

Universidad de Concepción

Facultad de Ciencias Biológicas

Programa de Magíster en Biotecnología Molecular

**“Análisis de la Comunidad Bacteriana Endófito en la
Embriogénesis Somática de *Pinus radiata*”**

Tesis para optar al grado de Magíster en Biotecnología Molecular

Francisco Javier Guic Neira

Concepción - Chile - 2025

Profesor Guía: Rodrigo Hasbún Zaror

Esta tesis ha sido realizada en el Departamento de Silvicultura de la Facultad de Ciencias Forestales de la Universidad de Concepción.



Profesores integrantes de la Comisión Evaluadora:

Profesor Dr. Rodrigo Hasbún

Facultad de Ciencias Forestales

Universidad de Concepción

Dra. Alisson Astuya Villalón

Facultad de Ciencias Naturales y Oceanográficas

Universidad de Concepción

Dr. Mauricio Schoebitz Cid

Guía Facultad de Agronomía

Universidad de Concepción

Dr. Óscar Toro Núñez

Guía Facultad de Ciencias Naturales y Oceanográficas

Universidad de Concepción



I.	RESUMEN:	VII
II.	INTRODUCCIÓN:	8
	1. Biología de la embriogénesis somática	8
	1.1. Concepto y relevancia biotecnológica	8
	1.2. Etapas del proceso embriogénico	9
	1.3. Factores críticos de la embriogénesis somática.....	11
	1.4. Limitaciones específicas de la embriogénesis somática en <i>Pinus radiata</i>	12
	2. Microbioma endófito: un actor subestimado en la embriogénesis somática	14
	2.1. Evidencia de interacciones planta-microbioma en cultivos <i>in vitro</i>	14
	2.2. Mecanismos de acción de los microorganismos endófitos en la embriogénesis somática	16
	2.3. Dinámica del microbioma endófito y su influencia en la competencia embriogénica .	19
	3. Avances metodológicos para el estudio del microbioma embriogénico	21
	3.1. Amplicones ribosomales 16S y 16S-ITS-23S.....	21
	3.2. Contaminación por ADN del hospedador: un reto crítico	22
	3.3. Nanopore: resolución taxonómica y portabilidad	26
III.	HIPÓTESIS DE TRABAJO:.....	30
	Hipótesis.....	30
IV.	OBJETIVOS:	31
	Objetivo general:.....	31
	4. Objetivos específicos:.....	31



4.1. Optimización de la extracción de ADN y PCR:	31
4.2. Análisis de la diversidad y composición de la microbiota bacteriana:	31
4.3. Identificación y funcionalización del microbiota núcleo ("core"):.....	32
V. METODOLOGÍA:	33
5. Optimización de la extracción de ADN:.....	33
5.1. Material vegetal:	33
5.2. Líneas celulares de embriogénesis somática	33
5.3. Selección de partidores in sílico para PCR:	34
5.4. Extracción de ADN.....	36
6. Optimización de las condiciones de PCR.....	37
6.1. Evaluación inicial con partidores 27F-1492R	37
6.2. Optimización con partidores 27F-2241R para el operón 16S-ITS-23S	38
7. Secuenciación por Oxford nanopore	39
7.1. Preparación de librerías de secuenciación	39
7.2. Datos crudos y preprocesamiento al análisis taxonómico	40
7.3. Flujo de trabajo wf-metagenomics por EPI2ME	41
7.4. Flujo de trabajo personalizado basado en RESCUE.....	42
7.5. Filtrado de secuencias mitocondriales.....	43
7.6. Análisis taxonómico	44
7.7. Diversidad alfa.....	46
7.8. Diversidad beta	47
8. Identificación y caracterización de especies bacterianas núcleo ("core"):.....	48
9. Análisis funcional del microbiota núcleo ("core").....	49



VI. RESULTADOS:.....	50
10. Optimización de la extracción de ADN y la amplificación por PCR:.....	50
10.1. Líneas celulares de embriogénesis somática:.....	50
10.2. Selección de partidores in sílico para PCR:	52
11. Extracción de ADN.....	55
12. Optimización de las condiciones de PCR:	56
13. Secuenciación por Oxford Nanopore:.....	59
14. Primera secuenciación (prueba piloto):.....	59
14.1. Preparación de Librerías de secuenciación.....	59
14.2. Análisis directo con flujo de trabajo wf-metagenomics	60
15. Análisis taxonómico con Minimap2	63
15.1. Preprocesamiento de datos basado en el pipeline RESCUE	63
15.2. Análisis taxonómico	66
16. Segunda secuenciación por Oxford Nanopore (corrida completa).	68
16.1. Preparación de Librerías de secuenciación.....	68
16.2. Preprocesamiento de datos basado en el pipeline RESCUE	69
16.3. Análisis taxonómico	73
Composición taxonómica y dominancia del núcleo microbiano	73
Diversidad alfa.....	75
Diversidad beta	78
16.4. Identificación de especies bacterianas núcleo ("core"):.....	79
16.5. Análisis funcional y relevancia biológica.....	81
VII. DISCUSIÓN.	84



17. Optimización de la extracción de ADN y la amplificación por PCR:.....	84
18. Primera secuenciación Oxford nanopore (prueba piloto):	86
18.1. Flujo <i>wf-metagenomics</i> (EPI2ME)	86
18.2. Pipeline basado en RESCUE (filtrado + Minimap2)	87
18.3. Efecto de los bloqueadores PNA.....	88
19. Segunda secuenciación por Oxford Nanopore (corrida completa).	89
19.1. Análisis taxonómico	89
19.2. Identificación de Especies Bacterianas Núcleo ("Core").....	91
19.3. Análisis funcional y relevancia biológica.....	92
VIII. CONCLUSIÓN:	96
IX. MATERIAL SUPLEMENTARIO:.....	99
Capacidad madurativa según familias	99
Secuenciación Sanger.....	101
Secuenciación por Oxford Nanopore.....	102
X. BIBLIOGRAFÍA:	108

I. RESUMEN:

El proyecto se centró en caracterizar la microbiota bacteriana de líneas embriogénicas somáticas de *Pinus radiata* mediante secuenciación de lecturas largas (Nanopore) del operón 16S–ITS–23S y análisis bioinformático con la base rEGEN. Primero, se optimizó un protocolo de extracción de ADN usando columnas y bloqueadores PNA, consiguiendo amplicones de alta integridad y mínima co-amplificación vegetal. A continuación, se compararon flujos wf-metagenomics y un pipeline adaptado de RESCUE, seleccionando rEGEN por su balance entre calidad de mapeo y recuperación de diversidad. En la corrida completa, se identificó un microbioma núcleo de ~15–17 especies con prevalencia $\geq 95\%$, dominado por *Sphingomonas sanxanigenens*, *Kordiimonas spp.* y *Candidatus Liberibacter spp.*, cuyas funciones metabólicas, detoxificación de fenoles, síntesis de exopolisacáridos y producción de fitohormonas o sideróforos, sugieren un apoyo activo al desarrollo embrionario *in vitro*. Aunque no se hallaron diferencias significativas en diversidad alfa ni beta entre líneas de alta y baja maduración, el estudio establece una base metodológica y funcional para ensayos futuros de metagenómica shotgun, transcriptómica y de inoculación con consorcios microbianos que optimicen la embriogénesis somática en coníferas.

II. INTRODUCCIÓN:

1. Biología de la embriogénesis somática

1.1. Concepto y relevancia biotecnológica

La embriogénesis somática (ES) representa un hito en biotecnología vegetal al permitir la regeneración de plantas completas a partir de células somáticas, evitando la variabilidad genética de la reproducción sexual. Descrita inicialmente en *Daucus carota* por Steward et al (1958), este proceso se fundamenta en la totipotencia celular, capacidad inherente de las células vegetales para reprogramar su identidad bajo condiciones in vitro (X. Yang & Zhang, 2010). A diferencia de la embriogénesis cigótica, la ES replica las etapas morfológicas del desarrollo embrionario mediante la activación de redes transcripcionales específicas, lo que ha posibilitado su uso en clonación masiva de genotipos élite y estudios de desarrollo temprano (Elhiti et al., 2013).

En coníferas como *P. radiata*, la ES adquiere relevancia estratégica debido a su ciclo de vida prolongado (10–20 años para madurez reproductiva) y las limitaciones de métodos tradicionales de propagación clonal (Lelu-Walter et al., 2016). Su aplicación no solo facilita la conservación de germoplasma mediante criopreservación (Ballesteros et al., 2024), sino que también ofrece una plataforma para investigar mecanismos moleculares del desarrollo en especies

leñosas. Sin embargo, su implementación industrial enfrenta desafíos críticos, como la pérdida progresiva de capacidad embriónica durante subcultivos prolongados y la baja eficiencia en la etapa de maduración (Montalbán et al., 2010).

1.2. Etapas del proceso embriónico

El proceso de ES en *P. radiata* (Figura 1) se organiza en cuatro fases secuenciales:

- **Inducción:** Células competentes (e.g., tejido nucelar de embriones cigóticos inmaduros) responden a señales químicas como auxinas, desdiferenciándose en masas proembrionarias (PEMs). La eficiencia de esta etapa depende de factores genéticos y ajustes en el medio de cultivo, como la concentración de nitrógeno y reguladores de crecimiento (Montalbán et al., 2012).
- **Multipliación:** Las PEMs proliferan en medios con balances hormonales específicos (alto 2,4-D y citocininas), manteniendo un estado indiferenciado.
- **Maduración:** Bajo estrés osmótico (ej. PEG-4000) y reducción de auxinas, las PEMs sincronizan su desarrollo hacia embriones somáticos estructurados (etapas cotiledonares). Sin embargo, en *P. radiata*, esta fase

enfrenta cuellos de botella, como la pérdida de capacidad embriogénica tras subcultivos prolongados (Montalbán et al., 2010).

- **Germinación:** Activación de programas genéticos para crecimiento radical y epicotilar, culminando en plántulas autótrofas (Arnold et al., 2002).

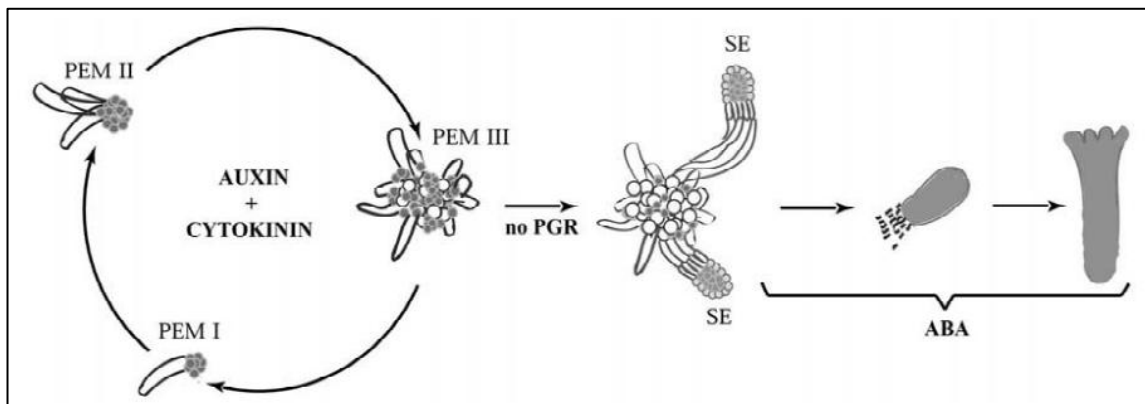


Figura 1 Representación esquemática de la vía de desarrollo de la embriogénesis somática en *Picea abies* (figura tomada de Arnold et al., 2002)

Este proceso depende de la interacción dinámica entre factores endógenos (ej. expresión de genes *SERK* y *WUSCHEL*, asociados a la pluripotencia) y exógenos (reguladores de crecimiento, nutrientes). Desequilibrios en etapas críticas, como la acumulación de especies reactivas de oxígeno (ROS) o alteraciones en la señalización hormonal, pueden generar aberraciones morfológicas o reversión a callo no embriogénico (De-la-Peña et al., 2015).

1.3. Factores críticos de la embriogénesis somática

Las auxinas, particularmente el 2,4-D, actúan como reguladores epigenéticos clave en la inducción de la ES. Estas hormonas promueven la descompactación de la cromatina mediante la activación de histona acetiltransferasas (HATs), facilitando la expresión de genes embriogénicos como WUSCHEL y SERK (Wang et al., 2020). Estudios recientes demuestran que el 2,4-D induce cambios rápidos en la accesibilidad de la cromatina, activando una red transcripcional jerárquica que incluye factores como LEC2, regulador maestro de genes de patrón embrionario temprano (e.g., WOX2 y WOX3) (F.-X. Wang et al., 2020).

El efecto de las auxinas es espaciotemporalmente restringido: concentraciones elevadas (5–10 μM) inducen PEMs, mientras que su reducción gradual (<1 μM) es crítica para la maduración. Esta dinámica está modulada por interacciones entre vías de señalización del ácido indolacético (AIA) y mecanismos epigenéticos, como la metilación de ADN y modificaciones postraduccionales de histonas (Wójcikowska et al., 2020). Por ejemplo, la inhibición de histona deacetilasas (HDACs) con tricostatina A (TSA) induce ES incluso en ausencia de auxinas exógenas, evidenciando el papel central de la reprogramación epigenética en la totipotencia (Wójcikowska et al., 2020).

1.4. Limitaciones específicas de la embriogénesis somática en *Pinus radiata*

A pesar de ser la conífera más plantada a nivel global (≈ 4 millones de hectáreas en Chile), *P. radiata* enfrenta desafíos únicos en ES:

Restricción genética y fenológica: Solo embriones cigóticos inmaduros (20–30 días post-polinización) en estadios 2–4 son inductores eficientes, debido a su mayor capacidad de desdiferenciación (Montalbán et al., 2012). Esta ventana temporal estrecha limita la recolección de explantes viables (Montalbán et al., 2016). Los embriones cigóticos inmaduros (20–30 días post-polinización) en estadios 2–4 son inductores eficientes, debido a su mayor capacidad de desdiferenciación celular y respuesta a señales auxínicas (Montalbán et al., 2012). Esta ventana temporal estrecha limita la recolección de explantes y exige sincronización precisa con el desarrollo del embrión cigótico (Montalbán et al., 2016).

Pérdida de competencia embriogénica: Cultivos prolongados (>6 meses) muestran una reducción del 70% en capacidad de maduración, asociada a

- Acumulación de ROS que alteran la estabilidad genómica (Klimaszewska et al., 2009)
- Cambios epigenéticos en genes clave (LEC2, WOX)

Dependencia de criopreservación: Aunque efectiva para mantener líneas embriogénicas, esta estrategia incrementa costos operativos en un 30% y no resuelve la pérdida de calidad post-descongelación (Ballesteros et al., 2024).

Estas limitaciones sugieren la presencia de factores dinámicos subestimados, como las comunidades bacterianas endófitas. En *Arabidopsis thaliana*, ciertos taxones bacterianos (e.g., *Methylobacterium*) modulan la producción de auxinas y reducen el estrés oxidativo mediante síntesis de ACC-desaminasa (Hardoim et al., 2015), mecanismos potencialmente aplicables para optimizar la ES en *P. radiata*. Comparar líneas de alta (>600 embriones/g) y baja (<200 embriones/g) capacidad de maduración es crucial para identificar si el microbioma endófito contribuye diferencialmente a la estabilidad embriogénica en subcultivos prolongados. En este contexto, la integración de ajustes en el medio de cultivo, control epigenético y modulación microbiana podría revolucionar los protocolos actuales (Figura 2), donde la embriogénesis somática se inserta como etapa crítica en los programas de silvicultura clonal (Park et al., 2016).

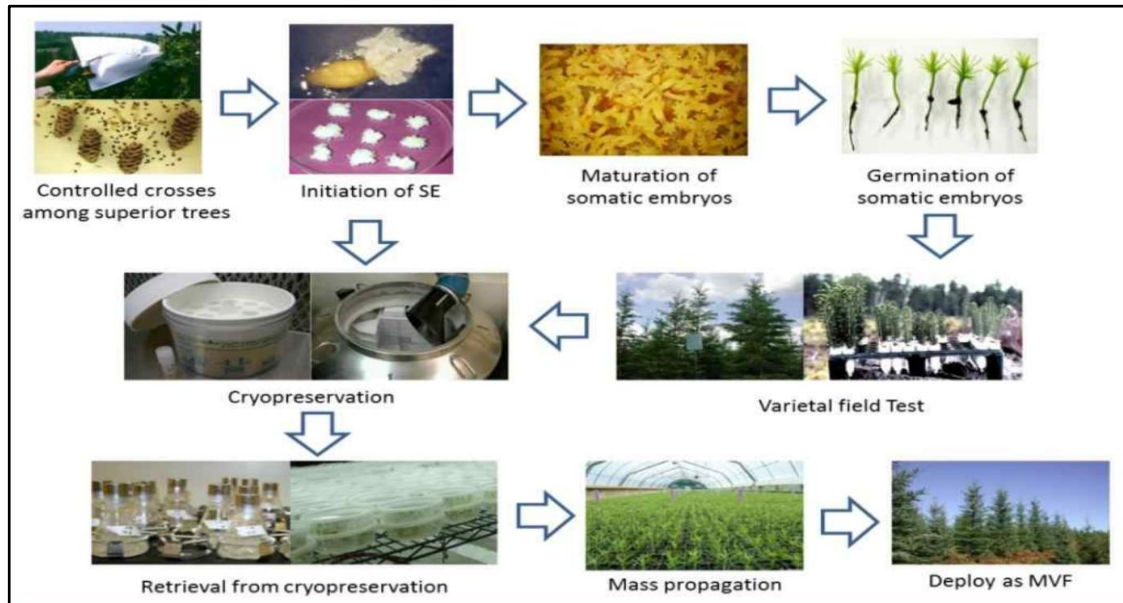


Figura 2 Esquema del proceso de silvicultura clonal por embriogénesis somática en coníferas (figura tomada de Park et al., 2016).

2. Microbioma endófito: un actor subestimado en la embriogénesis somática

2.1. Evidencia de interacciones planta-microbioma en cultivos *in vitro*

Aunque tradicionalmente se asumía que los cultivos de tejidos vegetales eran estériles, estudios recientes revelan que las líneas celulares embriogénicas albergan comunidades microbianas endófitas heredadas del explante parental, capaces de modular procesos críticos del desarrollo (Thomas & Franco, 2021). Por ejemplo, en micro cultivos de *Eucalyptus urophylla*, la metagenómica ha identificado bacterias endófitas como *Enterobacter spp.* y *Paenibacillus polymyxa* que promueven la síntesis de auxinas y la movilización de nutrientes,

mejorando la regeneración in vitro (De França Bettencourt et al., 2021). Thomas (2022) demostró que las bacterias intracelulares transmitidas verticalmente a través de embriones (“Cytobacts”) son un fenómeno ubicuo en plantas vasculares, influyendo en procesos metabólicos clave durante el desarrollo temprano. En coníferas, aunque no existen reportes directos de “Cytobacts” en *P. radiata*, estudios en *Picea glauca* han detectado comunidades bacterianas endófitas en tejidos embriogénicos, sugiriendo una posible asociación mutualista (Klimaszewska et al., 2009).

La relación simbiótica entre bacterias endófitas y células vegetales es un fenómeno extendido, incluso en cultivos supuestamente axénicos. En micro plantas de piña (*Ananas comosus*) y orquídeas, microscopía electrónica ha revelado bacterias intracelulares en espacios intercelulares de tejidos asintomáticos, independientemente del laboratorio o protocolo de cultivo (Esposito-Polesi et al., 2017). Estos hallazgos, respaldados por análisis de secuenciación 16S rRNA, indican que las comunidades bacterianas persisten de forma latente durante años en cultivos in vitro, influyendo en procesos metabólicos clave (Thomas & Franco, 2021). Esta relación puede facilitar la síntesis de metabolitos clave mediante distintos mecanismos de interacción planta-endófito (Figura 3; Brader et al., 2014)

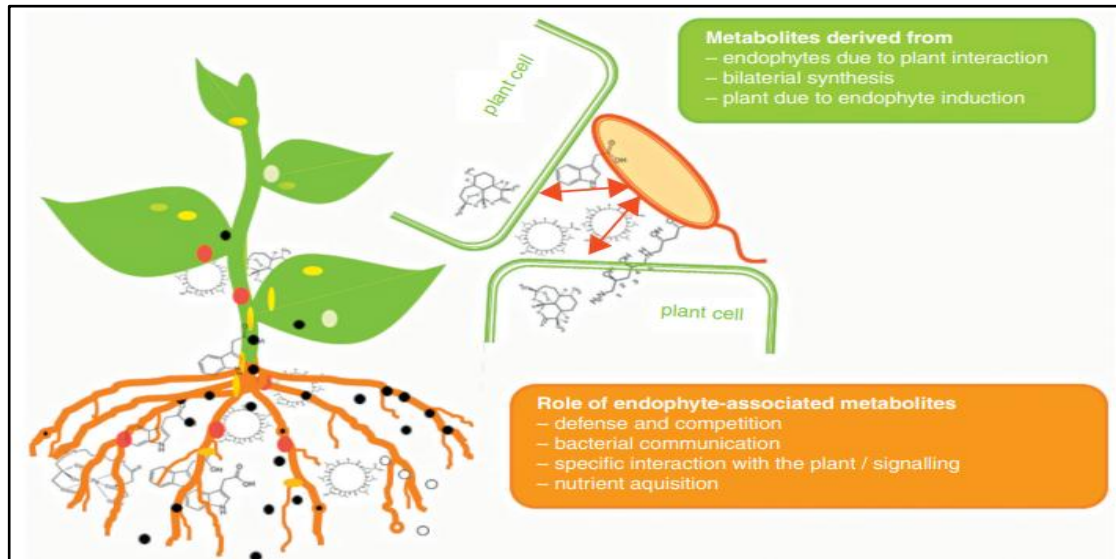


Figura 3 Esquema que muestra los diferentes tipos de interacciones planta-endófito que conducen a la síntesis de metabolitos. La relación simbiótica entre las bacterias endófitas y las células huésped también se ha demostrado en cultivos de callos y células de plantas, donde se han encontrado bacterias intracelulares denominadas "Cytobacts" (Thomas & Franco, 2021) Estas bacterias mantienen una asociación mutualista dinámica con las células huésped, lo que sugiere una posible interacción beneficiosa para la ES. Como se muestra en la Ilustración 3, esta interacción puede conducir a la síntesis de metabolitos a través de diferentes tipos de interacciones planta-endófito (figura tomada de Brader et al., 2014).

2.2. Mecanismos de acción de los microorganismos endófitos en la embriogénesis somática

Los microorganismos endófitos influyen en la embriogénesis somática (ES) a través de mecanismos complejos y altamente interconectados, que abarcan desde la modulación de vías de señalización molecular hasta la regulación del balance hormonal y la protección contra estrés oxidativo. Uno de los ejes centrales de su acción es la activación de la familia de genes "somatic embryogenesis receptor kinase" (SERK), que codifican proteínas quinasa transmembrana esenciales para la adquisición de competencia embriogénica.

La dimerización inducida por ligandos y la subsiguiente fosforilación de dominios intracelulares desencadenan cascadas de señalización que regulan tanto el desarrollo como las respuestas defensivas. Evidencia experimental en especies como *Arabidopsis thaliana* y *Theobroma cacao* respalda el papel conservado de SERK en la embriogénesis, mientras que su silenciamiento génico, como se observó en lechuga transgénica, reduce drásticamente la capacidad embriogénica (Santos & Aragão, 2009).

Además de la regulación de SERK, los endófitos modulan la producción y señalización de fitohormonas, creando un microambiente óptimo para la inducción y maduración de embriones. Bacterias como *Pseudomonas fluorescens* y *Sphingomonas spp.* sintetizan auxinas (AIA) y giberelinas, que promueven la expansión celular en tejidos embrionarios y estructuras como cotiledones y raíces. Este efecto se complementa con la regulación de otras hormonas, como citoquininas, ácido abscísico y etileno, que en conjunto coordinan la división y diferenciación celular durante la ES. La capacidad de los endófitos para equilibrar estas señales no solo favorece el desarrollo embrionario, sino que también fortalece la tolerancia al estrés, un factor crítico para la viabilidad de los tejidos embriogénicos (Chen et al., 2017).

Otro mecanismo clave es la mitigación del estrés oxidativo, que amenaza la estabilidad de las células en proliferación activa durante la ES. Los endófitos

inducen sistemas antioxidantes en el hospedero, como enzimas detoxificadoras y moléculas como glutatión, que neutralizan especies reactivas de oxígeno (ROS). Esta protección es particularmente relevante en etapas tempranas de la embriogénesis, donde el alto metabolismo oxidativo incrementa la susceptibilidad al daño celular. Estudios en plantas medicinales demuestran que la colonización por endófitos activa respuestas adaptativas que preservan la homeostasis redox, facilitando así la transición de células somáticas a embriogénicas (Pandey et al., 2023).

Los endófitos ejercen un control fino sobre la expresión génica del hospedero, como se ha observado con la regulación positiva de familias de transportadores de metales en *Sedum alfredii* inoculado con *Pseudomonas fluorescens* (Chen et al., 2017). Aunque este ejemplo no está directamente ligado a la ES, ilustra el potencial de los endófitos para reprogramar redes génicas, un principio que probablemente se extiende a genes asociados con la embriogénesis. La interacción entre estas vías señalización molecular, balance hormonal, protección redox y regulación transcripcional subraya el papel integral de los endófitos como moduladores de la ES, ofreciendo perspectivas promisorias para aplicaciones biotecnológicas en propagación vegetal.

En coníferas como *P. radiata*, los microorganismos endófitos modulan la embriogénesis somática al influir en la síntesis de fitohormonas y la



detoxificación de compuestos fenólicos acumulados en subcultivos (Pirttilä et al., 2000; Berg et al., 2014). En Chile, donde *P. radiata* es una especie forestal clave (INFOR, 2020), estudios preliminares sobre microbiomas rizosféricos sugieren que bacterias promotoras del crecimiento vegetal, tradicionalmente denominadas PGPR (*Plant Growth-Promoting Rhizobacteria*) mejoran la tolerancia al estrés abiótico en plantaciones (Gómez et al., 2020; Richardson et al., 2022). Sin embargo, dado que en este estudio se analizan comunidades endofíticas, el término más adecuado es el de PGPB (*Plant Growth-Promoting Bacteria*) o **endófitos promotores del crecimiento vegetal** (PGPE), que engloban a los taxones con funciones análogas dentro de los tejidos embrionarios.

2.3. Dinámica del microbioma endófito y su influencia en la competencia embriogénica

La dinámica del microbioma endófito y su influencia en la competencia embriogénica dependen de la estabilidad de ciertas comunidades microbianas a lo largo de distintas condiciones *in vitro*. En este contexto, el microbioma núcleo se define como el conjunto de taxones bacterianos consistentemente presentes en al menos el 95% de las muestras de un hospedero, independientemente de variaciones ambientales o genotípicas, según el análisis de metabarcoding del operón 16S-ITS-23S en este estudio, lo que

sugiere un rol potencial en la fisiología de *P. radiata* (adaptado de Shade & Handelsman, 2012). La plasticidad del microbioma endófito desempeña un papel clave en la regulación de la embriogénesis somática (ES), modulando procesos metabólicos a través de la interacción entre un núcleo persistente y una fracción variable que responde a factores como pH, nutrientes y cambios fisiológicos del hospedante (Kandel et al., 2017).

En *P. radiata*, la pérdida de competencia embriogénica en subcultivos prolongados, como observado en inoculaciones experimentales con *Fusarium circinatum*, se asocia preliminarmente con desplazamientos en la composición microbiana, afectando la homeostasis redox mediante la alteración de metabolitos como glutatión y enzimas antioxidantes (Sánchez, 2018). Esta disfunción se refleja en una disminución estimada del 70% en la conversión de embriones somáticos a plántulas viables. Estudios en especies relacionadas, como *Eucalyptus urophylla*, indican que la dominancia de Proteobacterias y Firmicutes facilita la movilización de nutrientes y la producción de fitohormonas (e.g., ácido indolacético), incrementando hasta un 40% la formación de brotes en medios con benziladenina (Kandel et al., 2017; Martínez-Ruiz et al., 2005). A nivel molecular, endófitos en *P. radiata*, incluyendo hongos como *Alternaria* y bacterias como *Herbaspirillum seropedicae*, reducen la acumulación de peróxido de hidrógeno (H_2O_2) en un 60% y regulan la señalización hormonal, optimizando la diferenciación celular (Kandel et al., 2017; Sánchez, 2018).



Sin embargo, la pérdida de diversidad microbiana en subcultivos prolongados altera la síntesis de enzimas como quitinasas y β -1,3-glucanasas, esenciales para la remodelación de paredes celulares. Estrategias como la caracterización metagenómica en tiempo real y el uso de antibióticos (e.g., terramicina a 40 mg/100 ml, reduciendo contaminación externa en un 90%) permiten aislar comunidades endófitas funcionales (Martínez-Ruiz et al., 2005). Además, inductores como el ácido salicílico podrían estabilizar la competencia embriogénica al enriquecer comunidades como Proteobacterias (Kandel et al., 2017). Estos hallazgos sugieren un mutualismo dinámico, donde el microbioma núcleo y su plasticidad coevolucionan con el hospedero, adaptándose a perturbaciones y modulando la eficiencia embriogénica.

3. Avances metodológicos para el estudio del microbioma embriogénico

3.1. Amplicones ribosomales 16S y 16S-ITS-23S

La amplificación y secuenciación del operón ribosomal completo (16S–ITS–23S, \approx 4 500 pb) proporciona un nivel de resolución filogenética muy superior al alcanzable con fragmentos parciales de 16S rRNA (\sim 1 500 pb). Al incorporar las regiones ITS y 23S, que presentan mayor variabilidad que el 16S solo, se incrementa la discriminación a nivel de especie e incluso de cepa, permitiendo diferenciar taxones con secuencias casi idénticas en el 16S, como *Escherichia*

coli y *Shigella spp.* (Martijn et al., 2019). Además, el uso de plataformas de lectura larga (PacBio, ONT) en una sola reacción evita los sesgos de ensamblaje de fragmentos cortos y reduce las ambigüedades filogenéticas asociadas a bases de datos incompletas. Srinivas et al. (2025) demostraron que la elección de pares de cebadores y la plataforma de secuenciación no introduce sesgos significativos, mientras que la cobertura completa del operón aumenta la proporción de lecturas clasificables a niveles de especie y cepa en más de un 25 % respecto al 16S parcial.

Este enfoque ha mostrado su utilidad en estudios de microbioma complejos, incluidos suelos forestales y rizosferas de coníferas, donde la detección de taxones minoritarios y variantes intraespecíficas es crucial para entender la ecología microbiana asociada a la salud vegetal (Martijn et al., 2019). En el caso de *P. radiata*, la secuenciación 16S–ITS–23S permitiría caracterizar con mayor detalle las comunidades endófitas embriogénicas, detectar cepas correlacionadas con la competencia embrionaria y minimizar la pérdida de información por contaminación hospedadora. De este modo, la estrategia RRN ofrece un marco robusto para avanzar en la comprensión de la interacción planta–microbio en sistemas de embriogénesis somática.

3.2. Contaminación por ADN del hospedador: un reto crítico

La extracción de ADN microbiano de tejidos de *P. radiata* en cultivos embriogénicos somáticos enfrenta desafíos significativos que justifican la optimización del protocolo. La alta proporción de ADN cloroplástico y mitocondrial, que puede alcanzar hasta el 99% del ADN total debido a la multicopia de organelos en coníferas (Lefèvre et al., 2020a; Song & Xie, 2020), genera co-amplificación no deseada de secuencias homólogas al 16S rRNA bacteriano durante el metabarcoding, enmascarando señales de comunidades microbianas minoritarias. Esta interferencia se agrava por inhibidores vegetales como fenoles y polisacáridos, abundantes en estos tejidos, que degradan el ADN microbiano y reducen la eficiencia de la PCR (Azmat et al., 2012; Demeke & Jenkins, 2010). A diferencia del ADN nuclear, menos abundante, estas secuencias organelares interfieren significativamente con la amplificación de secuencias bacterianas, reduciendo la sensibilidad de los métodos tradicionales de secuenciación (Fitzpatrick et al., 2018; Sonett et al., 2024). Por ejemplo, en la filósfera de arroz (*Oryza sativa*), las secuencias del huésped constituyeron el 99.4% de los amplicones antes de aplicar métodos de depleción (Song & Xie, 2020). Para mitigar estos problemas, Song & Xie (2020) desarrollaron el método Cas-16S-seq, que utiliza CRISPR/Cas9 para eliminar secuencias de cloroplasto y mitocondria durante la preparación de bibliotecas de amplicones, reduciendo la fracción de ADN del huésped en muestras de raíz de arroz del 63.2% al 2.9% y en filósfera del 99.4% al 11.6%, sin introducir sesgos en comunidades

bacterianas de suelo. Sin embargo, su aplicación en coníferas como *P. radiata* aún requiere desarrollo y validación, destacando la necesidad de métodos optimizados como columnas y PNA clamps para obtener ADN de alto peso molecular y reducir la co-amplificación vegetal, permitiendo la amplificación de operones completos (~4.2 kb) del 16S-ITS-23S.

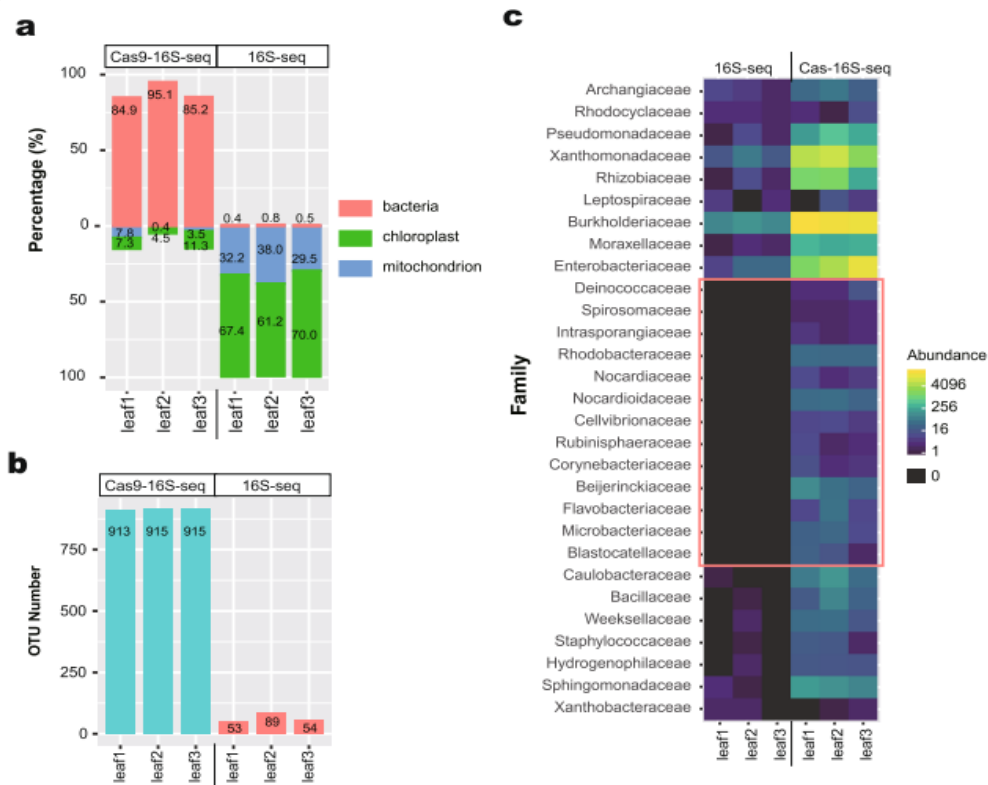


Figura 4 Agotamiento simultáneo de genes 16S rRNA mitocondriales y cloroplásticos del huésped en muestras de filósfera mediante Cas-16S-seq. Se utilizaron dos ARNg (mt-ARNg780 y cp-ARNg578) dirigidos al amplicón 515F-806R (región V4) en Cas-16S-seq para tres muestras de filósfera. La tabla de OTU está enrarecida a 50.000 lecturas. (a) Fracciones de genes de ARNr 16S bacterianos, mitocondriales y de cloroplastos en muestras de filósfera. (b) Comparación del número de OTU con abundancia mediana > 1 en cada muestra de filósfera (figura tomada de Song & Xie, 2020).

Alternativamente, los **bloqueadores de hibridación (PNA clamps)** dirigidos a secuencias conservadas del huésped han demostrado eficacia. Lefèvre et al.,

(2020) implementaron PNA específicos en hojas de *Pinus taeda*, reduciendo 20 veces la co-amplificación de ADN vegetal y aumentando las lecturas bacterianas de 53 a 2323 por muestra, con un incremento del triple en OTUs detectados. No obstante, la eficacia depende críticamente de la especificidad del diseño: incluso un único desajuste en la secuencia del PNA reduce su rendimiento (Fitzpatrick et al., 2018). Esto subraya la necesidad de personalizar los bloqueadores según el huésped, utilizando alineamientos de secuencias cloroplásticas y mitocondriales de amplio espectro taxonómico (Lefèvre et al., 2020b). En insectos herbívoros, Hanshew et al., (2013) optimizaron primers como 799F-mod3, mod6 y mod7, logrando reducir la contaminación por cloroplastos de >99% a menos del 1% en muestras de *Lymantria dispar* y *Azteca constructor*.

Aunque los PNA clamps han mostrado resultados prometedores, persisten limitaciones en tejidos con alta carga de ADN hospedador. Flörl & Bokulich, (2025) demostraron que concentraciones óptimas de 0.25 μM reducen >95% del ADN mitocondrial y 88% del cloroplástico en muestras de mosto de uva, sin afectar la diversidad alfa (índices de Shannon, $P = 0.553$) o beta (PERMANOVA, $R^2 = 0.003$, $P = 0.056$) (figura 5). Sin embargo, en células embriogénicas de coníferas, hasta el 11% de lecturas cloroplásticas persisten incluso con bloqueadores (Flörl & Bokulich, 2025). Además, estos protocolos requieren ajustes en las condiciones de PCR, como temperaturas de

alineamiento reducidas (55°C) y ciclos limitados (25 ciclos), para evitar artefactos de amplificación (Lefèvre et al., 2020a).

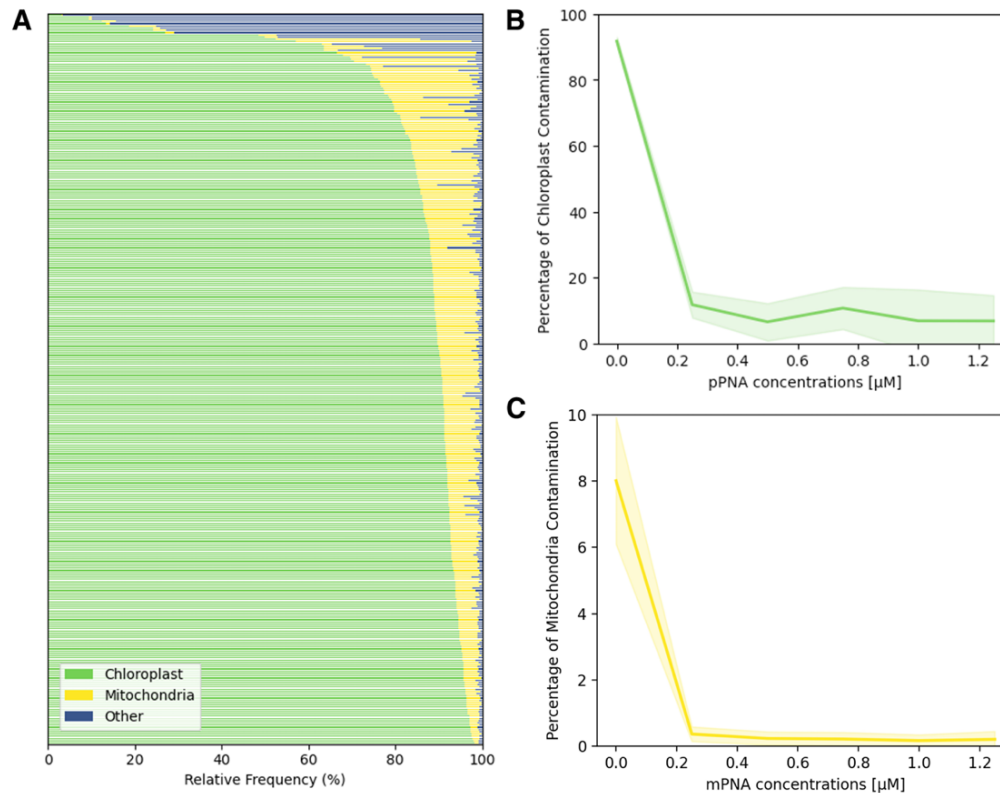


Figura 5 La pérdida de lecturas asignadas a cloroplastos y mitocondrias en los datos de secuenciación del gen 16S rRNA es sustancial (A), pero puede mitigarse mediante el uso de bloqueadores de ácido nucleico peptídico (PNA) (B y C). A: Un gráfico de barras taxonómicas con datos de secuenciación del gen 16S rRNA de 385 muestras de mosto de uva muestra que, en promedio, el 98.5% de las lecturas se atribuyeron a ADN del hospedador vegetal. Esto resultó en una profundidad de lectura insuficiente para un análisis confiable de la diversidad bacteriana, ya que más de la mitad de las muestras tuvieron menos de 120 características tras el filtrado. Pruebas posteriores aplicando B, bloqueadores de PNA para cloroplastos, y C, bloqueadores de PNA para mitocondrias, demostraron que una baja concentración de estos redujo de manera confiable el porcentaje de lecturas perdidas por ADN del hospedador vegetal. (Flörl & Bokulich, 2025).

3.3. Nanopore: resolución taxonómica y portabilidad

La secuenciación completa del gen 16S rRNA mediante Oxford Nanopore (ONT) ofrece una resolución taxonómica superior a nivel de especie en comparación con plataformas de lecturas cortas, como Illumina, al abarcar todas las regiones hipervariables (V1-V9) y minimizar los sesgos asociados a la amplificación con primers específicos (Cuscó et al., 2017). Estudios recientes destacan su capacidad para detectar taxones minoritarios (<0.01% de abundancia) y filos poco representados en bases de datos, como *Lentisphaerae*, que suelen omitirse con métodos tradicionales (Cuscó et al., 2017). Por ejemplo Szoboszlay (2023) demostraron que, utilizando la química Kit 12 (>99% de precisión), ONT clasificó entre el 82.1% y 97.4% de las lecturas a nivel de especie en muestras fecales, superando ampliamente a Illumina (30.8–53.8%). Sin embargo, la precisión taxonómica depende críticamente de la completitud de las bases de referencia, ya que hasta el 45% de las especies en comunidades modelo no pudieron clasificarse debido a la ausencia de secuencias completas de 16S rRNA en repositorios públicos (Cuscó et al., 2017).

El desarrollo de herramientas bioinformáticas especializadas, como el pipeline PRONAME (Dubois et al., 2024a), ha abordado limitaciones históricas de ONT, como la alta tasa de error (~12% en lecturas 2D). PRONAME integra corrección de errores mediante *Medaka*, *clustering* adaptativo y bases de datos curadas (Silva138, Greengenes2 y rEGEN-B), logrando una precisión del 99.5–99.7% en secuencias consenso, comparable a Illumina (Dubois et al., 2024a). Además, su

portabilidad (empaquetado en Docker) y compatibilidad con QIIME2 facilitan su aplicación en entornos con recursos limitados. Aunque PacBio ofrece mayor precisión en la identificación de isoformas transcriptómicas (Weirather et al., 2017), ONT supera a esta plataforma en portabilidad y flexibilidad experimental. El dispositivo MinION™ permite secuenciación *in situ*, evitando la degradación de muestras durante el transporte, y muestra menor ruido técnico que Illumina, con diferencias significativamente menores entre réplicas ($p < 0.0001$) (Szoboszlay et al., 2023). Estas ventajas son críticas para estudios de microbiomas complejos, donde la detección de taxones raros y la estimación precisa de diversidad son prioritarias.

Las mejoras recientes en la química de secuenciación (Kit 14, precisión Q20+) y herramientas como PRONAME han mitigado limitaciones históricas, posicionando a ONT como una opción óptima para caracterizar comunidades microbianas con alta resolución especie específica. Este avance metodológico, combinado con estrategias de depleción de ADN hospedador (e.g., PNA clamps), proporciona un marco robusto para explorar la relación entre la microbiota endófito y la competencia embriogénica en *P. radiata*.

En esta tesis, aplicaremos técnicas de secuenciación de amplicones bacterianos, enfocadas en el operón 16S-ITS-23S mediante plataformas de lecturas largas como Oxford Nanopore, para explorar la composición

taxonómica del microbioma asociado a líneas celulares embriogénicas somáticas juveniles de *P. radiata*. Sin embargo, este análisis enfrenta un desafío técnico inherente a la biología vegetal: la alta abundancia de ADN organelar (cloroplástico y mitocondrial), que puede representar hasta el 99% del material genético total en tejidos vegetales, debido a la estructura multicopiada de estos organelos (Lefèvre et al., 2020a; Song & Xie, 2020). Estas secuencias organelares, que contienen genes 16S rRNA homólogos a los bacterianos, interfieren con la amplificación específica de señales microbianas, enmascarando comunidades endófitas minoritarias y reduciendo la sensibilidad del metabarcoding tradicional. Nuestro enfoque, optimizado con bloqueadores PNA y pipelines bioinformáticos personalizados, permitirá un análisis exhaustivo de la diversidad taxonómica y la identificación del microbioma núcleo (core), proporcionando una visión inicial sobre las funciones de los microorganismos en la interacción planta-bacteria durante la embriogénesis somática in vitro.

III. HIPÓTESIS DE TRABAJO:

La embriogénesis somática (ES) en *P. radiata* enfrenta desafíos significativos, como la pérdida de la capacidad de maduración durante los subcultivos prolongados (Klimaszewska et al., 2009; Pullman et al., 2015). El microbioma asociado juega un papel esencial en el desarrollo de las plantas (Berg et al., 2014), y ciertos microorganismos están presentes en el proceso de micropropagación de *Pinus spp.* (Pirttilä et al., 2000; Ganley et al., 2015). Sin embargo, la identificación y caracterización de estos microorganismos beneficiosos y su impacto en la capacidad de maduración en cultivos embriogénicos somáticos siguen siendo áreas de investigación inexploradas. El microbioma núcleo se define como el conjunto de taxones microbianos que están consistentemente presentes en una especie a lo largo de distintas condiciones ambientales, y cuya presencia sugiere un rol funcional importante en la fisiología o adaptación del hospedero (Shade & Handelsman, 2012). Por lo tanto surge la siguiente pregunta de investigación: "¿Qué características taxonómicas (composición y diversidad alfa y beta) diferencian al microbioma núcleo asociado a líneas embriogénicas somáticas juveniles de *P. radiata* con alta y baja capacidad de maduración?"

Hipótesis

Las líneas embriogénicas somáticas juveniles de *Pinus radiata* con alta capacidad de maduración presentan un microbioma núcleo con una diversidad taxonómica (alfa y beta) significativamente mayor y composición enriquecida en bacterias endófitas promotoras del crecimiento vegetal (≥ 10 % de abundancia relativa), en comparación con líneas de baja capacidad embriogénica.

IV. OBJETIVOS:

Objetivo general:

Evaluar la diversidad, composición y estructura funcional del microbioma núcleo bacteriano (prevalencia ≥ 95 %) asociado a líneas embriogénicas de *Pinus radiata* con alta y baja capacidad de maduración, mediante secuenciación Nanopore del operón 16S–ITS–23S y análisis taxonómico-funcional con la base rEGEN.

4. Objetivos específicos:

4.1. Optimización de la extracción de ADN y PCR:

Fundamento: La alta proporción de ADN cloroplástico/mitocondrial y los inhibidores vegetales complican la recuperación de ADN microbiano de calidad.

Objetivo: Optimizar la extracción de ADN y amplificación del operón 16S–ITS–23S mediante columnas y PNA clamps, asegurando ADN microbiano de alto peso molecular y reduciendo co-amplificación vegetal, para permitir la caracterización comparativa del microbioma entre líneas embriogénicas de alta y baja capacidad de maduración.

4.2. Análisis de la diversidad y composición del microbiota bacteriano:

Fundamento: La diversidad alfa y beta pueden reflejar la influencia del microbioma en la capacidad embriogénica.

Objetivo: Analizar la diversidad alfa (Shannon, Observed) y beta (Bray-Curtis, PCoA) del microbioma bacteriano, comparando líneas de alta y baja capacidad

de maduración e identificando taxones clave asociados a la eficiencia embriogénica.

4.3. Identificación y funcionalización del microbiota núcleo ("core"):

Fundamento: La prevalencia se define como el porcentaje de muestras en las que un taxón es detectado. En este trabajo, se considerará como microbioma núcleo aquel conjunto de taxones con prevalencia $\geq 95\%$, es decir, presentes de manera consistente en prácticamente todas las líneas embriogénicas evaluadas. Estos taxones núcleo, al estar ampliamente distribuidos y mantenerse estables entre condiciones experimentales, son candidatos para modular funciones esenciales para el desarrollo embrionario.

Objetivo: Identificar el microbioma núcleo (taxones con prevalencia $\geq 95\%$ y abundancia promedio $\geq 1\%$) en líneas embriogénicas de alta y baja maduración, evaluando su potencial rol funcional en el desarrollo embrionario y su asociación con la capacidad de maduración.

V. METODOLOGÍA:

4.4. Optimización de la extracción de ADN:

4.5. Material vegetal:

Los megagametofitos inmaduros de *P. radiata* fueron recolectados de un huerto de cruzamiento controlado ubicado en la región del Bío-Bío, Chile, propiedad de la empresa ARAUCO S.A. Tras la colecta, los megagametofitos se almacenaron en frío a 4 °C en bolsas de polietileno perforadas hasta su procesamiento. Se seleccionaron megagametofitos hembras verdes de 17 meses de edad de 6 familias de *P. radiata* polinizadas de manera controlada, asegurando que el estado fisiológico del embrión cigótico estuviera en condiciones óptimas (Montalbán et al., 2012).

4.6. Líneas celulares de embriogénesis somática

Las semillas sin embriones fueron disecadas y se eliminaron sus cubiertas. Los megagametofitos se esterilizaron según el protocolo de Montalbán et al. (2016) y se colocaron en medio de iniciación Glitz suplementado con Gelrite, auxina 2,4-D y citoquinina BA. Se utilizaron 800 megagametofitos en total, distribuidos en 10 placas de Petri.

Las masas embrionarias (ME) proliferantes se separaron y se cultivaron en medio de proliferación fresco cada 2 semanas hasta la maduración. Después de tres



subcultivos, las ME en crecimiento se consideraron líneas celulares embriogénicas (ECL) establecidas. Se mantuvieron 50 ECL por subcultivo en medio de iniciación Glitz con mayor concentración de Gelrite. Al finalizar el tercer subcultivo, se evaluó la capacidad de maduración de las ECL en base a N° Emb Tot/g.

4.7. Selección de partidores in silico para PCR:

Se realizó una búsqueda exhaustiva en bases de datos científicas (PubMed, Google Scholar, Web of Science) utilizando términos como “16S full length primers”, “16S primer set”, “full 16S-ITS-23S sequencing”, “16S rRNA”, “Endophytes”, “Oxford nanopore 16S sequencing” y “16S-ITS-23S oxford nanopore”. Como referencia inicial, se identificó el estudio de Klindworth et al., (2013), que contiene un amplio listado de partidores. Esto facilitó la identificación de publicaciones relevantes que cumplieran con los criterios de inclusión: (1) pares de partidores diseñados para amplificar regiones ≥ 1500 pb del gen 16S rRNA (V1-V9) o ~ 4300 pb del operón 16S-ITS-23S y (2) secuencias únicas no redundantes (i.e., sin repeticiones exactas de pares forward/reverse). Las secuencias listadas en la Tabla 1 corresponden exclusivamente a las regiones de unión al ADN bacteriano, sin incluir adaptadores de secuenciación (e.g., barcodes o colas para nanopore), los cuales se añaden en etapas posteriores a la amplificación por PCR.

Tabla 1 Pares de partidores evaluados para la amplificación de regiones largas del 16S rRNA

N°	Región	Forward (5'→3')	Reverse (5'→3')	Longitud (pb)	Referencia
1	V1-V9 (16S rRNA)	ACGGGNGGCWGCAG	GACTACHVGGGTATCTAA	~1500	(Hlangwani et al., 2023)
2	V1-V9 (16S rRNA)	AGRGTGGATYMTGGCTCAG	GGYTACCTGTTACGACTT	~1500	(Schloss et al., 2016)
3	V1-V9 (16S rRNA)	GAGTTTGATCMTGGCTCAG	TACCTTGTACGACTT	~1500	(Knafo et al., 2023)
4	V1-V9 (16S rRNA)	AGAGTTTGATCMTGGCTCAG	CGGTACCTGTTACGACTT	~1500	(Waechter et al., 2023)
5	V1-V9 (16S rRNA)	AGRGTGGATYMTGGCTCAG	CGGYTACCTGTTACGACTT	~1500	(Matsuo et al., 2021)
6	V1-V9 (16S rRNA)	AGRGTGGATYMTGGCTCAG	RGYTACCTGTTACGACTT	~1500	(Notario et al., 2023)

Tabla 2 Pares de partidores evaluados para la amplificación de regiones del operón 16S-ITS-23S

N°	Región	Forward (5'→3')	Reverse (5'→3')	Longitud (pb)	Referencia
7	16S-ITS-23S rRNA	AGRGTGGATYMTGGCTCAG	TACCTTGTACGACTT	~4,300	(Stevens et al. 2023)
8	16S-ITS-23S rRNA	TTTCTGTTGGTGCTGATATTGC	ACTTGCCTGTCGCTCTATCTTC	~4,300	(Olivier et al., 2023)
9	16S-ITS-23S rRNA	AGRGTGGATYHTGGCTCAG	ACCRCCCCAGTHRAACT	~4,300	(Seol et al., 2022)
10	16S-ITS-23S rRNA	AGRGTGGATYMTGGCTCAG	CGACATCGAGGTGCCAAAC	~4,500	(Lengrand et al., 2024)
11	16S-ITS-23S rRNA	CAGCMGCCGCGTAA	CCRAMCTGTCTCACGACG	~4,500	(Srinivas et al., 2025b)
12	16S-ITS-23S rRNA	GCTGGATCACCTCTTTCT	GGTACTAGATGTTTCAGTTCC	~4,300	(Stoppioni et al., 2023)
13	16S:27F - 23S:2241R	AGRGTGGATYHTGGCTCAG	ACCRCCCCAGTHAACT	~4,300	(Cusco et al., 2018)
14	16S:27F - 23S:2428R	AGRGTGGATYHTGGCTCAG	CCRAMCTGTCTCACGACG	~4,300	(Walsh et al., 2024)
15	16S:519F - 23S:2428R	CAGCMGCCGCGTAA	CCRAMCTGTCTCACGACG	~4,300	(Martijn et al., 2019)

Para evaluar la cobertura y especificidad bacteriana de los partidores dirigidos al gen 16S rRNA (Tabla 1), se utilizó la herramienta TestPrime con la base de datos SILVA SSU r138.1 (Quast et al., 2012). En el caso de los partidores del operón 16S-ITS-23S (Tabla 2), dado que la base de datos SILVA no contiene el operón completo, se evaluaron los primers forward utilizando la base SSU y los reverse

con la base LSU de forma independiente, siguiendo recomendaciones previas para ensayos *in silico* de cobertura.

Adicionalmente, se empleó PrimerMap para simular la amplificación *in silico* de los partidores con mayor cobertura con 0, 1 y 2 mismatches, con el objetivo de evaluar su especificidad frente a secuencias de cloroplasto y mitocondria de *pinus ssp.* Se utilizaron como referencia *Pinus taeda* chloroplast, complete genome (>KY964286.1) y *Pinus Taeda* mitochondrion, complete genome (>NC_039746.1).

4.8. Extracción de ADN

La extracción de ADN de alta calidad a partir de callos embriogénicos de *P. radiata* enfrentó desafíos técnicos significativos debido a la alta proporción de ADN vegetal (cloroplástico/mitocondrial) y la presencia de inhibidores de PCR. Inicialmente se evaluaron múltiples métodos:

1. **Protocolo LEV (CTAB/SDS):** Se implementó con modificaciones (esferas de tungsteno, buffer API-SDS/PVP 4%, basado en columnas), pero requirió optimización en pretratamientos.
2. **Kits comerciales:** Se testearon *PureLink Microbiome*, *Quick-DNA Plant/Seed Miniprep* y *ZymoBIOMICS*, variando parámetros de molienda (esferas cerámicas/tungsteno, tiempos 0-5 min) y masas de muestra (50-200 mg).

Tras evaluar la eficiencia mediante electroforesis y amplificación PCR, se adoptó como estrategia principal el uso de **ADN extraído por BIOFOREST**, obtenido mediante un kit comercial basado en columnas (equivalente funcional a *DNeasy Plant Mini Kit*), optimizado para tejidos vegetales complejos.

Las muestras procesadas con este método (32 clones: 15 alta/17 baja capacidad madurativa) mostraron idoneidad para amplificación de regiones largas (16S-ITS-23S) y secuenciación Nanopore.

5. Optimización de las condiciones de PCR

5.1. Evaluación inicial con partidores 27F-1492R

La optimización de las condiciones de PCR se abordó mediante un proceso iterativo que comenzó con el par de partidores 27F-1492R dirigidos a la región V1-V9 del gen 16S rRNA. Inicialmente se evaluaron parámetros clave como temperatura de alineamiento (51-55°C), concentraciones de PNA (0-3 µM) y condiciones de clamping (68-76°C) para maximizar la especificidad bacteriana y minimizar la co-amplificación de ADN vegetal. Estos ensayos fueron monitoreados mediante electroforesis y secuenciación Sanger.

5.2. Optimización con partidores 27F-2241R para el operón 16S-ITS-23S

La transición a los partidores 27F-2241R (~4.2 kb) requirió adaptaciones significativas en el protocolo. Inicialmente, se evaluó la polimerasa GoTaq® G2 Flexi con tiempos de extensión estándar (90 s/ciclo), observándose amplificación incompleta y bandas difusas. Basado en recomendaciones para fragmentos largos (1 min/kb), se incrementó el tiempo de extensión a 270 s/ciclo, lo que evidenció las limitaciones de esta polimerasa para productos >4 kb.

La sustitución por Q5® High-Fidelity DNA Polymerase, diseñada para amplificación de fragmentos largos (>20 kb), resolvió estos problemas. Se realizaron pruebas experimentales evaluando el número de ciclos de amplificación (25 y 30 ciclos) y la temperatura de alineamiento (55°C, 57°C y 59°C). Estas pruebas se ejecutaron bajo condiciones estandarizadas que incluían una extensión prolongada de 270 segundos por ciclo utilizando la polimerasa Q5® High-Fidelity, manteniendo constante la concentración asimétrica de primers (0.8 µM para el cebador forward y 0.4 µM para el reverse).

Las condiciones finales incluyeron: desnaturalización inicial (98°C × 3 min), ciclos de amplificación (35) con desnaturalización (98°C × 30 s), alineamiento (60°C × 30 s) y extensión extendida (72°C × 150 s) a un volumen final de 10µL. Este enfoque metodológico se aplicó consistentemente al ADN extraído proporcionado

por BIOFOREST, estableciendo las bases para los análisis posteriores de microbiota.

6. Secuenciación por Oxford nanopore

6.1. Preparación de librerías de secuenciación

En la primera secuenciación (prueba piloto) se trabajó con ocho muestras representativas de cuatro clones de *P. radiata*, multiplexadas con los códigos 1 a 8. Los códigos 1–4 correspondieron a muestras tratadas con PNAs (1 μM) y los códigos 5–8 a los controles sin PNAs. Las bibliotecas de secuenciación se prepararon mediante el kit SQK-NBD114.24 (Oxford Nanopore Technologies). Para cada una, se partió de amplicones del operón 16S–ITS–23S, cuantificados por Qubit y purificados con perlas AMPure XP; solo se aceptaron bibliotecas con concentraciones $\geq 20.5 \text{ ng}/\mu\text{L}$).

Dado que el protocolo SQK-LSK114 requiere 200 fmol de amplicón por muestra (equivalente a aproximadamente 130 ng para un producto de 1 kb) en un volumen máximo de 11.5 μL , se utilizó la herramienta NEBioCalculator (New England Biolabs) para convertir las concentraciones en $\text{ng}/\mu\text{L}$ a fmol, de modo de calcular con precisión los volúmenes necesarios de ADN a incorporar en cada librería. La mezcla final para la reparación de extremos y dA-tailing fue llevada a 15 μL totales, compuesta por 11.5 μL de amplicón, 1 μL de DCS, 1.75 μL de End-Prep Buffer y 0.75 μL de End-Prep Enzyme Mix. Este paso se realizó utilizando el



módulo NEBNext Ultra II End Repair/dA-Tailing, incubando a 20 °C por 5 minutos seguido de 65 °C por 5 minutos. Posteriormente, se ligaron los adaptadores y códigos de barras nativos V14 (25 °C × 15 min, 80 °C × 10 s). Todos los productos purificados fueron validados mediante Qubit Fluorometer (Thermo Fisher Scientific). La corrida de secuenciación se ejecutó durante 3 horas en un dispositivo MinION Mk1C utilizando un flowcell Flongle R10.4.1 (~1400 poros activos)

En la segunda secuenciación (corrida completa) se extendió la misma metodología de preparación de librerías, esta vez aplicándola a 19 muestras de ADN genómico: siete de alta capacidad madurativa (IB 285, OAC 111, MWB 161, MWB 163, MWB 213 × 2 réplicas, OAL 101) y doce de baja (IB 278 × 2, IB 288, IB 291, OAC 103, OAC 106, MWB 160, MWB 162, MWB 165, MWB 204, OAL 104 × 2). Tras purificar y cuantificar (criterios idénticos a la prueba piloto) se ejecutó la corrida de secuenciación por 48 h, se reutilizó el mismo flowcell Flongle R10.4.1 (~600 poros activos).

6.2. Datos crudos y preprocesamiento al análisis taxonómico

Los datos crudos obtenidos mediante secuenciación Nanopore (Flongle R10.4.1, Kit SQK-NBD114.24) fueron procesados mediante dos estrategias bioinformáticas para el análisis metataxonómico de amplicones largos (~2.4 kb). En ambos casos, las señales eléctricas fueron convertidas en secuencias

nucleotídicas (basecalling) utilizando el modelo SUP implementado en Dorado v0.7.8.3, generando archivos FASTQ automáticamente demultiplexados por barcode.

6.3. Flujo de trabajo wf-metagenomics por EPI2ME

La primera estrategia de análisis consistió en la implementación directa del flujo de trabajo wf-metagenomics desarrollado por Oxford Nanopore Technologies, diseñado específicamente para el análisis de comunidades microbianas. El flujo de trabajo wf-metagenomics desarrollado por Oxford Nanopore Technologies incluye dos alternativas principales para el procesamiento de lecturas: (i) Kraken2, un clasificador taxonómico que asigna secuencias a taxones mediante comparación de k-mers contra bases de datos predefinidas, y (ii) Minimap2, un mapeador de secuencias que alinea lecturas de ADN o ARN frente a una base de referencia, permitiendo identificar el grado de similitud y cobertura. En este estudio, ambos enfoques se aplicaron de manera complementaria: Kraken2 fue empleado para obtener asignaciones taxonómicas rápidas y globales, mientras que Minimap2 permitió validar dichas asignaciones mediante alineamientos directos. La comparación de resultados entre ambos métodos se consideró como un ejercicio de complementariedad, con el fin de

seleccionar la estrategia que maximizara la resolución taxonómica y la calidad de los alineamientos en este sistema biológico.

Tabla 3 Configuraciones evaluadas para optimización (Secuenciación 1)

Nombre del flujo	Base de datos	Longitud (pb)	Identidad mínima	Clasificador
wf-metagenomics_01	Standard-8	Sin filtro	90%	Kraken2
wf-metagenomics_09	ncbi_16s_18s_28s_ITS	2000–5000	90%	Kraken2
wf-metagenomics_12	ncbi_16s_18s_28s_ITS	Sin filtro	85%	Minimap2

6.4. Flujo de trabajo personalizado basado en RESCUE

Se desarrolló un flujo de trabajo personalizado basado en el pipeline RESCUE (RRN-operon Enabled Species-level Classification Using EMU) (Petrone et al., 2023), optimizado para lecturas largas y orientado a maximizar la precisión taxonómica. Las lecturas generadas por Dorado fueron filtradas mediante NanoFilt v2.8.0, reteniendo solo aquellas con Q-score promedio ≥ 9 y longitud entre 2000 y 5000 pb. Para cada muestra (barcodes 1–8), se concatenaron los archivos provenientes de distintas corridas, y se calcularon estadísticas descriptivas (número de lecturas, longitud media, mínima, máxima y mediana) como control de calidad. El trimming se realizó con Cutadapt v1.18, exigiendo la presencia del barcode forward al inicio y del reverse (en forma complementaria) al final, con hasta un 5 % de error permitido. Estos pasos de filtrado y trimming se

aplicaron únicamente a las rutinas de análisis de 16S mediante RESCUE como parte del preprocesamiento de lecturas crudas antes de la clasificación taxonómica, y no se aplicaron a los análisis de wf-metagenomics.

Para detectar y eliminar posibles contaminaciones endógenas, se diseñó un protocolo en dos fases usando una base de datos local con secuencias organelares de *Pinus taeda* y *P. radiata* (mitocondrial y cloroplastidial). Primero, se seleccionó aleatoriamente el 10 % de las lecturas con seqkit sample, se convirtieron a FASTA y se alinearon con BLASTn v2.14.1. Se consideraron contaminantes aquellas lecturas con identidad ≥ 99 % y longitud ≥ 1000 pb. Luego, el alineamiento se aplicó al total de lecturas, eliminando las secuencias contaminantes mediante seqkit grep para generar archivos FASTQ depurados. Se elaboró un reporte por muestra, indicando el número de lecturas evaluadas y el porcentaje de alineamientos significativos frente a cada secuencia de referencia.

6.5. Filtrado de secuencias mitocondriales

Con el fin de garantizar la precisión del análisis taxonómico bacteriano y excluir posibles contaminaciones endógenas, se implementó un enfoque en dos etapas para la detección y eliminación de lecturas de origen mitocondrial derivadas del hospedero *P. radiata*. Como referencia se utilizó la secuencia completa del

genoma mitocondrial de *Pinus taeda* (NC_039746.1), dada su alta conservación y disponibilidad pública.

En la primera etapa, se seleccionó aleatoriamente un 10 % del total de lecturas procesadas según el punto 6.4 por muestra, sin aplicar filtrado previo, utilizando seqkit sample. Estas lecturas se convirtieron a formato FASTA y se alinearon mediante BLASTn (BLAST+ v2.13.0) contra una base de datos local construida a partir del genoma mitocondrial de referencia. Se consideraron lecturas mitocondriales aquellas con identidad $\geq 99\%$ y longitud de alineamiento ≥ 1000 pb.

En la segunda etapa, se aplicó un filtrado exhaustivo sobre el 100 % de las lecturas procesadas según el punto 6.4. Cada archivo de lectura fue alineado nuevamente mediante BLASTn utilizando los mismos parámetros estrictos ($\geq 99\%$ de identidad, ≥ 1000 pb de cobertura). Las lecturas que cumplieron ambos criterios fueron descartadas del conjunto de datos, asegurando así la eliminación efectiva de secuencias mitocondriales.

6.6. Análisis taxonómico

Una vez completada la eliminación de lecturas mitocondriales derivadas del hospedero, se procedió al análisis taxonómico de las muestras empleando Minimap2. El diseño fue observacional, con procesamiento *in silico* de datos de secuenciación de lectura larga generados mediante tecnología Oxford Nanopore.

El análisis bioinformático se inició alineando las lecturas filtradas contra bases de datos taxonómicas preconstruidas: EMU, rEGEN y GROND, en una versión compatible con la infraestructura computacional para evitar errores de alineamiento. Se empleó un umbral mínimo de abundancia de 0.00001 para descartar especies no significativas.

Durante la primera secuenciación, se utilizaron las tres bases de datos mencionadas con el objetivo de comparar su rendimiento en cuanto a resolución taxonómica, diversidad y calidad de alineamiento. En contraste, para la segunda secuenciación, se optó por emplear únicamente la base de datos rEGEN, dada su mayor capacidad para aprovechar lecturas largas del operón 16S-ITS-23S y su mejor desempeño en la detección de diversidad microbiana.

El análisis se diseñó para comparar la detección y abundancia de especies bacterianas entre condiciones experimentales (presencia/ausencia de PNA y capacidad madurativa alta/baja), con el fin de evaluar el efecto de estas variables sobre la composición bacteriana detectada. Para ello, se calcularon las diferencias promedio en abundancia relativa y absoluta entre grupos, seleccionándose las diez especies con mayor contraste. La elección de este número se fundamentó en que, al considerar un mayor conjunto (20 o 30 especies), la mayoría presentaba abundancias absolutas cercanas o inferiores a 10 lecturas, o relativas menores al 0.1 %, lo que reducía la solidez estadística y

la relevancia biológica de los resultados. Posteriormente, se aplicó el test no paramétrico de Mann–Whitney U, ya que las abundancias no seguían una distribución normal y los valores correspondían a taxones seleccionados, no a réplicas biológicas independientes. Bajo estas condiciones aumentar el número de especies no habría mejorado el poder estadístico al tratarse de taxones de muy baja abundancia. Los análisis de alineamiento se realizaron utilizando minimap2 v2.28 (Li, 2018) en el entorno Python v2.6.6. Los análisis estadísticos posteriores se efectuaron en Python v3.12.5, empleando las librerías pandas v2.2.3 (McKinney, 2010) y SciPy v1.15.3 (Virtanen et al., 2020), obteniéndose valores de p para cada especie evaluada mediante el test no paramétrico de Mann–Whitney U. Se priorizaron aquellas especies que presentaron diferencias significativas ($p < 0.05$) tanto en abundancia absoluta como relativa, reforzando la interpretación biológica de los resultados.

Finalmente, se generaron visualizaciones gráficas para ilustrar las diferencias en abundancia de las especies seleccionadas, separando los grupos según la condición PNA y la capacidad madurativa

6.7. Diversidad alfa

Para estimar la diversidad interna de las comunidades bacterianas en cada muestra, se calcularon múltiples índices de diversidad alfa a partir de las matrices de abundancia relativas por especie. Los cálculos se realizaron en Python,

utilizando la biblioteca scikit-bio (v0.5.5; The scikit-bio development team, 2020), e incluyeron métricas de riqueza observada (Observed), diversidad de Shannon, estimador de Chao1 y el índice de Simpson. Las matrices de abundancia relativas se obtuvieron a partir de los archivos SAM generados por Minimap2 (v2.28; Li, 2018). Para cada muestra, se contaron las lecturas alineadas a cada especie y se dividieron por el total de lecturas asignadas, generando así abundancias relativas utilizadas en los cálculos de diversidad alfa. Los datos se agruparon según la condición de capacidad madurativa (Alta o Baja) de cada muestra. La distribución de valores de cada índice se visualizó mediante diagramas de caja (boxplots), lo que permitió identificar tendencias globales entre condiciones. No se aplicó rarefacción previa, ya que métricas como Shannon, Simpson y Chao1 son robustas ante diferencias moderadas en la profundidad de secuenciación (Chao et al., 2014; Gotelli & Colwell, 2001).

6.8. Diversidad beta

Para evaluar diferencias en la composición microbiana entre los grupos experimentales, se calculó la diversidad beta utilizando la distancia de Bray-Curtis sobre matrices de abundancia relativas y absolutas por especie. A partir de estas matrices, se aplicaron análisis de varianza permutacional (PERMANOVA) con 999 permutaciones para tres niveles de agrupación: por Familia (IB, OAC, MWB, OAL), por condición de capacidad madurativa (Alta,



Baja), y por la combinación de ambas variables (Grupo). Cada nivel se analizó mediante comparaciones independientes; no se implementó un diseño anidado ni se evaluaron interacciones entre factores, de modo que cada PERMANOVA cuantificó la contribución específica de un factor a la disimilitud entre muestras. Los análisis se realizaron utilizando el módulo permanova de scikit-bio (v0.5.5; The scikit-bio development team, 2020) en Python. Para visualizar la distribución de las muestras en el espacio multivariado definido por las distancias de Bray-Curtis, se generaron gráficos de ordenación PCoA (Análisis de Coordenadas Principales) mediante scikit-bio y las librerías Matplotlib (v3.10.3) y Seaborn (v0.13.2) en Python.

7. Identificación y caracterización de especies bacterianas núcleo ("core"):

En este estudio, se definió como prevalencia de una especie la proporción de muestras de un mismo grupo en las que la especie está presente (abundancia relativa > 0), un indicador de consistencia o constancia en la comunidad. La abundancia relativa promedio corresponde a la media de las proporciones de cada especie calculadas sobre el total de lecturas por muestra, reflejando su contribución media al microbioma.

La identificación del microbioma núcleo bacteriano asociado a cada condición de capacidad madurativa (Alta o Baja) se basó en estos dos criterios cuantitativos complementarios:

- Prevalencia $\geq 0,95$: especie presente en al menos el 95 % de las muestras de un mismo grupo.
- Abundancia relativa promedio $\geq 0,01$: especie cuya abundancia relativa promedio supera el 1 %.

Para calcular estas métricas, los archivos CSV de abundancias por especie generados a partir de los archivos SAM de Minimap2 se cargaron en Python (v3.8), utilizando las librerías pandas (v1.3; McKinney, 2010) y NumPy (v1.21; Harris et al., 2020). Primero se construyó una matriz de conteos absolutos y luego se normalizó por el total de lecturas de cada muestra para obtener la matriz de abundancia relativa. A continuación, se calculó la prevalencia de cada especie como el número de muestras con abundancia relativa mayor a cero, y la abundancia relativa promedio como la media de las proporciones correspondientes. Finalmente, se aplicaron los umbrales establecidos para delimitar el conjunto de especies núcleo en cada condición, obteniendo la lista con sus abundancias absolutas respectivas según condición.

8. Análisis funcional del microbiota núcleo ("core")

A las especies identificadas como microbiota núcleo (prevalencia $\geq 95\%$, abundancia relativa promedio $> 1\%$) se les asignaron funciones biológicas mediante la integración de información procedente de bibliografía especializada, genomas de referencia y la base de datos funcional FAPROTAX (v1.2; Louca et al., 2018). Para la obtención de literatura relevante y evidencia experimental se emplearon herramientas basadas en inteligencia artificial y búsqueda académica: Elicit.org, Scite.ai y Research Rabbit, mientras que los genomas de referencia se consultaron en NCBI Genomes/RefSeq.

Cada taxón fue clasificado según su potencial impacto en embriogénesis somática en una de cuatro categorías funcionales: beneficioso, neutro, perjudicial o variable. El análisis combinó anotación automática de rasgos metabólicos, como rutas de degradación de compuestos, producción de fitohormonas o síntesis de polisacáridos protectores, con verificación manual de información experimental documentada en publicaciones revisadas por pares y descripciones taxonómicas de genomas de referencia.

VI. RESULTADOS:

9. Optimización de la extracción de ADN y la amplificación por PCR:

9.1. Líneas celulares de embriogénesis somática:

A partir de un grupo inicial de 60 líneas celulares embriogénicas (ECL) inducidas a partir de megagametofitos inmaduros de *P. radiata*., evaluamos la capacidad de maduración en función de la eficiencia de producción de embriones. El análisis cuantitativo de embriones normales, anormales y totales por gramo (Emb. Norm/g, Emb. Anorm/g, Emb. Tot/g) reveló una variabilidad estadísticamente significativa entre clones (Figura suplementaria 1). La producción de embriones totales varió desde $12,5 \pm 12,5$ Emb. Tot/g (MWB162) hasta $1.260,4 \pm 102,8$ Emb Tot/g (MWB168), demostrando una variación de 100 veces entre linajes.

Para identificar modelos contrastantes que permitan estudiar los mecanismos de maduración, seleccionamos 32 clones de seis familias (MWB, OAC, IB, MUZ, OAL, MSU) basándonos en sus valores de Emb. Tot/g (Figura 6). Los criterios de selección fueron:

- Alta capacidad madurativa: >600 Emb. Tot/g (n=15 clones)
- Baja capacidad madurativa: <200 Emb. Tot/g (n=17 clones)

En la figura suplementaria 2 se incluye un gráfico detallado de la capacidad madurativa de las ECL seleccionadas y utilizadas tanto para la primera secuenciación como para la segunda secuenciación por Oxford Nanopore.

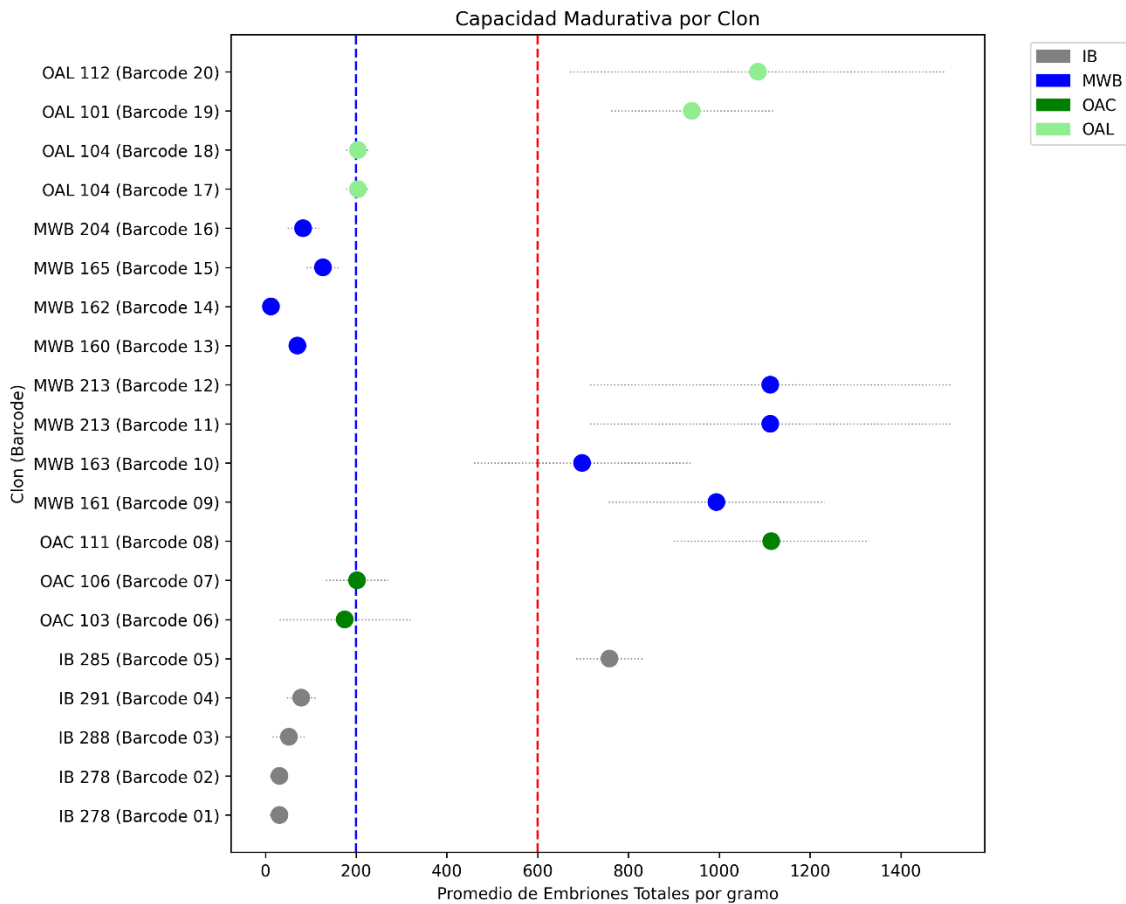


Figura 6 Clones seleccionados para el ensayo de secuenciación. Los colores representan las diferentes familias clonales: IB (gris), MWB (azul), OAC (verde) y OAL (verde claro). Se muestra la capacidad madurativa de 20 clones ordenados por código de barcode. En el eje horizontal se presenta el promedio de embriones totales por gramo. Cada valor corresponde a N = 3 repeticiones biológicas, con el promedio de 2 réplicas técnicas por placa. Las réplicas técnicas se muestran en filas consecutivas para facilitar la comparación. Las barras horizontales punteadas indican la desviación estándar entre repeticiones biológicas. Las líneas verticales demarcan los rangos de capacidad madurativa alta (>600) y baja (<200).

9.2. Selección de partidores in silico para PCR:

La evaluación in silico de siete combinaciones de partidores dirigidos al gen 16S rRNA mediante *TestPrime* y la base de datos SILVA reveló

diferencias importantes en la cobertura taxonómica bacteriana (Tabla 4). Los partidores 1, 3 y 6 destacaron con coberturas superiores al 69% sin mismatches y hasta 87.5% permitiendo dos mismatches. En contraste, el partidador 4 mostró la cobertura más baja con 0 mismatches (16.5%), aunque esta aumentó significativamente al permitir errores, alcanzando un 76.0% con 2 mismatches.

Tabla 4 valuación *in silico* de partidores dirigidos al gen 16S rRNA (V1–V9) mediante TestPrime (SILVA v138.1). Se muestra el porcentaje de cobertura bacteriana obtenido para los partidores 1 al 6, permitiendo 0, 1 y 2 mismatches durante el alineamiento.

c	Primer name and Sequence		Test Prime Silva		
	N°	Forward Primer	Reverse Primer	0 Mismatch	1 Mismatch
Cobertura				Cobertura	Cobertura
1	CCT ACGGGNGGCWGCAG	GACTACHVGGGTAT CTAATCC	Match: 79.2%	Match: 85.3%	Match: 87.5%
2	AGRGTTTGATYMTGGCTCAG	GGYTACCTTGTTACGACTT	Match: 68.6%	Match: 77.0%	Match: 79.2%
3	GAGTTTGATCMTGGCTCAG	TACCTTGTTACGACTT	Match: 74.9%	Match: 81.6%	Match: 83.2%
4	AGAGTTTGATCMTGGCTCAG	CGGTTACCTTGTTACGACTT	Match: 16.5%	Match: 69.0%	Match: 76.0%
5	AGRGTTYGATYMTGGCTCAG	CGGYTACCTTGTTACGACTT	Match: 51.5%	Match: 74.3%	Match: 78.1%
6	AGRGTTYGATYMTGGCTCAG	RGYTACCTTGTTACGACTT	Match: 69.9%	Match: 77.2%	Match: 79.2%

Para los partidores dirigidos al operón completo 16S-ITS-23S, se evaluaron forward y reverse por separado sobre las bases de datos SSU y LSU, respectivamente (Tabla 5). Los partidores forward del 16S presentaron coberturas entre 59.7% y 70.1% sin mismatches, mientras que los reverse del 23S superaron el 86%, alcanzando hasta 94.7% con dos mismatches. Estos resultados respaldan la idoneidad de los partidores seleccionados para capturar eficientemente la diversidad bacteriana a través del operón ribosomal completo.

Tabla 5 Resultados de cobertura *in silico* de partidores dirigidos al operón ribosomal 16S–ITS–23S. Los partidores forward se evaluaron contra la base de datos SSU y los reverse contra la base de datos LSU

de SILVA v138.1, utilizando la herramienta TestPrime. Se reporta el porcentaje de coincidencia con secuencias bacterianas permitiendo 0, 1 y 2 mismatches.

<i>Primer name and Sequence</i>			<i>Test Prime Silva</i>		
N°	Forward Primer	Reference Reverse Primer	0 Mismatch	1 Mismatch	2 Mismatch
			Cobertura	Cobertura	Cobertura
7	AGRGTTYGATYMTGGCTCAG	RGYTACCTTGTTACGACTT	Match: 69.9%	Match: 77.2%	Match: 79.2%
9	AGRGTTYGATYHTGGCTCAG	RGYTACCTTGTTACGACTT	Match: 70.1%	Match: 77.3%	Match: 79.5%
11	CAGCMGCCGCGGTAA	RGYTACCTTGTTACGACTT	Match: 59.7%	Match: 67.0%	Match: 78.2%
13	AGRGTTTGATYHTGGCTCAG	RGYTACCTTGTTACGACTT	Match: 69.4%	Match: 77.2%	Match: 79.5%
<i>Primer name and Sequence</i>			<i>Test Prime Silva</i>		
N°	Reference Forward Primer	Reverse Primer	0 Mismatch	1 Mismatch	2 Mismatch
			Cobertura	Cobertura	Cobertura
9	CYGAATGGGRVAACC	ACCRCCCAGTHRAACT	Match: 87.0%	Match: 88.9%	Match: 93.5%
10	CYGAATGGGRVAACC	CGACATCGAGGTGCCAAAC	Match: 86.7%	Match: 89.4%	Match: 94.0%
11	CYGAATGGGRVAACC	CCRAMCTGTCTCACGACG	Match: 86.4%	Match: 89.5%	Match: 94.7%
13	CYGAATGGGRVAACC	ACCRCCCAGTHAAACT	Match: 86.9%	Match: 88.9%	Match: 93.4%

De los seis pares de partidores seleccionados, tres dirigidos al gen 16S rRNA (pares 1, 3 y 6 de la Tabla 4) y tres al operón ribosomal completo 16S–ITS–23S (pares 9, 11 y 13 de la Tabla 5), el análisis con *PrimerMap* reveló que todos presentan sitios de unión en la secuencia cloroplastidial de *Pinus taeda*, con excepción de los partidores *reverse* de los pares 9 y 13 (“ACCRCCCAGTHRAACT” y “ACCRCCCAGTHAAACT”, respectivamente), los cuales no mostraron alineamiento con dicha secuencia. En cuanto a la mitocondria de *Pinus taeda*, ninguno de los partidores evaluados presentó afinidad de unión, lo que sugiere una baja probabilidad de amplificación inespecífica en este organelo. Se escogió el par número 13 (27f-2241r) para los análisis posteriores.

10. Extracción de ADN

La comparación de múltiples protocolos demostró que, aunque los métodos basados en CTAB/SDS (LEV) y los kits comerciales (ZymoBIOMICS, Quick-DNA Plant/Seed Miniprep) recuperaban, en algunos casos, concentraciones suficientes de ADN (según Qubit™), la calidad y reproducibilidad en callos embriogénicos de *P. radiata* fue insuficiente para amplificaciones largas (>2 kb). En particular, las electroforesis mostraron bandas débiles o muy difusas (Fig. 7A, B), con variaciones marcadas entre réplicas y clones, lo que dificultaba la estandarización de PCRs robustas.

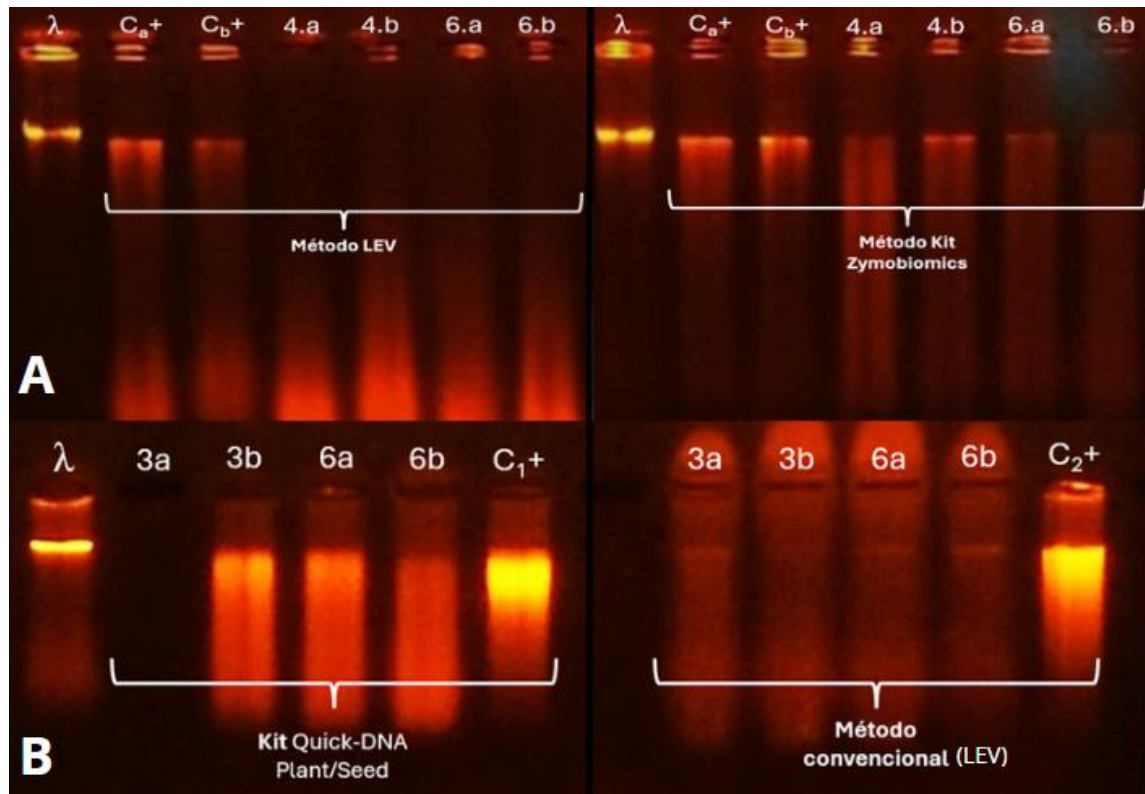


Figura 7 Comparación de métodos de extracción de ADN en tejidos embriogénicos de *Pinus radiata*.
 (A) Kit ZymoBIOMICS vs. método LEV. Marcador: λ -HindIII. Ca⁺/Cb⁺: controles positivos de megagametófitos; 4.a/4.b y 6.a/6.b: genotipos 4 y 6 con 1 y 2 esferas cerámicas; EmbA-B: embriones cigóticos (LEV). (B) Método LEV vs. Quick-DNA Plant/Seed Miniprep. C1/C2: controles positivos de embrión y megagametófitos; 3a/3b y 6a/6b: genotipos “3” y “6”

En contraste, el ADN extraído por Bioforest S.A. (protocolo basado en columnas tipo DNeasy Plant Mini Kit) presentó un perfil de alta integridad, con bandas nítidas y de alto peso molecular, y consistencia notable entre todos los clones evaluados, independientemente de su capacidad madurativa (Fig. 8).

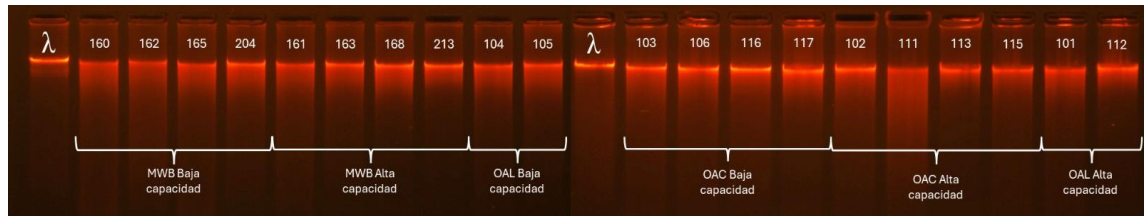


Figura 8 Electroforesis en gel de agarosa (1%), correspondiente a extracciones de ADN genómico de tejidos embriogénicos de *Pinus radiata* realizadas por Bioforest. Las muestras están agrupadas por línea (MWB, OAC, OAL) y capacidad madurativa (alta o baja).

11. Optimización de las condiciones de PCR:

Con el objetivo de optimizar la amplificación del operón ribosomal completo 16S–ITS–23S utilizando el par de partidores 27F–2241R, se evaluaron distintos parámetros de temperatura de alineamiento y número de ciclos. En primer lugar, se probaron temperaturas de alineamiento de 55 °C, 57 °C y 59 °C, combinadas con 25 y 30 ciclos de PCR. Los resultados se muestran en la Figura 9. A 59 °C y 30 ciclos se observó una banda principal nítida en el rango de 2 500–3 000 pb, tanto en el control positivo (*Pseudomonas rhizosphaerae*) como en el ADN

extraído desde muestras embriogénicas de *P. radiata*. Sin embargo, no se logró amplificar el fragmento esperado de aproximadamente 4 200–4 500 pb correspondiente al operón completo. Las bandas observadas podrían representar productos parciales de amplificación o regiones internas del operón.

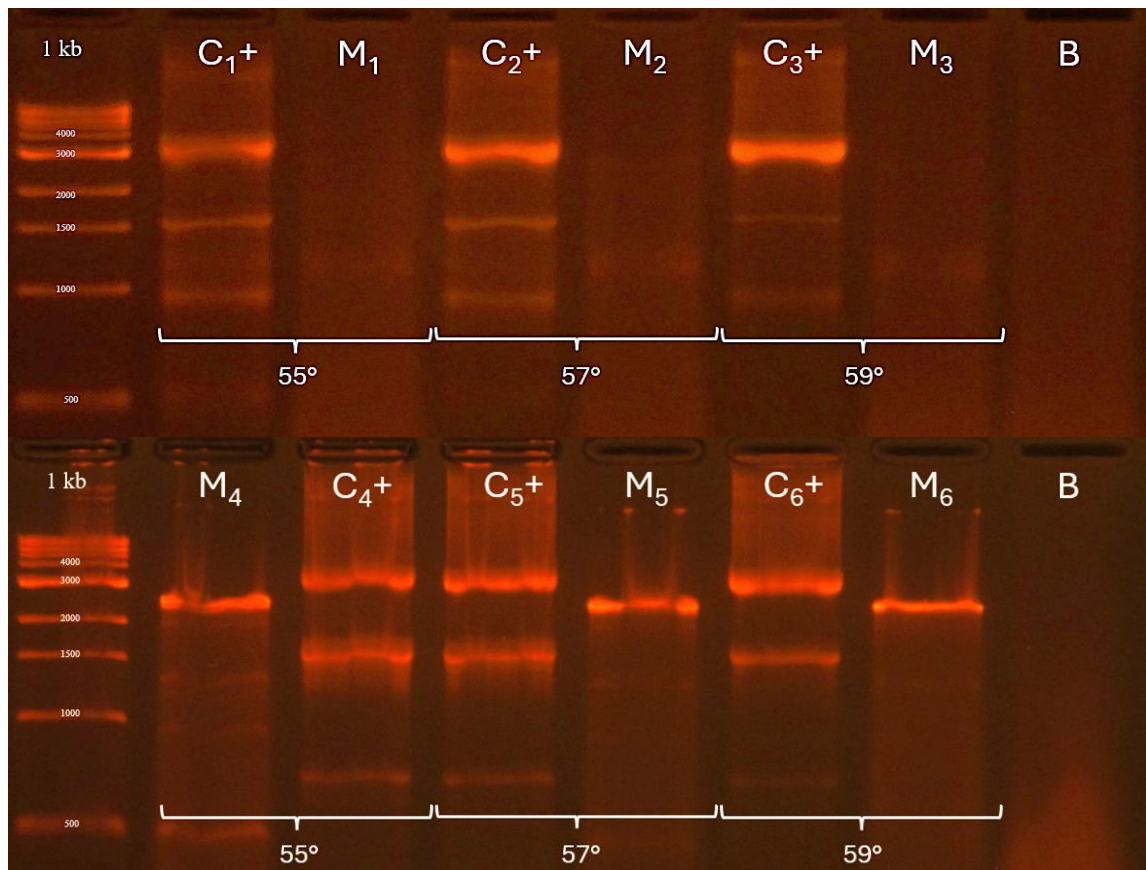


Figura 9 Optimización de la temperatura de alineamiento y número de ciclos para la amplificación del operón 16S–ITS–23S. Reacción de PCR utilizando el par de partidores 27F–2241R en presencia de ADN extraído desde una línea embriogénica de *Pinus radiata* (IB287, Bioforest) y un control positivo bacteriano (*Pseudomonas rhizosphaerae*, código 62). Se evaluaron tres temperaturas de alineamiento (55 °C, 57 °C, 59 °C) y dos condiciones de ciclos (25 y 30 respectivamente). M: ADN de muestra embriogénica; C+: control positivo; B: blanco sin ADN molde.

Posteriormente, se evaluó el efecto del uso de oligonucleótidos bloqueadores PNA (mPNA y pPNA, 2 μ M cada uno) para suprimir la amplificación de ADN vegetal, particularmente mitocondrial y cloroplastidial. Como se observa en la figura 10, sin el uso de PNA se detectaron múltiples productos inespecíficos, incluyendo bandas intensas superiores a 4 000 pb en los controles positivos. Asimismo, se observó un producto adicional cercano a los 2 500 pb, el cual podría representar una amplificación parcial del operón o contaminación organelar. Al incorporar PNA, se redujo la intensidad de las bandas secundarias, se atenuó el producto menor de 2 500 pb y se mejoró la especificidad, lo que sugiere un efecto positivo en la supresión de ADN vegetal contaminante.

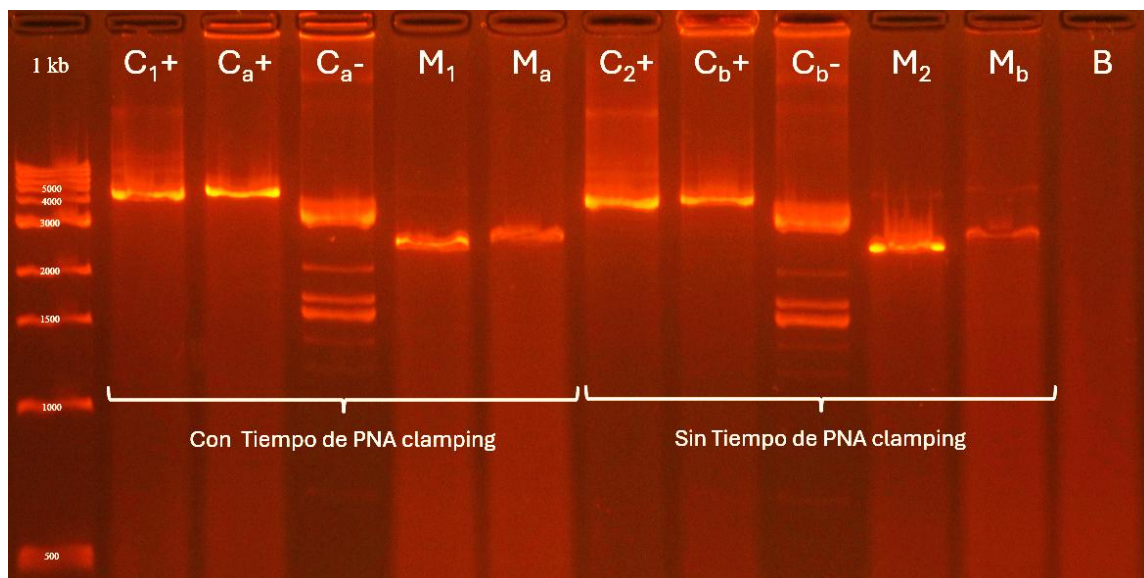


Figura 10 Efecto del uso de oligonucleótidos bloqueadores PNA en la especificidad de la amplificación del operón 16S–ITS–23S. Comparación de productos de PCR obtenidos con y sin adición de PNA (mPNA y pPNA, 2 μ M) en reacciones realizadas sobre ADN de *P. radiata* (IB287) y el control positivo bacteriano (*P. rhizosphaerae*). Se utilizó la misma condición de alineamiento (59 °C) y 30 ciclos de amplificación. M: ADN de muestra embriogénica; C+: control positivo; B: blanco sin ADN molde.

La secuenciación Sanger del amplicón de ~2400 pb reveló que no se obtuvo una amplificación íntegra de ADN cloroplástico ni mitocondrial (Tabla 6). En ambas muestras analizadas, únicamente el partidor forward (27F) presentó alineamiento parcial con la secuencia mitocondrial de *Pinus taeda* (NC_039746.1), con una longitud alineada considerablemente menor que la del amplicón observado en geles de agarosa. El partidor reverse (2241R) no mostró coincidencias significativas con dicha secuencia. Estos resultados sugieren que el producto amplificado corresponde predominantemente a ADN bacteriano y que la posible contaminación vegetal, si presente, es mínima y no representa un fragmento mitocondrial completo. La Figura Suplementaria 3 muestra el alineamiento visualizado mediante MSA Viewer (versión 1.26.0 Yachdav et al., 2016).

Tabla 6 Resultados de alineamiento de secuencias Sanger contra la mitocondria de *Pinus taeda*. Resultados del análisis BLAST de secuencias obtenidas por secuenciación Sanger del amplicón de ~2400 pb amplificado con los partidores 27F y 2241R. Se evaluó si las secuencias alineaban contra el genoma mitocondrial de *Pinus taeda* (accesión NC_039746.1). Se indica si existe alineamiento, la longitud alineada en pares de bases (pb), el porcentaje de identidad y cobertura. Solo el partidor forward (27F) presentó alineamientos significativos, pero en fragmentos mucho menores a la longitud total del amplicón observado en gel.

Muestra	Partidor	Alinea contra "NC_039746.1" <i>Pinus Taeda</i> mitochondrion	Longitud alineada (pb)	Identidad	Cobertura
M1	27f	Si	~470	98,06%	99,36%
M1	2241r	No	-	-	-
M2	27f	Sí	~600	99,54%	95,69%
M2	2241r	No	-	-	-

12. Secuenciación por Oxford Nanopore:

13. Primera secuenciación (prueba piloto):

13.1. Preparación de Librerías de secuenciación

En la prueba piloto de secuenciación, las concentraciones de ADN tras la reparación de extremos y dA-tailing (End-Prep) oscilaron entre 0,178 y 0,361 ng/μL. En seis de las ocho muestras analizadas fue posible alcanzar los 200 fmol requeridos utilizando entre 7,5 y 11,5 μL de ADN. En contraste, las muestras con menor concentración (105P y 105NP) no alcanzaron dicha cantidad, incluso empleando el volumen máximo disponible (11,5 μL), por lo que se procedió con una entrada reducida. La concentración del reactivo “Native Barcode” y de la librería final fue adecuada para la carga en flujo (ver Tabla 7).

Tabla 7 Concentraciones de ADN en distintas etapas del protocolo de preparación de librerías para la prueba piloto de secuenciación Nanopore.

Barcode	Muestra	Condición	Conc. PCR Purificación (ng/μL)	Conc. post End-Prep (ng/μL)	Native Barcode (ng/μL)	Librería final (ng/μL)
01	104P	Con PNA	52,2	25,4	—	—
02	105P	Con PNA	38,0	19,6	—	—
03	101P	Con PNA	54,0	31,4	—	—
04	112P	Con PNA	51,2	24,5	—	—
05	104P	Sin PNA	63,0	22,2	—	—
06	105P	Sin PNA	37,8	17,8	—	—
07	101P	Sin PNA	53,6	36,1	—	—
08	112P	Sin PNA	70,4	22,8	—	—
—	—	—	—	—	69,8	56,6

13.2. Análisis directo con flujo de trabajo wf-metagenomics

Se emplearon tres estrategias de clasificación taxonómica sobre ocho muestras secuenciadas mediante MinION (Flongle R10.4.1). Los resultados se resumen en la Figura 11. El análisis con **Kraken2 + base de datos Standard-8** generó 607,833 lecturas, con alta proporción de no clasificadas en muestras tratadas con



PNAs (38.27%) frente a las no tratadas (34.58%), y mayor en clones de alta capacidad madurativa (39.07%). Se identificó una comunidad bacteriana central de 12 especies, dominada por *Bacillus altitudinis*, *Rhodococcus qingshengii*, *Enterococcus faecium* y *Novosphingobium decolorationis* (35.6%). *Escherichia coli* fue la especie más abundante (13.72%). *Lactiplantibacillus* y *Pseudomonas* se restringieron a muestras con PNAs, mientras *Providencia* y *Stenotrophomonas* predominaron en las no tratadas. La diversidad Shannon fue ligeramente mayor sin PNAs, pero no se observaron diferencias significativas asociadas a la capacidad madurativa, aunque clones de baja capacidad tendieron a mayor diversidad efectiva. Con **Kraken2 + base de datos NCBI_16S_18S_28S_ITS**, se clasificaron 340,594 lecturas, con una proporción de no clasificadas <0.7%. Se identificaron 55 especies, con dominio absoluto de *Caldovatus sediminis* (64.2%), y un núcleo persistente de 15 especies (incluyendo *Aureimonas fodinaquatilis*, *Novosphingobium decolorationis*, *Xanthobacter aminoxidans* y *Sphingobium wenxiniae*) que en conjunto representaron >85% de la abundancia. Las comunidades fueron poco diversas (Shannon: 1.60–1.82, Pielou ≤ 0.44), altamente dominadas (Berger-Parker ≥ 0.61), sin diferencias relevantes por tratamiento ni por capacidad madurativa. El análisis con **Minimap2 + NCBI_16S_18S_28S_ITS** generó 617,882 lecturas, con una proporción variable de no clasificadas (27.38%–55.09%). Se identificó una comunidad central de 19 especies bacterianas y fúngicas dominada por *Sphingomonas sanxanigenens*, S.

trueperi, *Gonapodya prolifera* y *Blastobotrys illinoisensis* (84.6%). Las muestras con PNAs presentaron mayor proporción fúngica (5.6% vs 3.5%), destacando *G. prolifera* (2.7% vs 1.6%), mientras que *Sphingomonadaceae* fue dominante en ambos grupos. No se detectaron diferencias notables en diversidad asociadas a la capacidad madurativa: los clones de baja capacidad presentaron mayor riqueza (511.5 vs 499) y Shannon (2.04 vs 1.89), pero todos mostraron baja equitatividad (Pielou 0.25–0.35) y fuerte dominancia.

Comparación de workflows metagenómicos (Z-score por métrica)

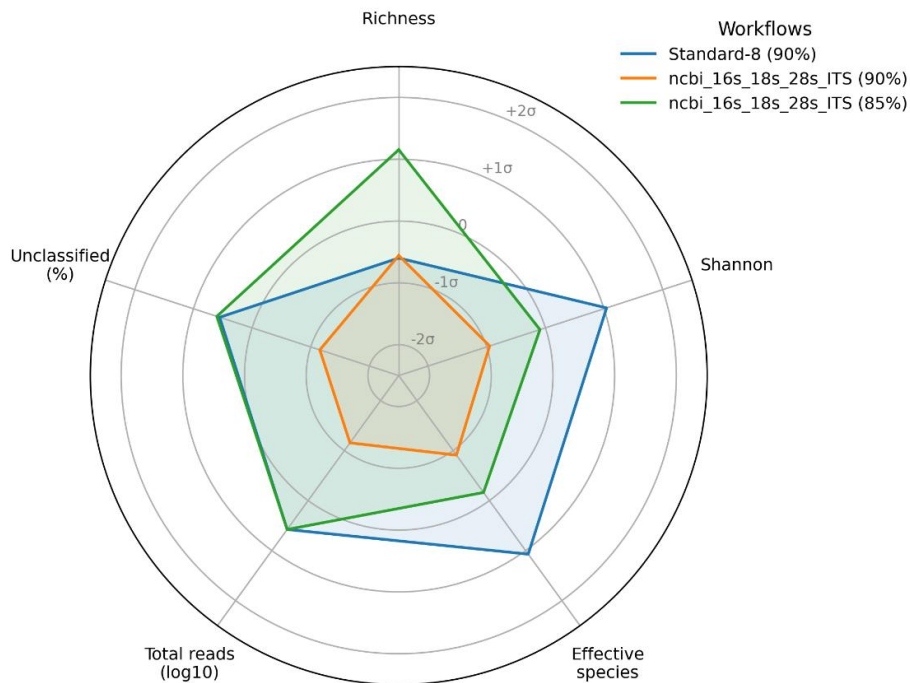


Figura 11 Comparación de métricas de diversidad y eficiencia entre workflows metagenómicos. Análisis de Richness, diversidad de Shannon, número efectivo de especies, reads totales (log10) y porcentaje de reads no clasificados, normalizados por Z-score. Se evaluaron tres flujos: (1) Standard-8 (90% identidad, Kraken2), (2) ncbi_16s_18s_28s_ITS (90% identidad, Kraken2), y (3) ncbi_16s_18s_28s_ITS (85% identidad, Minimap2). Los valores indican desviaciones estándar (σ) relativas al promedio de cada métrica.

14. Análisis taxonómico con Minimap2

14.1. Preprocesamiento de datos basado en el pipeline RESCUE

El conjunto de lecturas crudas generado mediante secuenciación Oxford Nanopore mostró alta variabilidad en longitud y calidad, con un promedio de 77.677 lecturas por muestra y una longitud media de 1.703 pb. Tras el filtrado por calidad (Q-score ≥ 9 , longitud entre 2000–5000 pb), se retuvo en promedio el 56.6 % de las lecturas, mejorando tanto la longitud promedio (2.446 pb) como la calidad (Q promedio: 63.43; %Q ≥ 15 : 85.43 %). Las etapas de trimming de barcodes y primers redujeron progresivamente el número de lecturas (41.460 y 37.299 lecturas promedio, respectivamente), manteniendo una calidad elevada y distribuciones de tamaño consistentes. Finalmente, tras eliminar lecturas con alineamientos significativos frente a secuencias organelares del hospedero, se conservaron en promedio 36.450 lecturas por muestra, con una longitud media de 2.323 pb y Q-score promedio de 63.77 (tabla 8).

Los alineamientos indicaron la presencia consistente de secuencias mitocondriales del hospedero (*Pinus taeda*), con un promedio general de 7.47 %



de lecturas clasificadas como contaminantes (≥ 99 % de identidad y ≥ 1000 pb de longitud), mientras que no se detectaron coincidencias con secuencias cloroplásticas ni con fragmentos mitocondriales de *P. radiata*. La proporción de contaminación se mantuvo estable entre condiciones experimentales: en clones de alta capacidad madurativa se obtuvo 7.51 % con PNAs y 7.53 % sin PNAs, mientras que en clones de baja capacidad se registró 7.46 % con PNAs y 7.40 % sin PNAs. La variación fue mínima (desviación estándar global = 0.071 %; rango = 7.38–7.57 %). El análisis estadístico mediante t-test independiente no detectó diferencias significativas entre grupos (alta vs. baja capacidad: $t(6) = 0.21$, $p = 0.84$; con vs. sin PNAs: $t(6) = -0.14$, $p = 0.89$), lo que confirma que **el uso de PNAs no tuvo un efecto relevante en la reducción del ADN mitocondrial residual.**

Tabla 8 Estadísticas básicas de secuencias crudas obtenidas de la primera secuenciación calculadas para cada uno de los archivos. fastq en cada etapa del preprocesamiento bioinformático. Los valores reflejan la cantidad de lecturas, su distribución en longitud y el Q score promedio.

Proceso	Lecturas totales promedio	Tamaño promedio	Mediana Tamaño	Tamaño mín. prom	Tamaño max. prom	Q promedio	%Q>=12	%Q>=15
Concat	77 677	1 703	2 328	71	18 301	62.81	88.10	84.36
Qfilt	43 945	2 446	2 446	2 000	4 972	63.43	89.02	85.43
Trim_bc	41 460	2 399	2 401	2 000	4 930	63.62	89.43	85.89
Trim_pri	37 299	2 323	2 314	1 206	4 833	63.98	90.26	86.82
Mit_filt	36 450	2 323	2 313	1 221	4 827	63.77	89.91	86.38

14.2. Análisis taxonómico

Se evaluó la capacidad de tres bases de datos taxonómicas (EMU, GROND y rEGEN) para clasificar lecturas de secuenciación del operón ribosomal 16S-ITS-23S mediante alineamientos con Minimap2. El análisis comparativo reveló diferencias notables en la calidad de los alineamientos, la riqueza taxonómica recuperada y la composición microbiana inferida (tabla 9).

La base de datos EMU, optimizada para secuencias de Nanopore, presentó una asignación con baja calidad de mapeo (más del 95 % de las lecturas mostraron valores de MAPQ ≤ 1), lo que indica ambigüedad en el alineamiento, posiblemente por la presencia de secuencias similares entre taxones cercanos. A pesar de esta limitación, EMU permitió recuperar perfiles microbianos consistentes entre muestras, con un número estimado de especies relativamente estable (360–369 por muestra) y una dominancia clara de *Sphingomonas sanxanigenens*, que representó aproximadamente el 50 % de las lecturas asignadas en cada clon. La composición microbiana fue similar entre grupos de capacidad madurativa, diferenciándose principalmente en las abundancias relativas de taxones secundarios, lo que sugiere que EMU puede capturar diferencias ecológicas finas aun con lecturas ambiguas (Figura Suplementaria 4).

En contraste, **la base GROND** mostró una excelente calidad de mapeo, con más del 94 % de las lecturas clasificadas con valores de MAPQ > 10 . Sin embargo,

esta mayor especificidad vino acompañada de una reducción sustancial en la diversidad detectada, con solo entre 93 y 144 especies por muestra. Las comunidades estuvieron fuertemente dominadas por un solo taxón, *Kordiimonas sp016764715*, que explicó más del 95 % de las lecturas en todas las muestras. Este patrón se reflejó también en índices de diversidad alfa bajos (Shannon entre 0.26 y 0.51), lo que sugiere una clasificación muy restrictiva con pobre resolución para microbiotas complejas o poco representadas en la base de datos (Figura Suplementaria 5).

La base rEGEN, construida con énfasis en diversidad bacteriana del suelo, ofreció el desempeño más equilibrado. Aunque presentó una calidad de mapeo intermedia (con un porcentaje variable de lecturas con MAPQ >1 según el clon), mostró la mayor longitud media de alineamiento (1372–1420 pb) y recuperó una diversidad taxonómica superior, con 309 a 405 especies por muestra y valores de diversidad alfa elevados (Shannon ~3.45). La composición microbiana fue comparable a la obtenida con EMU, con *S. sanxanigenens* como especie dominante en todas las muestras, pero con una contribución más distribuida entre otros taxones como *Zymomonas mobilis*, *Kordiimonas sp.* y representantes del género *Rhodococcus*, lo que resultó en perfiles comunitarios más equilibrados (Figura Suplementaria 6).

Tabla 9 Comparación de desempeño de tres bases de datos taxonómicas utilizadas para la clasificación de lecturas del operón ribosomal 16S-ITS-23S mediante Minimap2. Se muestran promedios por base de datos considerando los 8 barcodes secuenciados.

Métrica	EMU	GROND	rEGEN
% lecturas con MAPQ > 1	1.2 %	94.5 %	57.3 %
Longitud media de alineamiento	1238.3 pb	1306.4 pb	1396.0 pb
N° especies promedio por muestra	363.5	119.5	357.8
Índice de Shannon promedio	2.35	0.41	3.43
Taxón dominante	<i>S. sanxanigenens</i>	<i>Kordiimonas</i> sp.	<i>S. sanxanigenens</i>
% lecturas del taxón dominante	50.3 %	96.8 %	30.7 %

Dado su desempeño más equilibrado entre calidad de alineamiento, diversidad taxonómica recuperada y representatividad ecológica, se seleccionó la base de datos rEGEN para el análisis principal de la segunda secuenciación, que incluyó 19 barcodes.

15. Segunda secuenciación por Oxford Nanopore (corrida completa).

15.1. Preparación de Librerías de secuenciación

Las 20 muestras amplificadas fueron purificadas y cuantificadas, obteniéndose concentraciones entre 16,8 y 42,8 ng/μL, tal como se detalla en la Tabla 10. Para alcanzar una masa uniforme de entrada de 200 fmol, se ajustaron los volúmenes de carga entre 7,20 y 18,33 μL. Debido a su baja concentración, tres muestras quedaron excluidas del proceso de alicuotado. Posteriormente, tras el tratamiento enzimático de extremos (End-Prep), las concentraciones oscilaron entre 13,5 y

30,6 ng/μL, observándose una ligera disminución respecto a la etapa anterior. Finalmente, la mezcla de ADN “Native Barcode” alcanzó una concentración de 86,6 ng/μL, mientras que la librería final se ajustó a 75,8 ng/μL en un volumen total de 15 μL, garantizando una entrada adecuada para la secuenciación mediante Nanopore

Tabla 10 Concentraciones de ADN en distintas etapas del protocolo de preparación de librerías para la segunda secuenciación Nanopore.

Barcode	Muestra	Condición	Conc. PCR Purificación (ng/μL)	Conc. post End-Prep (ng/μL)	Native Barcode (ng/μL)	Librería final (ng/μL)
01	IB 278	Baja	42,8	20,0	—	—
02	IB 278	Baja	42,8	18,2	—	—
03	IB 288	Baja	30,8	28,2	—	—
04	IB 291	Baja	31,8	26,2	—	—
05	IB 285	Alta	20,5	27,0	—	—
06	OAC 103	Baja	35,4	30,6	—	—
07	OAC 106	Baja	29,0	21,6	—	—
08	OAC 111	Alta	27,4	24,4	—	—
09	MWB 161	Alta	30,6	23,2	—	—
10	MWB 163	Alta	28,0	22,6	—	—
11	MWB 213	Alta	40,4	18,8	—	—
12	MWB 213	Alta	40,4	13,5	—	—
13	MWB 160	Baja	27,0	30,4	—	—
14	MWB 162	Baja	29,2	26,4	—	—
15	MWB 165	Baja	24,4	17,9	—	—
16	MWB 204	Baja	26,6	23,4	—	—
17	OAL 104	Baja	34,4	22,0	—	—
18	OAL 104	Baja	34,4	21,6	—	—
19	OAL 101	Alta	34,8	18,6	—	—
20	OAL 112	Alta	37,2	20,8	—	—
—	—	—	—	—	86,6	75,8

15.2. Preprocesamiento de datos basado en el pipeline RESCUE

La segunda corrida de secuenciación generó un total de 591,8 k lecturas crudas (928,4 Mb; N50 = 2,42 kb). Luego del basecalling, 484,3 k lecturas (665,5 Mb)

superaron el umbral mínimo de calidad ($Q \geq 9$), y tras el filtrado por calidad y tamaño, se retuvieron 118,7 k lecturas “Pass” (121,4 Mb). Estas lecturas presentaron una longitud promedio de 2,4 kb y Q-scores superiores a 24, adecuándose al diseño de los amplicones.

Se procesaron 19 conjuntos de lecturas correspondientes a muestras de *P. radiata*, con un promedio de 23 373 lecturas por muestra. Las longitudes medias oscilaron entre 1 022 y 1 705 pb (medianas de 687 a 2 381 pb), con rangos de tamaño de 66 a 27 181 pb. En conjunto, las lecturas finales presentaron una longitud promedio de 1 404 pb (mediana de 1 162 pb) y un rango promedio de 79 a 10 939 pb.

Durante la etapa exploratoria, se analizó un 10 % aleatorio de las lecturas por muestra mediante alineamiento BLASTn (≥ 99 % identidad; $\geq 1\ 000$ pb) contra el genoma mitocondrial de *Pinus taeda* (NC_039746.1), detectándose entre 2,65 % (barcode 02) y 4,95 % (barcode 06) de lecturas mitocondriales (promedio $3,48 \pm 0,59$ %). Posteriormente, se aplicó el mismo criterio al conjunto completo, eliminándose entre 112 y 508 lecturas por muestra (2,95 %–3,56 %). En total, se descartaron 5 697 lecturas, obteniéndose finalmente 169 338 lecturas limpias para análisis taxonómico. Las estadísticas detalladas de cada etapa del preprocesamiento se presentan en la Tabla



11, donde se incluyen la evolución del número de lecturas, su longitud y calidad promedio.



Tabla 11 Estadísticas básicas de secuencias crudas obtenidas de la segunda secuenciación calculadas para cada uno de los archivos. fastq en cada etapa del preprocesamiento bioinformático. Los valores reflejan la cantidad de lecturas, su distribución en longitud y el Q score promedio.

Proceso	Reads totales promedio	Tamaño promedio	Mediana Tamaño	Tamaño mín. prom	Tamaño max. prom	Q promedio	%Q>=12	%Q>=15
Concat	22 283	1 366	1 023	78	10 173	60.91	84.58	80.06
Qfilt	10 854	2 458	2 443	2 004	4 967	62.28	86.72	82.62
Trim_bc	10 738	2 410	2 398	2 001	4 857	62.50	87.21	83.17
Trim_pri	8 540	2 331	2 314	1 859	4 792	62.86	88.09	84.19
Mit_filt	8 090	2 330	2 313	1 859	4 792	62.74	87.89	83.89

15.3. Análisis taxonómico

Composición taxonómica y dominancia del núcleo microbiano

Las muestras con alta capacidad madurativa presentaron, en general, una mayor abundancia absoluta de lecturas, destacando casos como el barcode 12 (MWB 213, alta) con 15.282 lecturas. En contraste, las muestras de baja capacidad madurativa tendieron a tener menos lecturas, con valores mínimos como el barcode 6 (OAC 103, baja) con solo 486 lecturas. Aunque existen excepciones, esta tendencia sugiere una posible relación entre carga microbiana y capacidad madurativa.

Más allá del número total de lecturas, se identificaron patrones consistentes en la composición microbiana. Las tres especies más abundantes —*Sphingomonas sanxanigenens*, *Candidatus Liberibacter sp.* y *Kordiimonas sp.*— estuvieron presentes de forma recurrente en la mayoría de las muestras, y representaron una fracción considerable del total de lecturas observadas en muchos barcodes (figura 12). Si bien la proporción relativa de cada especie varió entre muestras, su recurrencia en los perfiles taxonómicos más abundantes sugiere que podrían formar parte de un núcleo común o persistente del microbioma embriogénico. Esta dominancia también se evidencia en el gráfico de abundancia relativa (figura suplementaria 7), donde dichas especies explican una proporción significativa del contenido bacteriano por muestra, con porcentajes que en el caso de

Sphingomonas supera el 20% del total. Así, la combinación de abundancia absoluta y relativa proporciona una visión robusta de la carga microbiana y de la estructura comunitaria predominante, permitiendo identificar qué taxones contribuyen según la condición fisiológica y la pertenencia familiar de cada línea embriogénica.

Para complementar la evaluación de la composición microbiana, se analizaron lecturas representativas del barcode 12 mediante visualización en IGV, comparando la influencia del filtrado por calidad en la cobertura y estabilidad de los alineamientos. Las lecturas filtradas con umbral de calidad $Q \geq 5$ mostraron reducción de clipping y alineamientos más confiables en regiones clave del amplicón, tanto en los primeros 1.500 pb como en la región media/final (figura suplementaria 4). En contraste, las lecturas filtradas únicamente por longitud y cobertura, con calidad mínima $Q = 1$, presentaron niveles significativos de soft clipping y menor estabilidad en la alineación, aunque permitieron confirmar la presencia de secuencias en las mismas regiones del amplicón (figura suplementaria 5)

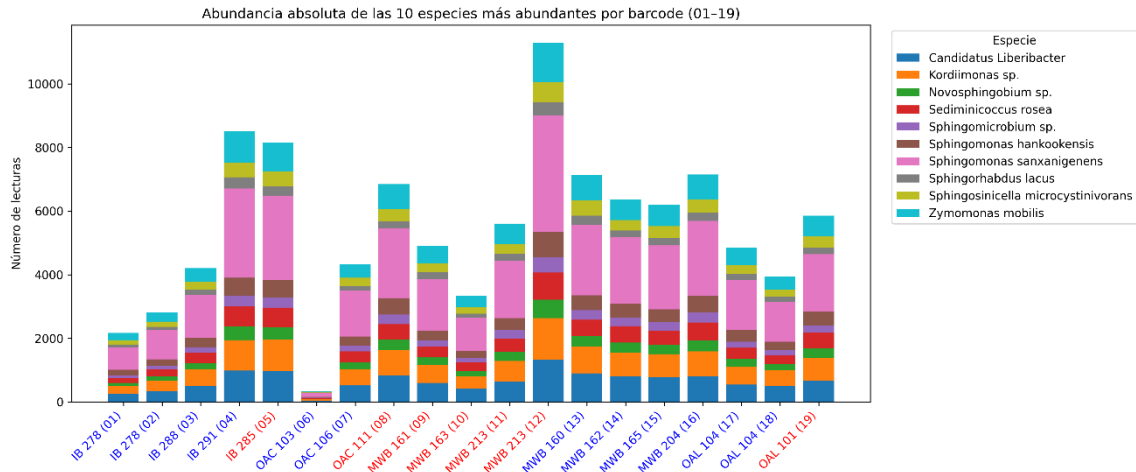


Figura 12 Abundancia absoluta de las 10 especies bacterianas más abundantes por barcode (01-19). Barras apiladas muestran el número de lecturas asignadas a cada especie en cada muestra, etiquetadas con el código de barcode y la familia taxonómica. Los diferentes colores representan las especies dominantes en el microbioma. El texto del eje X está coloreado según la condición fisiológica de las muestras (rojo = alta capacidad madurativa, azul = baja capacidad madurativa). Dentro de las barras, se indican en porcentaje (una cifra decimal) las contribuciones de las tres especies con mayor abundancia global.

Diversidad alfa

Se evaluó la diversidad alfa de las comunidades bacterianas en muestras embriogénicas de *P. radiata* clasificadas según su capacidad madurativa (Alta o Baja), utilizando los índices Observed (riqueza observada), Shannon (riqueza y uniformidad), Chao1 (riqueza estimada) y Simpson (uniformidad) (figura 13). Las muestras con alta capacidad madurativa presentaron, en general, una mayor riqueza de especies (mediana Observed = 143 frente a 145 en Baja), aunque con una alta variabilidad interna. Los índices Shannon y Simpson mostraron valores similares entre condiciones, lo que sugiere una distribución comparable en términos de diversidad y dominancia de especies. El índice Chao1 mostró una



tendencia a mayor riqueza estimada en la condición Alta (mediana ~233 vs. ~185), aunque esta diferencia no alcanzó significancia estadística.

Los análisis estadísticos mediante la prueba de Mann-Whitney U no revelaron diferencias significativas entre condiciones para ninguno de los índices evaluados: Observed (U=53.00, p=0.3738), Shannon (U=49.00, p=0.5918), Chao1 (U=63.00, p=0.0831) y Simpson (U=50.00, p=0.5358).

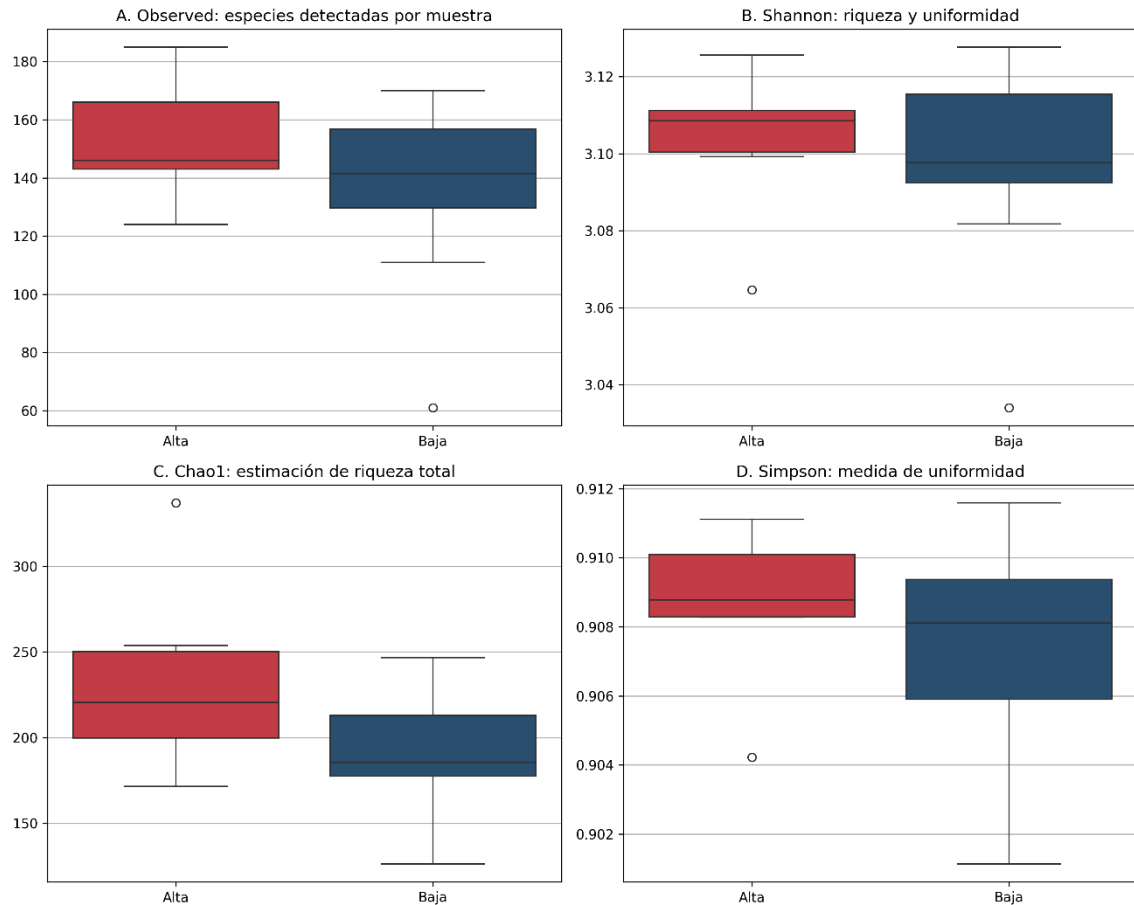


Figura 13 Comparación de diversidad alfa entre condiciones de alta y baja capacidad madurativa en líneas embriogénicas de *Pinus radiata*. (A) Observed: número de especies bacterianas detectadas por muestra. (B) Shannon: índice que combina riqueza y uniformidad de especies (C) Chao1: estimación de la riqueza total esperada, basada en la frecuencia de especies raras (observadas una o dos veces). (D) Simpson: medida de uniformidad, interpretada como la probabilidad de que dos lecturas seleccionadas al azar correspondan a la misma especie. En cada panel, las cajas representan el rango intercuartílico (Q1–Q3), la línea horizontal marca la mediana y los bigotes se extienden hasta 1.5 veces el rango intercuartílico (IQR). Los puntos fuera de este rango se muestran como valores atípicos. Las muestras correspondientes a clones con alta capacidad madurativa se ubican a la izquierda, y aquellas con baja capacidad madurativa, a la derecha en cada gráfico.

Diversidad beta

La evaluación de la diversidad beta mediante análisis de ordenación PCoA basado en la distancia Bray-Curtis mostró una superposición marcada entre las muestras, sin separación clara según las variables de Familia ni Condición (Figura 14). Los dos primeros ejes explican conjuntamente cerca del 25% de la variabilidad total en la composición bacteriana, reflejando una complejidad estructural que no se reduce a pocos componentes.

Al comparar cuantitativamente las distancias microbianas, la distancia media dentro del grupo de alta capacidad fue de 0.062, mientras que en el grupo de baja capacidad fue ligeramente mayor, con 0.067. La distancia media entre los grupos alta y baja fue intermedia, con un valor de 0.064. Estos valores indican que la variabilidad interna de cada grupo es similar a la distancia observada entre ellos, lo que refuerza la ausencia de una diferenciación clara en la estructura microbiana según la capacidad madurativa.

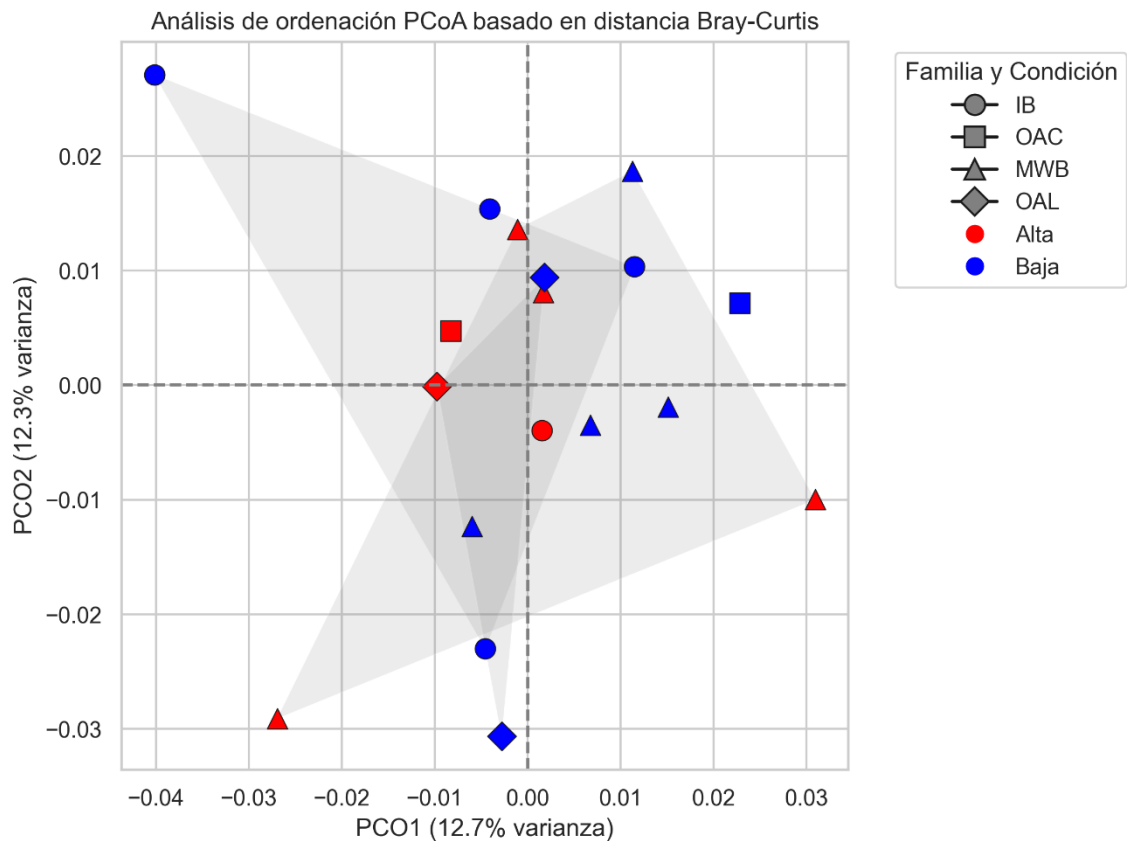


Figura 14 Análisis de ordenación PCoA basado en la distancia Bray-Curtis de la composición bacteriana en muestras de *Pinus radiata*. Se presentan tres gráficos correspondientes a datos de abundancia relativa organizados en una sola fila: (a) PCoA coloreado según Familia, mostrando los grupos experimentales IB, OAC, MWB y OAL; (b) PCoA coloreado por condición de capacidad madurativa, diferenciando muestras de capacidad Alta y Baja; y (c) PCoA coloreado por el grupo combinado que integra Familia y Condición. Las líneas grises discontinuas indican el eje cero en cada dimensión principal (PC1 y PC2). Los ejes PC1 y PC2 representan aproximadamente 12.7 % y 12.3 % de la disimilitud total entre muestras, respectivamente. La visualización muestra una superposición considerable entre las muestras, sugiriendo una estructura microbiana globalmente similar entre grupos.

15.4. Identificación de especies bacterianas núcleo ("core"):

Se identificaron 17 especies bacterianas núcleo en *P. radiata* que cumplen con criterios de **prevalencia $\geq 95\%$ y abundancia relativa promedio $> 1\%$** en al menos una de las condiciones de capacidad madurativa (Alta y Baja) (Figura 15). Estas especies mostraron una alta estabilidad y presencia consistente en ambos

grupos, destacándose *Sphingomonas sanxanigenens* como la más abundante, con valores similares de abundancia relativa (23.4 % en Alta y 23.8 % en Baja) y prevalencia total (100 %). Otros taxones significativos incluyen *Candidatus Liberibacter* y *Kordiimonas* sp., que también presentaron alta prevalencia y abundancia comparable entre condiciones.

El análisis por género reveló que el grupo *Sphingomonas* domina la comunidad núcleo, concentrando cuatro especies, seguido por géneros como *Candidatus*, *Zymomonas* y *Sediminicoccus*. La abundancia relativa agregada por género corrobora esta tendencia, con *Sphingomonas* representando aproximadamente un tercio de la comunidad bacteriana núcleo en ambos estados de capacidad madurativa. Estos resultados reflejan una estructura microbiana núcleo estable y compartida, sin diferencias significativas en la composición bacteriana entre las condiciones evaluadas.

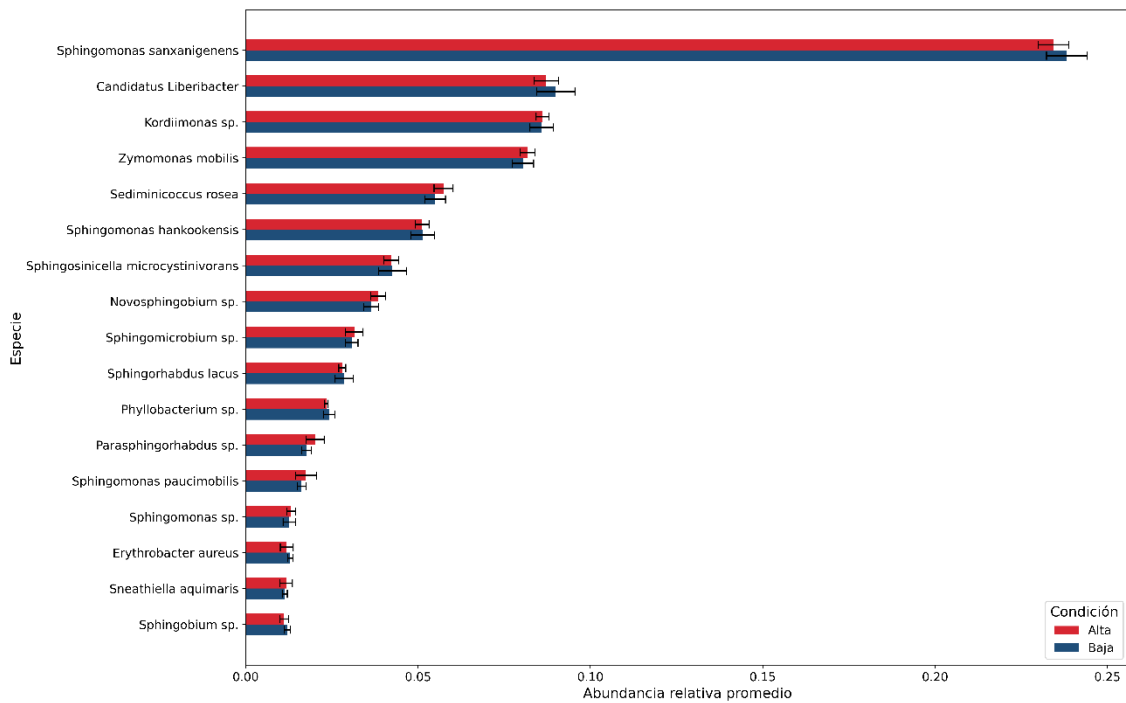


Figura 15 Comparación de la abundancia relativa promedio de las especies núcleo entre condiciones de capacidad madurativa (Alta vs. Baja). El gráfico presenta barras horizontales para las especies núcleo, ordenadas de mayor a menor según su abundancia en la condición Alta. Cada especie está representada con dos barras que muestran la abundancia relativa promedio en las condiciones Alta (rojo) y Baja (azul), con barras de error que indican la desviación estándar entre muestras.

15.5. Análisis funcional y relevancia biológica

En el conjunto de 17 especies núcleo (prevalencia $\geq 95\%$, abundancia $> 1\%$), se asignaron funciones a partir de bibliografía y genomas de referencia, clasificando cada taxón según su potencial impacto en la embriogénesis somática: beneficioso (●), neutro (○), perjudicial (●) o variable (●). Del total, el 29% (5/17) mostró funciones predominantemente beneficiosas, por ejemplo, rutas de degradación de fenantreno y fitohormonas en *Novosphingobium spp.*, síntesis de polisacáridos protectores en *Sphingomonas sanxanigenens*, y degradación de

microcistinas en *Sphingosinicella microcystinivorans*. Un 29% (5/17) presentó un perfil neutro, sin evidencias de interacción directa con el embrión, incluyendo *Erythrobacter aureus*, *Sediminicoccus rosea* y *Sphingorhabdus lacus*.

El 18% de las especies núcleo (3/17) exhibió funciones potencialmente perjudiciales, principalmente licuefacción de paredes celulares o producción de metabolitos tóxicos: *Candidatus Liberibacter* por patogenicidad sistémica, *Sneathiella aquimaris* por disrupción de simbiosis, y *Sphingomonas paucimobilis* por enzimas pectinolíticas. El 24% restante (4/17) mostró un perfil variable (●), como *Zymomonas mobilis* (producción de etanol, fitotóxico en altas dosis) y *Sphingomonas hankookensis* (efecto cepa-dependiente).

Al agrupar por género, *Sphingomonas* concentró el mayor número de especies núcleo (4/17) con funciones mayoritariamente protectoras (● en 50% de sus miembros: *S. sanxanigenens* y *S. sp.* beneficiosas), pero incluyó un 25% de especies perjudiciales (*S. paucimobilis*). Géneros como *Novosphingobium* y *Phyllobacterium* aportaron exclusivamente funciones beneficiosas (● 100%), mientras que taxones monotípicos (*Sediminicoccus*, *Kordiimonas*) fueron clasificados como neutros (○) o variables (●). Estos resultados revelan un "núcleo" funcional equilibrado, donde el 47% de los taxones (beneficiosos + variables con potencial positivo) favorecen la embriogénesis, con al menos un 18–29% (3–5 especies) identificadas como PGPB que podrían optimizar la



eficiencia de maduración en *P. radiata* mediante modulación hormonal y protección contra estrés (Moreno-Valencia et al., 2024), frente a un 18% con impacto adverso.

VII. DISCUSIÓN.

16. Optimización de la extracción de ADN y la amplificación por PCR:

La extracción de ADN bacteriano de muestras vegetales suele presentar problemas de rendimiento y pureza. En nuestro caso, los métodos basados en CTAB/SDS y algunos kits comerciales generaron ADN de baja calidad en callos somáticos de *P. radiata*, con bandas débiles o difusas en gel y marcada variabilidad entre réplicas, lo que dificultaba la amplificación de fragmentos largos. Este comportamiento es consistente con estudios previos, donde se reporta que la co-extracción de compuestos vegetales (como polisacáridos y polifenoles) y la alta relación planta: bacteria complican la obtención de ADN en tejidos de plantas (Sahu et al., 2012; Q. Wang et al., 2021). Estos contaminantes reducen los ratios A260/A230 (0.57–1.11) e inhiben enzimas como la Taq polimerasa. En contraste, el protocolo basado en columnas (DNeasy Plant Mini Kit, Bioforest) produjo ADN de alta integridad y peso molecular elevado en todos los clones evaluados, con perfiles nítidos y reproducibles. No se observaron diferencias significativas en la calidad del ADN entre líneas embriogénicas de alta o baja capacidad madurativa, lo que indica que la variabilidad en maduración no sesga la comparación de sus microbiotas.

En la optimización de la PCR para amplificar el operón bacteriano 16S–ITS–23S, se enfrentaron desafíos técnicos clave relacionados con la longitud del fragmento

y la co-amplificación de ADN vegetal. La polimerasa Taq convencional mostró limitaciones para amplificar fragmentos mayores a 4 kb, generando bandas difusas e incompletas, lo que se resolvió al emplear la polimerasa de alta fidelidad Q5®, capaz de producir amplicones robustos de $\approx 2.5\text{--}3$ kb bajo condiciones optimizadas (Cline et al., 1996). Este enfoque está respaldado por estudios que demuestran que las enzimas con corrección de errores son esenciales para la amplificación precisa de fragmentos largos. Además, la co-amplificación de ADN cloroplástico y mitocondrial del hospedador representó otro obstáculo crítico, el cual se superó mediante la incorporación de oligonucleótidos PNA específicos y el uso de partidores como el 27f y 2241r, diseñados para evitar la co-amplificación mitocondrial y cloroplástica *in silico* (Beckers et al., 2017; Cusco et al., 2018).

La estrategia de partidores selectivos demostró ser altamente efectiva, tal como reportan Fitzpatrick et al. (2018) y Lundberg et al. (2013) quienes encontraron que los bloqueadores de PNA reducen drásticamente la amplificación de ADN del hospedador (hasta en un 99%) en muestras vegetales. Estos ajustes permitieron obtener un producto de PCR predominantemente bacteriano, con mínima contaminación de ADN vegetal, proporcionando así una herramienta robusta para caracterizar comunidades microbianas asociadas a plantas. Además, la capacidad de amplificar regiones largas, como el operón 16S–ITS–23S, ofrece una mayor resolución taxonómica, permitiendo identificar bacterias a nivel de especie con mayor precisión que con el gen 16S solo. Este protocolo

optimizado constituye una plataforma metodológica sólida para la caracterización precisa de la microbiota asociada a *P. radiata*, facilitando análisis metagenómicos con resolución a nivel de especie.

17. Primera secuenciación Oxford nanopore (prueba piloto):

17.1. Flujo *wf-metagenomics* (EPI2ME)

El análisis mediante *wf-metagenomics* permitió una clasificación rápida, pero reveló un marcado sesgo hacia taxones dominantes y altas tasas de lecturas no asignadas (38-40% en muestras complejas), limitando su capacidad para detectar diversidad microbiana poco representada o novedosa (Forry et al., 2024; Navgire et al., 2022). Esta limitación se atribuye a su dependencia de bases de datos preconfiguradas y enfoques basados en k-mers, que priorizan microorganismos bien caracterizados mientras ignoran especies raras o no catalogadas. Además, la ausencia de efectos consistentes del tratamiento con PNA o de la capacidad madurativa en las métricas globales demuestra que este enfoque estandarizado carece de la sensibilidad necesaria para identificar variaciones sutiles en la estructura de las comunidades microbianas o para reducir eficazmente la interferencia del ADN del hospedador (Larrouy et al., 2024).

La baja resolución taxonómica observada -particularmente con la base de datos NCBI_16S_18S_28S_ITS- y los bajos índices de equidad (Shannon $\approx 1.6-1.8$)

reflejan una limitación fundamental: estas bases de datos están optimizadas para regiones 16S/18S/28S, mientras que nuestro estudio amplificó el operón extendido 16S-ITS-23S (~4.2-4.5 kb). Esta discrepancia explica por qué wf-metagenomics mostró menor eficiencia en la clasificación, ya que las regiones ITS y 23S no están adecuadamente representadas en las bases de referencia estándar (Forry et al., 2024; Navgire et al., 2022). Los resultados coinciden con reportes previos donde el uso de bases diseñadas para 16S subestima la diversidad cuando se analizan regiones extendidas, generando dominancia artificial de taxones bien caracterizados (como *Firmicutes* y *Proteobacteria*) mientras se pierde información de otros grupos microbianos.

17.2. Pipeline basado en RESCUE (filtrado + Minimap2)

Para aprovechar la riqueza de información del operón 16S–ITS–23S sin seguir íntegramente los pipelines originarios, adoptamos sus bases de datos —GROND, rEGEN y EMU— dentro de nuestro propio flujo de filtrado y alineamiento con Minimap2 (Dubois et al., 2024b; Petrone et al., 2023; Walsh et al., 2024). GROND, una colección curada de operones completos diseñada para distinguir especies muy emparentadas (p.ej. *Escherichia coli* vs *Shigella*), mostró excelente especificidad (MAPQ>10 en >94% de lecturas) pero clasificó un número limitado de taxones, reflejando su solidez para especies bien documentadas y sus limitaciones ante microbiomas no convencionales. EMU, optimizada

originalmente para ONT, asignó cientos de especies pero presentó mapeos mayoritariamente ambiguos (95% con $MAPQ \leq 1$), lo que redujo la confianza en cada asignación.

En contraste, rEGEN combinó los puntos fuertes de ambos: proporcionó una longitud de alineamiento robusta (~ 1400 pb) y una diversidad elevada (Shannon ≈ 3.43), equilibrando sensibilidad y calidad de mapeo. Al emplear rEGEN, recuperamos perfiles más representativos, con un núcleo dominado por *Sphingomonas sanxanigenens* pero con contribuciones sustanciales de géneros minoritarios, sin sacrificar la certeza de la clasificación. Por este balance único, optamos por rEGEN en la segunda secuenciación, convencidos de que su base de datos aporta la máxima resolución posible dentro de un pipeline personalizado.

17.3. Efecto de los bloqueadores PNA

Aunque los bloqueadores PNA se introdujeron con la expectativa de reducir drásticamente la co-amplificación de ADN mitocondrial y cloroplástico, la prueba piloto mostró que la proporción de lecturas hospedador ($\sim 7.5\%$) y la estructura comunitaria resultante apenas variaron entre muestras con y sin PNA. Este hallazgo indica que, en el contexto de nuestra metodología de filtrado y secuenciación larga, el impacto de los PNAs fue limitado en términos cuantitativos.

Sin embargo, dado que la inclusión de PNA no perjudicó la cantidad ni la calidad de las lecturas bacterianas y podría ejercer beneficios no medidos, por ejemplo, en variantes de baja abundancia o en situaciones con mayor carga de hospedador (Fitzpatrick et al., 2018; Lundberg et al., 2013), se decidió mantenerlos en la segunda corrida. De este modo, se preserva la consistencia metodológica y se conserva la posibilidad de capturar efectos más sutiles en futuros análisis.

18. Segunda secuenciación por Oxford Nanopore (corrida completa).

18.1. Análisis taxonómico

La estructura de las comunidades bacterianas detectadas mediante el pipeline personalizado con la base rEGEN confirmó la existencia de un microbioma dominante en tejidos embriogénicos de *P. radiata*. A diferencia de la prueba piloto, donde *Sphingomonas sanxanigenens* dominaba prácticamente toda la comunidad microbiana, en la corrida completa esta especie compartió protagonismo con otros taxones frecuentes como *Kordiimonas* spp. y *Candidatus Liberibacter* spp. Esta distribución más equitativa de la abundancia relativa indica una comunidad más equilibrada, representativa del microbioma natural de los tejidos embriogénicos de *P. radiata*.

La aparente homogeneidad entre líneas de alta y baja capacidad madurativa sugiere que el ambiente embriogénico impone un filtro ecológico fuerte,

favoreciendo un conjunto limitado de especies adaptadas y conservando ciertos taxones núcleo por compatibilidad funcional o ecológica con el hospedero (Bulgarelli et al., 2013; Vandenkoornhuyse et al., 2015). Así, la comunidad detectada refleja no solo mayor diversidad efectiva entre taxones dominantes, sino también un equilibrio funcional que podría ser relevante para la maduración somática.

En comparación con otras coníferas, nuestros hallazgos son consistentes con patrones donde comunidades microbianas dominadas por pocos géneros (usualmente Proteobacteria, como *Sphingomonas* o *Novosphingobium*) se repiten en semillas, brotes o polen (Armstrong et al., 2024; Carrell & Frank, 2015). Este núcleo reducido parece estar estrechamente relacionado con la capacidad de colonización endófito, tolerancia a compuestos fenólicos y potencial de interacción con el metabolismo vegetal. La presencia recurrente de géneros como *Sphingomonas*, *Novosphingobium* y *Rhodococcus* sugiere capacidad para tolerar y transformar compuestos fenólicos y derivados de lignina, rasgos ampliamente documentados en estas familias bacterianas (Gu et al., 2024; Kamimura et al., 2017; Masai et al., 2007; Perez et al., 2021). Esto apoya la idea de que variaciones en el contenido de lignina durante la maduración embrionaria podrían favorecer la persistencia de estos taxones núcleo.

Así, el uso de rEGEN no solo permitió detectar con mayor sensibilidad especies previamente subrepresentadas en el piloto, sino que también reafirmó la

existencia de un “núcleo” bacteriano asociado a la embriogénesis somática de *Pinus*, paralelo al observado en tejidos reproductivos o aéreos de la misma especie.

18.2. Identificación de Especies Bacterianas Núcleo ("Core")

La identificación de un conjunto reducido de especies presentes en todas las muestras, independientemente del clon o su capacidad madurativa, respalda firmemente la existencia de un microbioma núcleo en líneas embriogénicas de *P. radiata*. La consistencia en la detección de especies como *Sphingomonas sanxanigenens*, *Candidatus Liberibacter spp.* y *Kordiimonas spp.* sugiere que estas bacterias poseen rasgos funcionales esenciales que favorecen su persistencia en el ambiente embriogénico. Este tipo de asociación ha sido documentada en diversas plantas leñosas, donde ciertas bacterias endófitas son capaces de establecer relaciones mutualistas o comensales estables, tolerar condiciones extremas y modular procesos fisiológicos clave del hospedero (Compant et al., 2019). Además, varios estudios en embriogénesis somática vegetal señalan que el microbioma núcleo tiende a estar formado por bacterias con capacidades metabólicas versátiles, incluyendo fijación de nitrógeno, degradación de compuestos fenólicos y producción de fitohormonas como auxinas o citoquininas (Berg et al., 2020; Santoyo et al., 2016). Aunque este estudio no confirmó experimentalmente tales funciones en los aislamientos detectados, la recurrencia de géneros conocidos por estas propiedades en el

núcleo embriogénico —como *Rhodococcus*, *Sphingomonas* o *Zymomonas*— refuerza la hipótesis de un posible rol funcional adaptado al microambiente del cultivo embriogénico somático. Así, la caracterización de este microbioma núcleo no solo define un punto de partida para estudios funcionales, sino también una oportunidad para el diseño racional de consorcios bacterianos aplicados a biotecnología forestal.

18.3. Análisis funcional y relevancia biológica

El análisis funcional de las especies núcleo detectadas en los cultivos embriogénicos de *P. radiata* revela un perfil metabólico centrado en la degradación de compuestos vegetales tóxicos, la producción de exopolisacáridos protectores y la síntesis de metabolitos secundarios beneficiosos, funciones clave para el soporte del desarrollo embrionario in vitro (Flemming & Wingender, 2010; Kaur et al., 2022; Masai et al., 2007; Soundar Raju et al., 2020). En particular, varios géneros recurrentes en nuestras muestras (p. ej. *Sphingomonas*, *Novosphingobium*, *Rhodococcus*) poseen rutas metabólicas descritas para degradar o transformar derivados de lignina y otros compuestos aromáticos, lo que explicaría su persistencia en un microambiente enriquecido en fenoles. ((Gu et al., 2024; Kamimura et al., 2017; Masai et al., 2007)).

Sphingomonas sanxanigenens, la especie dominante, destaca por la producción del polisacárido sanxan, cuyas propiedades gelificantes podrían

contribuir a estabilizar la matriz extracelular de los embriones y mejorar la cohesión del biofilm microbiano en torno a tejidos en desarrollo (Huang et al., 2016; Huang et al., 2009; Zhou et al., 2025). Especies del mismo grupo se han asociado además a la inducción de rutas de defensa secundarias (p. ej. flavonoides) en tejidos vegetales, lo que puede favorecer interacciones simbióticas y protección frente a estrés (Mazoyon et al., 2023). ***Novosphingobium spp.* y otras *Sphingomonadaceae*** detectadas son conocidas por degradar hidrocarburos aromáticos y por sintetizar metabolitos de interacción (sideróforos, fitohormonas) que facilitan la detoxificación del medio y potencialmente promueven procesos de diferenciación celular (Boss et al., 2022; Krishnan et al., 2017; Rangjaroen et al., 2017; Rodriguez-Conde et al., 2016; F. Wang et al., 2022). Estos rasgos bioquímicos están ampliamente descritos en la literatura sobre catabolismo bacteriano de aromáticos y transformación de lignina (Cappelletti et al., 2020; Kamimura et al., 2017).

La bibliografía indica además que la biosíntesis y acumulación de lignina y compuestos fenólicos aumentan en fases avanzadas del desarrollo embrionario y de la maduración de semillas/embriones, originando cambios químicos tisulares que pueden funcionar como un **filtro selectivo** para la microbiota asociada (Fang et al., 2020; Vimal et al., 2024). En este contexto, la persistencia de bacterias con rutas de detoxificación aromática (p. ej. *Sphingomonadaceae*, *Rhodococcus*) en líneas con alta tasa de maduración resulta biológicamente coherente: estos

microorganismos estarían mejor adaptados para tolerar o transformar los compuestos fenólicos emergentes, manteniendo así un equilibrio funcional en el microambiente embriogénico (Gu et al., 2024; Masai et al., 2007).

Más allá de la detoxificación, un subconjunto de los taxa núcleo presenta funciones clásicas de bacterias promotoras del crecimiento vegetal (PGPB): producción de fitohormonas (principalmente auxinas tipo IAA y, en algunos casos, citoquininas), síntesis de sideróforos y solubilización de nutrientes. Estas funciones son de clara relevancia para el embrión in vitro: las auxinas y citoquininas regulan división celular, elongación y diferenciación, procesos directamente relacionados con la formación y maduración del tejido embrionario, mientras que los sideróforos microbianos aumentan la disponibilidad de hierro en medios con limitaciones de quelación, favoreciendo enzimas dependientes de Fe y rutas anabólicas críticas para el desarrollo (Lurthy et al., 2020; Orozco-Mosqueda et al., 2023; Timofeeva et al., 2022). En particular, el hierro es cofactor de numerosas deshidrogenasas y oxidorreductasas implicadas en metabolismo energético y biosíntesis de compuestos estructurales; por tanto, la actividad sideróforo-mediada podría ser un soporte indirecto pero esencial para la maduración somática en cultivo (Lurthy et al., 2020; Timofeeva et al., 2022).

Por tanto, proponemos que la **actividad funcional** de estos miembros núcleo más que la mera abundancia relativa podría modular diferencias fenotípicas en

la eficiencia de maduración entre líneas. Es decir, la presencia y actividad de PGPB productores de IAA, citoquininas o sideróforos podría acelerar o estabilizar procesos de maduración aun cuando la estructura taxonómica global sea similar entre líneas (Etesami, 2025; Jakubowska et al., 2025; Poria et al., 2022). No obstante, en nuestro estudio estas funciones se infirieron a partir de perfiles taxonómicos y anotaciones funcionales; la evidencia causal requerirá ensayos dirigidos, por ejemplo: determinación cuantitativa de IAA y citoquininas en cultivos bacterianos aislados (HPLC/LC-MS), ensayo de sideróforos (CAS assay), co-cultivos controlados con líneas de diferente potencial de maduración y experimentos de inoculación/curado para evaluar efectos sobre tasa y calidad de maduración (Orozco-Mosqueda et al., 2023; Timofeeva et al., 2022).

Finalmente, la detección de taxones potencialmente fitopatógenos (p. ej. *Candidatus Liberibacter* spp.) subraya la necesidad de mantener una visión equilibrada: mientras ciertos miembros del núcleo podrían sostener la maduración por vías beneficiosas, otros constituyen un riesgo latente que podría comprometer rutas metabólicas y defensas del hospedero mediante efectores (C. Yang & Ancona, 2022). Esto refuerza la importancia de combinar análisis taxonómicos con ensayos funcionales y cribados de seguridad antes de proponer consorcios bacterianos aplicados. En conjunto, los resultados posicionan al microbioma núcleo como un actor plausible en la maduración somática y definen

líneas experimentales claras para validar su rol funcional en biotecnología forestal (Ajijah et al., 2023; Cappelletti et al., 2020).

VIII. CONCLUSIÓN:

La presente tesis caracterizó la microbiota bacteriana asociada a líneas embriogénicas somáticas de *Pinus radiata* mediante secuenciación Nanopore del operón 16S–ITS–23S, análisis con la base rEGEN y la optimización experimental de extracción de ADN y PNA clamps. Aplicando el criterio operativo de microbioma núcleo (prevalencia ≥ 95 %; abundancia promedio ≥ 1 %) se delimitó un conjunto taxonómico recurrente, con *Sphingomonas sanxanigenens*, *Kordiimonas* spp. y *Candidatus Liberibacter* spp. entre los más representativos; los análisis de diversidad alfa y beta (Shannon, Observed, Bray-Curtis, PCoA) no mostraron diferencias significativas entre líneas de capacidad madurativa alta vs baja, por lo que la hipótesis original, que las líneas de alta capacidad presentarían mayor diversidad y enriquecimiento en bacterias promotoras del crecimiento, queda refutada en las condiciones y con los umbrales aplicados. En consecuencia, se cumplieron los objetivos de optimización metodológica (4.1) y de identificación del microbioma núcleo (4.3), mientras que el objetivo de demostrar una asociación diferencial de la diversidad con la capacidad de maduración (4.2) no obtuvo soporte empírico.

Estos resultados aportan evidencia consistente sobre la existencia de un núcleo bacteriano asociado a los callos embriogénicos y delimitan claramente las limitaciones del abordaje amplicónico para inferir actividad biológica; por ello las conclusiones funcionales deben considerarse provisionales hasta su validación. Como paso intermedio y crucial se propone realizar secuenciación del genoma completo (WGS) del ADN total del callo embriogénico: a pesar de la presencia masiva de ADN organelar vegetal, la WGS permite detectar y caracterizar genes funcionales microbianos y reconstruir rutas metabólicas, recuperar ensamblajes metagenómicos y genomas reconstruidos mediante binning (MAGs) que facilitan resolución a nivel de genoma/cepa, identificar elementos móviles y genes de interés (p. ej. biosíntesis de EPS, rutas de degradación de compuestos fenólicos, genes de hormonas y sideróforos) y generar referencias que habiliten el diseño de qPCR/dPCR dirigidos y sondas para estudios cuantitativos y de expresión. Estas ventajas mitigarán las limitaciones del análisis amplicónico y proporcionarán la base necesaria para inferencias funcionales más robustas, siempre considerando **estrategias de enriquecimiento y control computacional** de secuencias hospedadoras para maximizar la recuperación microbiana.

A partir de lo anterior, los próximos pasos experimentales propuestos son: (i) repetir la secuenciación del amplicón 16S–ITS–23S focalizando específicamente en la familia MWB y comparando líneas de capacidad madurativa alta vs baja,



pero empleando una metodología de extracción especializada distinta a la utilizada en este estudio por Bioforest, con el objetivo explícito de asegurar la recuperación de bacterias intracelulares presentes en los callos; (ii) efectuar la secuenciación del genoma completo del ADN total del callo embriogénico para abordar aspectos funcionales y generar datos que permitan diseñar y validar ensayos dirigidos; y (iii) proyectar análisis funcionales mediante inoculación con bacterias promotoras del crecimiento que sean cultivables, evaluando su capacidad para mantener o mejorar la habilidad embriogénica a lo largo de subcultivos en ensayos controlados que incluyan controles gnotobióticos y cuantificación molecular (qPCR) de los inóculos y genes funcionales. Estas etapas integradas permitirán pasar de la descripción composicional a la validación funcional y a la evaluación de aplicaciones biotecnológicas con fundamento experimental.

IX. MATERIAL SUPLEMENTARIO:

Capacidad madurativa según familias

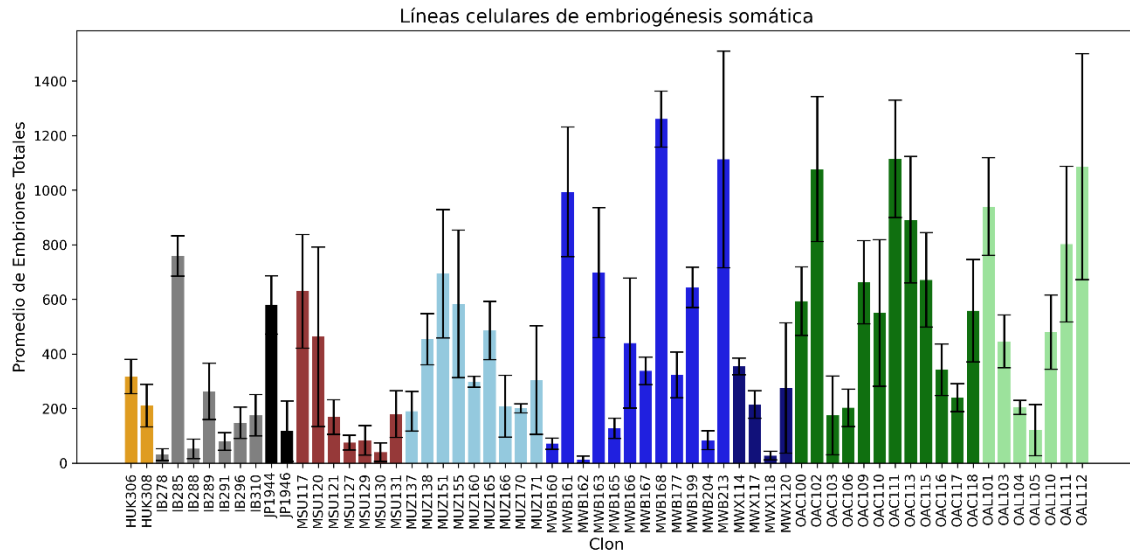


Figura Suplementaria 1. Capacidad de maduración de líneas celulares embriogénicas (ECL) de *Pinus radiata* evaluada mediante la producción promedio de embriones normales y totales por gramo (Emb Norm/g y Emb Tot/g) en 60 clones. Los datos representan el promedio de 3 repeticiones biológicas con 2 réplicas técnicas cada una; las barras de error indican la desviación estándar. Los colores identifican las familias clonales (MWB, IB, OAC, OAL, entre otras). Se observa una amplia variabilidad entre clones, con embriones totales que oscilan entre $12,5 \pm 12,5$ (clon MWB162) y $1.260,4 \pm 102,8$ (clon MWB168), reflejando una diferencia de más de 100 veces en la capacidad madurativa.

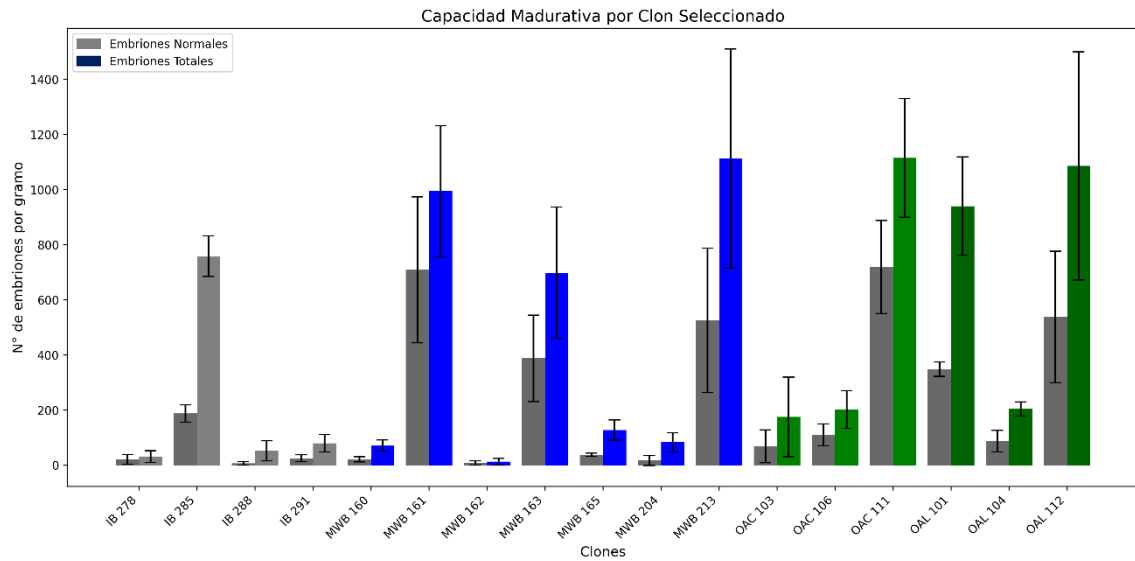


Figura Suplementaria 2 Capacidad de maduración de líneas celulares embriogénicas (ECL) seleccionadas de *Pinus radiata*, evaluada por el número promedio de embriones normales y totales por gramo de masa embrionaria (Emb Norm/g y Emb Tot/g). Se muestran 18 clones representativos pertenecientes a distintas familias clonales (MWB, IB, OAC, OAL), cuyos colores distintivos permiten su diferenciación en el gráfico. Los valores corresponden al promedio de tres repeticiones biológicas con dos réplicas técnicas cada una; las barras de error representan la desviación estándar. Se observa una marcada variabilidad entre los clones analizados, con valores de embriones totales que varían desde $12,5 \pm 12,5$ (MWB 162) hasta $938,5 \pm 38,4$ (MWB 165), reflejando diferencias significativas en la capacidad madurativa dentro del grupo seleccionado.

Secuenciación Sanger

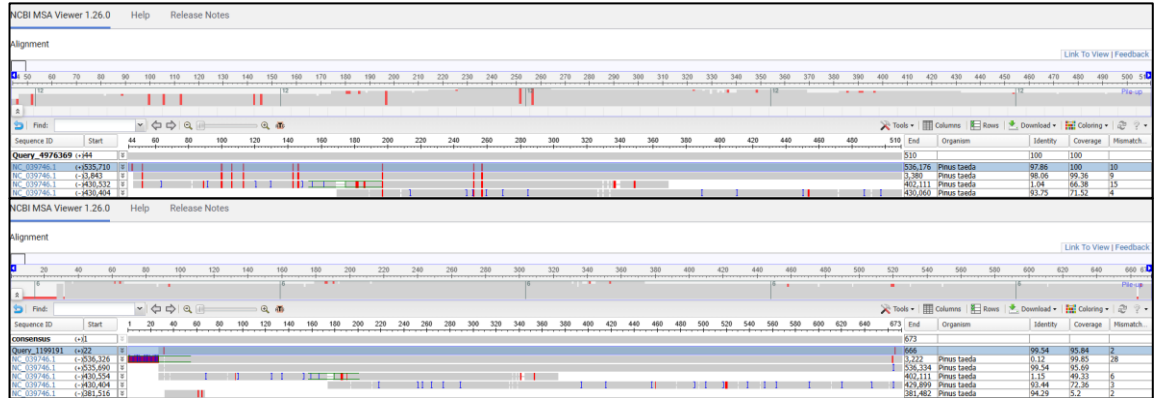


Figura Suplementaria 3 Alineamientos BLAST de secuencias Sanger contra la mitocondria de *Pinus taeda* visualizados con MSA viewer. Visualización de los resultados de alineamiento mediante BLAST entre las secuencias obtenidas por Sanger y el genoma mitocondrial de *Pinus taeda* (NC_039746.1).

Secuenciación por Oxford Nanopore.



Figura Suplementaria 4 Visualización de alineamientos representativos del barcode 12 en IGV, filtrados por longitud ≥ 1000 pb, cobertura ≥ 95 % y calidad $Q \geq 5$. (a) Lectura mapeada en los primeros 1.500 pb de CP006644.1_amplicon1, mostrando reducción de clipping y alineamiento estable (MAPQ = 41, Base = A, QV 17). (b) Lectura mapeada en la región media/final del amplicón (NZ_CP006644.1_amplicon3: 3.295–4.578 pb), con alineamiento confiable y menor pérdida de secuencia por clipping (MAPQ = 16, Base = A, QV 32).



Figura Suplementaria 5 Visualización de alineamientos representativos del barcode 12 en IGV, filtrados únicamente por longitud ≥ 1000 pb y cobertura ≥ 95 %, con calidad mínima $Q = 1$. (a) Lectura mapeada en los primeros 1.500 pb de AP018711.1_amplicon1, mostrando alto nivel de soft clipping al inicio y al final de la lectura, lo que refleja menor estabilidad del alineamiento (MAPQ = 1, Base = C, QV 10). (b) Lectura mapeada en la región media/final del amplicón (AP029021.1_amplicon1: 3.132–4.414 pb), con similar nivel de clipping y menor confianza en las bases alineadas (MAPQ = 1, Base = G, QV 42).

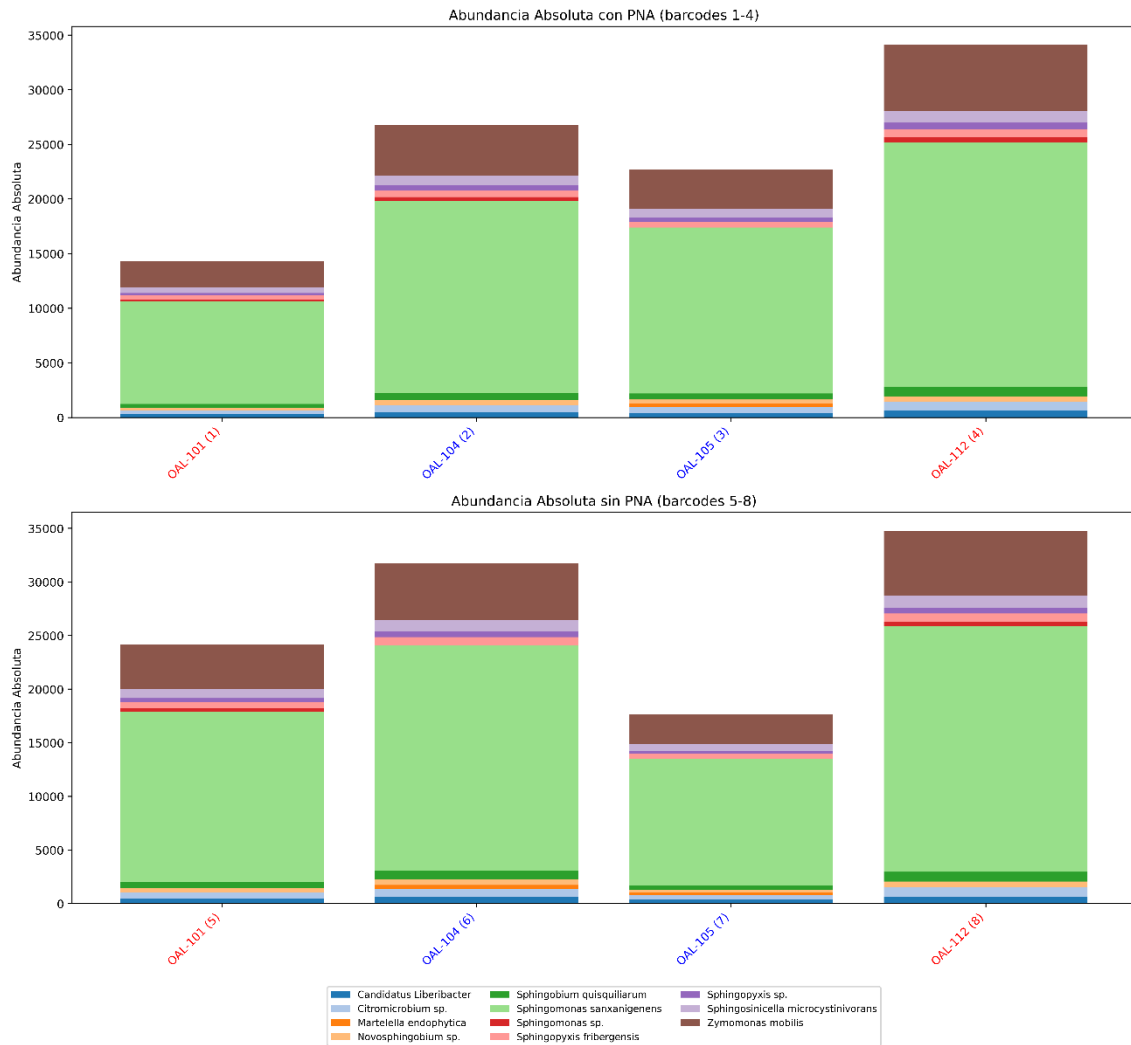


Figura Suplementaria 6 Abundancia absoluta de las 10 especies bacterianas más representativas en muestras de clones de muestras embriogénicas de *Pinus radiata* con y sin presencia de PNAs. Los gráficos muestran barras apiladas que representan la contribución individual de cada especie a la comunidad microbiana total en cada muestra, identificada por su código de barcode y nombre del clon. Las etiquetas en el eje x están coloreadas según la capacidad madurativa del clon (rojo para alta capacidad, azul para baja). El panel superior corresponde a las muestras con PNA (barcodes 1 a 4), mientras que el inferior agrupa las muestras sin PNA (barcodes 5 a 8). La clasificación taxonómica a nivel de especie se obtuvo mediante alineamiento de las lecturas con minimap2 usando la base de datos EMU.

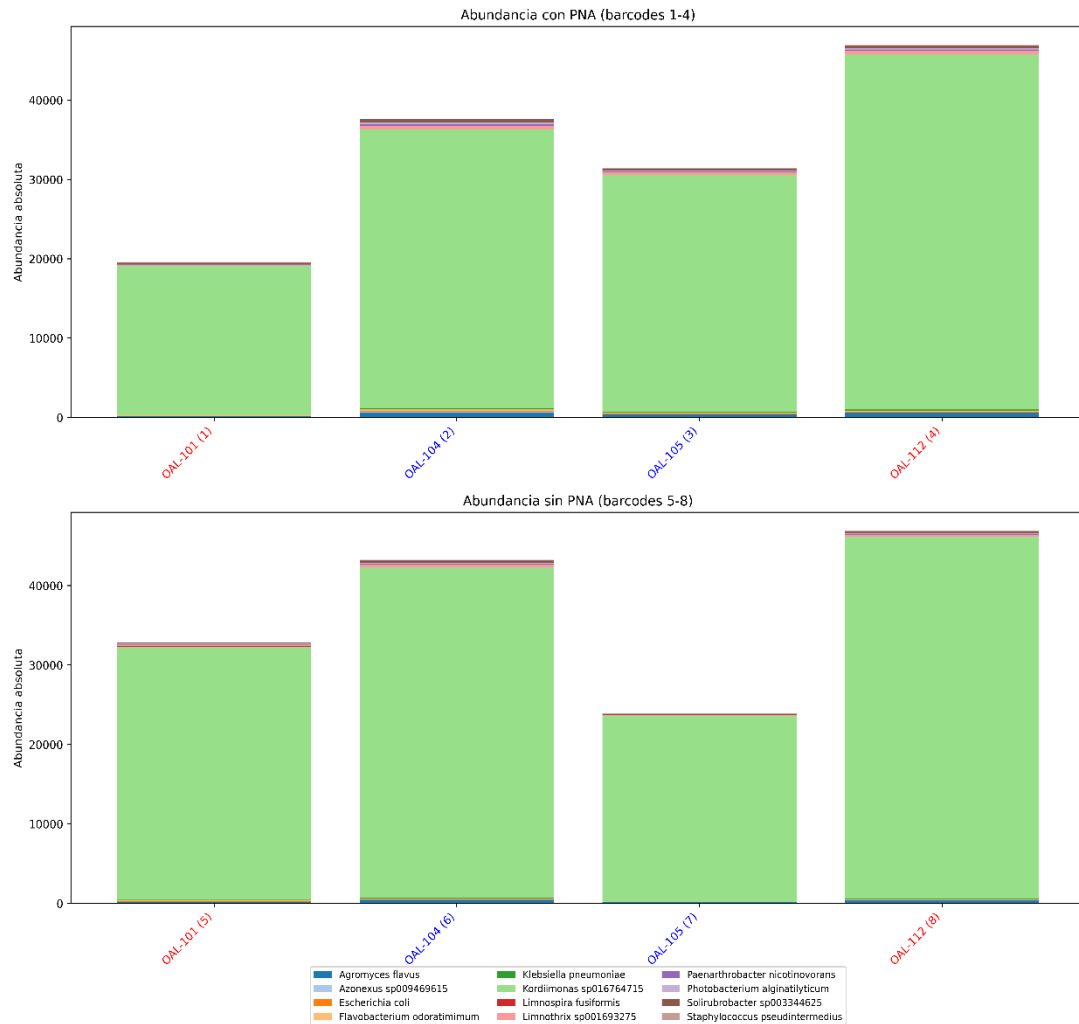


Figura Suplementaria 7 Abundancia absoluta de las 10 especies bacterianas más representativas en muestras de clones de muestras embriogénicas de *Pinus radiata* con y sin presencia de PNAs. Los gráficos muestran barras apiladas que representan la contribución individual de cada especie a la comunidad microbiana total en cada muestra, identificada por su código de barcode y nombre del clon. Las etiquetas en el eje x están coloreadas según la capacidad madurativa del clon (rojo para alta capacidad, azul para baja). El panel superior corresponde a las muestras con PNA (barcodes 1 a 4), mientras que el inferior agrupa las muestras sin PNA (barcodes 5 a 8). La clasificación taxonómica a nivel de especie se obtuvo mediante alineamiento de las lecturas con minimap2 usando la base de datos GROND

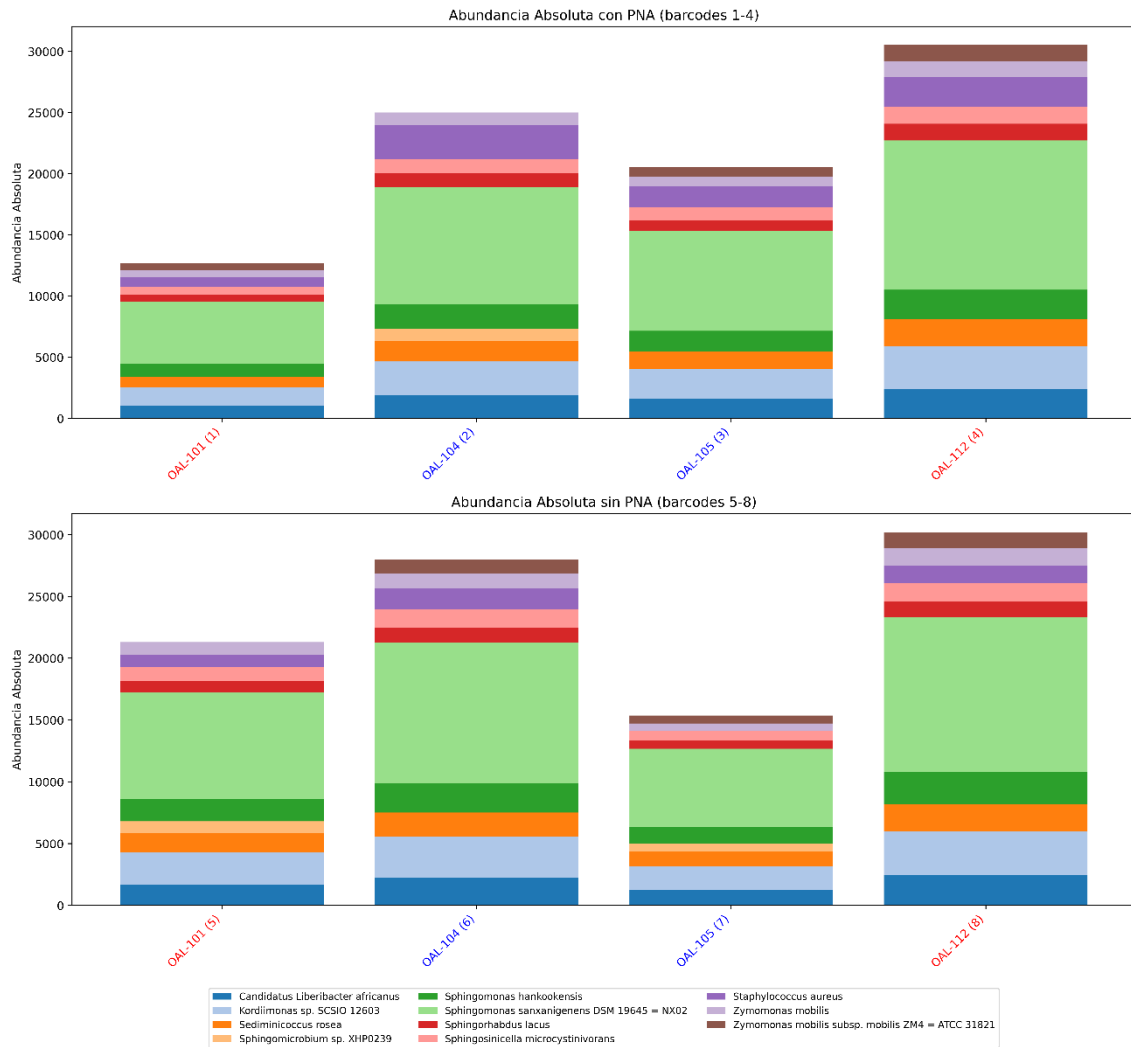


Figura Suplementaria 8 Abundancia absoluta de las 10 especies bacterianas más representativas en muestras de clones de muestras embriogénicas de *Pinus radiata* con y sin presencia de PNAs. Los gráficos muestran barras apiladas que representan la contribución individual de cada especie a la comunidad microbiana total en cada muestra, identificada por su código de barcode y nombre del clon. Las etiquetas en el eje x están coloreadas según la capacidad madurativa del clon (rojo para alta capacidad, azul para baja). El panel superior corresponde a las muestras con PNA (barcodes 1 a 4), mientras que el inferior agrupa las muestras sin PNA (barcodes 5 a 8). La clasificación taxonómica a nivel de especie se obtuvo mediante alineamiento de las lecturas con minimap2 usando la base de datos rGEN.

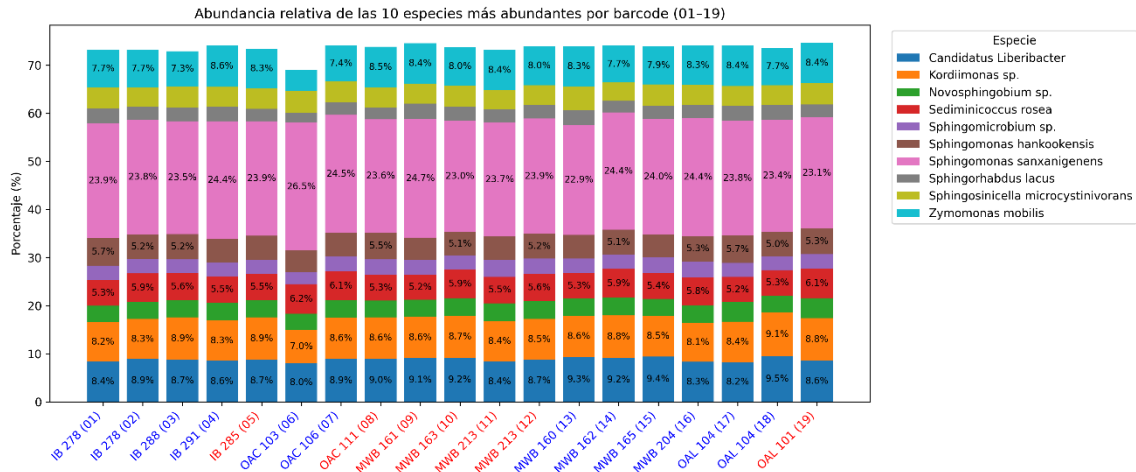


Figura Suplementaria 9 Composición relativa de las 10 especies bacterianas más abundantes en muestras embriogénicas de *Pinus radiata*. Gráfico de barras apiladas que muestra la proporción relativa de lecturas asignadas a las diez especies bacterianas más abundantes en cada muestra individual (barcodes 01-19). Cada barra representa una muestra y está coloreada según la contribución proporcional de cada especie al total de lecturas clasificadas. Los valores porcentuales están indicados dentro de cada segmento. Las etiquetas del eje X están codificadas por color de acuerdo con la condición de la muestra: azul para clones de baja capacidad madurativa y rojo para clones de alta capacidad madurativa.

X. BIBLIOGRAFÍA:

Ajjah, N., Fiodor, A., Pandey, A. K., Rana, A., & Pranaw, K. (2023). Plant Growth-Promoting Bacteria (PGPB) with Biofilm-Forming Ability: A Multifaceted Agent for Sustainable Agriculture. *Diversity*, 15(1), 112. <https://doi.org/10.3390/d15010112>

Armstrong, C., Ganasamurthy, S., Wigley, K., Mercier, C., & Wakelin, S. (2024). The microorganisms and metabolome of *Pinus radiata* Pollen. *Environmental Microbiome*, 19(1). <https://doi.org/10.1186/s40793-024-00656-4>

Arnold, S., Sabala, I., Peter, B., Dyachok, J., & Filonova, L. (2002). Developmental pathways of somatic embryogenesis. *Plant Cell, Tissue and Organ Culture*, 69, 233-249. <https://doi.org/10.1023/A:1015673200621>

Azmat, M. A., Khan, I. A., Cheema, H. M. N., Rajwana, I. A., Khan, A. S., & Khan, A. A. (2012). Extraction of DNA suitable for PCR applications from mature leaves of *Mangifera indica* L. *Journal of Zhejiang University SCIENCE B*, 13(4), 239-243. <https://doi.org/10.1631/jzus.B1100194>

Ballesteros, D., Martínez, M. T., Sánchez-Romero, C., Montalbán, I. A., Sales, E., Moncaleán, P., Arrillaga, I., & Corredoira, E. (2024). Current status of the cryopreservation of embryogenic material of woody species. *Frontiers in Plant Science*, 14, 1337152. <https://doi.org/10.3389/fpls.2023.1337152>

Beckers, B., Op De Beeck, M., Weyens, N., Boerjan, W., & Vangronsveld, J. (2017). Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome*, 5(1). <https://doi.org/10.1186/s40168-017-0241-2>

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., ...

- Schlöter, M. (2020). Microbiome definition re-visited: Old concepts and new challenges. *Microbiome*, 8(1), 103. <https://doi.org/10.1186/s40168-020-00875-0>
- Boss, B. L., Wanees, A. E., Zaslow, S. J., Normile, T. G., & Izquierdo, J. A. (2022). Comparative genomics of the plant-growth promoting bacterium *Sphingobium* sp. Strain AEW4 isolated from the rhizosphere of the beachgrass *Ammophila breviligulata*. *BMC Genomics*, 23(1). <https://doi.org/10.1186/s12864-022-08738-8>
- Brader, G., Compant, S., Mitter, B., Trognitz, F., & Sessitsch, A. (2014). Metabolic potential of endophytic bacteria. *Current Opinion in Biotechnology*, 27, 30-37. <https://doi.org/10.1016/j.copbio.2013.09.012>
- Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology*, 64(1), 807-838. <https://doi.org/10.1146/annurev-arplant-050312-120106>
- Cappelletti, M., Presentato, A., Piacenza, E., Firrincieli, A., Turner, R. J., & Zannoni, D. (2020). Biotechnology of *Rhodococcus* for the production of valuable compounds. *Applied Microbiology and Biotechnology*, 104(20), 8567-8594. <https://doi.org/10.1007/s00253-020-10861-z>
- Carrell, A. A., & Frank, A. C. (2015). Bacterial endophyte communities in the foliage of coast redwood and giant sequoia. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.01008>
- Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84(1), 45-67. <https://doi.org/10.1890/13-0133.1>



Chen, B., Luo, S., Wu, Y., Ye, J., Wang, Q., Xu, X., Pan, F., Khan, K. Y., Feng, Y., & Yang, X. (2017). The Effects of the Endophytic Bacterium *Pseudomonas fluorescens* Sasm05 and IAA on the Plant Growth and Cadmium Uptake of *Sedum alfredii* Hance. *Frontiers in Microbiology*, 8, 2538. <https://doi.org/10.3389/fmicb.2017.02538>

Cline, J. (1996). PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, 24(18), 3546-3551. <https://doi.org/10.1093/nar/24.18.3546>

Compant, S., Samad, A., Faist, H., & Sessitsch, A. (2019). A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *Journal of Advanced Research*, 19, 29-37. <https://doi.org/10.1016/j.jare.2019.03.004>

Cusco, A., Catozzi, C., Vines, J., Sanchez, A., & Francino, O. (2018). *Microbiota profiling with long amplicons using Nanopore sequencing: Full-length 16S rRNA gene and whole rrn operon*. *Microbiology*. <https://doi.org/10.1101/450734>

Cuscó, A., Viñes, J., D'Andrea, S., Riva, F., Casellas, J., Sánchez, A., & Francino, O. (2017). *Using MinION™ to characterize dog skin microbiota through full-length 16S rRNA gene sequencing approach*. *Microbiology*. <https://doi.org/10.1101/167015>

De França Bettencourt, G. M., Degenhardt, J., Dos Santos, G. D., Vicente, V. A., & Soccol, C. R. (2021). Metagenomic analyses, isolation and characterization of endophytic bacteria associated with *Eucalyptus urophylla* BRS07-01 in vitro plants. *World Journal of Microbiology and Biotechnology*, 37(10), 164. <https://doi.org/10.1007/s11274-021-03127-x>

De-la-Peña, C., Nic-Can, G. I., Galaz-Ávalos, R. M., Avilez-Montalvo, R., & Loyola-Vargas, V. M. (2015). The role of chromatin modifications in somatic

embryogenesis in plants. *Frontiers in Plant Science*, 6.
<https://doi.org/10.3389/fpls.2015.00635>

Demeke, T., & Jenkins, G. R. (2010). Influence of DNA extraction methods, PCR inhibitors and quantification methods on real-time PCR assay of biotechnology-derived traits. *Analytical and Bioanalytical Chemistry*, 396(6), 1977-1990.
<https://doi.org/10.1007/s00216-009-3150-9>

Dubois, B., Delitte, M., Lengrand, S., Bragard, C., Legrève, A., & Debode, F. (2024a). PRONAME: A user-friendly pipeline to process long-read nanopore metabarcoding data by generating high-quality consensus sequences. *Frontiers in Bioinformatics*, 4, 1483255. <https://doi.org/10.3389/fbinf.2024.1483255>

Dubois, B., Delitte, M., Lengrand, S., Bragard, C., Legrève, A., & Debode, F. (2024b). PRONAME: A user-friendly pipeline to process long-read nanopore metabarcoding data by generating high-quality consensus sequences. *Frontiers in Bioinformatics*, 4. <https://doi.org/10.3389/fbinf.2024.1483255>

Elhiti, M., Stasolla, C., & Wang, A. (2013). Molecular regulation of plant somatic embryogenesis. *In Vitro Cellular & Developmental Biology - Plant*, 49(6), 631-642.
<https://doi.org/10.1007/s11627-013-9547-3>

Esposito-Polesi, N. P., De Abreu-Tarazi, M. F., De Almeida, C. V., Tsai, S. M., & De Almeida, M. (2017). Investigation of Endophytic Bacterial Community in Supposedly Axenic Cultures of Pineapple and Orchids with Evidence on Abundant Intracellular Bacteria. *Current Microbiology*, 74(1), 103-113.
<https://doi.org/10.1007/s00284-016-1163-0>

Etesami, H. (2025). The dual nature of plant growth-promoting bacteria: Benefits, risks, and pathways to sustainable deployment. *Current Research in Microbial Sciences*, 9, 100421. <https://doi.org/10.1016/j.crmicr.2025.100421>



Fang, L., Xu, X., Li, J., Zheng, F., Li, M., Yan, J., Li, Y., Zhang, X., Li, L., Ma, G., Zhang, A., Lv, F., Wu, K., & Zeng, S. (2020). Transcriptome analysis provides insights into the non-methylated lignin synthesis in *Paphiopedilum armeniacum* seed. *BMC Genomics*, *21*(1), 524. <https://doi.org/10.1186/s12864-020-06931-1>

Fitzpatrick, C. R., Lu-Irving, P., Copeland, J., Guttman, D. S., Wang, P. W., Baltrus, D. A., Dlugosch, K. M., & Johnson, M. T. J. (2018). Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome*, *6*(1), 144. <https://doi.org/10.1186/s40168-018-0534-0>

Flemming, H.-C., & Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology*, *8*(9), 623-633. <https://doi.org/10.1038/nrmicro2415>

Flörl, L., & Bokulich, N. A. (2025). Is the Juice Worth the Squeeze? Evaluating Peptide Nucleic Acid Blockers for Reducing Plant DNA Contamination in 16S rRNA Gene Sequencing. *PhytoFrontiers™*, *5*(1), 17-21. <https://doi.org/10.1094/PHYTOFR-11-24-0124-SC>

Forry, S. P., Servetas, S. L., Kralj, J. G., Soh, K., Hadjithomas, M., Cano, R., Carlin, M., Amorim, M. G. D., Auch, B., Bakker, M. G., Bartelli, T. F., Bustamante, J. P., Cassol, I., Chalita, M., Dias-Neto, E., Duca, A. D., Gohl, D. M., Kazantseva, J., Haruna, M. T., ... Jackson, S. A. (2024). Variability and bias in microbiome metagenomic sequencing: An interlaboratory study comparing experimental protocols. *Scientific Reports*, *14*(1). <https://doi.org/10.1038/s41598-024-57981-4>

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*(4), 379-391. <https://doi.org/10.1046/j.1461-0248.2001.00230.x>

Gu, J., Qiu, Q., Yu, Y., Sun, X., Tian, K., Chang, M., Wang, Y., Zhang, F., & Huo, H. (2024). Bacterial transformation of lignin: Key enzymes and high-value



products. *Biotechnology for Biofuels and Bioproducts*, 17(1), 2. <https://doi.org/10.1186/s13068-023-02447-4>

Hanshew, A. S., Mason, C. J., Raffa, K. F., & Currie, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *Journal of Microbiological Methods*, 95(2), 149-155. <https://doi.org/10.1016/j.mimet.2013.08.007>

Hardoim, P. R., Van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., Döring, M., & Sessitsch, A. (2015). The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiology and Molecular Biology Reviews*, 79(3), 293-320. <https://doi.org/10.1128/MMBR.00050-14>

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362. <https://doi.org/10.1038/s41586-020-2649-2>

Hlangwani, E., Abrahams, A., Masenya, K., & Adebo, O. A. (2023). Analysis of the bacterial and fungal populations in South African sorghum beer (umqombothi) using full-length 16S rRNA amplicon sequencing. *World Journal of Microbiology and Biotechnology*, 39(12), 350. <https://doi.org/10.1007/s11274-023-03764-4>

Huang, H., Wu, M., Yang, H., Li, X., Ren, M., Li, G., & Ma, T. (2016). Structural and physical properties of sanxan polysaccharide from *Sphingomonas sanxanigenens*. *Carbohydrate Polymers*, 144, 410-418. <https://doi.org/10.1016/j.carbpol.2016.02.079>

Huang, H.-D., Wang, W., Ma, T., Li, G.-Q., Liang, F.-L., & Liu, R.-L. (2009). *Sphingomonas sanxanigenens* sp. Nov., isolated from soil. *INTERNATIONAL*

JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY, 59(4), 719-723. <https://doi.org/10.1099/ijs.0.000257-0>

Jakubowska, Z., Gradowski, M., & Dobrzyński, J. (2025). Role of plant growth-promoting bacteria (PGPB) in enhancing phenolic compounds biosynthesis and its relevance to abiotic stress tolerance in plants: A review. *Antonie van Leeuwenhoek*, 118(9), 123. <https://doi.org/10.1007/s10482-025-02130-8>

Kamimura, N., Takahashi, K., Mori, K., Araki, T., Fujita, M., Higuchi, Y., & Masai, E. (2017). Bacterial catabolism of lignin-derived aromatics: New findings in a recent decade: Update on bacterial lignin catabolism. *Environmental Microbiology Reports*, 9(6), 679-705. <https://doi.org/10.1111/1758-2229.12597>

Kandel, S., Joubert, P., & Doty, S. (2017). Bacterial Endophyte Colonization and Distribution within Plants. *Microorganisms*, 5(4), 77. <https://doi.org/10.3390/microorganisms5040077>

Kaur, J., Mudgal, G., Chand, K., Singh, G. B., Perveen, K., Bukhari, N. A., Debnath, S., Mohan, T. C., Charukesi, R., & Singh, G. (2022). An exopolysaccharide-producing novel *Agrobacterium pusense* strain JAS1 isolated from snake plant enhances plant growth and soil water retention. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-25225-y>

Klimaszewska, K., Noceda, C., Pelletier, G., Label, P., Rodriguez, R., & Lelu-Walter, M. A. (2009). Biological characterization of young and aged embryogenic cultures of *Pinus pinaster* (Ait.). *In Vitro Cellular & Developmental Biology - Plant*, 45(1), 20-33. <https://doi.org/10.1007/s11627-008-9158-6>

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1-e1. <https://doi.org/10.1093/nar/gks808>

Knafo, M., Rezenman, S., Idan, T., Elgart, M., Dagan, S., Reich, Z., Kapon, R., Sade, D., & Shental, N. (2023). *CoSMIC - A hybrid approach for large-scale, high-resolution microbial profiling of novel niches*. *Molecular Biology*. <https://doi.org/10.1101/2023.09.03.556087>

Krishnan, R., Menon, R. R., Likhitha, Busse, H.-J., Tanaka, N., Krishnamurthi, S., & Rameshkumar, N. (2017). *Novosphingobium pokkali* sp nov, a novel rhizosphere-associated bacterium with plant beneficial properties isolated from saline-tolerant pokkali rice. *Research in Microbiology*, 168(2), 113-121. <https://doi.org/10.1016/j.resmic.2016.09.001>

Larrouy, J. L., Ridgway, H. J., Dhimi, M. K., & Jones, E. E. (2024). Improvement in Microbiota Recovery Using Cas-9 Digestion of Mānuka Plastid and Mitochondrial DNA. *Microbial Ecology*, 87(1). <https://doi.org/10.1007/s00248-024-02436-6>

Lefèvre, E., Gardner, C. M., & Gunsch, C. K. (2020a). A novel PCR-clamping assay reducing plant host DNA amplification significantly improves prokaryotic endo-microbiome community characterization. *FEMS Microbiology Ecology*, 96(7), fiaa110. <https://doi.org/10.1093/femsec/fiaa110>

Lefèvre, E., Gardner, C. M., & Gunsch, C. K. (2020b). A novel PCR-clamping assay reducing plant host DNA amplification significantly improves prokaryotic endo-microbiome community characterization. *FEMS Microbiology Ecology*, 96(7). <https://doi.org/10.1093/femsec/fiaa110>

Lelu-Walter, M.-A., Klimaszewska, K., Miguel, C., Aronen, T., Hargreaves, C., Teyssier, C., & Trontin, J.-F. (2016). Somatic Embryogenesis for More Effective Breeding and Deployment of Improved Varieties in *Pinus* spp.: Bottlenecks and Recent Advances. En V. M. Loyola-Vargas & N. Ochoa-Alejo (Eds.), *Somatic Embryogenesis: Fundamental Aspects and Applications* (pp. 319-365). Springer International Publishing. https://doi.org/10.1007/978-3-319-33705-0_19

Lengrand, S., Dubois, B., Pesenti, L., Debode, F., & Legrève, A. (2024). Humic substances increase tomato tolerance to osmotic stress while modulating vertically transmitted endophytic bacterial communities. *Frontiers in Plant Science*, 15, 1488671. <https://doi.org/10.3389/fpls.2024.1488671>

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>

Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., Ackermann, M., Hahn, A. S., Srivastava, D. S., Crowe, S. A., Doebeli, M., & Parfrey, L. W. (2018). Function and functional redundancy in microbial systems. *Nature Ecology & Evolution*, 2(6), 936-943. <https://doi.org/10.1038/s41559-018-0519-1>

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, 10(10), 999-1002. <https://doi.org/10.1038/nmeth.2634>

Lurthy, T., Cantat, C., Jeudy, C., Declerck, P., Gallardo, K., Barraud, C., Leroy, F., Ourry, A., Lemanceau, P., Salon, C., & Mazurier, S. (2020). Impact of Bacterial Siderophores on Iron Status and Ionome in Pea. *Frontiers in Plant Science*, 11, 730. <https://doi.org/10.3389/fpls.2020.00730>

Martijn, J., Lind, A. E., Schön, M. E., Spiertz, I., Juzokaite, L., Bunikis, I., Pettersson, O. V., & Ettema, T. J. G. (2019). Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. *Environmental Microbiology*, 21(7), 2485-2498. <https://doi.org/10.1111/1462-2920.14636>

Martínez-Ruiz, R., Azpíroz-Rivero, H. S., Rodríguez-De La O, J. L., Cetina-Alcalá, V. M., Gutiérrez-Espinosa, M. A., & Sahagún-Castellanos, J. (2005). MICROPROPAGACIÓN CLONAL IN VITRO EN EUCALYPTUS GRANDIS Y E.

UROPHYLLA. *Ra Ximhai*, 111-130.
<https://doi.org/10.35197/rx.01.01.2005.08.RM>

Masai, E., Katayama, Y., & Fukuda, M. (2007). Genetic and Biochemical Investigations on Bacterial Catabolic Pathways for Lignin-Derived Aromatic Compounds. *Bioscience, Biotechnology, and Biochemistry*, 71(1), 1-15.
<https://doi.org/10.1271/bbb.60437>

Matsuo, Y., Komiya, S., Yasumizu, Y., Yasuoka, Y., Mizushima, K., Takagi, T., Kryukov, K., Fukuda, A., Morimoto, Y., Naito, Y., Okada, H., Bono, H., Nakagawa, S., & Hirota, K. (2021). Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC Microbiology*, 21(1), 35. <https://doi.org/10.1186/s12866-021-02094-5>

Mazoyon, C., Hirel, B., Pecourt, A., Catterou, M., Gutierrez, L., Sarazin, V., Dubois, F., & Duclercq, J. (2023). *Sphingomonas sediminicola* Is an Endosymbiotic Bacterium Able to Induce the Formation of Root Nodules in Pea (*Pisum sativum* L.) and to Enhance Plant Biomass Production. *Microorganisms*, 11(1), 199. <https://doi.org/10.3390/microorganisms11010199>

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56-61.
<https://doi.org/10.25080/Majora-92bf1922-00a>

Montalbán, I. A., De Diego, N., & Moncaleán, P. (2010). Bottlenecks in *Pinus radiata* somatic embryogenesis: Improving maturation and germination. *Trees*, 24(6), 1061-1071. <https://doi.org/10.1007/s00468-010-0477-y>

Montalbán, I. A., De Diego, N., & Moncaleán, P. (2012). Enhancing initiation and proliferation in *radiata* pine (*Pinus radiata* D. Don) somatic embryogenesis through seed family screening, zygotic embryo staging and media adjustments. *Acta Physiologiae Plantarum*, 34(2), 451-460. <https://doi.org/10.1007/s11738-011-0841-6>

Montalbán, I. A., García-Mendiguren, O., & Moncaleán, P. (2016). Somatic Embryogenesis in *Pinus* spp. En M. A. Germana & M. Lambardi (Eds.), *In Vitro Embryogenesis in Higher Plants* (Vol. 1359, pp. 405-415). Springer New York. https://doi.org/10.1007/978-1-4939-3061-6_21

Moreno-Valencia, F. D., Plascencia-Espinosa, M. Á., Morales-García, Y. E., & Muñoz-Rojas, J. (2024). Selection and Effect of Plant Growth-Promoting Bacteria on Pine Seedlings (*Pinus montezumae* and *Pinus patula*). *Life*, *14*(10), 1320. <https://doi.org/10.3390/life14101320>

Navgire, G. S., Goel, N., Sawhney, G., Sharma, M., Kaushik, P., Mohanta, Y. K., Mohanta, T. K., & Al-Harrasi, A. (2022). Analysis and Interpretation of metagenomics data: An approach. *Biological Procedures Online*, *24*(1). <https://doi.org/10.1186/s12575-022-00179-7>

Notario, E., Visci, G., Fosso, B., Gissi, C., Tanaskovic, N., Rescigno, M., Marzano, M., & Pesole, G. (2023). Amplicon-Based Microbiome Profiling: From Second- to Third-Generation Sequencing for Higher Taxonomic Resolution. *Genes*, *14*(8), 1567. <https://doi.org/10.3390/genes14081567>

Olivier, S. A., Bull, M. K., Strube, M. L., Murphy, R., Ross, T., Bowman, J. P., & Chapman, B. (2023). Long-read MinION™ sequencing of 16S and 16S-ITS-23S rRNA genes provides species-level resolution of Lactobacillaceae in mixed communities. *Frontiers in Microbiology*, *14*, 1290756. <https://doi.org/10.3389/fmicb.2023.1290756>

Orozco-Mosqueda, Ma. D. C., Santoyo, G., & Glick, B. R. (2023). Recent Advances in the Bacterial Phytohormone Modulation of Plant Growth. *Plants*, *12*(3), 606. <https://doi.org/10.3390/plants12030606>

Pandey, P., Tripathi, A., Dwivedi, S., Lal, K., & Jhang, T. (2023). Deciphering the mechanisms, hormonal signaling, and potential applications of endophytic



microbes to mediate stress tolerance in medicinal plants. *Frontiers in Plant Science*, 14, 1250020. <https://doi.org/10.3389/fpls.2023.1250020>

Park, Y.-S., Beaulieu, J., & Bousquet, J. (2016). Multi-varietal forestry integrating genomic selection and somatic embryogenesis. *Vegetative Propagation of Forest Trees*, 302, 302-332.

Perez, J. M., Kontur, W. S., Gehl, C., Gille, D. M., Ma, Y., Niles, A. V., Umana, G., Donohue, T. J., & Noguera, D. R. (2021). Redundancy in Aromatic O - Demethylation and Ring-Opening Reactions in *Novosphingobium aromaticivorans* and Their Impact in the Metabolism of Plant-Derived Phenolics. *Applied and Environmental Microbiology*, 87(8), e02794-20. <https://doi.org/10.1128/AEM.02794-20>

Petrone, J. R., Rios Glusberger, P., George, C. D., Milletich, P. L., Ahrens, A. P., Roesch, L. F. W., & Triplett, E. W. (2023). RESCUE: A validated Nanopore pipeline to classify bacteria through long-read, 16S-ITS-23S rRNA sequencing. *Frontiers in Microbiology*, 14, 1201064. <https://doi.org/10.3389/fmicb.2023.1201064>

Poria, V., Dębiec-Andrzejewska, K., Fiodor, A., Lyzohub, M., Ajjiah, N., Singh, S., & Pranaw, K. (2022). Plant Growth-Promoting Bacteria (PGPB) integrated phytotechnology: A sustainable approach for remediation of marginal lands. *Frontiers in Plant Science*, 13, 999866. <https://doi.org/10.3389/fpls.2022.999866>

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590-D596. <https://doi.org/10.1093/nar/gks1219>

Rangjaroen, C., Sungthong, R., Rerkasem, B., Teaumroong, N., Noisangiam, R., & Lumyong, S. (2017). Untapped Endophytic Colonization and Plant Growth-Promoting Potential of the Genus *Novosphingobium* to Optimize Rice Cultivation.

Microbes and Environments, 32(1), 84-87.

<https://doi.org/10.1264/jsme2.me16112>

Rodríguez-Conde, S., Molina, L., González, P., García-Puente, A., & Segura, A. (2016). Degradation of phenanthrene by *Novosphingobium* sp. HS2a improved plant growth in PAHs-contaminated environments. *Applied Microbiology and Biotechnology*, 100(24), 10627-10636. <https://doi.org/10.1007/s00253-016-7892-y>

Sahu, S. K., Thangaraj, M., & Kathiresan, K. (2012). DNA Extraction Protocol for Plants with High Levels of Secondary Metabolites and Polysaccharides without Using Liquid Nitrogen and Phenol. *ISRN Molecular Biology*, 2012, 1-6. <https://doi.org/10.5402/2012/205049>

Sánchez, E. T. (2018). *Trabajo Fin de Máster: Estudio de la diversidad fúngica en cinco especies de pino inoculadas con Fusarium circinatum*.

Santos, M. O., & Aragão, F. J. L. (2009). Role of SERK genes in plant environmental response. *Plant Signaling & Behavior*, 4(12), 1111-1113. <https://doi.org/10.4161/psb.4.12.9900>

Santoyo, G., Moreno-Hagelsieb, G., Del Carmen Orozco-Mosqueda, Ma., & Glick, B. R. (2016). Plant growth-promoting bacterial endophytes. *Microbiological Research*, 183, 92-99. <https://doi.org/10.1016/j.micres.2015.11.008>

Schloss, P. D., Jenior, M. L., Koumpouras, C. C., Westcott, S. L., & Highlander, S. K. (2016). Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ*, 4, e1869. <https://doi.org/10.7717/peerj.1869>

Seol, D., Lim, J. S., Sung, S., Lee, Y. H., Jeong, M., Cho, S., Kwak, W., & Kim, H. (2022). Microbial Identification Using rRNA Operon Region: Database and Tool for Metataxonomics with Long-Read Sequence. *Microbiology Spectrum*, 10(2), e02017-21. <https://doi.org/10.1128/spectrum.02017-21>

Shade, A., & Handelsman, J. (2012). Beyond the Venn diagram: The hunt for a core microbiome. *Environmental Microbiology*, 14(1), 4-12. <https://doi.org/10.1111/j.1462-2920.2011.02585.x>

Sonett, D., Brown, T., Bengtsson-Palme, J., Padilla-Gamiño, J. L., & Zaneveld, J. R. (2024). Organelles in the ointment: Improved detection of cryptic mitochondrial reads resolves many unknown sequences in cross-species microbiome analyses. *ISME Communications*, 4(1), ycae114. <https://doi.org/10.1093/ismeco/ycae114>

Song, L., & Xie, K. (2020). Engineering CRISPR/Cas9 to mitigate abundant host contamination for 16S rRNA gene-based amplicon sequencing. *Microbiome*, 8(1), 80. <https://doi.org/10.1186/s40168-020-00859-0>

Soundar Raju, C., Aslam, A., Thangadurai, D., Sangeetha, J., Kathiravan, K., & Shajahan, A. (2020). Indole acetic acid (IAA) producing endophytic bacteria on direct somatic embryogenesis and plant regeneration of *Exacum travancoricum* Bedd. *Vegetos*, 33(4), 690-702. <https://doi.org/10.1007/s42535-020-00159-w>

Srinivas, M., Walsh, C. J., Crispie, F., O'Sullivan, O., Cotter, P. D., Van Sinderen, D., & Kenny, J. G. (2025a). Evaluating the efficiency of 16S-ITS-23S operon sequencing for species level resolution in microbial communities. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-024-83410-7>

Srinivas, M., Walsh, C. J., Crispie, F., O'Sullivan, O., Cotter, P. D., Van Sinderen, D., & Kenny, J. G. (2025b). Evaluating the efficiency of 16S-ITS-23S operon sequencing for species level resolution in microbial communities. *Scientific Reports*, 15(1), 2822. <https://doi.org/10.1038/s41598-024-83410-7>

Steward, F. C., Mapes, M. O., & Mears, K. (1958). GROWTH AND ORGANIZED DEVELOPMENT OF CULTURED CELLS. II. Organization in Cultures Grown from Freely Suspended Cell. *American Journal of Botany*, 45(10), 705-708. <https://doi.org/10.1002/j.1537-2197.1958.tb10599.x>



Stoppani, N., Colussi, S., Pastorino, P., Prearo, M., Sciuto, S., Altinok, I., Öztürk, R. Ç., Ture, M., Vela, A. I., Blanco, M. D. M., Kotzamanidis, C., Bitchava, K., Malousi, A., Fariano, L., Volpatti, D., Acutis, P. L., & Fernández-Garayzábal, J. F. (2023). 16S-23S rRNA Internal Transcribed Spacer Region (ITS) Sequencing: A Potential Molecular Diagnostic Tool for Differentiating *Lactococcus garvieae* and *Lactococcus petauri*. *Microorganisms*, 11(5), 1320. <https://doi.org/10.3390/microorganisms11051320>

Szoboszlay, M., Schramm, L., Pinzauti, D., Scerri, J., Sandionigi, A., & Biazzo, M. (2023). Nanopore Is Preferable over Illumina for 16S Amplicon Sequencing of the Gut Microbiota When Species-Level Taxonomic Classification, Accurate Estimation of Richness, or Focus on Rare Taxa Is Required. *Microorganisms*, 11(3), 804. <https://doi.org/10.3390/microorganisms11030804>

The scikit-bio development team. (2020). *scikit-bio: A Bioinformatics Library for Data Scientists, Students, and Developers* (Versión 0.5.5) [Software]. <http://scikit-bio.org>

Thomas, P., & Franco, C. M. M. (2021). Intracellular Bacteria in Plants: Elucidation of Abundant and Diverse Cytoplasmic Bacteria in Healthy Plant Cells Using In Vitro Cell and Callus Cultures. *Microorganisms*, 9(2), 269. <https://doi.org/10.3390/microorganisms9020269>

Thomas, P., Rajendran, T. P., & Franco, C. M. M. (2022). Cytobacts: Abundant and Diverse Vertically Seed-Transmitted Cultivation-Recalcitrant Intracellular Bacteria Ubiquitous to Vascular Plants. *Frontiers in Microbiology*, 13, 806222. <https://doi.org/10.3389/fmicb.2022.806222>

Timofeeva, A. M., Galyamova, M. R., & Sedykh, S. E. (2022). Bacterial Siderophores: Classification, Biosynthesis, Perspectives of Use in Agriculture. *Plants*, 11(22), 3065. <https://doi.org/10.3390/plants11223065>

Vandenkoornhuysen, P., Quaiser, A., Duhamel, M., Le Van, A., & Dufresne, A. (2015). The importance of the microbiome of the plant holobiont. *New Phytologist*, 206(4), 1196-1206. <https://doi.org/10.1111/nph.13312>

Vimal, S. R., Singh, J. S., Kumar, A., & Prasad, S. M. (2024). The plant endomicrobiome: Structure and strategies to produce stress resilient future crop. *Current Research in Microbial Sciences*, 6, 100236. <https://doi.org/10.1016/j.crmicr.2024.100236>

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261-272. <https://doi.org/10.1038/s41592-019-0686-2>

Waechter, C., Fehse, L., Welzel, M., Heider, D., Babalija, L., Cheko, J., Mueller, J., Pöling, J., Braun, T., Pankuweit, S., Weihe, E., Kinscherf, R., Schieffer, B., Luesebrink, U., Soufi, M., & Ruppert, V. (2023). Comparative analysis of full-length 16S ribosomal RNA genome sequencing in human fecal samples using primer sets with different degrees of degeneracy. *Frontiers in Genetics*, 14, 1213829. <https://doi.org/10.3389/fgene.2023.1213829>

Walsh, C. J., Srinivas, M., Stinear, T. P., Van Sinderen, D., Cotter, P. D., & Kenny, J. G. (2024). GROND: A quality-checked and publicly available database of full-length 16S-ITS-23S rRNA operon sequences. *Microbial Genomics*, 10(6). <https://doi.org/10.1099/mgen.0.001255>

Wang, F., Wei, Y., Yan, T., Wang, C., Chao, Y., Jia, M., An, L., & Sheng, H. (2022). *Sphingomonas* sp. Hbc-6 alters physiological metabolism and recruits beneficial rhizosphere bacteria to improve plant growth and drought tolerance. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.1002772>

Wang, F.-X., Shang, G.-D., Wu, L.-Y., Xu, Z.-G., Zhao, X.-Y., & Wang, J.-W. (2020). Chromatin Accessibility Dynamics and a Hierarchical Transcriptional Regulatory Network Structure for Plant Somatic Embryogenesis. *Developmental Cell*, 54(6), 742-757.e8. <https://doi.org/10.1016/j.devcel.2020.07.003>

Wang, Q., Shen, X., Qiu, T., Wu, W., Li, L., Wang, Z., & Shou, H. (2021). Evaluation and application of an efficient plant DNA extraction protocol for laboratory and field testing. *Journal of Zhejiang University-SCIENCE B*, 22(2), 99-111. <https://doi.org/10.1631/jzus.b2000465>

Weirather, J. L., De Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., & Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, 6, 100. <https://doi.org/10.12688/f1000research.10571.2>

Wójcikowska, B., Wójcik, A. M., & Gaj, M. D. (2020). Epigenetic Regulation of Auxin-Induced Somatic Embryogenesis in Plants. *International Journal of Molecular Sciences*, 21(7), 2307. <https://doi.org/10.3390/ijms21072307>

Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., Lewis, S. E., Rost, B., & Goldberg, T. (2016). MSAViewer: Interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics*, 32(22), 3501-3503. <https://doi.org/10.1093/bioinformatics/btw474>

Yang, C., & Ancona, V. (2022). An Overview of the Mechanisms Against “Candidatus Liberibacter asiaticus”: Virulence Targets, Citrus Defenses, and Microbiome. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.850588>

Yang, X., & Zhang, X. (2010). Regulation of Somatic Embryogenesis in Higher Plants. *Critical Reviews in Plant Sciences*, 29(1), 36-57. <https://doi.org/10.1080/07352680903436291>



Zhou, Y., Hu, J., Zhong, Y., & Nie, S. (2025). Promoting efficient synthesis and customization of sphingans based on metabolic engineering and synthetic biology strategies. *Carbohydrate Polymers*, 363, 123734.
<https://doi.org/10.1016/j.carbpol.2025.123734>