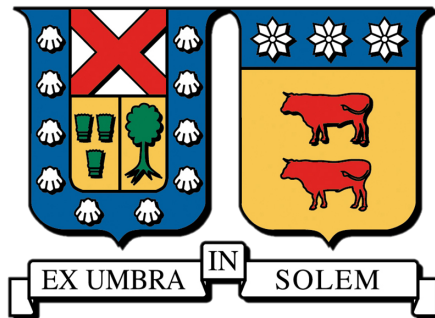


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



**IMPLEMENTACIÓN DE HERRAMIENTAS DE
ANÁLISIS FILOGENÉTICO PARA APLICACIÓN
INTEGRAL DE GENÓMICA DE HONGOS**

MARCELO ANDRE GARCÍA RAMOS

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO
CIVIL TELEMÁTICO**

PROFESOR GUÍA:

NICOLÁS TORRES

PROFESOR CORREFERENTE:

ESLI LOBAINA

OCTUBRE 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título; ☐ Tesis de Postgrado;

Título del trabajo: Implementación de herramientas de análisis filogenético para aplicación integral de genómica de hongos

Nombre del candidato(a): Marcelo Andre García Ramos

Carrera / Grado: Ingeniería Civil Telemática

Campus: Casa Central Valparaíso ; Departamento: Electrónica

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Nicolás Torres Rudloff, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

☒ El trabajo **NO** contiene información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

☐ 6 meses; ☐ 12 meses; ☐ 2 años; ☐ 3 años; ☐ 5 años; ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 20/10/2025; Firma: _____

Estudiante o Candidato(a):

Fecha: 20-10-2025

; Firma: _____

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Agradecimientos

Me gustaría agradecer primeramente a Dios, que me ha acompañado durante toda mi etapa universitaria y ha sido más que bueno y fiel. Solo Él sabe por todo lo que he pasado, pero gracias a Él y su misericordia puedo continuar cada mañana.

También a los miembros de la Iglesia Misionera Pentecostés, la cual me ha acogido como a una familia, me ha ayudado durante todo el proceso y me han hecho sentir amado.

Además a mis hermanas Maryland y Karenina, han sido de gran ayuda para mi, no solo económica, sino que han creído en mi desde pequeño, me han dado fuerzas y consejos invaluable.

Agradecer igualmente a mis amigos del colegio, sin ellos probablemente no hubiera continuado muchas veces, son fundamentales para mi salud mental y han estado para mi en mis peores momentos.

Finalmente, a cada persona que hizo que mi vida universitaria fuera menos pesada, y aportó con un granito de arena a mi vida, los que quieren verme triunfar, esto también es para ustedes.

Resumen

En campo de la investigación de hongos se hace cada vez más relevante debido a sus aplicaciones en la industria, la salud y para combatir el cambio climático. Es por esto, que existe una necesidad creciente de herramientas que permitan analizar los genomas de esto. Actualmente existen diversas herramientas de análisis, pero muchas de estas son de una estructura simple, sin interfaz, y no se integra con otras herramientas, restando tiempo de investigación para dedicar al aprendizaje de ellas, o incluso de manejo de computación avanzada como sería un lenguaje de programación o el uso de la consola. Por esto mismo, nuestra propuesta es una aplicación web integral y automatizada para el análisis genómico de hongos, llamada FungiTG. En base a esto, presentaremos el trabajo realizado en el módulo de análisis filogenético, las herramientas utilizadas PyANI y OrthoANI para la comparación genómica y los resultados obtenidos al probar las herramientas.

Keywords: Filogenético, herramienta, Python, ANI, genoma, MLSA, dendrograma, Matriz, identidad, NCBI, BLAST, FungiTG, integración, especies, comparación, ADN, secuencia.

Índice de figuras

4.1. Flujo general del módulo de análisis filogenético, mostrando la secuencia de etapas desde la descarga de datos hasta la obtención de resultados visuales.	12
4.2. Flujo del programa entradas y salidas.	15
4.3. Flujo del programa entradas y salidas.	16
5.1. Dendrograma obtenido de BlastN.	20
5.2. Matriz de identidad obtenido de BlastN.	21
5.3. Dendrograma de identidad obtenido de OrthoANI.	22
5.4. Dendrograma de identidad obtenido de OrthoANI.	23
5.5. Dendrograma obtenido de BlastN.	25
5.6. Matriz de identidad obtenido de BlastN.	26
5.7. Dendrograma de identidad obtenido de OrthoANI.	27
5.8. Dendrograma de identidad obtenido de OrthoANI.	28
5.9. Dendrograma obtenido de BlastN.	30
5.10. Matriz de identidad obtenido de BlastN.	31
5.11. Dendrograma de identidad obtenido de OrthoANI.	32
5.12. Dendrograma de identidad obtenido de OrthoANI.	33
6.1. Cluster obtenido a partir del mismo estudio.	35

Índice de cuadros

3.1. Comparación de herramientas para análisis genómico y filogenético . .	11
5.1. Lista de genomas completos utilizados de diversas familias	19
5.2. Lista de genomas completos para distintas especies.	24
5.3. Lista de genomas completos de distintas cepas de <i>Pseudomonas aeru-</i> <i>ginosa</i>	29

Índice general

1. Introducción	4
1.1. Objetivos	5
1.1.1. Objetivo General	5
1.1.2. Objetivos Específicos	5
1.2. Estructura	6
2. Marco Teórico	7
3. Estado del Arte	9
4. Desarrollo de la plataforma	12
4.1. Flujo de la plataforma	12
4.2. Análisis Filogenético	13
4.3. BlastN	14
4.3.1. Prerrequisitos	14
4.3.2. Procedimiento de implementación	14
4.4. OrthoANI	17
5. Resultados	18
5.1. Escenario 1: Cepas de distintas familias de hongos	19
5.1.1. BlastN	20
5.1.2. OrthoANI	22
5.2. Escenario 2: Cepas de misma familia, diversas especies	24

5.2.1. BlastN	25
5.2.2. OrthoANI	27
5.3. Escenario 3: Cepas de misma especie, diversas cepas	29
5.3.1. BlastN	30
5.3.2. OrthoANI	32
6. Conclusiones	34
7. Anexos	36
7.1. Anexo A: Códigos en Python y archivos utilizados	36
7.2. Anexo B: Repositorio del programa FungiTG	36

Capítulo 1

Introducción

En el campo de la biotecnología, el análisis genómico y filogenético de los hongos es esencial, aunque presenta una complejidad significativa debido al manejo de grandes volúmenes de datos y la necesidad de integrar múltiples herramientas bioinformáticas. Este análisis tiene un impacto profundo en diversos sectores, incluyendo la agricultura, la medicina y la biorremediación, puesto que permite entender mejor la diversidad, evolución y potencial de estos organismos.

Actualmente, existe una carencia de herramientas integrales y accesibles que permitan a los investigadores realizar análisis genómicos y filogenéticos de hongos de manera eficiente y simplificada. Los investigadores a menudo deben utilizar múltiples herramientas independientes, generalmente sin una interfaz visual, lo que puede ser laborioso, prolongado y requiere conocimientos sobre línea de comandos o inclusive lenguajes de programación para su ejecución. Esto dificulta el flujo de trabajo y puede limitar el avance en las investigaciones.

En esta propuesta, nos enfocaremos en generar una solución integral, incluyendo diversas técnicas de análisis filogenético, como también su interfaz para una mejor manipulación del usuario y finalmente un despliegue de resultados visuales para que los investigadores (principalmente del Centro de Biotecnología Daniel Alkalay Lowitt de la UTFSM) puedan enfocar sus esfuerzos a la investigación.

1.1. Objetivos

1.1.1. Objetivo General

Proponer una solución integral que facilite la carga, análisis y visualización de datos genómicos, utilizando tecnologías modernas y herramientas actualizadas para el análisis filogenético.

1.1.2. Objetivos Específicos

- **01:** Diseñar un sistema que permita la carga de archivos de secuenciación en formatos estándar, garantizando su calidad e integridad.
- **02:** Implementar herramientas de análisis filogenético basadas en algoritmos existentes que aseguren precisión y eficiencia en los resultados a partir de los archivos cargados.
- **03:** Desarrollar representaciones visuales de los análisis obtenidos, principalmente utilizando dendrogramas y matrices de identidad, que sirvan de apoyo para la interpretación de los datos.
- **04:** Integrar los componentes desarrollados en una interfaz web amigable, que permita un acceso intuitivo y eficiente para los usuarios.

1.2. Estructura

Este documento se estructura de la siguiente forma: En el Capítulo 2 se revisará el desafío propuesto, la importancia que tiene y también las hipótesis y supuestos que conlleva el desarrollo de nuestro proyecto.

En el Capítulo 3, se presentarán todas las herramientas similares que ya existen publicadas, también el contexto de sus soluciones y como resuelven cada uno de los desafíos que ellos se proponen y la utilidad de estos en nuestro desarrollo.

El en Capítulo 4 se abordará el desarrollo de la plataforma, con especial detalle en el análisis filogenético, presentaremos las herramientas, arquitectura, flujo de trabajo y también el proceso de creación de esta sección.

En el Capítulo 5 se verán los resultados del proceso trabajado, cómo interpretarlos y cual es la función que cumplirán para el usuario.

Finalmente en el Capítulo 6, se discutirá el proceso previo de desarrollo, las proyecciones y también las conclusiones sobre los objetivos propuestos en este escrito.

Capítulo 2

Marco Teórico

La genómica fúngica ha emergido como un campo de estudio esencial en la biología molecular y la bioinformática, debido a la importancia de los hongos en los ecosistemas, la industria y la salud humana. Sin embargo, el análisis genómico de estos organismos presenta desafíos significativos, como la complejidad de sus genomas y la gran diversidad genética entre especies. En este contexto, el proyecto FungiTG se concibe como una plataforma integral que facilita el análisis genómico de hongos a través de herramientas especializadas y una interfaz web accesible.

El análisis filogenético ha sido un pilar en el estudio de los hongos, ya que permite comprender su evolución, diversidad y relaciones dentro del reino Fungi y con otros organismos. Cabe destacar el rol fundamental de los hongos en procesos ecológicos como la descomposición y la simbiosis, que los convierte en actores clave en el equilibrio de los ecosistemas. Estudiar la variedad de especies existentes resulta esencial, pues al conocer sus comportamientos y relaciones es posible generar productos útiles para diversas áreas de la sociedad, con especial énfasis en desafíos actuales como el cambio climático, la producción agrícola y el desarrollo de medicamentos. La filogenética ha permitido identificar linajes antiguos, clasificar nuevas especies y explorar cómo los hongos se han adaptado a diferentes hábitats y nichos ecológicos.

Este análisis puede complementarse con herramientas tecnológicas modernas que

aumenten la rapidez, la precisión y la calidad en la interpretación de los resultados. Actualmente, el uso de estas herramientas permite a los investigadores desarrollar estudios y publicaciones de manera constante y eficaz.

El objetivo principal de FungiTG es proporcionar a la comunidad científica una solución unificada que abarque desde la anotación genómica hasta el análisis filogenético. Esta solución debe contemplar la carga de archivos desde las principales fuentes de datos internacionales, como GenBank de NCBI, y el uso de técnicas de análisis como Average Nucleotide Identity (ANI), que permite medir el grado de similitud genómica entre dos especies. Este enfoque puede aplicarse al genoma completo o mediante Multilocus Sequence Analysis (MLSA), técnica que fracciona el genoma para realizar comparaciones específicas. Finalmente, la visualización de datos debe facilitarse a través de representaciones gráficas, como árboles filogenéticos que reflejen la relación entre especies, o matrices que representen el grado de similitud, lo que contribuye a un entendimiento más profundo de los resultados.

Capítulo 3

Estado del Arte

El primer trabajo destacado es el estudio de Kubicek en 2011 [1], que analiza la evolución de los genomas de las especies más comunes de *Trichoderma*. Este estudio reveló cómo estos hongos, utilizados en agricultura y biotecnología, se diversificaron hace aproximadamente 66 millones de años, mostrando la evolución de capacidades de microparasitismo y oportunismo ambiental. La investigación subraya la importancia de los análisis genómicos comparativos para comprender la biología y la evolución de organismos cruciales en la agricultura y biotecnología.

En un trabajo complementario, Hoff y Stanke en 2011 [2] presentan **AUGUSTUS**, una herramienta para la predicción de genes en genomas individuales, que es fundamental para el análisis filogenético detallado. Esta herramienta permite una anotación precisa y confiable de secuencias genómicas, facilitando la construcción de árboles filogenéticos y el estudio de relaciones evolutivas. *FungiTG* incorpora herramientas como **Braker3**, que se basa en AUGUSTUS, para mejorar la anotación estructural de los genomas de hongos.

En el estudio hecho por Richter y Rosello-Mora en 2009 [3] se propone el uso de ANI para la comparación de genomas, como método de reemplazo de DNA-DNA Hybridization (DDH), el cual ha sido utilizado durante 5 décadas. Los resultados del estudio son prometedores debido a que ANI permite un umbral de precisión del 95 %

aproximadamente. Adicional a eso, se estudió una aplicación llamada JSpecies que permite realizar estos estudios con una interfaz amigable. Y finalmente, también se comenta que no es necesario usar ANI con el genoma completo sino basta con usar al menos el 20 % del genoma completo para hacer la comparación.

Siguiendo la línea del estudio anterior, un grupo de investigadores desarrollaron PyANI como herramienta libre para su uso en Python. [4] la cual utilizaron para realizar la secuenciación genómica de Enterobacteriaceae de pudrición blanda (SRE) los cuales son patógenos que afectan a los cultivos. Esto con el objetivo de revisar si existen inconsistencias taxonómicas y la temprana detección de estos agentes en los alimentos, utilizando PyANI como principal comparador.

Cuadro 3.1: Comparación de herramientas para análisis genómico y filogenético

Herramienta	Objetivo / Aplicación	Limitaciones
AUGUSTUS [2]	Predicción de genes en genomas individuales; permite anotación estructural precisa y construcción de árboles filogenéticos.	Requiere conocimientos avanzados en bioinformática; orientado a usuarios con experiencia técnica.
Braker3	Basado en AUGUSTUS, optimiza la anotación estructural de genomas de hongos dentro de <i>FungiTG</i> .	Necesita integrarse con otras herramientas para análisis filogenético completo.
ANI — JSpecies [3]	Comparación genómica como alternativa al método DDH; permite medir similitud genómica incluso con fracciones del genoma ($\geq 20\%$).	Umbral de precisión depende de la calidad del genoma; JSpecies limitado a ciertos formatos.
PyANI [4]	Implementación en Python para calcular ANI; usada en estudios de <i>Enterobacteriaceae</i> para detección temprana de patógenos.	Precisión afectada en genomas completos poco alineados; requiere conocimientos en Python.
OrthoANI	Comparación de genes ortólogos; genera matrices y dendrogramas con mayor fidelidad que ANI tradicional.	Computacionalmente más exigente; depende de anotaciones previas de calidad.

Capítulo 4

Desarrollo de la plataforma

4.1. Flujo de la plataforma

Para la construcción de la plataforma, se definirá el flujo de la información desde la captura hasta la entrega. La entrada al programa de análisis filogenético serán los datos genómicos, principalmente en archivo FASTA o FNA. Posteriormente, ingresarán al programa de análisis filogenético el cual explicaremos más adelante a detalle.

Y como salida queremos obtener un resultado visual en forma de gráficos, los cuales tendrán dos formatos principales, el primero es un dendrograma para analizar las similitudes filogenéticas y se pueda establecer el orden taxonómico de las especies estudiadas. El segundo tipo de gráfico, será un diagrama matricial que representará el porcentaje de similitud entre las especies participantes, lo cual será de gran ayuda para entender las relaciones de distintas especies, o incluso entre la misma especie.

El flujo está presentado en la figura 4.1.



Figura 4.1: Flujo general del módulo de análisis filogenético, mostrando la secuencia de etapas desde la descarga de datos hasta la obtención de resultados visuales.

Para el desarrollo de las herramientas, se escogieron dos existentes para el análisis filogenético: BlastN, específicamente la librería y función de ANIb (ANI integrado con BLAST). Y OrthoANI.

Para la construcción de ambas presentadas anteriormente, se escogió Python como lenguaje de programación, la distribución de Anaconda para crear el ambiente de trabajo, y específicamente la IDE de jupyter que nos permite trabajar en distintos bloques, separando el proceso de las herramientas y generando módulos del proceso. Adicionalmente, se utilizó el SO Linux, específicamente Ubuntu como distribución para la correcta integración del trabajo.

4.2. Análisis Filogenético

Como se mencionó anteriormente se utilizarán librerías referidas a ANI (Average Nucleotide Identity). El cual se define como una métrica que mide la similitud promedio a nivel de secuencia de ADN entre dos genomas completos.

Esta medida se basa en la comparación par a par de regiones homólogas (u ortólogas) compartidas entre ambos genomas y expresa el porcentaje promedio de identidad nucleotídica observado en dichas regiones alineadas. A diferencia de la identidad puntual de genes individuales, el ANI ofrece una aproximación global al valor de similitud entre genes y es el estándar para la delimitación de especies. Obedece la siguiente ecuación:

$$ANI = \frac{\sum_{i=1}^n (Len_i \times ID_i)}{\sum_{i=1}^n Len_i} \quad (4.1)$$

donde:

- Len_i es la longitud del fragmento alineado i (en pares de bases),
- ID_i es la identidad en términos de nucleótidos del fragmento i (en fracción decimal),

- n es el número total de alineamientos válidos.

Los valores de ANI y los niveles de relación taxonómica están estrechamente correlacionados. De manera general, valores superiores al 95–96 % indican que los genomas pertenecen a la misma especie, mientras que valores entre 85–94 % corresponden a organismos del mismo género, y valores inferiores al 80 % reflejan relaciones más distantes, a nivel de familia o superior.

4.3. BlastN

4.3.1. Prerrequisitos

Para la herramienta de análisis filogenético que utiliza ANIBLAST (En específico BlastN), debemos tener instalado previamente en nuestro entorno el paquete de herramientas BLAST+ (Basic Local Alignment Search Tool) disponible desde la plataforma de NCBI, esto nos permitirá hacer la comparación de secuencias entre dos archivos diferentes.

4.3.2. Procedimiento de implementación

El primer paso de implementación es utilizar la plataforma FungiGT para descargar los archivos genómicos, se observa en la imagen siguiente la interfaz de descarga y adquisición de datos, donde el usuario puede escribir el código NCBI para descargar directamente los archivos. Adicionalmente, se pueden descargar múltiples secuencias siempre y cuando existan dentro de la plataforma NCBI utilizando separacion entre comas (sin espacio). De manera interna (a nivel de código) se utilizará ANIb (ANI BLAST) para generar una base de datos de los archivos FNA. En el desarrollo del código, al tener almacenados de forma local los archivos con extensión .FASTA - .FNA, tomamos todos los que existen almacenados en la misma carpeta, idealmente que hayan pasado por un proceso de anotación y calidad en la parte previa de la aplicación

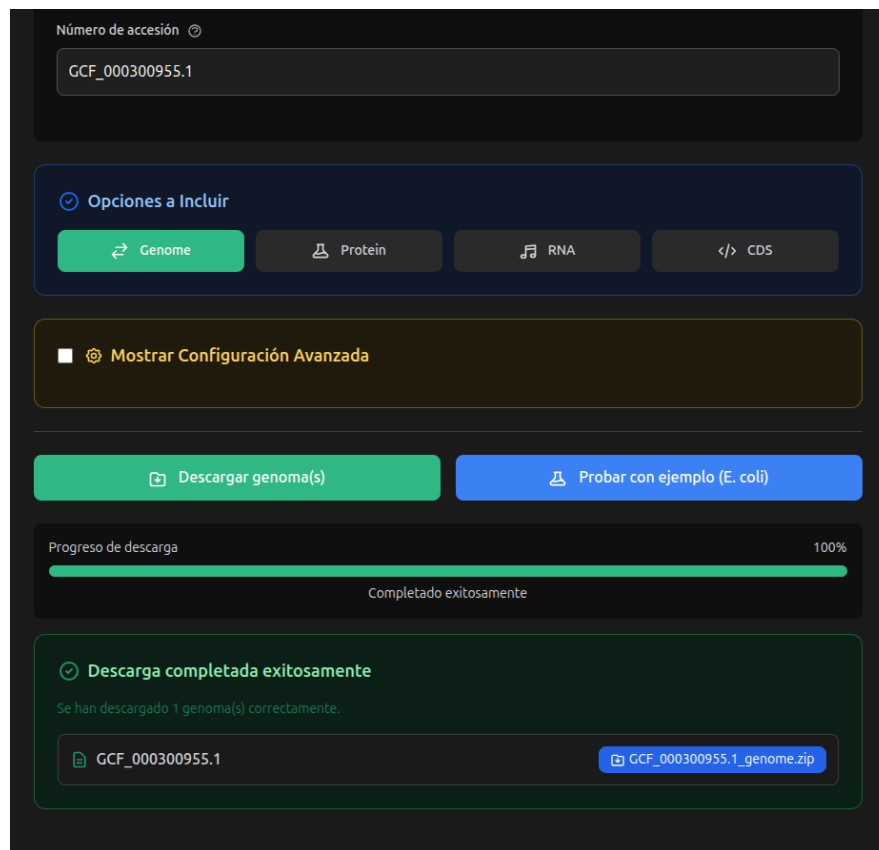


Figura 4.2: Flujo del programa entradas y salidas.

FungiTG, los cuales al procesarse se generará una base de datos para cada uno de ellos. Esta base de datos generará archivos con las siguientes extensiones: .ndb, .nhr, .nin, .njs, .not, .nsq, .ntf, .nto, las cuales serán un archivo por cada especie (cada uno de los archivos .FASTA). Lo siguiente, será utilizar BLASTN para realizar la comparación entre los archivos de la base de datos y los archivos FASTA originales. Esta comparación también incluye comparaciones entre la base de datos y el archivo de la misma especie para verificar la precisión de la herramienta. Estas comparaciones serán almacenadas en archivos blast_tab, las que obtendrán la identidad promedio de los genomas comparados. Si no hay identidad promedio entre ambas especies contrastadas, el archivo que se genera es vacío interpretándose un valor de ANI de cero.

Finalmente, al obtener todos los archivos con los resultados de comparación se pueden obtener tres salidas posibles, la primera es un dendrograma que permite ver las dis-

tancias (también referidas como las diferencias genéticas) entre las especies revisadas, útil para situaciones en donde queremos obtener una categorización de las especies, en especial para especies desconocidas familiarmente. La segunda salida es una matriz cuadrada que muestra el valor de identidad promedio entre las especies analizadas, Estas van desde el cero a cien dependiendo de la coincidencia genética entre las especies. Y finalmente como tercera salida es la unión entre el dendrograma y la matriz permitiendo revisar ambos gráficos en un mismo lugar.

Todas estas salidas han sido desplegadas utilizando las librerías matplotlib, scipy, pandas y seaborn, librerías de python que permiten el manejo e impresión de los datos.

En la plataforma, en la sección de análisis podemos realizar el procedimiento siguiendo la figura siguiente: En donde se permite la carga de los archivos previamente



Figura 4.3: Flujo del programa entradas y salidas.

bajados en la plataforma y guardar los archivos en las carpetas para facilitar su uso. Como también utilizarlos de manera multiple para la comparación. Para los resultados de este informe, no se utilizó la interfaz para agilizar los procesos de resultados y obtener imágenes de mejor resolución.

4.4. OrthoANI

Para el análisis filogenético que emplea OrthoANI, necesitamos la librería de BLAST+ previamente instalada igual que en el método anterior para realizar la comparación de secuencias. El primer paso es igual al anterior, pero en esta herramienta necesitamos la lectura de los archivos, en la cual para efectos prácticos ocuparemos la librería `subprocess`, para utilizar OrthoANI previamente instalado en el entorno a través de instrucciones por comando. Utilizamos los archivos previamente almacenados a través de la interfaz de FungiTG, los cuales como mencionamos deben estar previamente anotados y haber pasado por el proceso de calidad. Estos archivos serán inmediatamente llamados por OrthoANI para el proceso de diferenciación genética. OrthoANI es un proceso de comparación el cual se encarga de capturar los genes ortólogos, los cuales son aquellos segmentos de genes que son de especies diferentes, pero que tienen un ancestro en común, esto aplica principalmente a fragmentos funcionales (como por ejemplo funciones de metabolismo). Estas comparaciones de los genes ortólogos se guardarán en un archivo con extensión `.csv` el cual tendrá tres columnas, las primeras dos con los genes a comparar y la últimas con el valor de ANI calculado.

El paso final es la impresión de los resultados los cuales tienen dos posibles salidas. La primera es la matriz de ANI con el valor promedio de identidad al igual que en la herramienta anterior y también el árbol filogenético o dendrograma respectivo con las distancias entre especies (explicitado en la sección de resultados). Cabe destacar que esta es otra herramienta disponible en la plataforma FungiTG

Capítulo 5

Resultados

Para asegurarnos del buen funcionamiento del programa, postulamos tres escenarios distintos:

- El primero de 10 cepas que provengan de familias de hongos diferentes. Naturalmente se espera una baja similitud al compararlas entre ellas (entre 60 % – 80 %).
- El segundo de 10 cepas de la misma familia, pero de especies diferentes. Se proyecta una subida en torno a su similitud.
- El tercero 10 cepas distintas de la misma especie y familia. Se estima una similitud extremadamente alta

Estos escenarios son principalmente para detectar anomalías en la ejecución y en el funcionamiento de las herramientas presentadas, como también para mostrar un flujo de investigación para el usuario, en el caso de poseer datos de una cepa desconocida a nivel familias y de especie, será de gran ayuda que los tres elementos anteriores funcionen como es debido para detectar rápidamente la inscripción de la cepa donde corresponda.

Para todos los datos y campos de datos, se utilizó la aplicación FungiTG para la descarga y en el caso de ser necesario, la anotación correspondiente.

5.1. Escenario 1: Cepas de distintas familias de hongos

Cuadro 5.1: Lista de genomas completos utilizados de diversas familias

ID genoma NCBI	Nombre especie
GCF_000300955.1	<i>Agaricus bisporus</i>
GCF_000002855.3	<i>Aspergillus niger</i>
GCF_000142945.1	<i>Claviceps purpurea</i>
GCF_000816465.1	<i>Alternaria alternata</i>
GCF_000271645.1	<i>Trametes versicolor</i>
GCA_905123625.1	<i>Cortinarius glaucopus</i>
GCF_000146045.2	<i>Saccharomyces cerevisiae</i>
GCF_000182925.2	<i>Mucor circinelloides</i>
GCF_000146945.2	<i>Sclerotinia sclerotiorum</i>
GCF_000149925.1	<i>Puccinia graminis</i> f. sp. <i>tritici</i>

5.1.1. BlastN

Utilizando BlastN, se obtuvieron los siguientes resultados a partir de las salidas descritas en el Capítulo 4:

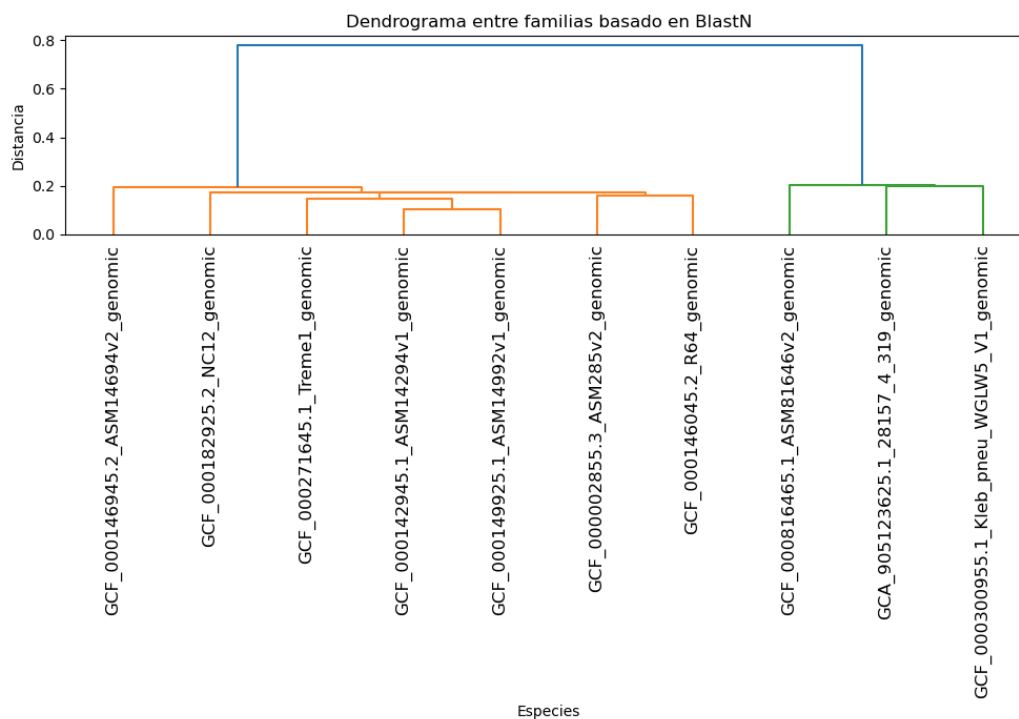


Figura 5.1: Dendrograma obtenido de BlastN.

Con respecto al gráfico anterior (figura 5.1) se puede apreciar las distancias entre las especies, se aprecia que la distancia máxima es de 0.8 entre dos especies, lo cual indica que están muy alejados en la evolución, mientras que en la ramificaciones siguientes es del orden del 0.2 y descendiendo. Además de la presencia de diversos grupos que comparten una línea evolutiva según el análisis.

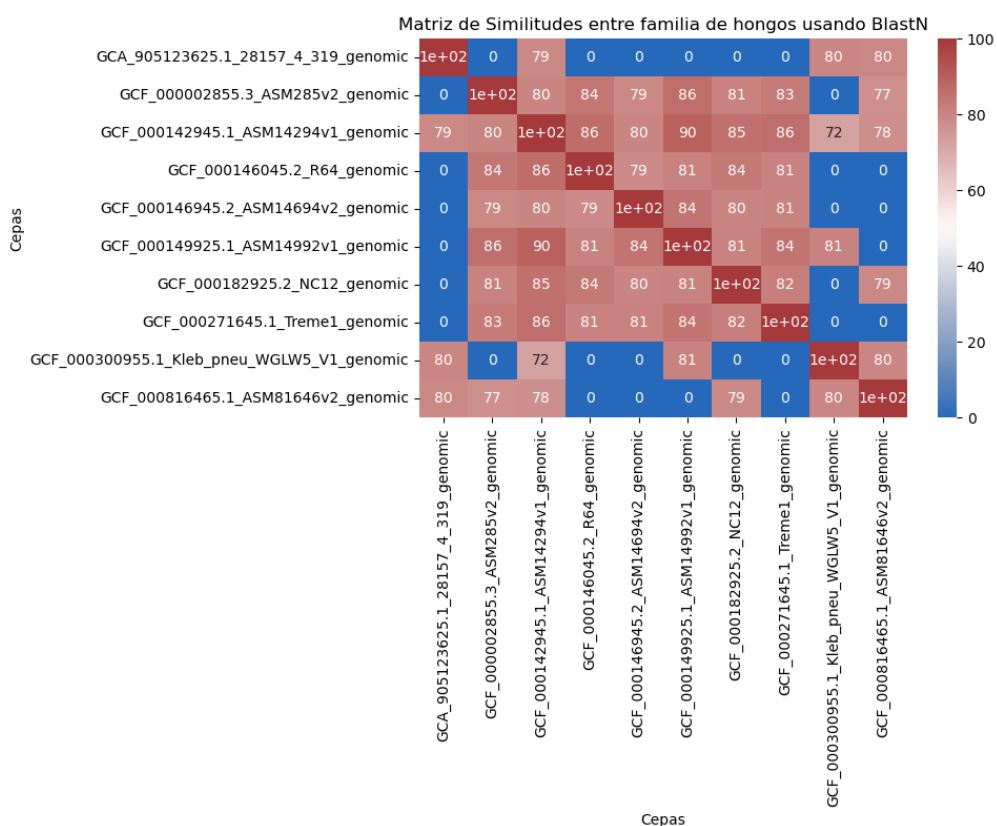


Figura 5.2: Matriz de identidad obtenido de BlastN.

Como se puede apreciar en la matriz (figura 5.2), Hay especies que no poseen una similitud alguna entre las familias, pero eso es causado por la herramienta ya que al pertenecer al mismo reino se postula una semejanza del 60 % aproximadamente o superior, lo cual indica un error de lectura por parte de la base de datos (anotaciones diferentes, o cepas muy distintas en los archivos) aún así la gran mayoría conserva una semejanza en su mayoría superiores a 80 % pero menor al 90 %

5.1.2. OrthoANI

Para OrthoANI se obtuvieron los siguientes resultados:

Primeramente, el dendrograma en la figura 5.3 el cual difiere del método anterior, ya que la distancia máxima corresponde a 0.4 arrojando que tienen similitudes evolutivas entre especies pero siguen siendo de distinto orden filogenético en este caso hay una distancia superior entre las especies más cercanas, aseverando que son de distintas familias sin llegar a cruzarse como especies.

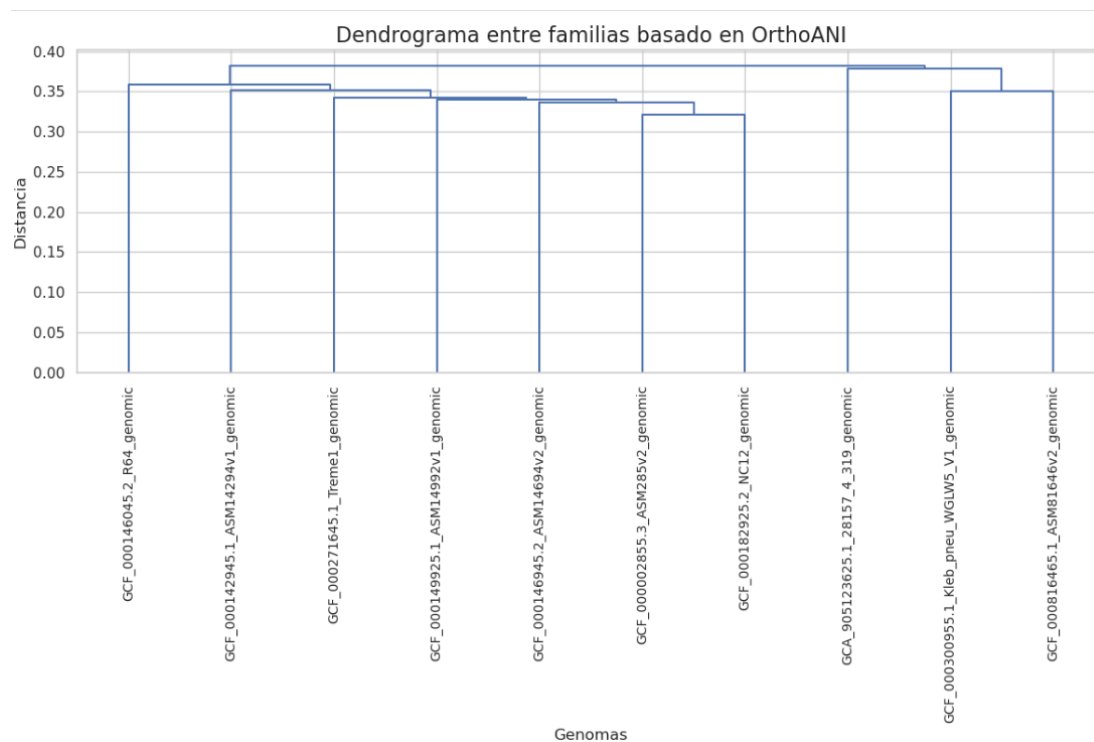


Figura 5.3: Dendrograma de identidad obtenido de OrthoANI.

Y como segunda salida, se obtiene la matriz de identidad utilizando los genes ortólogos, como se puede apreciar si mantiene uniformidad al comparar el genoma consigo mismo, lo cual podemos identificar mayor precisión, además de incluir valores ligeramente más bajos que la herramienta de BlastN. Se puede inferir que al utilizar cadenas del genoma de forma específica para la comparación, hay una comparación más fiel a la realidad. Ya que a pesar de ser valores bajos, conserva en su totalidad una semejanza superior al 59 %

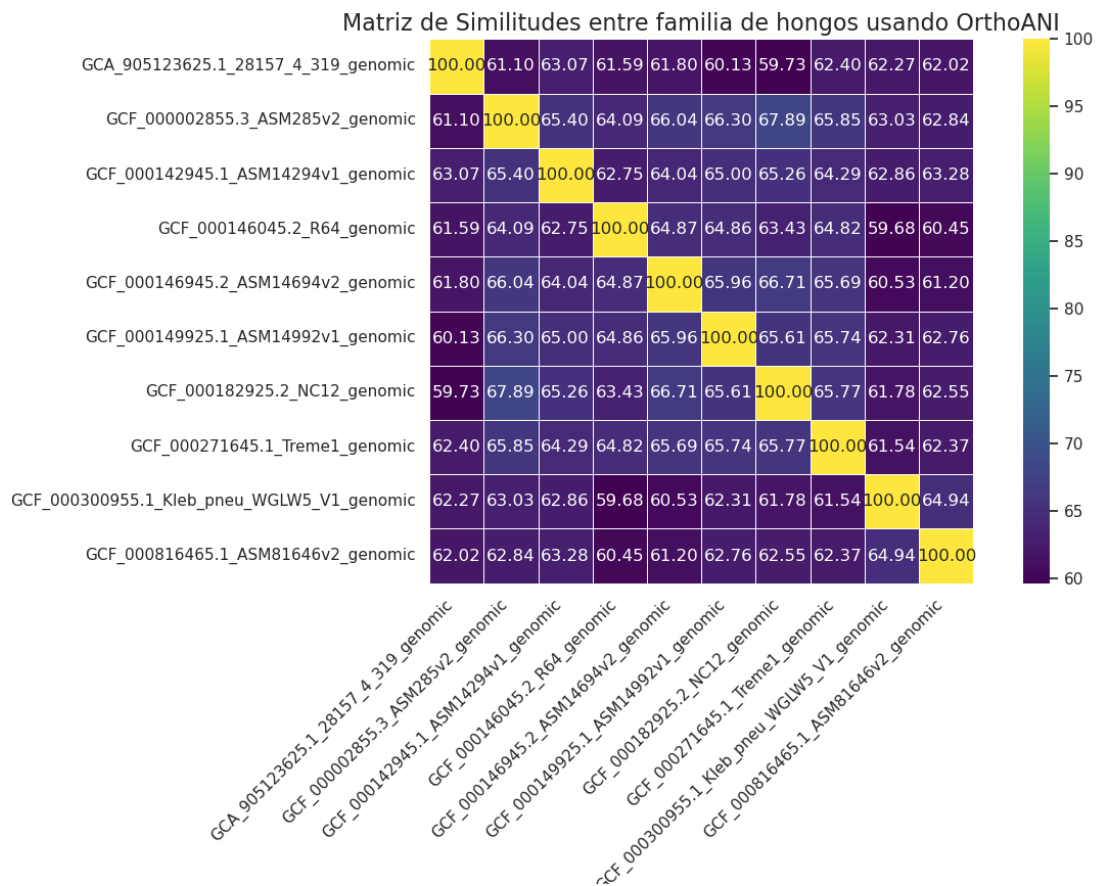


Figura 5.4: Dendrograma de identidad obtenido de OrthoANI.

5.2. Escenario 2: Cepas de misma familia, diversas especies

Cuadro 5.2: Lista de genomas completos para distintas especies.

ID genoma NCBI	Nombre especie
GCF_000006765.1	<i>Pseudomonas aeruginosa</i>
GCF_000009225.2	<i>Pseudomonas fluorescens</i>
GCF_000005845.2	<i>Pseudomonas putida</i>
GCF_000007805.1	<i>Pseudomonas syringae</i> pv. tomato
GCF_000011365.1	<i>Pseudomonas synxantha</i>
GCF_001439815.1	<i>Pseudomonas orientalis</i>
GCF_009659625.1	<i>Pseudomonas haemolytica</i>
GCF_008692035.1	<i>Pseudomonas marginalis</i>
GCF_001902145.1	<i>Pseudomonas fluorescens</i>
GCF_000421165.1	<i>Pseudomonas alcaligenes</i>

5.2.1. BlastN

Utilizando BlastN, se obtuvieron los siguientes resultados a partir de las salidas descritas en el Capítulo 4:

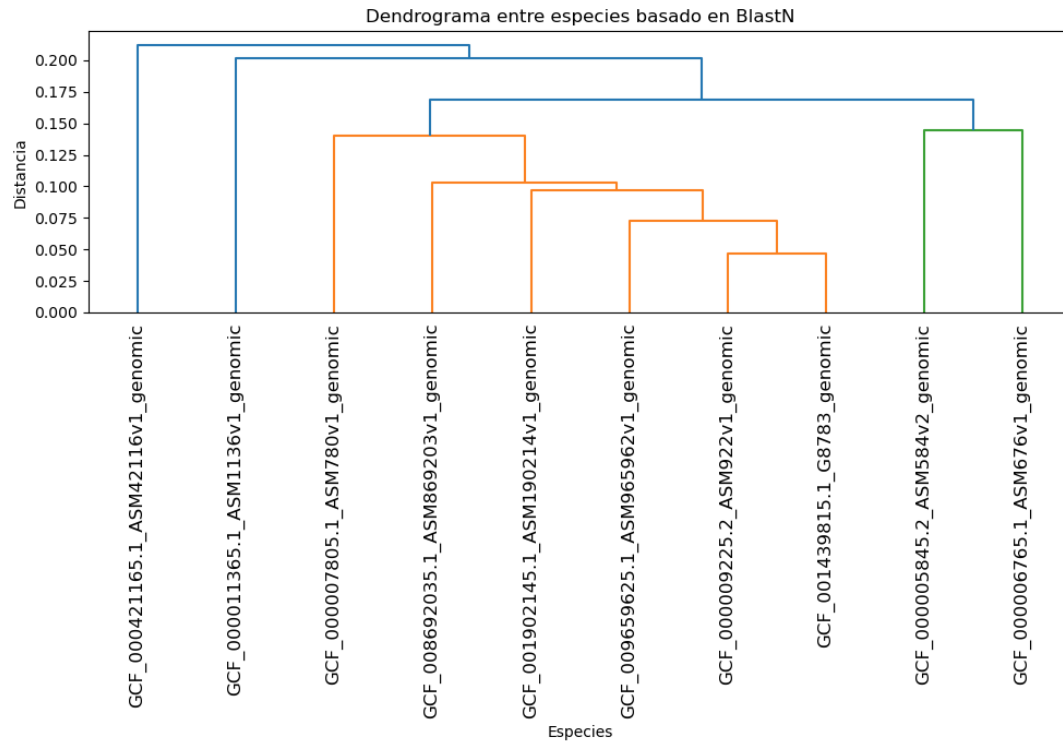


Figura 5.5: Dendrograma obtenido de BlastN.

Con respecto al gráfico anterior (figura 5.5) se puede apreciar las distancias entre las especies, podemos notar que en el análisis las distancias en referencia con el escenario anterior son mucho más cortas, y la asociación grupal es más alta, mostrando tres grupos evolutivos distintos. con distancia máxima 0.2

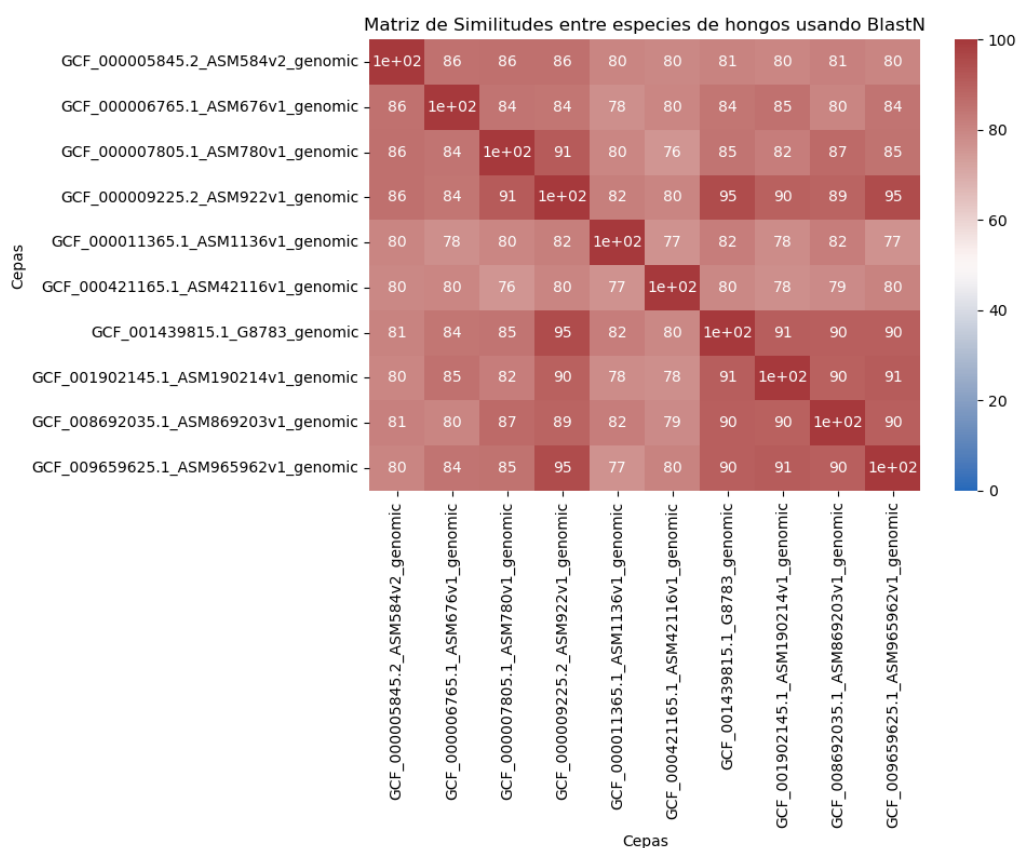


Figura 5.6: Matriz de identidad obtenido de BlastN.

Como se puede apreciar en la matriz (figura 5.6), al comparar entre especies tenemos valores mucho más altos que lo anterior, mostrando valores superiores al 75 % y en la inmensa mayoría de los casos sobre 80 %. Demostrando que entre especies si podemos categorizarlas y la herramienta funciona de manera precisa al generar comparaciones entre la misma familia

5.2.2. OrthoANI

Para OrthoANI se obtuvieron los siguientes resultados:

Primeramente, el dendrograma el cual difiere del método anterior ya que se identificaron solo dos grupos evolutivos, casi equitativamente, también tienen valores de distancia más alto, aún así se mantienen dentro de los márgenes de pertenecer a la misma familia (dentro del 0.4 de distancia) por lo cual es una aplicación correcta.

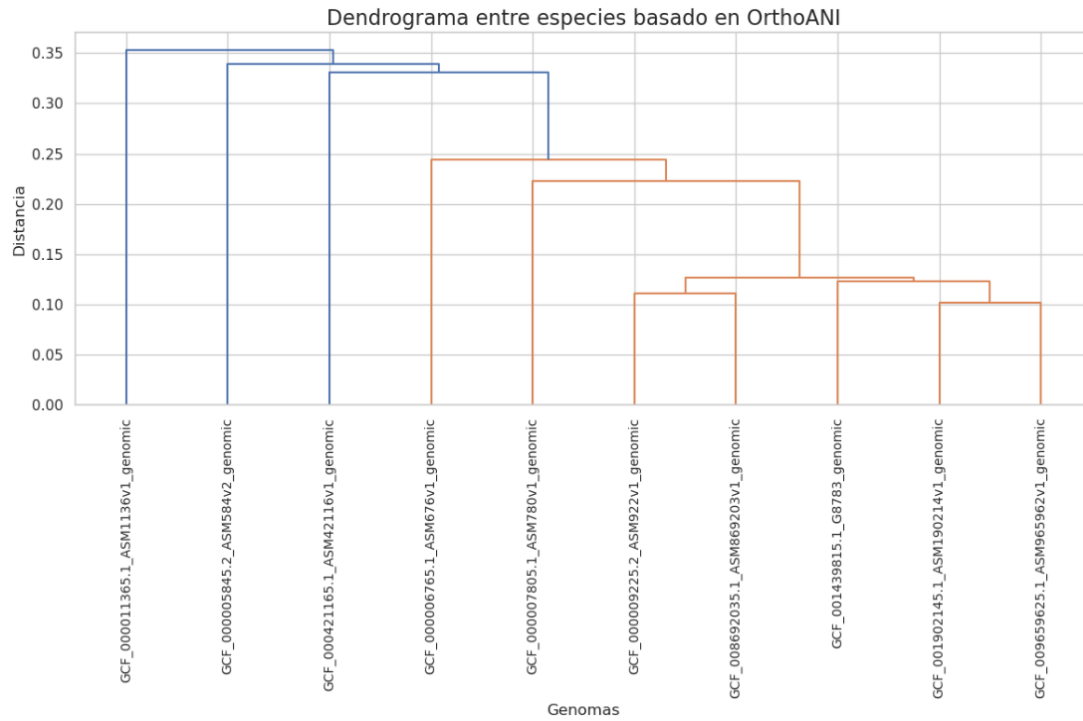


Figura 5.7: Dendrograma de identidad obtenido de OrthoANI.

Y como segunda salida, se obtiene la matriz de identidad utilizando los genes ortólogos, como se puede apreciar ha subido considerablemente el valor en muchos casos, también se nota la diferencia evolutiva en sus valores de semejanza. y el valor promedio es de 75 %. Asumiendo que es un análisis preciso se mantiene dentro del margen de ser de la misma familia, más valores bajos entre ellos demuestran que no son del mismo género ni comparten ancestro evolutivo.

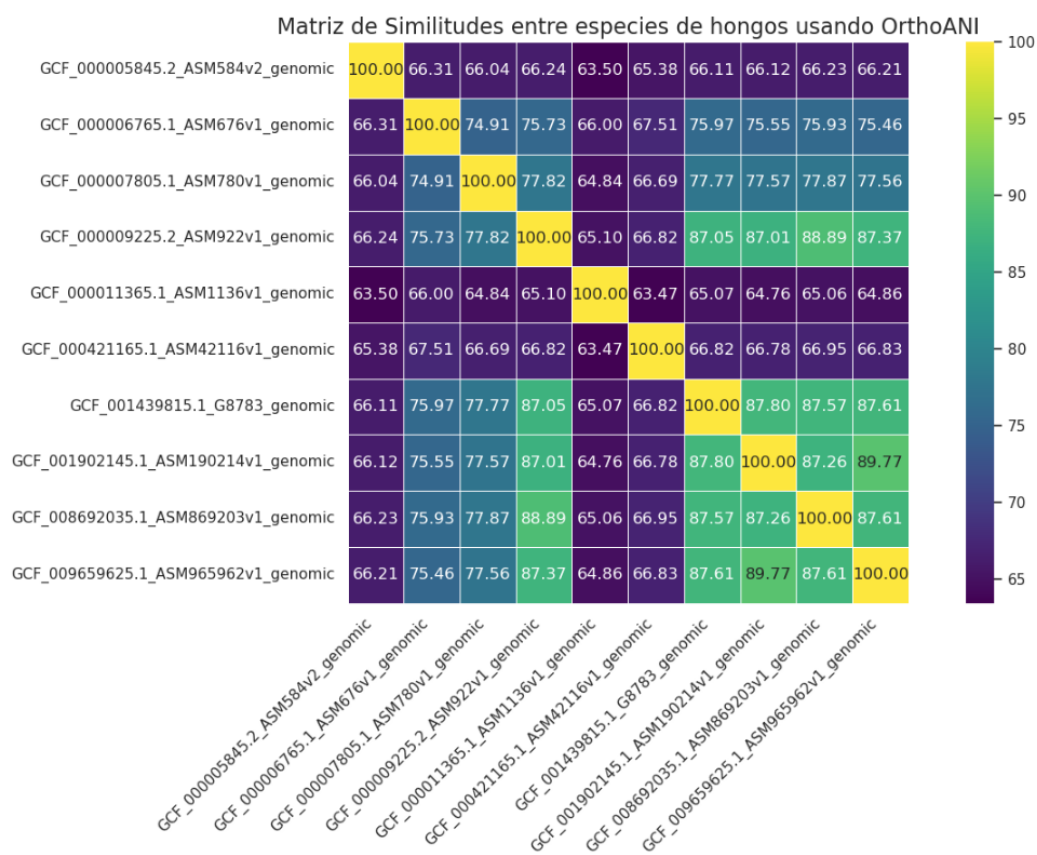


Figura 5.8: Dendrograma de identidad obtenido de OrthoANI.

5.3. Escenario 3: Cepas de misma especie, diversas cepas

Cuadro 5.3: Lista de genomas completos de distintas cepas de *Pseudomonas aeruginosa*.

ID genoma NCBI	Nombre de cepa
GCF_000006765.1	Pseudomonas aeruginosa PAO1
GCF_000014625.1	Pseudomonas aeruginosa UCBPP-PA14
GCF_000017205.1	Pseudomonas aeruginosa PA7
GCF_000026645.1	Pseudomonas aeruginosa LESB58
GCF_000271365.1	Pseudomonas aeruginosa DK2
GCF_000226155.1	Pseudomonas aeruginosa M18
GCF_001618925.1	Pseudomonas aeruginosa ATCC 27853
GCF_000284555.1	Pseudomonas aeruginosa NCGM2.S1
GCF_002968955.1	Pseudomonas aeruginosa strain PA34
GCF_001792835.1	Pseudomonas aeruginosa VRFPA04

5.3.1. BlastN

Utilizando BlastN, se obtuvieron los siguientes resultados a partir de las salidas descritas en el Capítulo 4:

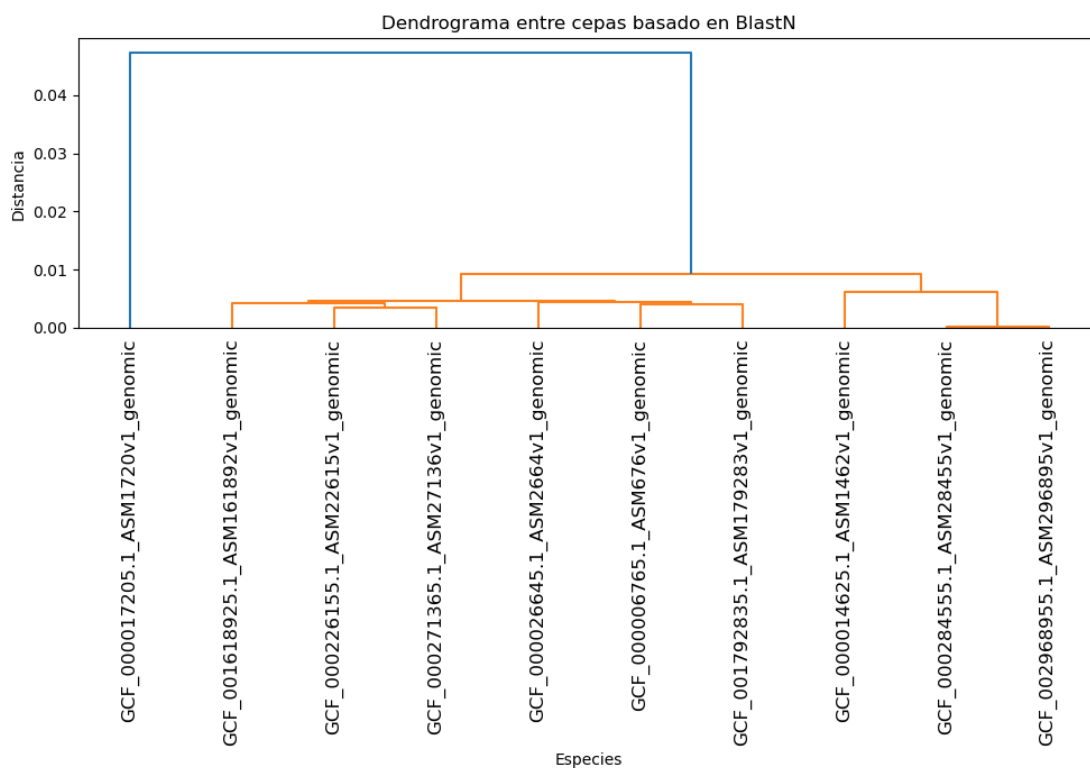


Figura 5.9: Dendrograma obtenido de BlastN.

Con respecto al gráfico anterior (figura 5.9) se puede apreciar las distancias entre las especies, podemos notar que en el análisis las distancias son significativamente menores, pasando de un orden decimal al centesimal en distancias, y la asociación grupal es más alta, mostrando que en su inmensa mayoría se tiene un grupo evolutivo en común a excepción de la primera cepa, que parece no compartir en demasía su linaje con el resto.

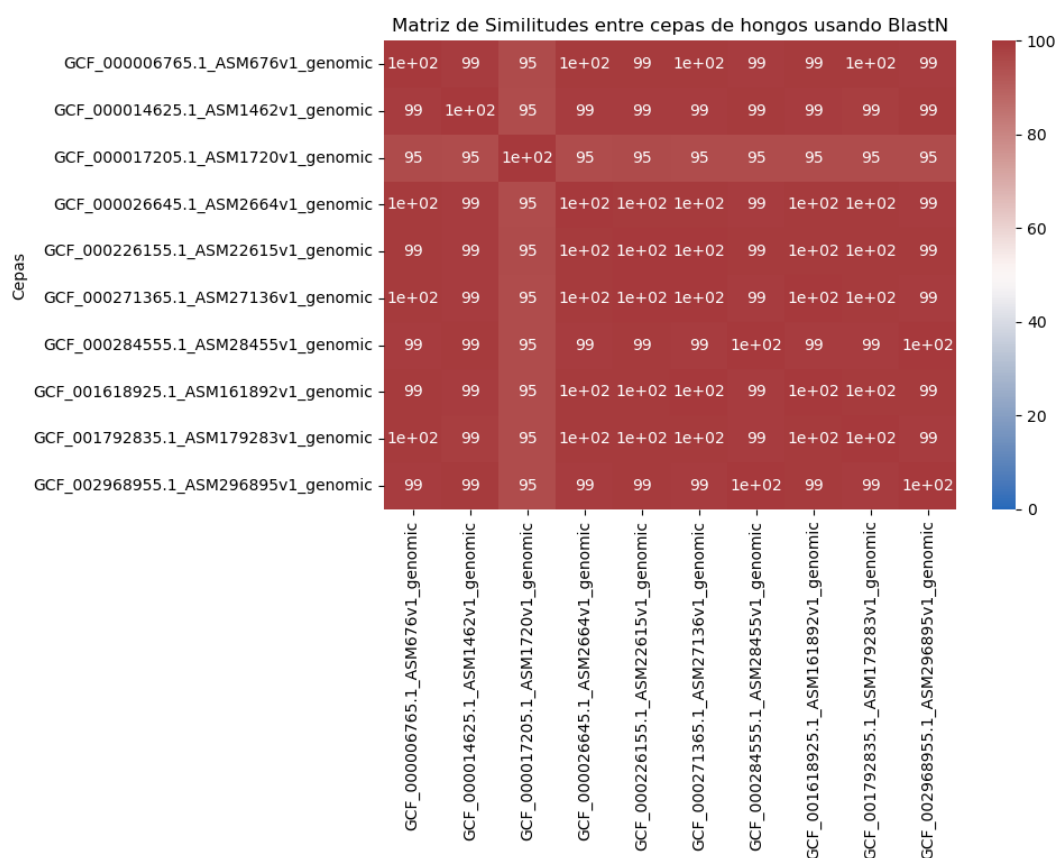


Figura 5.10: Matriz de identidad obtenido de BlastN.

Como se puede apreciar en la matriz (figura 5.6), al comparar entre especies se obtiene valores exageradamente alto, donde es prácticamente una semejanza total, esto es debido al gran parecido genético entre ellos al ser de la misma especie, y como vimos antes representan un linaje de origen en común, por lo cual las variaciones genéticas son significativamente menores.

5.3.2. OrthoANI

Para OrthoANI se obtuvieron los siguientes resultados:

Primeramente, el dendrograma el cual mejora significativamente sus valores de distancia, a una distancia máxima de 0.06, lo que puede interpretarse como que a pesar de que son la misma especie, hay variaciones evolutivas entre si, compartiendo un ancestro en común pero que los años de evolución han contribuido a separar las cepas.

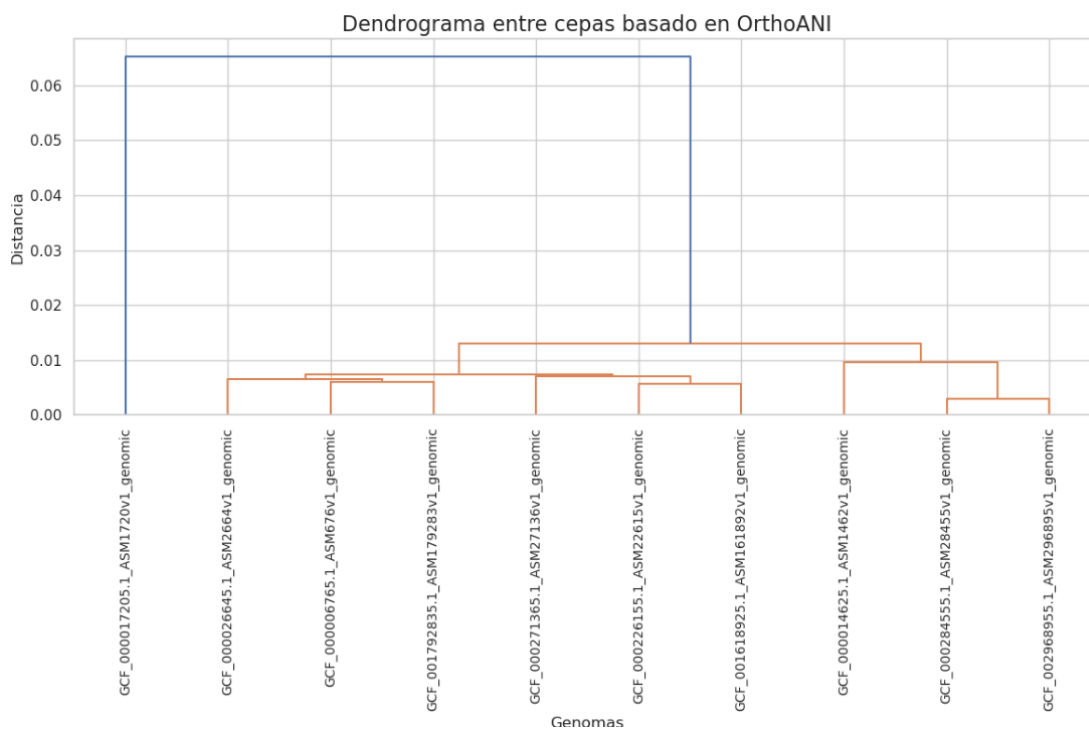


Figura 5.11: Dendrograma de identidad obtenido de OrthoANI.

Y como segunda salida, se obtiene la matriz de identidad utilizando los genes ortólogos, como se puede apreciar ha crecido considerablemente el valor en todos los casos, también se nota la diferencia evolutiva en sus valores de semejanza. y el valor promedio es de 98 %. Donde solo una especie a nivel genómico varía, teniendo resultados muy positivos y entregando más precisión que el método de BlastN.

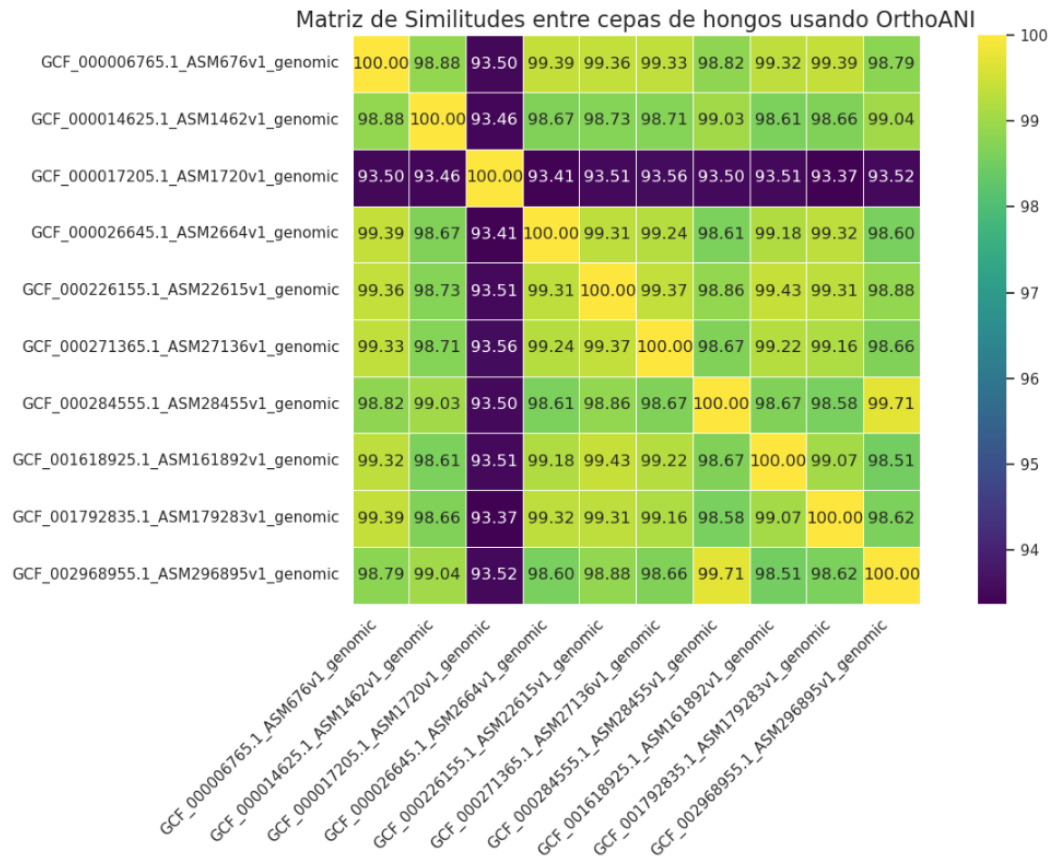


Figura 5.12: Dendrograma de identidad obtenido de OrthoANI.

Capítulo 6

Conclusiones

Como podemos observar en los resultados, existen diferencias menores respecto a ambos métodos, en BlastN notamos muchas más uniformidad respecto a los valores, como también de valores realistas, pero menos convincentes. Esto se evidencia al momento de generar la matriz de similitudes e incluso del primer dendrogramas entre familias, que refleja una gran diferencia con respecto a su matriz. Mientras que, en el caso de OrthoANI, los valores de las distancias.

Con esto podemos concluir en que la herramienta de OrthoANI parece la más precisa de ambas, debido a que compara fragmentos de genomas específicos, ideal para cadenas de genomas grandes, manteniendo mayor rigurosidad y precisión. Aún así la técnica de ANI integrada con BLAST no es necesariamente mala, puesto que podemos utilizar MLSA para reducir los espacios de las cadenas enfocándonos en ciertos cromosomas y aumentar la precisión. Si utilizamos MLSA, el uso de OrthoANI no es estrictamente necesario debido a que las cadenas ya están particionadas en sus cromosomas, reduciendo la cantidad de información y enfocándolo en funciones específicas, donde BlastN tomaría más relevancia.

Adicionalmente, BlastN puede ser configurado con mayor libertad que OrthoANI, permitiendo que la precisión, velocidad e incluso maneras de modelar los resultados cambien, haciéndolo más versátil a largo plazo, pudiéndose adaptar a diferentes situa-

ciones de comparación. También, destacar su impresionante velocidad de comparación, que al crear diversas bases de datos, permitió la generación de resultados mucho más rápida que OrthoANI

En el futuro, se proyecta la generación resultados con volúmenes de datos altos, incluyendo hasta cien o más datos de genomas para hacer investigaciones más reales. Como también utilizar otras herramientas que estén disponibles en el mercado, que representen un estudio y una versión más actual sobre la comparación genómica de hongos, y también explotar mucho más el potencial de la plataforma a través de la creación de servidores permitiendo tiempos de carga mínimos para los estudios. Además de integrar dendrograma con matriz para una visualización cómoda y más acorde al usuario, dando la opción de tenerlas de manera individual para sus publicaciones, un ejemplo sería el siguiente:

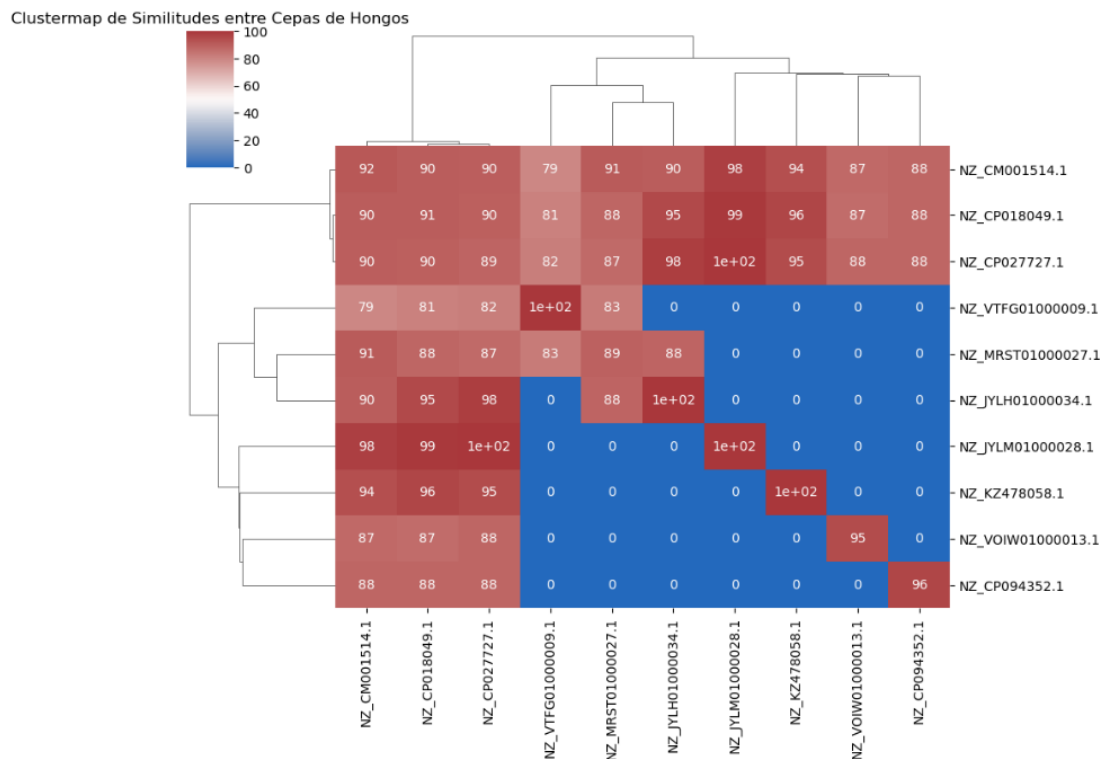


Figura 6.1: Cluster obtenido a partir del mismo estudio.

Capítulo 7

Anexos

7.1. Anexo A: Códigos en Python y archivos utilizados

El código fuente desarrollado se encuentra disponible en el repositorio de GitHub del autor [5].

7.2. Anexo B: Repositorio del programa FungiTG

El código fuente desarrollado se encuentra disponible en el repositorio de GitHub [6].

Bibliografía

- [1] C.P. Kubicek, A. Herrera-Estrella, V. Seidl-Seiboth, D.A. Martínez, I.S. Druzhinina, M. Thon, S. Zeilinger, S. Casas-Flores, B.A. Horwitz, P.K. Mukherjee, M. Mukherjee, L. Kredics, L.D. Alcaraz, A. Aerts, Z. Antal, L. Atanasova, M.G. Cervantes-Badillo, J. Challacombe, O. Chertkov, K. McCluskey, F. Coulpier, N. Deshpande, H. von Döhren, D.J. Ebbole, E.U. Esquivel-Naranjo, E. Fekete, M. Flippin, F. Glaser, E.Y. Gómez-Rodríguez, S. Gruber, C. Han, B. Henrissat, R. Hermosa, M. Hernández-Onate, L. Karaffa, I. Kosti, S. Le Crom, E. Lindquist, S. Lucas, M. Lübeck, P.S. Lübeck, A. Margeot, B. Metz, M. Misra, H. Nevalainen, M. Omann, N. Packer, G. Perrone, E.E. Uresti-Rivera, A. Salamov, M. Schmoll, B. Seiboth, H. Shapiro, S.A. Sukno, J.A. Tamayo-Ramos, D. Tisch, A. Wiest, H.H. Wilkinson, M. Zhang, P.M. Coutinho, C.M. Kenerley, E. Monte, S.E. Baker, and I.V. Grigoriev. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of trichoderma. *Genome Biology*, 12(4):R40, 2011.
- [2] K.J. Hoff and M. Stanke. Predicting genes in single genomes with augustus. *Current Protocols in Bioinformatics*, 65:e57, 2019.
- [3] Michael Richter and Ramon Rosselló-Móra. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, 106(45):19126–19131, 2009.
- [4] Leighton Pritchard, Rachel H. Glover, Sonia Humphris, John G. Elphinstone, and Ian K. Toth. Genomics and taxonomy in diagnostics for food security: soft-rotting

enterobacterial plant pathogens. *Anal. Methods*, 8:12–24, 2016.

- [5] Marcelo García. Implementación de herramientas de análisis filogenético para aplicación integral de genómica de hongos (código fuente). https://github.com/marcelo-garcia/FungiTG_filogenetica, 2025. Repositorio privado en GitHub correspondiente al código fuente utilizado en la memoria de titulación de la UTFSM.
- [6] Marcelo García, Farid Diaz, and Fabian Godoy. Fungitg, aplicación integral de genómica de hongos (código fuente). <https://github.com/mercelo13/FungiGT>, 2025. Repositorio privado en GitHub correspondiente al código fuente utilizado en la memoria de titulación de la UTFSM.