir la abundancia de los diversos grupos de organismos sin considerar la funcionalidad que estos puedan tener. Esto ha motivado el desarrollo de programas bioinformáticos que usan la información de bases de datos para predecir las funciones de un microbioma secuenciado. Por ende, el objetivo de este estudio es predecir las funciones de las comunidades de hongos asociadas al suelo, por medio de algoritmos que simplifiquen la información del secuenciamiento de siguiente generación y así facilitar su interpretación. Para crear el algoritmo se utilizó un set de datos de secuenciamiento de ITS proveniente de muestras de la rizósfera y el suelo de tres variedades de rosas. Se decidió utilizar el programa FUNGuild para la predicción de las funciones de los hongos de las muestras. El programa se creó en el lenguaje de programación R. El algoritmo resultante se implementó de manera exitosa, permitiendo generar los diferentes outputs para el set de datos utilizado. La información de estos outputs mostró ser de utilidad en cuatro aplicaciones discutidas que fueron el monitoreo de patógenos, la asistencia en el diagnóstico, el análisis de tratamientos con hongos benéficos y la formulación de hipótesis. Por otro lado, se encontraron limitaciones como la pérdida de taxones en las limpiezas, la asignación de múltiples gremios en ciertos taxones y la naturaleza del programa que limita su aplicación a un uso predictivo. A pesar de esto, se concluye que esta nueva herramienta es extremadamente útil para facilitar la interpretación de los datos provenientes del secuenciamiento de ITS, y así poder aplicar esta información en diversos contextos científicos y prácticos.

Palabras clave: predicción, FUNGuild, gremio, gremio funcional, bioinformática, microbioma, hongos, ASV

ABSTRACT

The recent boom in next-generation sequencing has facilitated its application in a large number of studies of the soil and plant-associated microbiome. However, many of these studies are limited to just describing the abundance of the various groups of organisms without considering the functionality these may have. This has motivated the development of bioinformatics programs that use database information to predict the functions of a sequenced microbiome. Therefore, the objective of this study is to predict the functions of soil-associated fungal communities by means of algorithms that simplify the next-generation sequencing information and thus facilitate its interpretation. An ITS sequencing dataset from rhizosphere and soil samples of three rose varieties was used to create the algorithm. It was decided to use the FUNGuild program for the prediction of fungal functions of the samples. The program was created in the R programming language. The resulting algorithm was successfully implemented, allowing to generate the different outputs for the data set used. The information from these outputs proved to be useful in four applications discussed, which were pathogen monitoring, diagnostic assistance, analysis of treatments with beneficial fungi and hypothesis formulation. On the other hand, limitations were found such as the loss of taxa in the cleanups, the assignment of multiple guilds in certain taxa and the nature of the program that limits its application to a predictive use. Nevertheless, it is concluded that this new tool is extremely useful to facilitate the interpretation of ITS sequencing data, and thus to be able to apply this information in diverse scientific and practical contexts.

Key words: prediction, FUNGuild, guild, functional group, bioinformatic, microbiome, fungi, ASV

TABLA DE CONTENIDO

Introducción	11
Métodos.....	14
Set de datos a analizar	14
Implementación de “FUNGuild”	15
Desarrollo del programa.....	16
Resultados	17
Paso 1: Acceso a las tablas de datos	17
Paso 2: Primera limpieza de taxones.....	17
Paso 3: Análisis con “FUNGuild”	18
Paso 4: Segunda limpieza: taxones no asignados y nivel de confianza	18
Paso 5: Adaptación de tablas para su análisis	19
Paso 6: Presentación de resultados.....	20
Discusión	21
Conclusiones	25
Tablas	26
Figuras	29
Referencias Bibliográficas.....	33
Anexos	36

ÍNDICE DE TABLAS

Tabla 1. Abundancias de gremio por cada muestra	26
Tabla 2. Taxones contenidos en el gremio “Patógeno de plantas”	26
Tabla 3. Resultados del análisis estadístico de los taxones contenidos en el gremio “Patógeno de plantas”	27
Tabla 4. Posibles aplicaciones del programa creado.....	28

ÍNDICE DE FIGURAS

Figura 1. Flujograma del programa creado en R	29
Figura 2. Muestra de las tablas para importación	30
Figura 3. Resultados de las limpiezas y procesamiento.....	31
Figura 4. Abundancia de gremios en función de la combinación factores	32

ÍNDICE DE ANEXOS

Anexo 1. Lista de gremios funcionales.....	36
Anexo 2. Paquetes utilizados por el programa	37
Anexo 3. Tabla de información generada por FUNGuild	38

INTRODUCCIÓN

En los años setenta se da uno de los mayores avances para la genética que corresponde al desarrollo de un método de secuenciamiento de ADN desarrollado por Sanger y Coulson (1978). A partir de este momento, el fundamento de la técnica de Sanger se siguió utilizando por más de 30 años para elaborar sistemas cada vez más poderosos que incrementaban la capacidad de secuenciamiento de la época (Schuster, 2007). Sin embargo, una vez terminado el “Proyecto Genoma Humano” en 2004, se concluyó que era necesario el desarrollo de técnicas más rápidas, con mayor capacidad y de menor precio que permitieran facilitar el secuenciamiento de ADN (Grada y Weinbrecht, 2013). Por esto, el Instituto Nacional de Investigación del Genoma Humano (National Human Genome Research Institute, NHGRI) empezó un programa de financiamiento lo que permitió el surgimiento de las técnicas de secuenciamiento de siguiente generación (Next Generation Sequencing, NGS). Estas nuevas técnicas ofrecen el beneficio de no requerir clonar los fragmentos en bacterias, la cantidad de reacciones que ocurren es mucho mayor y el producto del secuenciamiento se puede leer sin la necesidad de una electroforesis que revele los resultados (Van Dijk et al., 2014). Adicionalmente, estas reacciones ocurren en paralelo de manera que se pueden secuenciar millones y millones de muestras al mismo tiempo, reduciendo el costo y tiempo que toma secuenciar cada muestra (Reis-Filho, 2009).

Más aún en los últimos años, el bajo costo que se ha alcanzado con las tecnologías emergentes ha provocado un auge del uso del secuenciamiento de siguiente generación en diversos estudios de microbiomas de suelos. En el pasado, para analizar un microbioma se recurría a técnicas clásicas de cultivo para intentar extraer e identificar los microorganismos que componen a dicho microbioma. Sin embargo estos métodos solían ser sesgados, poco replicables y, en general, poco confiables para la caracterización de microorganismos. Además,

existen ciertas limitaciones para este enfoque como ciertos organismos o estructuras que son muy difíciles de cultivar como los biofilms u organismos que simplemente no pueden ser cultivados bajo condiciones normales de laboratorio (Hodkinson y Grice, 2015). Por estas razones, el secuenciamiento ha sido la técnica preferida en varios estudios de microbiomas ya que supera muchas de las limitaciones mencionadas al no requerir el cultivo de microorganismos, además de que toma menos tiempo y recursos.

De este modo, los beneficios que ofrece el secuenciamiento ha permitido desarrollar nuevas investigaciones capaces de determinar la composición, abundancia y diversidad de las poblaciones microbianas se los suelo. Esto es útil en particular para tener un conocimiento más comprensivo acerca de la influencia que tienen diversos factores en la forma en que se ensamblan las comunidades microbianas (Fierer, 2017). Sin embargo, muchos de estos estudios pasan por alto o no ahondan en el tema de la funcionalidad de dichas comunidades (Douglas et al., 2020). Esto ha motivado el desarrollo de programas bioinformáticos que permiten predecir las funciones que tendrían los microorganismos de una muestra utilizando los datos de secuenciamiento de marcadores genéticos como el 16S ARN o los ITS. Ejemplos de estos programas son “PICRUST2” (Douglas et al., 2020) y “FUNGuild” (Nguyen et al., 2016), diseñados para bacterias y hongos respectivamente. Estas herramientas tienen una utilidad fundamental al revelar información adicional y complementaria sobre una muestra lo cual ayuda a comprender mejor las funciones presentes en un microbioma (Zhu et al., 2019). Adicionalmente, los resultados obtenidos del secuenciamiento son tan extensos que se vuelven difíciles de comprender lo cual, en definitiva, expone la necesidad apremiante de encontrar formas de simplificar la información para facilitar su manejo e interpretación.

A partir de la problemática planteada, surge la idea de utilizar los programas bioinformáticos disponibles para predecir las funciones de microbiomas de hongos con el objetivo de simplificar y sintetizar la información, y con esto conseguir facilitar la

interpretación de los datos para su uso a nivel práctico y científico. Por lo tanto, el objetivo principal del siguiente trabajo de investigación es predecir las funciones específicas de las comunidades de hongos asociadas al suelo y a las plantas por medio de algoritmos que simplifiquen la información del secuenciamiento de siguiente generación. Para esto se va a implementar el programa “FUNGuild” dentro de un algoritmo de análisis de secuencias ITS de hongos. Luego se utilizará la información sobre gremios funcionales de hongos para identificar taxones específicos e inferir diferencias entre variables de un estudio. Por último, se plantea representar gráficamente los resultados obtenidos de los análisis de funcionalidad de hongos provenientes de datos de secuenciamiento de ITS de rosas. Este proyecto muestra relevancia el ámbito científico y agrícola ya que provee una nueva herramienta para el procesamiento e interpretación de datos de secuenciamiento de hongos.

MÉTODOS

Set de datos a analizar

Para probar y desarrollar el programa se utilizó un set de datos proveniente del secuenciamiento de ITSs, obtenidos de muestras de la rizósfera y el suelo de tres variedades de rosas injertadas y no injertadas. En la industria florícola se recurre a los injertos para producir diversas variedades de rosas, utilizando plantas ya enraizadas conocidas como patrones o “portainjertos” (“rootstocks”). Esta técnica se utiliza ya que se ha observado que, el rendimiento, número y calidad de rosas producidas es mejor en plantas injertadas que en las plantas normales, en especial cuando estas se siembran en tierra o en sustratos a base de tierra (Cabrera, 2002). Adicionalmente, el uso de diferentes patrones permite controlar y estandarizar factores como el vigor de la rosa, el área foliar, el rendimiento y calidad de la rosa injertada (Otiende et al., 2016). En este caso, se obtuvieron muestras del patrón “Natal Briar” que es un portainjertos muy utilizado en el Ecuador (Cabascango, 2008). Las demás muestras se obtuvieron de dos variedades no injertadas llamadas “Freedom” y “Explorer”. Para las muestras de suelo se recogió material de la tierra cercana a las raíces de la plantas. En el caso de las muestras de rizósfera, estas se obtuvieron extrayendo la planta, retirando el exceso de tierra adherida a las raíces y luego realizando un lavado para obtener el material pegado a la raíz de la planta. Se obtuvieron 4 réplicas para cada tipo de muestra utilizando 4 plantas diferentes. Para la extracción de ADN y preparación de librerías se utilizó el protocolo descrito por Finkel y colaboradores (2019). Se amplificó la región ITS 1 y se secuenció con el instrumento Illumina MiSeq. Los amplicones se procesaron usando el paquete DADA2 (Callahan et al., 2016) en R y se obtuvo la taxonomía utilizando el clasificador “mothur” (Schloss et al., 2006) con la base de datos UNITE a manera de referencia. De este

procesamiento se obtuvieron dos archivos típicos que son el archivo de conteos en formato “.tsv” y el de taxonomía en formato “.taxonomy”.

Implementación de “FUNGuild”

Se propuso utilizar el algoritmo “FUNGuild” desarrollado por Nguyen y colaboradores (2016). El propósito de este programa es clasificar taxones de hongos diferentes en categorías “ecológicamente relevantes” llamadas gremios funcionales. Se entiende como un gremio funcional a las especies de hongos que consumen los mismos recursos (Koch et al., 2019). Dichos gremios también se pueden entender como la función que cumplen dichos hongos en su ambiente lo cual es relevante para conocer la funcionalidad de las comunidades de hongos de un microbioma. La razón por la que el programa utiliza esta clasificación es que los gremios son una forma muy útil de resumir y agrupar información muy compleja sobre un microbioma de hongos, en unidades ecológicas que son más comprensibles y manejables. Esta perspectiva difiere de las estrategias comúnmente utilizadas para describir la composición microbiana que son la diversidad y abundancia lo que, en cuestión, facilita el realizar comparaciones entre comunidades incluso cuando estas no poseen los mismos tipos de hongos ya que realiza la clasificación por modo trófico en lugar de taxonomía (Nguyen et al., 2016).

Profundizando un poco más acerca del funcionamiento del programa, básicamente lo que realiza es tomar la taxonomía de un hongo y asignarla a diferentes gremios funcionales con base a la información contenida en la base de datos creada para el proyecto la cual se aloja en la página de GitHub del proyecto (<https://github.com/UMNFuN/FUNGuild>). El programa realiza tres distinciones principales de gremios que son los patótrofos, saprótrofos y simbiótrofos. A la primera clasificación de patótrofos se asignan los hongos que obtienen los nutrientes dañando las células hospedadores. Los hongos que reciben nutrientes descomponiendo materia se clasifican en saprótrofos. Finalmente, se clasifican en el grupo de simbiótrofos a aquellos hongos que obtienen nutrientes al intercambiar recursos con la

célula hospedadora (Nguyen et al., 2016). Una vez que se han clasificado los taxones dentro de estas tres clasificaciones principales, el programa los va a asignar a cualquiera de los gremios funcionales que se detallan en el Anexo 1.

Desarrollo del programa

El programa fue desarrollado en el lenguaje de programación R (versión 4.0.4) usando la terminal de “R Studio” (versión 1.4.1106). Durante todo el programa se utilizaron en total de 22 paquetes que se muestran en el Anexo 2. El programa parte de los archivos de conteos y de taxonomía y además requiere de un archivo metadata en formato “.tsv” el cual contiene toda la información correspondiente a las muestras. El flujo de trabajo del programa se diseñó con base a los objetivos específicos planteados.

RESULTADOS

A continuación se describe el funcionamiento del algoritmo creado, que consta de seis pasos principales como se indica en el flujoograma de la Figura 1.

Paso 1: Acceso a las tablas de datos

Para esto, el programa accede desde el directorio de trabajo a los archivos de conteos, taxonomía y metadata, que se muestran en la Figura 2. En primer lugar, el archivo de conteos consta de una primera columna con los nombres de las muestras y varias columnas independientes que corresponden a los ASVs encontrados en la muestra con su respectivo conteo. Se entiende a un ASV como una secuencia representativa única entre todos los amplicones, producto del proceso de denoising realizado por DADA2. Por lo tanto esta tabla muestra cuantas veces se contó cada ASV en cada muestra. La tabla de taxonomía muestra los ASVs en la primera columna y la taxonomía asignada por “mothur” en la segunda columna. Finalmente el archivo metadata muestra en una primera columna los nombres de las muestras de igual forma que lo hace el archivo de conteos. En las columnas posteriores se registra la información sobre cada muestra lo cual, para el set de datos utilizado, corresponden a la variedad a la que pertenecen las muestras y la fracción de la que fueron obtenidas.

Paso 2: Primera limpieza de taxones

Siguiendo con el objetivo de la implementación del “FUNGuild”, se procede a realizar una primera limpieza. El propósito de este paso es eliminar todos los taxones que no pertenecen al reino “Fungi”. En este caso se eliminan los ASVs cuya taxonomía contenga patrones no deseados como: “mitochondria”, “chloroplast”, “Eukaryota”, “unknown”, “Fungi-unclassified”, entre otros. De esta manera se descartan los taxones que no son de interés y así se disminuye el tiempo de procesamiento del programa. El programa exporta la tabla resultante de la limpieza en un archivo de nombre “1_asv_1clean_RFUNGuild.txt”.

Paso 3: Análisis con “FUNGuild”

En este paso culmina el primer objetivo específico que consiste en realizar la predicción de los grupos funcionales para cada taxón insertado. Para esto se aplica el paquete de R llamado “FUNGuildR” creado por Ferneaux y Song (2021) que corresponde a una adaptación del programa originalmente creado en el idioma Python por Nguyen y colaboradores (2016). En el proceso, se inserta la tabla previamente filtrada y el paquete acude a la base datos de “FUNGuild” para asignar la información pertinente a cada taxón que se pueda encontrar en los registros. Esta información se adjunta como columnas al final de la tabla original como se observa en el Anexo 3, las cuales describen características como el gremio, los rasgos, modo trófico y la referencia de dicha información. Los resultados crudos de la asignación se exportan en un segundo archivo de nombre “2_raw_output_FUNGuild.txt”.

Paso 4: Segunda limpieza: taxones no asignados y nivel de confianza

Algo que se aprecia claramente en el Anexo 3 es que varios taxones no recibieron asignación lo que se ve como valores “NA” en las columnas de información. Por lo tanto, el siguiente paso, consiste en una segunda limpieza donde se eliminan estos taxones sin asignación. Adicionalmente, también se filtran los taxones según su nivel de confianza. El programa predictivo “FUNGuild” establece tres niveles de confianza para las asignaciones que son “Highly Probable”, “Probable” y “Possible”. La primera asignación es la de mayor confianza y se establece cuando la información de referencia proviene de fuentes primarias que permiten establecer, con absoluta seguridad, que dicho hongo cumple la función especificada. El segundo es un nivel intermedio de confianza, que se establece cuando la información sobre un hongo no proviene de fuentes primarias. El último nivel es el de menor confianza y se establece cuando la información provista no ha sido probada o cuando existen estudios contradictorios respecto a estos datos. En la práctica y creación del programa se observó que, eliminando este último nivel de confianza, se consiguen datos más relevantes y concisos para

los análisis por lo que para el ejemplo se programó al algoritmo para eliminar todos los taxones con este nivel de confianza. En este sentido, se recomienda dejar este parámetro por defecto, sin embargo cada usuario puede elegir cuáles niveles de confianza conservar. Al final de esta etapa se obtiene una tabla totalmente limpia, la cual es exportada por el programa con el nombre “3.1_asv_2clean_RFUNGuild.txt”.

Paso 5: Adaptación de tablas para su análisis

Para este paso, primero se observan los resultados de las limpiezas en la Figura 3. La sección A de esta figura muestra el número de ASVs resultantes de cada etapa. Para el ejemplo, se comenzó con 7936 taxones, valor que se redujo a tan solo 2958 después de la primera limpieza. Al final quedan 1632 taxones que corresponde a aquellos que sí recibieron asignación. A continuación, en la sección B se observa que de los 1632 asignados, 491 se eliminan por el nivel de confianza en su asignación, de forma que luego de todas las limpiezas quedaron solo 1141 ASVs. La gráfica C muestra la información sobre la asignación de gremio de los taxones sobrantes. Esto proviene de que varios de los taxones no caen exclusivamente en una sola clasificación de gremios, sino que puede pertenecer a 2 o más gremios al mismo tiempo. Por lo tanto, lo que se puede interpretar de la gráfica C es que una mayoría de los taxones se les asignó un solo gremio. Como resumen se establece que, en promedio, a cada ASV se le asignó 1.902 gremios funcionales. Esta figura es exportada por el programa con el nombre “3.2_cleaning_stats” en formato “.png” y en formato “.tiff” a conveniencia del usuario.

A continuación, el programa va a crear una tabla que resume las abundancias de cada gremio según las muestras. Para esto, el programa filtra los ASVs según la asignación de gremio, creando una lista que contendrá los ASVs filtrados. A continuación, el programa suma los conteos para cada muestra lo que resulta en la generación de la Tabla 1, la cual se exporta en un archivo llamado “4_guild_counts.tsv”. Esta tabla muestra la síntesis de los conteos de cada gremio con relación a las muestras. De igual forma, el programa implementa pasos similares a

los anteriores pero esta vez para extraer los taxones pertenecientes a cada gremio. Es decir, va a filtrar los ASVs según su gremio y en esta ocasión va crear tablas que muestren los taxones que componen a cada gremio funcional. Cabe mencionar que estos taxones se colapsan hasta el nivel de género para poder resumir la información. El producto de este paso se muestra en la Tabla 2, que se exporta como un archivo Excel llamado “5_composition_guilds_by_taxons.xlsx”, el cual posee una pestaña por cada gremio. De esta forma, cada usuario pueda navegar entre estas pestañas viendo los taxones que se asignan dentro de cada gremio junto con su abundancia.

Finalmente, el programa procede a realizar la inferencia estadística por cada taxón tomando en cuenta las variables de interés. Para esto utiliza la Tabla 2 generada previamente y se realiza un ANOVA en conjunto con una prueba de Tukey para cada taxón único encontrado. Esta figura permite observar si existen diferencias significativas en la abundancia de un taxón con respecto a las variables de interés. Para el ejemplo, esta acción implicó que el programa realice 630 ANOVAs independientes para generar el archivo Excel de nombre “7_anova_by_taxons.xlsx”, el cual se ve como se muestra en la Tabla 3.

Paso 6: Presentación de resultados

Cumpliendo con el último objetivo, el programa elabora un heatmap a partir de los datos de la Tabla 1. Esta gráfica muestra, según la intensidad de color, la abundancia relativa de las muestras con respecto a las variables de interés. También aplica una prueba ANOVA para cada gremio de interés de forma que se obtienen las letras de significancia de Tukey lo que permite comparar si existe diferencias significativas en la abundancia de gremios con relación a las variables. Esta gráfica se exporta con el nombre “8_heat_map” en formato “.png” y “.tiff”.

DISCUSIÓN

Se logró implementar el algoritmo de manera exitosa y obtener los diferentes productos para el set de datos utilizado. Este presentó beneficios como que está totalmente automatizado lo cual facilita su uso ya que el usuario solo debe definir los parámetros deseados en la primera sección y el programa va a ejecutar todos los pasos de manera autónoma. En este sentido, el algoritmo permite customizar los análisis que se desean realizar sin la necesidad de modificar el código ya escrito. Otro beneficio observado es la rapidez del programa, tomando alrededor de 7 minutos en procesar el ejemplo mostrado, desde la carga de tablas hasta los resultados finales. Por lo tanto, cualquier usuario puede procesar los datos de un microbioma de hongos en menos de 10 minutos y obtener una serie de archivos y gráficas de fácil interpretación que detallan la funcionalidad de los datos insertados. Así pues, con el set de datos de rosas se logró obtener dos productos principales. El primero es el archivo “.xlsx” que muestra la abundancia de diferentes géneros de hongos, separados según su gremio (Tabla 3). El segundo es el heatmap que muestra la abundancia relativa de cada gremio según las variables de interés (Figura 4). Estos dos archivos pueden ser usados para diferentes aplicaciones como el monitoreo de patógenos, la asistencia en el diagnóstico de enfermedades con base a sintomatología, el análisis de tratamientos con hongos benéficos y la formulación de hipótesis en el ámbito práctico y científico (Tabla 4). La primera aplicación consisten en el monitoreo de la abundancia diversos patógenos conocidos con respecto a las variables de interés. La utilidad de estos monitores es que permiten conocer los taxones patógenos que se encuentran abundantes en un microbioma y así se pueden tomar decisiones acertadas para su tratamiento. Como ejemplo para este set de datos, en la Tabla 3 se observa que el género *Alternaria* sp. está enriquecido en la variedad “Natal Briar”. Esto es importante considerar ya que este género de hongos es conocido por causar lesiones en los pétalos de las rosas que se evidencian como

manchas negras (Abbas et al., 2017). Utilizando esta información, se pueden aplicar tratamientos preventivos a los cultivos de esta variedad de rosas y así evitar pérdidas. Una opción propuesta por Mazur y colaboradores (2019) es la aplicación de hongos micorrízicos, los cuales mejoran significativamente la resistencia de las plantas ante *Alternaria alternata* que es la principal especie causante de enfermedades.

La segunda aplicación consiste en el uso de la información de abundancia para agilizar el diagnóstico de enfermedades con base a su sintomatología. Esto permite deducir cuáles taxones podrían estar causando la enfermedad y rechazar aquellos que no son relevantes, lo cual es particularmente útil al momento de diagnosticar enfermedades que están asociadas a varios hongos, algo que es común en las rosas (Ghosh y Shamsi, 2014). Dado que en este estudio se recolectaron plantas sanas, no es posible relacionar los datos directamente con cierta sintomatología. No obstante, si se hallaran indicios de alguna enfermedad como la pudrición blanda en las rosas, se puede acceder al documento que se muestra en el Anexo 3 y filtrar los taxones según los síntomas que producen al infectar una planta, que en este caso correspondería a “Soft Rot”. Entre los taxones resultantes se puede encontrar el género *Torula*. A continuación se puede encontrar la abundancia de este taxón en la pestaña de “Patógenos del plantas” dentro del documento de la Tabla 4, donde se puede ver que la abundancia de *Torula* es muy baja en todas las muestras. Por lo tanto, en el diagnóstico se puede descartar el género *Torula* como posible causante de síntomas de pudrición blanda ya que su abundancia es muy baja, lo cual permite ahorrar recursos y reducir el tiempo de diagnóstico.

Otra aplicación consiste en el análisis de la efectividad de tratamientos con hongos benéficos. Por medio de la Figura 4 se determinó que la variedad de la rosa influye sobre la abundancia de micorrizas arbusculares, que son hongos con un papel fundamentales en la defensa y protección contra insectos herbívoros y otros patógenos (Frew et al., 2021). Se observó que la variedad “Freedom” presenta una alta abundancia de estos hongos mientras que

“Natal Briar” presenta una baja abundancia. Este factor debe ser considerado al momento de aplicar tratamientos con micorrizas ya que un tratamiento diseñado para una variedad podría no ser efectivo en otras variedades. Por ejemplo, si queremos aplicar el tratamiento propuesto en la primera aplicación para combatir *Alternaria*, es posible que dicho tratamiento aplicado en “Freedom” sea más efectivo que en “Natal Briar”.

La última aplicación relevante del programa es para generar hipótesis sobre el microbioma en cuestión. Como ejemplo, se puede cuestionar por qué razón las variedades reclutan diferente número de micorrizas arbusculares y cuales son los mecanismos que rigen sobre este fenómeno. De igual forma, se podría hipotetizar que la razón por la que la variedad “Natal Briar” tiene mayor abundancia de *Alternaria* (Tabla 3) es porque tiene menor abundancia de micorrizas (Figura 4). Como se observa, la información de diferentes resultados se puede utilizar para formular hipótesis científicas.

Si bien el algoritmo presenta varios beneficios, a lo largo del desarrollo también se pudieron observar algunas limitaciones. Primero se observó que una gran parte de los taxones iniciales se eliminan con las limpiezas (Figura 3 A). En el ejemplo se inició con 7936 taxones y al final quedaron tan solo 1141 taxones. La razón de esta gran pérdida es que, debido a que el marcador ITS está presente en otros taxones además de hongos (Rivers et al., 2018), el secuenciamiento va a generar varias lecturas provenientes de organismos como plantas, nemátodos e insectos, las cuales se deben descartar. Adicionalmente, en la segunda limpieza se eliminan muchos taxones de interés que no recibieron asignación, producto de la ausencia de información en la base de datos lo que limita la profundidad con la que se puede caracterizar el microbioma. En la misma Figura 3 A se observa que de los 2958 taxones de hongos analizados, el 44.8% no recibieron asignación lo cual es comparable a los resultados obtenidos por Nguyen y colaboradores (2016) quienes obtuvieron un porcentaje de no asignados de entre 41% al 59% utilizando FUNGuild con diferentes sets de datos. A pesar de esto, la base de datos

del proyecto está en constante crecimiento lo que podría mejorar el nivel de asignación a largo plazo.

Otra limitación que se encuentra en el programa es la asignación de varios gremios a un mismo ASV. Esto provoca que algunos ASVs se figuren en 2 o más listas de conteos al mismo tiempo. La consecuencia de esto es que se van a sobreestimar las abundancias de algunos gremios. Esta limitación proviene de la complejidad funcional que puede tener algunos hongos (Wagg et al., 2019). Aquí también se encuentra otro problema que consiste en que, dependiendo del set de datos con el que se trabaje, algunos hongos tendrán una asignación irrelevante o incorrecta. Para demostrar esto se puede tomar de ejemplo a *Trichoderma* el cual es un organismo endófito comúnmente encontrado en las raíces de plantas donde se ha observado que puede presentar actividad en contra de patógenos (Hanada et al., 2010). Este género también puede ser patógeno al causar pudrición blanda en algunas plantas (Seehan et al., 1975) no obstante, no hay reporte de este hecho en rosas. A pesar de esto, en los resultados de la Tabla 3 igual se puede encontrar este taxón en el gremio de patógenos de plantas. Por lo tanto, este tipo de incongruencias van a ocurrir dependiendo de las muestras con las que se esté trabajando. La última limitación radica en la naturaleza predictiva de los programas como “FUNGuild”. Dentro de los suelos, la biodiversidad es muy elevada y el funcionamiento de los microbiomas está liderado por interacciones de competencia, intercambio y antagonismo entre todos los microorganismos presentes (Saleem et al., 2019). Estas relaciones son extremadamente complejas por lo que es imposible conocer si los resultados obtenidos de este programa representan fielmente la realidad del microbioma. Debido a esto, no se debe usar “FUNGuild” como una herramienta confirmatoria, sino como una herramienta para generar hipótesis con la información obtenida sobre la funcionalidad de hongos. A pesar de esto, “FUNGuild” ha demostrado su utilidad para facilitar la interpretación de la información de secuenciamiento de ITS, especialmente en el contexto agrícola y científico.

CONCLUSIONES

Mediante el programa se logró predecir las funciones del microbiomas de hongos de rosas utilizando el algoritmo FUNGuild. Este demostró ser muy útil para obtener clasificaciones ecológicamente relevantes e inferir las diferencias entre las variables de los datos utilizando métodos estadísticos. Finalmente, se presentaron estos resultados mediante gráficos y tablas de fácil interpretación para cada usuario. Por lo tanto, se cumplió con el objetivo principal de predecir la funcionalidad de comunidades de hongos y generar resultados de fácil interpretación que resumieron de forma sencilla los resultados de un secuenciamiento de ITS.

Se destacan beneficios de este programa como la automatización lo que facilita su uso, la rapidez en la obtención de resultados y la generación de resultados fáciles de interpretar. También se examinaron cuatro potenciales aplicaciones del programa que son: el monitoreo de microorganismos, asistencia en diagnóstico de enfermedades con base a sintomatología, análisis de tratamientos con hongos benéficos y la formulación de hipótesis en el ámbito práctico y científico. Por otro lado, se encontraron limitaciones como la pérdida de taxones en las limpiezas, la asignación de múltiples gremios en ciertos taxones y la naturaleza del programa que limita su aplicación a un uso predictivo.

En conclusión, se logró crear un programa capaz de analizar, procesar y resumir los datos del secuenciamiento de ITS de hongos, utilizando el algoritmo predictivo “FUNGuild”. Se probó la aplicabilidad de esta nueva herramienta en el ámbito práctico y científico lo cual corrobora la utilidad de este proyecto para interpretar los datos de secuenciamiento de ITS y obtener información valiosa sobre un microbioma de hongos.

TABLAS

Tabla 1. Abundancias de cada gremio por cada muestra. Aquí se observan los conteos de algunos gremios con los que trabaja el programa lo que representa su abundancia en la muestra.

	Plant Pathogen	Animal Pathogen	Bryophyte Parasite	Clavicipitaceous Endophyte	Fungal Parasite	Lichen Parasite	Algal Parasite
ROERh1	1112	1254	473	0	3722	205	473
ROERh2	3565	980	328	0	1250	189	886
ROERh3	888	965	108	0	901	41	104
ROERh4	1164	1455	222	0	930	67	222
ROESo1	1119	1203	282	0	1196	127	261
ROESo2	950	585	142	0	545	73	137
ROESo3	2085	1740	130	0	1363	279	127
ROESo4	1215	1084	225	0	1035	106	220
ROFRh1	468	522	59	0	2448	68	59
ROFRh2	1389	3177	517	0	3620	222	518
ROFRh3	1415	1502	385	0	3053	41	379
ROFRh4	5401	4246	1638	0	6805	310	632
ROFSO1	4800	544	4501	0	5493	91	4494
ROFSO2	2026	826	77	0	1669	243	74
ROFSO4	13788	3220	210	0	5939	430	188
RONRh1	1672	1439	86	0	644	619	86
RONRh2	3756	1130	73	0	1994	1070	98
RONRh3	7793	3821	3201	0	4709	4020	3201
RONRh4	5734	336	133	0	258	2821	133
RONSo1	3444	2346	226	0	2718	1462	245
RONSo3	3252	2437	119	0	2457	523	129

Tabla 2. Taxones contenidos en el gremio “Patógeno de plantas”. Se observan varios de los taxones que se clasificaron dentro de este gremio funcional con su respectivo conteo para 5 de las 22 muestras totales. La última columna muestra los nombres de los ASVs que pertenecen y contribuyen a los conteos de cada taxonomía.

Taxonomy	ROERh1	RONRh4	RONSo1	RONSo3	RONSo4	ASVs
Fungi;Ascomycota;Dothideomycetes;Botryosphaeria	0	0	0	7	0	ASV4364
Fungi;Ascomycota;Dothideomycetes;Capnodiales;M	160	64	243	71	162	"ASV164", "ASV1913", "ASV2872"
Fungi;Ascomycota;Dothideomycetes;Capnodiales;M	0	0	0	0	0	ASV2872
Fungi;Ascomycota;Dothideomycetes;Capnodiales;Te	0	0	0	4	0	ASV6654
Fungi;Ascomycota;Dothideomycetes;Pleosporales;C	0	0	0	0	0	"ASV4849", "ASV5256", "ASV1524", "ASV7351", "ASV5364"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;C	0	0	104	17	0	"ASV1524", "ASV7351", "ASV5364"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;D	0	0	0	0	0	ASV5364
Fungi;Ascomycota;Dothideomycetes;Pleosporales;D	3	0	0	5	0	"ASV3103", "ASV5232", "ASV1203", "ASV4166", "ASV1406", "ASV1975", "ASV3393", "ASV4742"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;D	27	33	30	56	92	"ASV1203", "ASV4166", "ASV1406", "ASV1975", "ASV3393", "ASV4742"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;D	7	9	12	23	35	"ASV1406", "ASV1975", "ASV3393", "ASV4742"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;D	20	0	10	0	0	"ASV3393", "ASV4742"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;Li	5	0	0	0	0	"ASV1819", "ASV5021", "ASV7175"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;Li	0	0	3	0	0	ASV7175
Fungi;Ascomycota;Dothideomycetes;Pleosporales;Li	54	27	14	35	38	"ASV1035", "ASV106", "ASV2826", "ASV3069"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;N	0	0	0	0	0	"ASV2826", "ASV3069"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;P	0	0	2	0	65	ASV1420
Fungi;Ascomycota;Dothideomycetes;Pleosporales;P	0	0	12	0	0	ASV4387
Fungi;Ascomycota;Dothideomycetes;Pleosporales;P	0	0	0	0	0	ASV7783
Fungi;Ascomycota;Dothideomycetes;Pleosporales;P	0	0	13	0	0	"ASV4234", "ASV7508"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;P	0	0	8	0	0	"ASV2015", "ASV6773", "ASV3673", "ASV5336"
Fungi;Ascomycota;Dothideomycetes;Pleosporales;P	0	0	0	5	0	"ASV3673", "ASV5336"

Tabla 3. Resultados del análisis estadístico de los taxones contenidos en el gremio “Patógeno de plantas”. Se muestran los taxones de cada gremio en función de las combinaciones de las variables, y se muestran los resultados de los ANOVAs realizados por cada taxón mediante las letras de significancia de la prueba de Tukey. Se señalaron dos taxones correspondientes a hongos patógenos de interés en la agricultura que son *Alternaria* sp. y *Botrytis* sp. De esta manera se puede observar si existen diferencias significativas entre las muestras con respecto a las variables de interés.

	Explorer/Rizósfera	Explorer/Suelo	Freedom/Rizósfera	Freedom/Suelo	Natal Briar/Rizósfera	Natal Briar/Suelo
Fungi.Ascomycota.Dothideomycetes.Pleosporales.Pleomassariaceae.unidenti	0 (a)	0 (a)	0 (a)	0 (a)	0 (a)	0.200686663775987 (a)
Fungi.Ascomycota.Dothideomycetes.Pleosporales.Pleosporaceae. <i>Alternaria</i>	1.68290170335339 (a)	1.74147071576043 (a)	1.94942309512739 (ab)	2.29871834491482 (b)	3.20680362141713 (b)	3.04862816813529 (b)
Fungi.Ascomycota.Dothideomycetes.Pleosporales.Pleosporaceae. <i>Bipolaris</i>	0.597733776775845 (a)	0 (a)	0 (a)	0.371314450768946 (a)	0.211274510003564 (a)	0.347130895052742 (a)
Fungi.Ascomycota.Dothideomycetes.Pleosporales.Pleosporaceae. <i>Curvularia</i>	0.194537812595911 (a)	0.269795311511906 (a)	0 (a)	0 (a)	0.150514997831991 (a)	0.433676665221327 (a)
Fungi.Ascomycota.Dothideomycetes.Pleosporales.Pleosporaceae. <i>Drechslera</i>	0.677317490243958 (a)	1.07993262358256 (a)	0.639075625191822 (a)	0 (a)	0.225772496747986 (a)	1.56659523714578 (a)
Fungi.Ascomycota.Dothideomycetes.Pleosporales.Torulaceae. <i>Dendryphion</i>	0 (a)	0 (a)	0 (a)	0 (a)	0 (a)	0.159040418239887 (a)
Fungi.Ascomycota.Dothideomycetes.Pleosporales.Torulaceae. <i>Torula</i>	1.34374204657481 (a)	1.4724878143205 (a)	0.53653200891956 (a)	1.13031170236779 (a)	0.286532008919559 (a)	1.42484243479868 (a)
Fungi.Ascomycota.Eurotiomycetes.Eurotiales.Aspergillaceae. <i>Penicillium</i>	0 (a)	0.238560627359831 (a)	0 (a)	0 (a)	0 (a)	0.501716659439969 (a)
Fungi.Ascomycota.Leotiomycetes.Erysiphales.Erysiphaceae. <i>Podosphaera</i>	0 (a)	0 (a)	0 (a)	0 (a)	0.433098439955742 (a)	0.736505004847544 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Chaetomellaceae. <i>Pilidium</i>	0 (a)	0.150514997831991 (a)	0 (a)	0 (a)	0 (a)	0 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Dermateaceae. <i>Mollisia</i>	0 (a)	1.05735646198017 (a)	0.498908798649387 (ab)	0 (b)	0 (b)	1.3169829047998 (b)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Calyptrozyma</i>	0 (a)	0.386017011087569 (a)	0 (a)	0 (a)	0 (a)	0 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Crocicreas</i>	0.150514997831991 (a)	0.389075625191822 (a)	0 (a)	0 (a)	0 (a)	0 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Fontanospora</i>	0 (a)	0 (a)	0.319688400238207 (a)	0 (a)	0 (a)	0 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Geniculospora</i>	0 (a)	0.119280313679916 (a)	0 (a)	0 (a)	0 (a)	0 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae.Helotiaceae_unclas	0 (a)	0.788682051860039 (a)	0.174742501084005 (a)	0.866990965228916 (a)	0 (a)	1.23966654595957 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Meliniomyces</i>	0 (a)	0.533384727092554 (a)	0 (a)	0 (a)	0 (a)	0 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Pseudoclathrosphaera</i>	0 (a)	0.418024464483929 (a)	0.87301540112815 (a)	0.653013797440365 (a)	0 (a)	0 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Scytalidium</i>	0.5 (a)	0.33055482368348 (a)	0.389075625191822 (a)	0.23299000144534 (a)	0 (a)	0.651414169813108 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Helotiaceae. <i>Tetracladium</i>	0 (a)	0.225772496747986 (a)	0.345052810427901 (a)	0.159040418239887 (a)	0 (a)	0.301029995663981 (a)
Fungi.Ascomycota.Leotiomycetes.Helotiales.Sclerotiniaceae. <i>Botrytis</i>	1.08341218912868 (a)	1.45445478240664 (a)	0.884139610642882 (a)	2.48501836635743 (a)	1.77918380614346 (a)	1.46799265455645 (a)
Fungi.Ascomycota.Pezizomycetes.Pezizales.Rhiziniaceae. <i>Rhizina</i>	0 (a)	0 (a)	0 (a)	0 (a)	0 (a)	0.534019997109321 (a)

Tabla 4. Posibles aplicaciones del programa creado. Se resumen las cuatro posibles aplicaciones del programa con los respectivos ejemplos en los datos obtenidos

Aplicación	Utilidad	Ejemplo en el estudio
Monitoreo de patógenos	Conocer la abundancia de hongos patógenos de interés bajo diversas condiciones que permitan tomar decisiones respecto a la necesidad de tratamientos preventivos para ciertas enfermedades	En la Tabla 3 se observó que el taxón <i>Alternaria</i> sp. está enriquecido en la variedad “Natal Briar”. Este hongo es conocido por causar lesiones en las rosas que se ven como manchas negras (Abbas et al., 2017). Por lo que se podría enfocar el uso de tratamientos preventivos para esta variedad específica. De igual forma, <i>Botrytis</i> sp. es un hongo patógeno de importancia para las rosas el cual no aparenta estar aumentado en ninguna de las muestras, lo que significa que se puede aplicar el mismo tratamiento en todas las variedades.
Asistencia en el diagnóstico con base a sintomatología	Analizar la abundancia de diversos hongos asociados a una enfermedad para agilizar el diagnóstico y ahorrar recursos	Para este estudio se tomaron muestras de plantas sanas por lo que no es posible relacionar los datos directamente con sintomatología. Sin embargo, si es que existiera alguna sintomatología, es posible acudir a la información de abundancias para deducir los posibles taxones causantes de la enfermedad o descartar aquellos que no son relevantes. Por ejemplo, si se encontraran rosas con sintomatología de pudrición blanda, se puede usar el archivo “3.1_asv_2clean_RGUNGuild.txt” (Anexo 3) para filtrar los taxones según sus rasgos (“trait”) y obtener los ASVs relacionados con la enfermedad de “Soft Rot”, encontrado géneros como <i>Torula</i> . Luego, en el archivo correspondiente a la Tabla 4, llamado “7_anova_by_taxons.xlsx”, se puede ver que la abundancia de <i>Torula</i> es muy baja por lo que se descarta este género del diagnóstico de la podredumbre blanda. La utilidad radica en que se logra ahorrar recursos y reducir el tiempo de diagnóstico, en especial cuando se trata con enfermedades asociadas a varios hongos, lo cual ocurre mucho en rosas (Ghosh y Shamsi, 2014).
Análisis de tratamientos con hongos benéficos	Determinar la influencia de diversos factores en la efectividad de los tratamientos con hongos benéficos.	Por medio de la Figura 4 se determinó que la variedad influye en la abundancia de micorrizas arbusculares en la rizósfera, de manera que la variedad “Freedom” presenta una abundancia significativamente mayor de micorrizas. Estos hongos cumplen un papel importante en la defensa contra insectos y patógenos (Frew et al., 2021) por lo que conocer la influencia de factores puede ayudar a implementar correctamente los tratamientos.
Formulación de hipótesis	La información puede ser utilizada para formular hipótesis sobre algún fenómeno de interés en el ámbito práctico y científico	Para el ejemplo, se observó que la variedad aparenta tener un efecto en el reclutamiento de ectomicorrizas y micorrizas arbusculares, que se evidencia en la Figura 4. Por lo tanto, se establece la hipótesis de que la variedad de las rosas determina el modo en que se articulan y se reclutan las comunidades microbianas en la rizósfera de la planta. Esto da paso a la elaboración de estudios que puedan probar esta hipótesis.

FIGURAS



Figura 1. Flujoograma del programa creado en R. Este resumen las etapas que el programa sigue al momento de procesar los datos.

A	B	C	D	E	F	G	H	I	J	K	
2	ROERh1	4001	11	426	1769	493	0	945	2291	21	2286
3	ROERh2	5698	45	325	1520	780	0	3255	5527	0	637
4	ROERh3	365	879	456	1108	669	5	89	1239	24	30
5	ROERh4	54	111	12447	1437	9256	2549	76	1106	19	8
6	ROESo1	767	125	435	5469	3399	0	232	2578	0	0
7	ROESo2	2371	1214	518	948	3961	34	412	952	0	0
8	ROESo3	894	313	496	2344	1936	0	221	1663	0	8
9	ROESo4	2532	1931	19341	543	4267	39	733	572	0	0
10	ROFRh1	90	2636	502	567	664	0	68	607	27	3
11	ROFRh2	4734	65	487	427	680	6560	2905	863	2598	10023
12	ROFRh3	140	246	992	250	370	19976	52	309	17162	0
13	ROFRh4	1881	38	556	3688	788	508	647	1593	543	7684
14	ROFSo1	13682	2399	238	158	1685	0	10346	239	555	0
15	ROFSo2	2142	5013	21	462	887	485	1198	506	3651	106
16	ROFSo4	8601	1081	235	445	881	0	7212	1618	120	27
17	RONRh1	194	0	122	2178	109	0	73	2834	1132	9
18	RONRh2	648	0	253	2415	119	162	375	2475	71	1651
19	RONRh3	525	0	238	6614	167	0	86	1011	662	673
20	RONRh4	1269	0	42	1741	25	0	452	127	0	280
21	RONSo1	409	0	117	572	0	0	114	238	17	10
22	RONSo3	404	350	144	2725	202	927	248	937	95	212

A	B	C	D	E	F	G	H	I	J	K	
1	ASV5556	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
2	ASV5557	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
3	ASV5558	Fungi(62);Fungi_unclassified(62);Fungi_unclassified(62);Fungi_unclassified(62);Fungi_unclassified(62);Fu									
4	ASV5559	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
5	ASV5560	Fungi(100);Basidiomycota(100);Agaricostilbomycetes(100);Agaricostilbales(100);Kondoaaceae(100);Kondoa(100);Kondoa_aeria(10									
6	ASV5561	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
7	ASV5562	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
8	ASV5563	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
9	ASV5564	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
10	ASV5565	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
11	ASV5566	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
12	ASV5567	Fungi(100);Ascomycota(100);Sordariomycetes(100);Myrmecridiales(95);Myrmecridiales_fam_Incertae_sedis(87);Atractospora(87)									
13	ASV5568	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
14	ASV5569	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
15	ASV5570	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
16	ASV5571	Fungi(91);Basidiomycota(73);Agaricomycetes(70);Agaricales(64);Agaricales_unclassified(64);Agaricales_unclassified(64);Agaricales									
17	ASV5572	unknown;unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_unclassified(0);unknown_uncl									
18	ASV5573	Fungi(100);Ascomycota(100);Sordariomycetes(97);Hypocreales(81);Hypocreales_unclassified(81);Hypocreales_unclassified(81);H									
19	ASV5574	Fungi(71);Fungi_unclassified(71);Fungi_unclassified(71);Fungi_unclassified(71);Fungi_unclassified(71);Fungi_unclassified(71);Fu									
20	ASV5575	Fungi(89);Ascomycota(82);Ascomycota_unclassified(82);Ascomycota_unclassified(82);Ascomycota_unclassified(82);Ascomycota_									
21	ASV5576	Fungi(100);Ascomycota(100);Leotiomycetes(97);Helotiales(97);Helotiales_unclassified(97);Helotiales_unclassified(97);Helotiales									
22	ASV5577	Fungi(100);Ascomycota(100);Sordariomycetes(100);Hypocreales(100);Cordycipitaceae(99);Beauveria(99);Beauveria_sp(99);									

A	B	C	D	
1	n_muestra	factor1	factor2	factor3
2	Muestra	Variedad	Fracción	Proveniencia
3	ROERh1	Explorer	Rizósfera	Rosa
4	ROERh2	Explorer	Rizósfera	Rosa
5	ROERh3	Explorer	Rizósfera	Rosa
6	ROERh4	Explorer	Rizósfera	Rosa
7	ROESo1	Explorer	Suelo	Rosa
8	ROESo2	Explorer	Suelo	Rosa
9	ROESo3	Explorer	Suelo	Rosa
10	ROESo4	Explorer	Suelo	Rosa
11	ROFRh1	Freedom	Rizósfera	Rosa
12	ROFRh2	Freedom	Rizósfera	Rosa
13	ROFRh3	Freedom	Rizósfera	Rosa
14	ROFRh4	Freedom	Rizósfera	Rosa
15	ROFSo1	Freedom	Suelo	Rosa
16	ROFSo2	Freedom	Suelo	Rosa
17	ROFSo4	Freedom	Suelo	Rosa
18	RONRh1	Natal Briar	Rizósfera	Rosa
19	RONRh2	Natal Briar	Rizósfera	Rosa
20	RONRh3	Natal Briar	Rizósfera	Rosa
21	RONRh4	Natal Briar	Rizósfera	Rosa
22	RONSo1	Natal Briar	Suelo	Rosa

Figura 2. Muestra de las tablas para importación. (A) tabla de conteos: se ven los nombres de las muestras en la primera columna y los conteos de cada ASV en las columnas posteriores. (B) tabla de taxonomía: muestra los nombres de ASVs en la primera columna y su taxonomía en la segunda columna. (C) metadata: se observan los nombres de las muestras en la primera columna y los diversos factores de la muestra en las columnas posteriores.

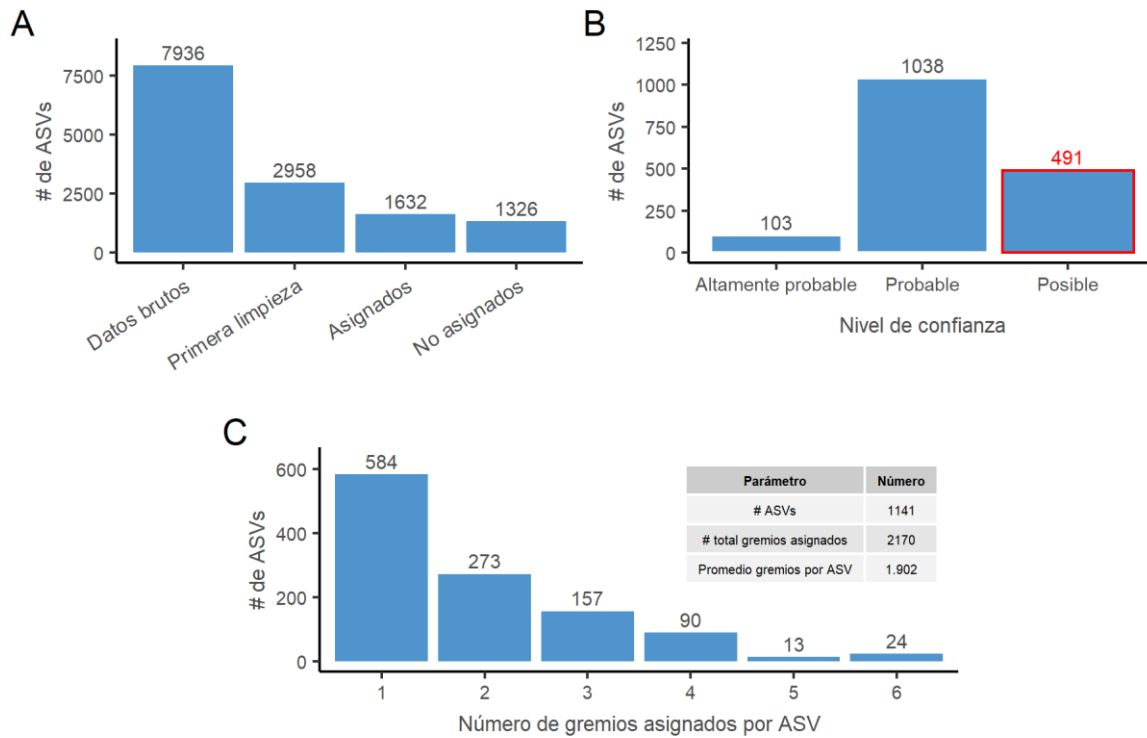


Figura 3. Resultados de las limpiezas y procesamiento. (A) se observan la cantidad de ASVs restantes luego de cada etapa. En el ejemplo se comenzó con 7636 como “Datos brutos”; luego de la primera limpieza sobraron 2968 de los cuales 1632 se asignaron en el análisis predictivo. (B) se muestra la calidad de asignación de los ASVs asignados, de los cuales 491 se van a eliminar como parte de la limpieza. (C) muestra el número de gremios asignados por ASV, producto de la redundancia donde un taxón puede tener asignado más de un gremio. Se observa que 584 ASVs recibieron una sola asignación de gremio y cada ASV recibió en promedio 1.902 asignaciones de gremio.

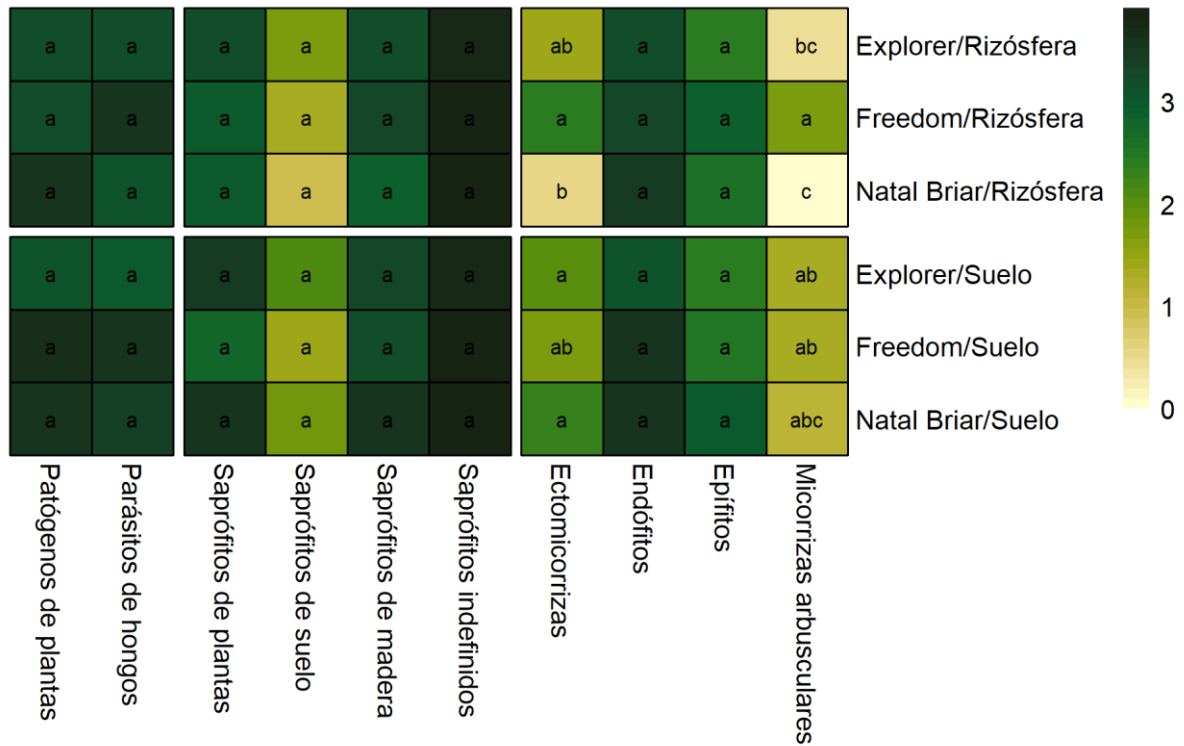


Figura 4. Abundancia de gremios en función de la combinación factores. En el heatmap se observa, con base a la intensidad de color, la abundancia relativa de cada gremio con relación a la combinación de los factores de interés. También se figuran las letras de significancia de la prueba de Tukey a partir de los ANOVAs realizados para cada gremio.

REFERENCIAS BIBLIOGRÁFICAS

- Abbas, M. F., Aziz-ud-Din, Rafique, K., Qadir, A., Qamar, M. I., Rafiq, M. y Gleason, M. L. (2017). First Report of Alternaria Black Spot of Rose Caused by *Alternaria alternata* in Pakistan. *Plant Disease*, *101*(9), 1676. doi:10.1094/pdis-04-17-0578-pdn
- Cabascango, W. (2008). Evaluación de sustratos para el enraizamiento de estaca de rosa (*rosa* sp.) del patrón natal brier en Otón – Pichincha [Tesis de grado, Universidad Politécnica Salesiana]. *Repositorio Institucional de la Universidad Politécnica Salesiana*. <https://dspace.ups.edu.ec/handle/123456789/6751>
- Cabrera, R. I. (2002). Rose yield, dry matter partitioning and nutrient status responses to rootstock selection. *Scientia Horticulturae*, *95*(1-2), 75–83. doi:10.1016/s0304-4238(02)00020-1
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. y Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods*, *13*(7), 581-583.
- Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C. y Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*. doi:10.1038/s41587-020-0548-6
- Furneaux, B., Song, Z. (2021). FUNGuildR: Look Up Guild Information for Fungi. *R package version 0.2.0.9000* [Software de computadora]. <https://github.com/brendanf/FUNGuildR>.
- Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, *15*(10), 579–590. doi:10.1038/nrmicro.2017.87
- Finkel, O. M., Salas-González, I., Castrillo, G., Spaepen, S., Law, T. F., Teixeira, P. J. P. L., Jones, C. J y Dangl, J. L. (2019). The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response. *PLOS Biology*, *17*(11), e3000534. doi:10.1371/journal.pbio.3000534
- Ghosh, A. y Shamsi, S. (2014). Fungal diseases of Rose plant in Bangladesh. *Journal of Bangladesh Academy of Sciences*, *38*(2), 225–233. doi:10.3329/jbas.v38i2.21347
- Grada, A. y Weinbrecht, K. (2013). Next-Generation Sequencing: Methodology and Application. *Journal of Investigative Dermatology*, *133*(8), 1–4. doi:10.1038/jid.2013.248
- Hanada, R. E., Pomella, A. W. V., Costa, H. S., Bezerra, J. L., Loguercio, L. L. y Pereira, J. O. (2010). Endophytic fungal diversity in *Theobroma cacao* (cacao) and *T. grandiflorum* (cupuaçu) trees and their potential for growth promotion and biocontrol of black-pod disease. *Fungal Biology*, *114*(11-12), 901–910. doi:10.1016/j.funbio.2010.08.006

- Hodkinson, B. P. y Grice, E. A. (2015). Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. *Advances in Wound Care*, 4(1), 50–58. doi:10.1089/wound.2014.0542
- Koch, E. B. A., Castaño-Meneses, G. y Delabie, J. H. C. (2019). El concepto de gremio: del feudalismo a la ecología de comunidades. *Acta Biológica Colombiana*, 24(2), 224–231. doi:10.15446/abc.v24n2.74726
- Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S. y Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, 20, 241–248. doi:10.1016/j.funeco.2015.06.006
- Mazur, S., Nadziakiewicz, M., Kurzawińska, H. y Nawrocki, J. (2019). Effectiveness of mycorrhizal fungi in the protection of juniper, rose, yew and highbush blueberry against *Alternaria alternata*. *Folia Horticulturae*, 31(1), 117-127.
- Otiende, M. A., Nyabundi, J. O. y Ngamau, K. (2016). Rose rootstocks position and auxins affect grafting take of 'Inca'. *Journal of Applied Horticulture*, 18(1).
- Reis-Filho, J. S. (2009). Next-generation sequencing. *Breast Cancer Research*, 11(S3). doi:10.1186/bcr2431
- Rivers, A. R., Weber, K. C., Gardner, T. G., Liu, S. y Armstrong, S. D. (2018). ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis. *F1000Research*, 7.
- Sanger, F. y Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing. *FEB*.
- Saleem, M., Hu, J. y Jousset, A. (2019). More Than the Sum of Its Parts: Microbiome Biodiversity as a Driver of Plant Growth and Soil Health. *Annual Review of Ecology, Evolution, and Systematics*, 50(1). doi:10.1146/annurev-ecolsys-110617-062605
- Seehann, G., Liese, W., Kess, B. y International Research Group on Wood Preservation. (1975). *List of fungi in soft-rot tests*. Princes Risborough, Eng: Princes Risborough Laboratory.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. O., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. y Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23), 7537-7541.
- Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nature Methods*, 5(1), 16–18. doi:10.1038/nmeth1156
- Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. y Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 418–426. doi:10.1016/j.tig.2014.07.001

- Wagg, C., Schlaeppli, K., Banerjee, S., Kuramae, E. E. y van der Heijden, M. G. A. (2019). Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications*, 10(1). doi:10.1038/s41467-019-12798-y
- Zhu, C., Miller, M., Lusskin, N., Mahlich, Y., Wang, Y., Zeng, Z. y Bromberg, Y. (2019). Fingerprinting cities: differentiating subway microbiome functionality. *Biology Direct*, 14(1). doi:10.1186/s13062-019-0252-y

ANEXOS

Anexo 1. Lista de gremios funcionales

Gremio principal	Gremio funcional
Patótrofos	Plant Pathogen, Animal Pathogen, Bryophyte Parasite, Clavicipitaceous Endophyte, Fungal Parasite, Lichen Parasite, Algal Parasite, Insect Parasite
Saprótrofos	Dung Saprotroph, Leaf Saprotroph, Plant Saprotroph, Soil Saprotroph, Wood Saprotroph, Litter Saprotroph, Undefined Saprotroph
Simbiótrofos	Ectomycorrhizal, Ericoid Mycorrhizal, Endophyte, Epiphyte, Lichenized, Arbuscular Mycorrhizal, Animal Endosymbiont, Endomycorrhizal

Anexo 2. Paquetes utilizados por el programa

Paquete	Referencia
devtools	Wickham, H., Hester, J., Chang, W. y Bryan, J. (2021). devtools: Tools to Make Developing R Packages Easier. <i>R package version 2.4.3</i> . https://CRAN.R-project.org/package=devtools
FUNGuildR	Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S. y Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. <i>Fungal Ecology</i> , 20, 241–248. doi:10.1016/j.funeco.2015.06.006
ggplot2	Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer - Verlag
BiocManager	Morgan, M. (2021). BiocManager: Access the Bioconductor Project Package Repository. <i>R package version 1.30.16</i> . https://CRAN.R-project.org/package=BiocManager
ggtree	Yu, G. (2020). Using ggtree to visualize data on tree-like structures. <i>Current Protocols in Bioinformatics</i> , 69:e96. doi: 10.1002/cpbi.96
ohchibi	Gonzalez, I. (2021). ohchibi: ohchibi. <i>R package version 0.0.0.9000</i>
data.table	Dowle, M. y Srinivasan, A. (2021). data.table: Extension of `data.frame`. <i>R package version 1.14.2</i> . https://CRAN.R-project.org/package=data.table
xcms	Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. y Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. <i>Analytical chemistry</i> , 78(3), 779-787.
dplyr	Wickham, H., François, R., Henry, L. y Müller, K. (2021). dplyr: A Grammar of Data Manipulation. <i>R package version 1.0.7</i> . https://CRAN.R-project.org/package=dplyr
dbplyr	Wickham, H., Girlich, M. y Ruiz, E. (2021). dbplyr: A 'dplyr' Back End for Databases. <i>R package version 2.1.1</i> . https://CRAN.R-project.org/package=dbplyr
tidyverse	Wickham et al., (2019). Welcome to the tidyverse. <i>Journal of Open Source Software</i> , 4(43), 1686. https://doi.org/10.21105/joss.01686
pals	Wright, K. (2021). pals: Color Palettes, Colormaps, and Tools to Evaluate Them. <i>R package version 1.7</i> . https://CRAN.R-project.org/package=pals
DSR	“First Last” (2021). DSR: Data sets for Data Science with R. <i>R package version 0.1</i> .
NLP	Hornik, K. (2020). NLP: Natural Language Processing Infrastructure. <i>R package version 0.2-1</i> . https://CRAN.R-project.org/package=NLP
ggpmisc	Aphalo, P. J. (2021). ggpmisc: Miscellaneous Extensions to 'ggplot2'. <i>R package version 0.4.4</i> . https://CRAN.R-project.org/package=ggpmisc
patchwork	Pedersen, T. L. (2020). patchwork: The Composer of Plots. <i>R package version 1.1.1</i> . https://CRAN.R-project.org/package=patchwork
openxlsx	Schauberger, P. y Walker, A. (2021). openxlsx: Read, Write and Edit xlsx Files. <i>R package version 4.2.4</i> . https://CRAN.R-project.org/package=openxlsx
janitor	Firke, S. (2021). janitor: Simple Tools for Examining and Cleaning Dirty Data. <i>R package version 2.1.0</i> . https://CRAN.R-project.org/package=janitor
grid	R Core Team (2021). <i>R: A language and environment for statistical computing</i> . R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ .
agricolae	de Mendiburu, F. (2021). agricolae: Statistical Procedures for Agricultural Research. <i>R package version 1.3-5</i> . https://CRAN.R-project.org/package=agricolae
gridExtra	Auguie, B. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. <i>R package version 2.3</i> . https://CRAN.R-project.org/package=gridExtra
multcompView	Graves, S., Piepho, H., Selzer, L. (2019). multcompView: Visualizations of Paired Comparisons. <i>R package version 0.1-8</i> . https://CRAN.R-project.org/package=multcompView

