UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

TAXONOMIA MULTIFASICA DE HONGOS ENDOFITOS DEL ORDEN HYPOCREALES EN RUBIACEAE DE BOSQUE NATURAL DE COSTA RICA

Tesis sometida a la consideración de la Comisión del Programa de Posgrado de Doctorado en Ciencias para optar al grado y título de Doctorado en Ciencias

EFRAIN ESCUDERO LEYVA

Ciudad Universitaria Rodrigo Facio, Costa Rica

2023

DEDICATORIA

A mi familia,

Patricia y Alejandro. Mis padres. Infinitas gracias por apoyarme en cada paso, en cada etapa. Sin importar la distancia o el tiempo, los llevo presentes a cada lugar con mucho amor.

Braulio, Gerardo, Daniel. Mis hermanos. Gracias por compartir esta vida, no importa el rumbo, siempre están ahí.

Yayo. Porque algún día este capítulo de mi vida te sirva de inspiración para que alcances los sueños más increíbles que tengas.

Algunas cosas pueden parecer nada y lo son todo. Hay que saber ver, aprender a apreciar lo menudo y a despreciar lo que sólo hace bulto. Nada que parece grande ni que reluce en exceso tiene gran validez. Lo bueno es aquello que sin grandes destellos lo llena todo.

Carmen Laforet.

AGRADECIMIENTO

El siguiente trabajo fue posible gracias al apoyo financiero de diversas entidades de la Universidad de Costa Rica: Vicerrectoría de la Investigación (VI), Sistemas de Estudios de Posgrado (SEP), Programa de Doctorado en Ciencias (PDC), Centro de Investigación en Productos Naturales (CIPRONA), Escuela de Biología.

Fondos adicionales fueron provistos gracias a la beca CeNAT-CONARE de 2017 a 2020, además de realizar parte experimental en el Centro Nacional de Innovaciones Biotecnológicas (CENIBiot).

Apoyo económico para asistencia a congresos, experimentos, materiales, entre otros fueron amablemente concedidos por el Centro de Desarrollo de Alternativas Orgánicas (CEDAO-CoopeTarrazú).

Agradezco profundamente a mi tutora Priscila Chaverri Echandi, *PhD* por su guía, enseñanza y todas las increíbles oportunidades brindadas a lo largo de estos años, Agradezco a mi comité de tesis, María del Milagro Montero Granados, *PhD* por sus lecciones, compañía y asistencia durante los ensayos en campo. Giselle Tamayo Castillo, *Dr. rer. Nat.* por su paciencia y conocimientos compartidos. A todas las valiosas mujeres en mi comité, gracias por inspirarme y por exigirme alcanzar siempre la excelencia más allá del trabajo.

Agradezco la ayuda experta y amistad brindadas por los profesores Max Chavarría, Víctor Vásquez, Lorena Hernández, Alicia Hernández. Con toda la gente que trabaja en CIPRONA y que amablemente me acompaño. Fanny Durán, por siempre ayudar con los trámites administrativos y hacerlos fáciles. Elida Ruballos, por la pronta ayuda en el envío de muestras. A todos en CIPRONA: Gracias por ser primer hogar durante estos años.

A las personas en CENIBiot, Randall Loaiza, Emmanuel Araya, Cristofer Orozco, mis agradecimientos por el apoyo y asistencias durante mi estadía.

Expresó mi gratitud con el personal de Coopetarrazú, Eduardo Alvarado y Jimmy Porras. Gracias por el apoyo y retroalimentación en los experimentos en campo y por mostrarme el maravilloso mundo del café. A mis amigos en CEDAO, muchas gracias por ayuda y amistad.

Agradezco profundamente a Laura Aldrich-Wolfe, *PhD* por brindarme la oportunidad de realizar el intercambio académico en su laboratorio en North Dakota State University (NDSU) y por continuar colaborando en Costa Rica y USA. A Laura y los amigos del laboratorio, Jeffrey, Elena, Ankita, Ethan; así como a los amigos de otros departamentos, Diel, Emma, Eglantina, Abby, infinitas gracias por su amistad y por brindar calor a mi alma durante el duro y no obstante, hermoso invierno.

Elena Vásquez, mi compañera, amiga, mi pareja. Mi editora más estricta. Gracias por siempre impulsarme a dar el esfuerzo final.

A mis amigos en Costa Rica. Jackie, Laura, Miguel, Tatiana, Cristinina, Pamela, Marvin, Rodrigo, Noelia, Andrés, Antonio, David, Karolina, Cinthia, Andrea, Ernesto, Josué, Stephanie, Blanca….. a todos ustedes, muchas gracias por apoyarme y por brindarme su amistad y sonrisas en los tiempos difíciles.

A mis amigos en México, la amistad supera fronteras y tiempo, Rubén, Ismael, Denis, Juan, Emmanuel, Cristóbal, Alfonso. Gracias por renovar mi espíritu en cada visita a casa.

A todas las personas que han estado aquí y allá para tenderme su mano, compartir historias y brindar sin importar la razón, GRACIAS.

Esta tesis fue aceptada por la Comisión del Programa de Doctorado en Ciencias de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Doctorado en Ciencias.

CATALINA SALAS DURAN (FIRMA)
Firmado digitalmente por CATALINA SALAS DURAN (FIRMA)
Fecha: 2023.11.22 10:01:25 -06'00'

Dra. Catalina Salas Durán
Representante de la Decana
Sistema de Estudios de Posgrado

PRISCILA CHAVERRI ECHANDI (FIRMA)
PERSONA FISICA, CPF-01-0765-0769.
Fecha declarada: 16/11/2023 04:24:41 PM
Esta representación visual no es fuente de confianza. Valide siempre la firma.

Dra. Priscila Chaverri Echandi
Directora de Tesis

GISELLE TAMAYO CASTILLO (FIRMA)
Digitally signed by GISELLE TAMAYO CASTILLO (FIRMA)
Date: 2023.11.17 08:38:32 -06'00'

Dra. Giselle Tamayo Castillo
Asesora

MARIA DEL MILAGRO GRANADOS MONTERO (FIRMA)
Firmado digitalmente por MARIA DEL MILAGRO GRANADOS MONTERO (FIRMA)
Fecha: 2023.11.20 17:12:37 -06'00'

Dra. María del Milagro Granados Montero
Asesora

MAX CHAVARRIA VARGAS (FIRMA)
Firmado digitalmente por MAX CHAVARRIA VARGAS (FIRMA)
Fecha: 2023.11.24 21:43:50 -06'00'

Dr. Max Chavarría Vargas
Representante de la Directora
Programa de Doctorado en Ciencias

EFRAIN ESCUDERO LEYVA (FIRMA)
Digitally signed by EFRAIN ESCUDERO LEYVA (FIRMA)
Date: 2023.11.15 20:42:04 -06'00'

Efraín Escudero Leyva
Candidato

REFERENCIAS DE LAS PUBLICACIONES

A continuación, se enlistan los artículos científicos que conforman el presente documento de tesis.

1. **Escudero-Leyva, E.**, Granados-Montero, MdM., Orozco-Ortíz, C., Araya-Valverde, E., Alvarado-Picado, E., Chaves-Fallas J.M, Aldrich-Wolfe, L. & Chaverri. P. **2023**. The endophytobiome of wild Rubiaceae as a source of antagonistic fungi against the American Leaf Spot of coffee (*Mycena citricolor*). Journal of Applied Microbiology. 134(5). https://doi.org/10.1093/jambio/lxad090
2. **Escudero-Leyva, E.**, Quirós-Guerrero, L., Vásquez-Chaves, V., Pereira-Reyes, R., Chaverri, P. & Tamayo-Castillo, G. **2023**. Differential volatile organic compound expression in the interaction of *Daldinia eschscholtzii* and *Mycena citricolor*. ACS Omega, 8(34), 31373–31388. https://doi.org/10.1021/acsomega.3c03865
3. Castro-Moretti, F., Cocuron, J-C., Castillo-Gonzales, H., **Escudero-Leyva, E.**, Chaverri, P., Guerreiro-Filho, O., Slot, C.J. & Alonso, A.P. **2023**. A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography mass spectrometry. Frontiers in Plant Science. 13:1057645. doi: 10.3389/fpls.2022.1057645
4. **Escudero-Leyva, E.**, Alfaro-Vargas, P., Muñoz-Arrieta, R., Charpentier-Alfaro, C., Granados-Montero, MdM., Valverde-Madrigal, K.S., Pérez-Villanueva, M., Méndez-Rivera, M., Rodríguez-Rodríguez, C.E., Chaverri, P. & Mora-Villalobos, J.A. **2022**. Tolerance and Biological Removal of Fungicides by Trichoderma Species Isolated from the Endosphere of Wild Rubiaceae Plants. Front. Agron. 3:772170. doi: 10.3389/fagro.2021.772170
5. Zuffa S., Schmid R., Bauermeister A., Gomes PWP., Caraballo-Rodriguez A.M., Abiead Y.E., Aron A.T., Gentry E.C., Zemlin J., Meehan M.J., Avalon N.E., Cichewicz R.H., Buzun E., Terrazas M.C., Hsu C., Oles R., Ayala A.V., Zhao J., Chu H., Kuijpers M.C.M., Jackrel S.L., Tugizimana F., Nephali L.P., Dubery I.A., Madala N.E., Moreira E.A., Costa-Lotufo L.V., Lopes N.P., Rezende-Teixeira P., Jimenez P.C., Rimal B., Patterson A.D., Traxler M.F., de Cassia Pessotti R., Alvarado-Villalobos D., Tamayo-Castillo G., Chaverri P., **Escudero-Leyva E**, Quiros-Guerrero L., Bory A.J., Joubert J., Rutz A., Wolfender J., Allard P., Sichert A., Pontrelli S., Pullman B.S., Bandeira N., Gerwick W.H., Gindro K., Massana-Codina J., Wagner B.C., Forchhammer K., Petras D., Aiosa N., Garg N., Liebeke M., Bourceau P., Kang K.B., Gadhavi H., de Carvalho L.P.S., dos Santos M.S., Pérez-Lorente A.I., Molina-Santiago C., Romero D., Franke R., Brönstrup M., de León A.V.P., Pope P.B., La Rosa S.L., Barbera G.L., Roager H.M., Laursen M.F., Hammerle F., Siewert B., Peintner U., Licona-Cassani C., Rodriguez-Orduña L., Rampler E., Hildebrand F., Koellensperger G., Schoeny H., Hohenwallner K., Panzenboeck L., Gregor R., O'Neill E.C., Roxborough E.T., Odoi J., Bale N.J., Ding S., Sinninghe Damsté J.S., Guan X.L., Cui J.J., Ju K., Silva D.B., Ribeiro Silva F.M., da Silva G.F., Koolen H.H.F., Grundmann C., Clement J.A., Mohimani H., Broders K., McPhail K.L., Ober-Singleton S.E., Rath C.M., McDonald D.,

Knight R., Wang M., Dorrestein P.C. A Taxonomically-informed Mass Spectrometry Search Tool for Microbial Metabolomics Data. doi:10.1101/2023.07.20.549584. PPR:PPR695223.

6. Montero-Vargas, M., Umaña-Jiménez, J.C., **Escudero-Leyva, E.** & P. Chaverri. **2020**. Análisis filogenético de secuencias ITS provenientes de hongos endófitos utilizando Inferencia Bayesiana Paralela de árboles con Exabayes. Revista Tecnología en Marcha, 33(5), Pág. 74-79. https://doi.org/10.18845/tm.v33i5.5079
7. Montero-Vargas, M., **Escudero-Leyva, E.**, Díaz-Valerio, S. & P. Chaverri. **2020**. Step-by-step pipeline for the ecological analysis of endophytic fungi using ITS nrDNA Data. Current Protocols in Microbiology. https://doi.org/10.1002/cpmc.96

Participación en conferencias.

Participación oral.

1. 2023. **Escudero-Leyva E**., Morera-Uribe, D., Rivera-Serracín, K., Sandí-Bolaños, C., Vásquez-Cháves, V., Tamayo-Castillo, G. y Chaverri, P. Diversidad de *Trichoderma* endófitas de Rubiaceae silvestre y su relación con el volatiloma. 2023. XI Congreso Latinoamericano de Micología. Ciudad de Panamá. Panamá.
2. 2021. Chaverri, P., Slot, J., Alonso, A.P., Castillo-Gonzalez, H., **Escudero-Leyva, E.**, Scott, K., Castro-Moretti, F., Alvarado-Picado, E., Granados-Montero, M.M. The mycobiome of wild Rubiaceae to improve the health of coffee plants. 2021. 28th Conference of the Association for the Science and Information on Coffee (ASIC). Montpellier, France.
3. 2021. **Escudero-Leyva E.**, Chaverri, P. Endophytic fungi: from natural forest to the coffee plantations. Researcher Iberoamerican Night.
4. 2019. **Escudero-Leyva E.**, Granados-Montero, M.M., Alvarado, E., Slot, J.C., Alonso, A.P., Chaverri, P. Endophytic *Trichoderma* species with fungicide tolerance. Mycological Society of America annual meeting. Minneapolis, MN, USA.

Poster.

1. 2022. Scott K., H. Castillo-Gonzalez, G. Valero-David, L. Slattery, **E. Escudero-Leyva**, P. Chaverri, J. Slot. Genome comparison of 45 fungal endophytes from Rubiaceae. 31st Fungal Genetics Conference. Asilomar, CA, USA.
2. 2018. **Escudero-Leyva, E.**, H. Castillo González, P. Juárez, M. D. M. Granados, E. Alvarado, P. Chaverri. A first look at culture-dependent endophytic fungal diversity of wild Rubiaceae in Costa Rica. International Mycological Congress. San Juan, Puerto Rico. Jul. 2018.
3. 2018. Castillo-González H., M. Steward, P. Juárez, **Escudero-Leyva, E.**, J. Slot, A. P. Alonso, P. Chaverri. Effect of the leaf developmental stage on the chemical and fungal endophytic composition in wild Rubiaceae. International Mycological Congress. San Juan, Puerto Rico. Jul. 2018.

TABLA DE CONTENIDO

RESUMEN

Los problemas fitosanitarios relacionados al café en Latinoamérica y otras áreas en el mundo, han registrado un aumento en la frecuencia y la severidad de las enfermedades producidas por hongos como *Hemileia vastatrix* ó *Mycena citricolor*. Los hongos endófitos son importantes simbiontes de plantas, conocidos por tener un amplio rango de papeles ecológicos, habitando los tejidos internos sin causar efectos negativos aparentes o inmediatos en la salud del hospedero. Basados en la hipótesis de que los hongos endófitos aislados de bosques prístinos pueden conferir protección a las plagas a través de diferentes mecanismos, el presente trabajo comprende el aislamiento de hongos endófitos de plantas silvestres de Rubiaceae de áreas naturales de Costa Rica para probarlos contra patógenos que afectan al café. Se presentó un protocolo para analizar datos ITS fúngicos, desde el preprocesamiento de calidad de secuencias hasta la clasificación taxonómica, asignación de rasgos funcionales y análisis filogenético, clasificando 1400 aislamientos de hongos endófitos de Rubiaceae silvestre en aproximadamente 800 especies. Para comenzar a generar una alternativa verde para el control de *M. citricolor*, se seleccionaron los hongos endófitos *Daldinia eschscholzi*i GU11N, *Nectria pseudotrichia* GUHN1, *Purpureocillium* aff. *lilacinum* CT24, *Sarocladium* aff. *kiliense* CT25, *Trichoderma rifaii* CT5, *T.* aff *crassum* G1C, *T.* aff. *atroviride* G7T, *T.* aff. *strigosellum* GU12 y *Xylaria multiplex* GU14T. Los aislamientos de *Trichoderma* spp. produjeron los mayores porcentajes de inhibición del crecimiento *in vitro*. Posteriormente, se verificó la colonización endofítica, seguida de ensayos de promoción y antagonismo del crecimiento *in planta*, demostrando que *Trichoderma* CT5 y G1C tienen potencial para promover el crecimiento de las plantas y antagonizar contra *M. citricolor*, reduciendo la incidencia, la severidad y previniendo la mortalidad de las plantas. Estos dos aislamientos posteriormente mostraron tolerancia a fungicidas, obteniendo que CT5 tiene capacidad de remoción de hasta 89% para clorotalonil, 46,4% para ciproconazol y 33,1% para trifloxistrobina. En el caso de la azoxistrobina, la mayor remoción (82,2%) se produjo por adsorción a la biomasa fúngica, al mismo tiempo que detoxifica la matriz acuosa. Posteriormente, se compararon los perfiles volátiles de GU11N con y sin la presencia del patógeno *M. citricolor* mediante ensayos sin contacto y de cultivo dual, utilizando métodos de GC-MS-HS-SPME y UHPLC-HRMS/MS para identificar compuestos no volátiles de extractos orgánicos de los micelios involucrados en la interacción. Los resultados mostraron más compuestos volátiles usando HS-SPME (39 componentes) que con la técnica HS (13 componentes), compartiendo solo 12 compuestos. Las pruebas estadísticas sugirieron que *D. eschscholtzii* inhibe el crecimiento de *M. citricolor* a través de la liberación de COVs que contienen una combinación de 1,8-dimetoxinaftaleno y compuestos terpénicos afectando las estructuras reproductivas del patógeno, por lo que se corroboraron los efectos nocivos del 1,8-dimetoxinaftaleno en una prueba *in vitro* confirmando el daño estructural con fotografías SEM. Los datos de UHPLC-HRMS/MS, revelaron un predominio de derivados de ácidos grasos entre los compuestos identificados. Como parte del estudio de las plantas de Rubiaceae de donde provenían los endófitos, se desarrolló un flujo de trabajo de alto rendimiento para identificar diversos fitoquímicos. El estudio de metabolitos secundarios se realizó en hojas de una variedad comercial de café y dos especies silvestres de Rubiaceae, siendo la cafeína, la trigonelina y la teobromina muy abundantes en las hojas de café, mientras que las moléculas relacionadas con la defensa química, como los fenilpropanoides, el ácido giberélico y el monolignol sinaldehído se encontraron más abundantemente en las hojas silvestres de Rubiaceae. Finalmente, se colaboró integrando datos taxonómicos y metabolómicos a la plataforma MicrobeMASST, una herramienta de búsqueda de espectrometría de masas con información taxonómica altamente curada para la anotación de metabolitos microbianos en experimentos de metabolómica no dirigidos. Esta plataforma contiene 60 000 monocultivos microbianos, permitiendo buscar espectros de MS/MS conocidos y desconocidos y vincularlos con sus respectivos productores microbianos a través de patrones de fragmentación de MS/MS.

ABSTRACT

The phytosanitary problems related to the coffee in Latin America and other producer regions around the globe, have experienced an increase in the frequency and severity of the diseases produced by fungi like *Hemileia vastatrix* or *Mycena citricolor*, in part, because of the climate change. The endophytic fungi are important plant symbionts, known because of their wide ecological guilds, inhabiting the internal tissues without causing immediate negative damages in their hosts. Based in the hypothesis that, endophytic fungi isolated from pristine areas can confer protection against pathogens through different mechanisms, the current work, includes the endophytic fungi isolation from wild Rubiaceae plants in natural areas in Costa Rica to test them against coffee pathogens. A protocol to analyze fungal ITS data is presented, including the steps from the pre-processing of raw sequences up to the taxonomical classification, ecological guilds and phylogenetic analyses. With this, 1400 endophytic isolates from wild Rubiaceae plants were classified in nearly 800 species. Later, to propose a green alternative against *M. citricolor*, the following fungi were selected *Daldinia eschscholzii* GU11N, *Nectria pseudotrichia* GUHN1, *Purpureocillium* aff. *lilacinum* CT24, *Sarocladium* aff. *kiliense* CT25, *Trichoderma rifaii* CT5, *T.* aff *crassum* G1C, *T.* aff. *atroviride* G7T, *T.* aff. *strigosellum* GU12 and *Xylaria multiplex* GU14T. From this, the isolates belonging to Trichoderma spp. obtained the better inhibition values on in vitro conditions. Later, thee endophytic colonization was confirmed and *Trichoderma* CT5 and G1C resulted with growth promotion and antagonistic capacities against *M. citricolor*, reducing incidence, severity and preventing mortality, in tests with coffee plantlets. These two isolates also resulted with tolerance to fungicides, where CT5 had a remotion capacity of 89% for clorotalonil, 46.4% for ciproconazol and 33.1% for trifloxistobine, according to available biomass. For azoxistrobine, the remotion (82.2%) occurred as a result of absorption to the fungal biomass, while detoxification of the matrix was a simultaneous process. In the third part of the results, the volatile profiling of GU11N with and in the absence of M. citricolor was done through contactless and in coculture experiments, using GC-MS-HS-SPME and UHPLC-HRMS/MS to identify the non-volatile compounds from the organic extractions from mycelia recovered from the interaction zone. Results showed more volatile compounds using HS-SPME (39 components) rather than HS (13 components), sharing only 12 compounds. The statistical tests suggested that *D. eschscholtzii* was inhibiting the growth of *M. citricolor* trough the release of VOCs containing a combination of 1,8-dimethoxynaphtalene and terpenic compounds, affecting the reproductive structures of the pathogen, this was confirmed with an-in vitro test, obtaining SEM photographs with evidence of the structural damage in the pathogen. The UHPLC-HRMS/MS data revealed many fatty acids among the determined compounds. As part of the study related to the Rubiaceae plants from where the endophytes were retrieved, a pipeline for identification and quantification of polyphenols was done. The workflow was optimized to extract over 40 families of phytochemicals and the developed metabolomic platform was able of identification and quantification of 184 polyphenols in 15 minutes. The study of secondary metabolites was performed in leaves of coffee and two wild species of Rubiaceae. Caffeine, trigonelline and theobromine were abundant in the coffee leaves, while compounds related to the chemical defense, like phenylpropanoids, gibberellic acid and monolignol sinaldehyde were found abundantly in the wild Rubiaceae. Finally, as part of the collaboration with the MicrobeMASST platform, the taxonomic and metabolomic data related to the endophytes was integrated. This is a tool for the search of mass spectrometry with highly curated taxonomic data for the annotation of microbial metabolites in non-targeted metabolomic experiments. This platform contains 60 000 microbial monocultures, allowing to search for known and unknown MS/MS spectra and link them with their current microbial producer's trough the MS/MS fragmentation patterns.

LISTA DE ABREVIATURAS

ADN	Ácido desoxirribonucleico.
CVOs	Compuestos volátiles orgánicos.
ITS	Subregión interna de transcripción.
PDA	Papa-Dextrosa-Agar.
PDB	Caldo-Papa-Dextrosa.
PCR	Reacción en cadena de la polimerasa.
RPB2	gen de la RNA polimerasa II.
TEF	Factor de traslación de la elongación.
GCMS	Cromatografía Gaseosa acoplada a Masas (Gas Chromatography Mass Spectrometry).
HS	Espacio de cabeza (Headspace).
SPME	Microextracción de Fase Sólida (Solid Phase Microextraction).
UHPLC-HMRMS/MS	Cromatografía Líquida de Ultra Alto Rendimiento acoplada a Masas (Ultra High Performance Liquid Chromatography Mass Spectrometry).

CAPÍTULO I. INTRODUCCIÓN

Los problemas fitosanitarios relacionados al café en Latinoamérica, así como en otras áreas productoras en el mundo, han registrado un aumento en la frecuencia y la severidad debido entre varias causas, a los efectos derivados del cambio climático como lo son las alteraciones en los ciclos de lluvia y sequía (Alpízar et al., 2020). Esto ha disparado la presencia de hongos patógenos como la roya (*Hemileia vastatrix* Berk. and Broome), la cual es considerada actualmente como la enfermedad causante de las mayores pérdidas de café en el mundo (Libert Amico et al., 2020; Talhinhas et al., 2017); o como el "ojo de gallo", enfermedad restringida al continente Americano causada por *Mycena citricolor* Berk. and Broome; la cual es particularmente severa durante períodos de lluvia prolongados, registrando pérdidas de hasta 60 millones de dólares solamente en Costa Rica (Granados-Montero et al., 2020). Aunado a otras plagas, la investigación científica enfocada en reducir el impacto en este importante cultivo puede considerarse como prioritaria no sólo en Costa Rica sino en todos los países productores como Brasil, Colombia y México.

Actualmente, el uso de agroquímicos como los fungicidas es una práctica ampliamente común ya que reduce la severidad de las plagas, sin embargo, existen efectos negativos relacionados a la contaminación del agua y suelo, así como alteraciones en la salud humana (Alpízar et al., 2020), lo que está generando presión en la transición hacia una agricultura con mejores prácticas por el ambiente (Lee et al., 2015). Algunas de las estrategias involucran la detección temprana de los patógenos, mejora en las varietales del café, uso de microorganismos como controladores biológicos y la aplicación de compuestos derivados de los mismos (Avelino et al., 2015). Dichas estrategias, han atraído la atención de los productores y consumidores, quienes se preocupan cada vez más por prácticas más verdes, especialmente aquellas que incluyen microorganismos con potencial antagonista (Pinto et al., 2014). No obstante, existen retos para la utilización intensiva de microbios, como la tolerancia a los agroquímicos o los factores ambientales (por ejemplo, radiación UV, exposición al viento, etc.) que podrían afectar la efectividad de estos en campo (Gharieb et

al., 2004). Dentro de las alternativas mejor estudiadas, los hongos endófitos conforman un grupo interesante de organismos con una alta diversidad por descubrir en los trópicos (Pujade-Renaud et al., 2019b).

La evidencia sugiere que los hongos endófitos son importantes simbiontes de plantas con registros que datan de más de 400 millones de años (Rodriguez et al., 2009); estos organismos son conocidos por tener un amplio rango de papeles ecológicos, habitando los tejidos internos de la planta sin causar efectos negativos aparentes o inmediatos en la salud del hospedero. Además, algunas especies endófitas han sido exitosamente inoculadas en plantas hospedero y puestas en pruebas de resistencia a la sequía, promoción del crecimiento, inducción en la tolerancia a metales pesados, entre otros (Barberis et al., 2021; Busby et al., 2016; Hubbard et al., 2014). El proceso de colonización puede ocurrir a través de aperturas naturales en las plantas, alcanzando la albura, hojas e incluso semillas (Bernardi-Wenzel et al., 2010). La trasmisión de hongos endófitos puede ocurrir de manera horizontal o bien, de la planta madre a la hija durante la formación de las semillas (trasmisión vertical), dependiendo de la planta hospedero y la Clase a la cual pertenece el endófito (Rodriguez et al., 2009).

La comparación de la diversidad de hongos endófitos entre plantas usadas en monocultivos y aquellas habitando bosques prístinos ha demostrado patrones de riqueza biológica que favorecen a aquellas especies de bosque (Evans et al., 2003; Sternhagen et al., 2020). Esto se ha observado en monocultivos de árboles del género *Hevea*, que, comparados con sus parientes silvestres, sufren una pérdida de riqueza en el endobioma (Gazis & Chaverri, 2010). De acuerdo con esto, la Familia Rubiaceae, donde se encuentra *Coffea arabica* L., es una Familia altamente diversa en los trópicos (Manns et al., 2012); y basados en la hipótesis de que los hongos endófitos aislados de bosques prístinos pueden conferir protección a las plagas a través de diferentes mecanismos, como el antagonismo, posterior a la colonización de los tejidos internos de la planta hospedera (De Souza et al., 2008a; Pujade-Renaud et al., 2019), y con datos que sostienen que aquellos endófitos provenientes de plantas filogenéticamente cercanas tendrán mayores rangos de colonización y protección, el presente trabajo comprende el aislamiento de hongos endófitos de plantas

silvestres de Rubiaceae de áreas naturales de Costa Rica para probarlos contra patógenos concernientes al café. Como parte del proyecto, se propuso la utilización de metabolómica no dirigida en ciertas especies de hongos seleccionadas para detectar los posibles compuestos involucrados en la actividad antagonista. Por ejemplo, se sabe que los endófitos son capaces de producir compuestos volátiles orgánicos (CVOs) con actividad biológica (Liarzi et al., 2016). Algunos efectos benéficos en las plantas producidos por los CVOs son: promoción del crecimiento, efectos antimicrobianos y antifúngicos (Morath et al., 2012; Wang et al., 2017). La producción de estos CVOs y otros metabolitos secundarios, derivan de relaciones diversas e intrincadas entre los endófitos y sus hospederos, resultando en profundas interacciones ecológicas (Flores-Reséndiz et al., 2021). Por ello, este trabajo también comprende el estudio de los metabolitos secundarios volátiles y no volátiles de ciertas especies de hongos y de especies de plantas de Rubiaceae silvestre, para futuras hipótesis sobre las relaciones intra e inter-especie.

CAPÍTULO II. HIPOTESIS Y OBJETIVOS

1. Hipótesis

Las especies de Rubiaceae en bosques prístinos de Costa Rica contienen especies fúngicas endófitas con potencial para controlar plagas a través de diversos mecanismos y colonizan los tejidos internos de las plantas de café gracias a su cercanía filogenética dentro de la Familia botánica. Al ocurrir la colonización, son observados efectos benéficos en las plantas de café.

2. Objetivo general

Resolver la clasificación taxonómica de hongos endófitos aislados plantas de Rubiaceae silvestre a través de un acercamiento multifásico y correlacionar sus propiedades antifúngicas contra problemas fitosanitarios del café.

2.1 Objetivos específicos

2.1.1 Aislar e identificar una colección de hongos endófitos de Rubiaceae silvestre de áreas naturales de Costa Rica.

2.2.2 Obtener una filogenia selecta de hongos con potencial como biocontroladores.

2.2.3 Determinar si los caracteres químicos (metabolitos secundarios) se correlacionan con la actividad antifúngica.

CAPÍTULO III. MATERIALES Y METODOS

3.1 Aislamiento, caracterización y pruebas de antagonismo de los hongos endófitos provenientes de Rubiaceae contra *Mycena citricolor*.

3.1.1 Sitios de estudio.

Se muestrearon plantas de Rubiaceae silvestre provenientes de bosques naturales de Costa Rica: Volcán Miravalles y Rincón de la Vieja en el Área de Conservación Guanacaste (ACG) (10.7477 N, 85.1612 W) y bosques privados en los alrededores de Río Claro en el Área del Pacifico Sur (PSA) (8.6389 N, 83.0914 W) en Costa Rica.

3.1.2 Aislamiento e identificación de hongos endófitos de Rubiaceae.

Los aislamientos se obtuvieron a partir de hojas de plantas de Rubiaceae silvestre, siguiendo los protocolos descritos en (Gazis & Chaverri, 2010). Los aislamientos se conservaron en glicerol 20% a –80 °C y a temperatura ambiente en el Centro de Investigación en Productos Naturales (CIPRONA) de la Universidad de Costa Rica. Estos aislamientos se cultivaron en placas Petri de 10 cm de diámetro con PDA (Difco Laboratories, USA). Después de siete días el micelio se cosechó de la superficie y se realizó la extracción de ADN con el kit PrepMan Ultra (Applied Biosystems, USA). Se realizaron PCR usando los primers ITS4 e ITS5 (Schoch et al., 2012), posteriormente se usaron los primers EF728Mf y EF1R para obtener la región TEF (Carbone & Kohn, 1999), finalmente, se amplificó la región RPB2 usando los primers RPB2-5F y RPB2-7CR (Liu et al., 1999). El proceso de edición, alineamiento, comparación con bases de datos y construcción de la filogenia se describen en (Montero-Vargas et al., 2020a) y (Montero Vargas et al., 2020b).

3.1.3 Pruebas de antagonismo *in vitro* contra *Mycena citricolor*.

Se seleccionaron nueve endófitos basados en la relación filogenética de los taxa con su habilidad antagonista: *Daldinia eschscholtzii* GU11N, *Nectria pseudotrichia* GUHN1, *Purpureocillium* aff. *lilacinum* CT24, *Sarocladium* aff. *kiliense* CT25, *Trichoderma rifaii* CT5, *T.* aff. *crassum* G1C, *T.* aff. atroviride G7T, *T.* aff. *strigosellum* GU12 y *Xylaria multiplex* GU14T. El patógeno *Mycena citricolor* se aisló de plantas de café infectadas en

San Marcos de Tarrazú. Se colectaron los pseudopileos y se transfirieron a placas Petri con medio PDA. Para los experimentos de confrontación *in vitro*, patógeno y endófito fueron colocados de manera yuxtapuesta a 5 mm del borde de una placa Petri y su crecimiento se midió cada 24 h por 5 días a temperatura ambiente hasta registrar el contacto. Se midieron rangos de crecimiento control para cada hongo y se calculó el porcentaje de inhibición.

3.1.4 Pruebas de antagonismo en planta.

Para este ensayo se trabajó con plántulas de café (*Coffea arabica* cv. caturra). Se utilizaron 20 plántulas por tratamiento y se trasladaron a contenedores herméticos previamente esterilizados con 700 g de sustrato Berger Custom Blend (Canadá). Los tratamientos fueron i) plantas con el endófito CT5; ii) plantas con el endófito G1C y iii) plantas sin endófito.

Cada una de las plántulas dentro de los tratamientos (i) y (ii) se inocularon con 15 mL de una suspensión de esporas a una concentración de 1×10^7 conidios/mL. Las plantas sin endófito se irrigaron únicamente con 15mL de agua destilada estéril. Para confirmar la colonización de endófitos en diferentes tejidos (tallos y hojas), se siguió el procedimiento detallado en (Escudero-Leyva et al., 2023).

El experimento de antagonismo se realizó utilizando bloques de 10 plantas por tratamiento, para un total de 30 plantas por tratamiento. Las plantas fueron infectadas con el patógeno *Mycena citricolor* en cada tratamiento. Cada semana, las plantas fueron monitoreadas para detectar síntomas de infección y signos del patógeno, y para mantener la humedad >90%. Después de 4 semanas, se midieron los diámetros de las lesiones causadas por *M. citricolor* (en mm). También se registraron otros parámetros, como el número de hojas infectadas (se inocularon tres hojas/planta con el patógeno), la defoliación (%), el número de pseudopilei y la mortalidad de las plantas.

3.1.5 Pruebas estadísticas.

Las variables medidas en el proceso de colonización de endófitos y el ensayo de antagonismo se sometieron a pruebas de normalidad y los datos se transformaron en consecuencia: transformación $\log^{10}$ para altura de planta y raíz cuadrada para longitud de raíz y número de hojas en el experimento de colonización; raíz cuadrada para hojas infectadas. Luego, se utilizaron modelos mixtos lineales generalizados (GLMM) para

inferir el efecto de los tratamientos utilizando el paquete lmer v.3.1.3 debido al diseño del modelo de parcelas divididas derivado del arreglo del experimento. Los modelos consideraron los bloques y contenedores como efectos aleatorios y tratamientos como el efecto fijo. La significación entre los tratamientos se determinó mediante la prueba post hoc Tukey HSD. Todas las pruebas y gráficos se realizaron utilizando el software R v.4.1.3 (R Core Team, 2020).

3.2 Tolerancia de los endófitos a fungicidas.

Para esta etapa se trabajó solo con los aislamientos pertenecientes al género *Trichoderma*: *Trichoderma rifaii* CT5, *T.* aff. *crassum* G1C, *T.* aff. atroviride G7T, *T.* aff. *strigosellum* GU12.

3.2.1 Pruebas en medio sólido.

Se ensayó la capacidad de crecimiento en medio sólido que contenía fungicidas residuales para cuatro aislamientos. Se probaron siete fungicidas de uso común en cafetales: azoxistrobina (sistémico), clorotalonil (no sistémico), ciproconazol (sistémico), propineb (sistémico), tolclofos-metil (no sistémico), trifloxistrobina (sistémico) y validamicina-A. (no sistémico). Los fungicidas se agregaron al PDA después de la esterilización de los medios para evitar su degradación. Se colocaron discos de micelio (~5 mm de diámetro) a 5 mm del borde de la placa. Como se observó tolerancia fúngica con las dosis completas, no se realizaron concentraciones ni diluciones adicionales. El crecimiento radial (mm) se midió cada 24 h durante 5 días con un fotoperíodo de 12 h a 25◦C ± 2.

3.2.2 Pruebas en medio líquido.

Para este ensayo, solo se usaron los aislamientos T1, T2 y T4 debido a su capacidad para producir abundantes conidios para la producción de biomasa. Se evaluó su capacidad de biodegradación frente a los ingredientes activos de los fungicidas azoxistrobina, clorotalonil, ciproconazol y trifloxistrobina. Se realizó un experimento independiente para cada fungicida con tres aislados por separado. Se agregaron fungicidas comerciales a los medios PDB de acuerdo con la dosis recomendada para el café. Los experimentos se realizaron por triplicado, incluyendo: (i) controles abióticos (sin inocular) para evaluar la

degradación por factores abióticos; (ii) controles muertos por calor (HKC; que contienen biomasa esterilizada en autoclave cultivada previamente en condiciones idénticas, pero sin el fungicida) para evaluar la adsorción; y (iii) control de crecimiento (biomasa fúngica viable sin fungicida) para determinar el peso seco final. Todos los cultivos se cultivaron durante 14 días en un agitador rotatorio (250 rpm) a 21°C en la oscuridad. Las pruebas analíticas, así como de citotoxicidad se detallan en (Escudero-Leyva et al., 2022).

3.2.3 Pruebas estadísticas.

Todas las estadísticas y gráficos se realizaron utilizando R Core Team (2020). El efecto de los fungicidas sobre los diferentes aislados de *Trichoderma* en fase sólida se analizó mediante un Análisis de Varianza (ANOVA) de una vía para cada fungicida. Los datos de biodegradación de cada fungicida se procesaron para generar un gráfico múltiple de concentración frente al tiempo y la variabilidad de la muestra se calculó como una desviación estándar (±SD).

3.3 Diferencias en el volatiloma de un endófito durante su interacción con el patógeno de café *Mycena citricolor*.

3.3.1 Experimento sin contacto para *Daldinia eschscholtzii* y *Mycena citricolor.*

Se siguió la técnica del "sándwich", descrita previamente (Liarzi et al., 2016), cultivando ambos hongos en cajas de Petri independientes de 90 mm con PDA. El inóculo consistió en discos de micelio (~5 mm) colocados a 5 mm del borde de la placa de Petri. Los discos de micelio de *Daldinia* se colocaron en la placa inferior y los de *Mycena citricolor* en la placa superior, frente a *Daldinia*. Se midió el crecimiento radial (mm) de *M. citricolor* cada 24 h durante 10 días. Se realizaron diez repeticiones para el experimento. Todas las placas se incubaron a 25ºC con ciclos de luz/oscuridad de 12h.

3.3.2 Experimento en co-cultivo.

Para este experimento se utilizaron cajas de Petri de 90 mm con 20 mL de PDA para cultivar los aislados de *Daldinia* y *Mycena* de forma independiente y en un experimento dual con ambos hongos ("co-cultivo") en la misma placa. Después de diez días, se recolectaron cinco discos de micelio del borde de cada cultivo independiente y tres de la

línea de confrontación del grupo experimental de cocultivo y se colocaron dentro de viales con tapa de rosca de 20 ml con tapones de PTFE/silicona durante 24 horas. h a 25ºC. La mitad de los micelios de los cultivos independientes y los micelios adyacentes de la línea de confrontación se recolectaron con un bisturí estéril y se almacenaron en viales de 20 ml prepesados y se liofilizaron inmediatamente durante 24 horas.

3.3.3 HS y HS-SPME-GCMS.

Los viales con el disco dentro se preequilibraron durante 15 min a 45 °C. Se expuso una fibra PDMS SPME de 100 µm al espacio libre de la muestra durante 15 min a la misma temperatura con agitación en el muestreador automático. Posteriormente, la fibra SPME se desorbió térmicamente a 250 °C durante 10 min en el inyector GC-MS. Las mediciones de compuestos volátiles se realizaron con un cromatógrafo de gases Trace 1310 (Thermo Scientific) con un sistema detector de masas de triple cuadrupolo TSQ 9000 (Agilent Technologies, EE. UU.), equipado con una columna capilar TG-5MS (30 m de longitud × 25 µm de diámetro interno × 1 espesor de la película, Thermo Scientific, EE. UU.). El inyector de cada muestra estaba en modo dividido con un radio dividido de 10 y el caudal del gas portador (helio, 99,999 % de pureza) era de 1,2 ml/min. El programa de temperatura se fijó de la siguiente manera: inicialmente 60°C durante 3 min, se aumentó a 200°C a razón de 10°C/min, luego se aumentó a 290°C a razón de 15°C/min, y finalmente, mantenido durante 2 min. El cromatógrafo de gases se ajustó al modo de ionización EI, flujo de 1 ml/min, rango de masas de 33-550 amu, línea de transferencia a 280 ºC y fuente de iones a 260 ºC.

3.3.4 Procesamiento de datos.

Los archivos provenientes de las corridas de GCMS se procesaron y analizaron usando los softwares MZMine (Smirnov et al., 2018), AMDIS y la plataforma GNPS (Global Natural Products Social Network) (Allegra et al., 2020). Los detalles de los parámetros y otras condiciones pueden consultarse en (Escudero-Leyva, Quirós-Guerrero, et al., 2023).

3.3.5 Análisis estadísticos para el dataset de GCMS.

La tabla de cuantificación en formato .csv obtenida de los experimentos HS y HS-SPME se normalizó por la suma de cada muestra para realizar pruebas ANOVA en R v.4.1.3 88. Debido a la mayor cantidad de características detectadas para el HS-SPME experimento, se realizaron las siguientes pruebas multivariadas aplicando una escala de Pareto: Análisis de Componentes Principales (PCA) usando los paquetes R FactoMineR v.2.4 89 y FactoExtra v.1.0.7 90 seguido de un Análisis Dependiente de Mínimos Cuadrados Parciales (PLS-DA) con el paquete mixOmics v.6.6.2 91 en R v.3.5.1. Se realizó un análisis secundario de PCA y PLSDA para los datos de HS-SPME utilizando solo las características detectadas también en el experimento de HS para comparar de manera justa las detecciones de ambas herramientas.

3.3.6 Actividad *in vitro* de los compuestos sintéticos contra estructuras de *M. citricolor*.

Se saturaron discos de papel filtro (0,5 mm de diámetro) con 1, 10 y 100 µL de la mezcla de terpenos (Supelco, CRM40937, EE. UU.) por duplicado. El 1,8-dimetoxinaftaleno se resuspendió en 509 µL de éter. Los discos de filtro se saturaron por triplicado con 10, 50 y 100 µL usando una jeringa de 100 µL. Los discos se secaron al aire y se colocaron en placas de Petri (6 cm de diámetro) junto con fragmentos de estructuras de *M. citricolor*; las placas se sellaron con Parafilm y se dejaron en temperatura ambiente (25 ºC). Después de 24 h, tres pseudopileos de cada tratamiento se retiraron al azar y se inocularon en cajas de Petri con PDA para evaluar la germinación y el crecimiento. El control negativo para el 1,8-dimetoxinaftaleno consistió en papeles de filtro saturados con 500 µL de éter y los de la mezcla de terpenos con la misma cantidad de metanol, ambos grupos de control se realizaron por triplicado. Los detalles de la obtención del 1,8-dimetoxinaftaleno aparecen en Escudero-Leyva *et al.*, 2023.

3.3.7 Análisis en UHPLC-HRMS.

Las muestras previamente liofilizadas se prepararon como se menciona en Escudero-Leyva *et al.*, 2023b. Los análisis se realizaron con un sistema UHPLC Vanquish Duo (Thermo, ScientificTM, Alemania) acoplado a un espectrómetro de masas Orbitrap Exploris 120 (Thermo ScientificTM, Alemania). El analizador de masas se calibró con una mezcla de

cafeína, metionina-arginina-fenilalanina-alanina-acetato (MRFA), dodecilsulfato de sodio, taurocolato de sodio y Ultramark 1621 en una solución de acetonitrilo/metanol/agua que contenía ácido fórmico al 1 % mediante inyección directa. El control de los instrumentos se realizó mediante el software Thermo Scientific Xcalibur 3.1 v. 4.6.67.17. Se adquirieron escaneos completos a una resolución de 60 000 FWHM (a m/z 200) y escaneos MS2 a 17 500 FWHM en el rango de 100 a 1000 m/z, con 1 microescaneo, tiempo (ms): 200 ms, una lente RF (%) :70; Estándar objetivo AGC. Los eventos de adquisición de escaneo de MS2 (dd-MS2) dependientes de los datos del centroide se realizaron en modo de descubrimiento, activados por la detección de Apex con una detección de activación (%): 75. Los 3 principales precursores abundantes (estado de carga 1 - 2) dentro de una ventana de aislamiento de 2 m/z se consideraron para el análisis MS/MS. La exclusión dinámica se fijó en 2 s. Se permitió una tolerancia de masa de ±10 ppm y el umbral de intensidad del precursor se fijó en 150 × 103. Para la fragmentación del precursor en modo HCD, se utilizó una energía de colisión normalizada del 30 %. La separación cromatográfica se realizó en una columna Waters BEH C18 (50 × 2,1 mm de d.i., 1,7 μm, Waters, Milford, MA, Estados Unidos) utilizando un gradiente lineal de 5–100 % de B durante 13,50 min, seguido de un paso isocrático a 100% B hasta 15.50 min. Las fases móviles fueron: (A) agua con 0,1% (Merck, grado HPLC-MS), ácido fórmico (Merck, grado HPLC-MS) y (B) acetonitrilo con 0,1% de ácido fórmico (Merck, grado HPLC-MS). El caudal se fijó en 500 μL/min, el volumen de inyección fue de 2 μL y la columna se mantuvo a 40 °C.

3.3.8 Procesamiento de datos UHPLC-HMRMS/MS.

Los archivos crudos se transformaron a un formato compatible con MZMine3 para su análisis. Los parámetros utilizados aparecen en Escudero-Leyva *et al*., 2023. De aquí se obtuvo un archivo cuantitativo que fue tratado con un script personalizado en Python para agrupar las variables acordes al Ion Identity Network. El mismo archivo cuantitativo se normalizó por la suma de cada muestra y se aplicó un escalado de Pareto para realizar un análisis de componentes principales (PCA) usando los paquetes R FactoMineR v.2.4 89 y FactoExtra v.1.0.7 90 en R v.4.1.3 88. Luego, se realizó un Análisis Dependiente de Mínimos Cuadrados Parciales (PLS-DA) con el paquete mixOmics v.6.6.2 91 en R v.3.5.1.

Los archivos cuantitativos y espectrales derivados de MZMine se subieron a la plataforma GNPS para generar la red molecular.

3.4 Identificación de polifenoles en plantas de café y especies de Rubiaceae afines.

3.4.1 Recolección de hojas.

Se realizaron análisis metabolómicos dirigidos y no dirigidos con hojas de café comercial (*Coffea arabica*) que fueron colectadas en San José, Costa Rica, en plantaciones de Coopetarrazú. Se recolectaron hojas silvestres de Rubiaceae de las dos especies *Isertia hankeana* y *Simira maxonii* en una finca privada de la región de Golfito (Puntarenas) de Costa Rica. Por cada especie de planta, tres hojas maduras fueron cortadas de cuatro árboles y mantenidas en hielo durante el transporte al laboratorio, una vez ahí, fueron liofilizadas. Para optimizar la extracción y el método LC-MS/MS, se usaron hojas maduras de la especie de café silvestre *Coffea liberica* var. *dewevrei* y *Coffea salvatrix*, ambas plantas cultivadas en condiciones de campo provenientes del Instituto Agronómico de Campinas (São Paulo, Brasil), las cuales siguieron el mismo procedimiento de liofilización.

3.4.2 Extracción de metabolitos secundarios.

La extracción de los metabolitos secundarios de las hojas se realizó después de triturar y pesar 10 mg de material foliar en polvo. Los metabolitos se extrajeron agregando 10 µL de ácido transcinámico-b,2,3,4,5,6-d6 1 mM y 490 µL de metanol al 100 %, seguido de molienda con una perla de metal de 5 mm a 30 Hz durante 5 min. usando un molino mezclador MM400 de Retsch (Haan, Alemania). Luego, los extractos se sonicaron a 35-40 °C durante 20 min y se centrifugaron a 9600 g durante 5 min a temperatura ambiente. Los sobrenadantes se transfirieron a tubos de microcentrífuga de 1,5 ml. Los sedimentos restantes se resuspendieron en 500 µL de metanol/agua (30:70, v/v), se sonicaron durante 20 min a 35-40 °C y se centrifugaron en las mismas condiciones que se mencionaron anteriormente. Los sobrenadantes se combinaron a los primeros y luego se filtraron 500 µL de extractos a través de dispositivos de filtración Amicon de 3 kDa (MilliporeSigma, Burlington, MA) a 14.000 g durante 60 min a temperatura ambiente. Los eluatos resultantes se almacenaron a -20 °C hasta el análisis LC-MS/MS.

3.4.3 Metabolómica no dirigida.

El análisis de los metabolitos se realizó mediante un Sistema de cromatografía líquida de ultra alto rendimiento Exion junto con un espectrómetro de masas de alta resolución TripleTOF6600+ de AB Sciex (Framingham, MA). El espectrómetro de masas se ajustó para escanear metabolitos desde m/z 100-1500 amu en modo negativo o positivo. Para la polaridad negativa, el voltaje de rociado de iones fue de 4500 V, el tiempo de acumulación fue de 100 ms, el potencial de desagrupamiento y la energía de colisión fueron de 60 V y 10 V, respectivamente. Los espectros MS/MS se adquirieron sobre m/z 30-1500 amu con un tiempo de acumulación de 25 mseg. Los parámetros como el potencial de desagrupamiento, la energía de colisión y la dispersión de la energía de colisión se establecieron en 60 V, 45 V y 15 V, respectivamente. Los parámetros para el modo positivo fueron muy similares a los del modo negativo excepto por el voltaje de rociado de iones y el potencial de desagrupamiento que fueron de 5000 V y 35 V, respectivamente. El tiempo de ciclo total fue de 0,65 seg. Los parámetros para la ionización de la fuente de electrospray tales como gas de cortina (nitrógeno), gas de nebulización, gas de calentamiento y la temperatura de la fuente se fijaron en 40 psi, 70 psi, 70 psi y 650 °C, respectivamente. Las condiciones de la fuente fueron las mismas para las polaridades negativa y positiva. Se entregó una solución de calibración positiva o negativa de ionización química a presión atmosférica (APCI) mediante un sistema de entrega de calibrante cada 5 muestras para corregir cualquier desviación de masa que pudiera ocurrir durante la ejecución. Los espectros de MS se adquirieron utilizando el software Analyst TF 1.8.1 (AB Sciex, Framingham, MA). Es importante señalar que la adquisición dependiente de datos (DDA) se ejecutó en modos negativo y positivo, en una mezcla compuesta por extractos CC, WR1 y WR2. Los iones precursores presentes en esta mezcla se usaron luego para la adquisición de ventanas secuenciales de todos los espectros de masas teóricos (SWATH-MS). Los detalles de la adquisición de datos en DDA y el procesamiento están disponibles en la publicación de (Castro-Moretti et al., 2023).

3.4.4 Metabolómica dirigida.

La detección y cuantificación de fitoquímicos fue realizado como se describió anteriormente por (Cocuron et al., 2019) con modificaciones relacionadas con el gradiente

de LC y el uso de monitoreo de reacción múltiple programado (sMRM). Los 184 fitoquímicos considerados en este estudio fueron optimizados uno a uno por infusión directa después de haber sido diluidos a 1 µM con una solución de acetonitrilo/agua (50:50; v/v) que contiene 0,1% de ácido acético como aditivo. El flujo para la infusión directa se ajustó a 10 µl/min y se determinaron parámetros como el potencial de desagrupamiento (DP), la energía de colisión (CE) y el potencial de salida celular (CXP) para los cinco iones de producto más abundantes derivados de cada ion precursor. La lista de los compuestos, así como las particularidades de la adquisición en el equipo aparecen en (Castro-Moretti et al., 2023).

3.4.5 Análisis estadísticos.

El análisis de componentes principales (PCA), el análisis discriminante de mínimos cuadrados parciales (PLS-DA), el mapa de calor y el análisis de componentes simultáneos ANOVA (ASCA), tanto para análisis no dirigidos como dirigidos, se realizaron después de la transformación logarítmica y el escalado automático, utilizando MetaboAnalyst 5.0 (Chong et al., 2018).

CAPÍTULO IV. RESULTADOS

4.1 Aislamiento y caracterización de hongos endófitos.

El espaciador de transcripción interno (ITS) del ADN ribosómico nuclear se acepta como marcador genético o código de barras de elección para la identificación de muestras fúngicas. Aquí, presentamos un protocolo para analizar datos ITS fúngicos, desde el preprocesamiento de calidad de secuencias sin procesar hasta la identificación de unidades taxonómicas operativas (OTU), clasificación taxonómica y asignación de rasgos funcionales. La canalización se basa en recopilaciones de datos bien establecidas y seleccionadas manualmente, a saber, la base de datos UNITE y el script FUNGuild. Como ejemplo, se analizaron datos ITS reales de hongos endófitos cultivables, brindando descripciones detalladas para cada paso, parámetro y análisis posterior, y finalizando con un análisis filogenético de las secuencias y roles ecológicos asignados. Este artículo constituye una guía completa para investigadores que están poco familiarizados con el análisis bioinformático de los pasos esenciales requeridos en estudios ecológicos adicionales de comunidades fúngicas. A partir de los resultados obtenidos con este flujo de trabajo, fue posible catalogar los 1400 aislamientos de hongos endófitos de Rubiaceae silvestre en aproximadamente 800 especies. La colección se encuentra en el CIPRONA, Universidad de Costa Rica.

Publicaciones:

A) Montero-Vargas, M., **Escudero-Leyva, E.**, Díaz-Valerio, S. & P. Chaverri. **2020**. Step-by-step pipeline for the ecological analysis of endophytic fungi using ITS nrDNA Data. Current Protocols in Microbiology. https://doi.org/10.1002/cpmc.96

B) Montero-Vargas, M., Umaña-Jiménez, J.C., **Escudero-Leyva, E.** & P. Chaverri. **2020**. Análisis filogenético de secuencias ITS provenientes de hongos endófitos utilizando Inferencia Bayesiana Paralela de árboles con Exabayes. Revista Tecnología en Marcha, 33(5), Pág. 74-79. https://doi.org/10.18845/tm.v33i5.5079

Step-by-Step Pipeline for the Ecological Analysis of Endophytic Fungi using ITS nrDNA Data

Maripaz Montero-Vargas,[1,5] Efraín Escudero-Leyva,[2,3] Stefani Díaz-Valerio,[1] and Priscila Chaverri[2,4]

[1]Colaboratorio Nacional de Computación Avanzada (CNCA), CeNAT-CONARE, San José, Costa Rica
[2]Escuela de Biología and Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, San José, Costa Rica
[3]Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, San José, Costa Rica
[4]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, Maryland
[5]Corresponding author: *mmontero@cenat.ac.cr*

The nuclear ribosomal DNA internal transcribed spacer (ITS) is accepted as the genetic marker or barcode of choice for the identification of fungal samples. Here, we present a protocol to analyze fungal ITS data, from quality preprocessing of raw sequences to identification of operational taxonomic units (OTUs), taxonomic classification, and assignment of functional traits. The pipeline relies on well-established and manually curated data collections, namely the UNITE database and the FUNGuild script. As an example, real ITS data from culturable endophytic fungi were analyzed, providing detailed descriptions for every step, parameter, and downstream analysis, and finishing with a phylogenetic analysis of the sequences and assigned ecological roles. This article constitutes a comprehensive guide for researchers that have little familiarity with bioinformatic analysis of essential steps required in further ecological studies of fungal communities.

Basic Protocol 1: Raw sequencing data processing
Support Protocol: Building a BLAST database
Basic Protocol 2: Obtaining information from databases
Basic Protocol 3: Phylogenetic analysis

Keywords: database • ecological guilds • fungal communities • fungi • ITS region • phylogenetics

How to cite this article:
Montero-Vargas, M., Escudero-Leyva, E., Díaz-Valerio, S., & Chaverri, P. (2020). Step-by-step pipeline for the ecological analysis of endophytic fungi using ITS nrDNA data. *Current Protocols in Microbiology*, *56,* e96. doi: 10.1002/cpmc.96

INTRODUCTION

This article presents a complete pipeline for analyzing fungal sequence data from the internal transcribed spacers (ITS) of the nuclear ribosomal DNA region (ITS) to guide mycologists and ecology researchers in taking their first steps into bioinformatic data processing (see Fig. 1). The protocols go from the adaptation of sequencing data formats to quality checking, trimming, and data unification, resulting in taxonomic

CURRENT PROTOCOLS
A Wiley Brand
Current Protocols in Microbiology e96, Volume 56
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpmc.96

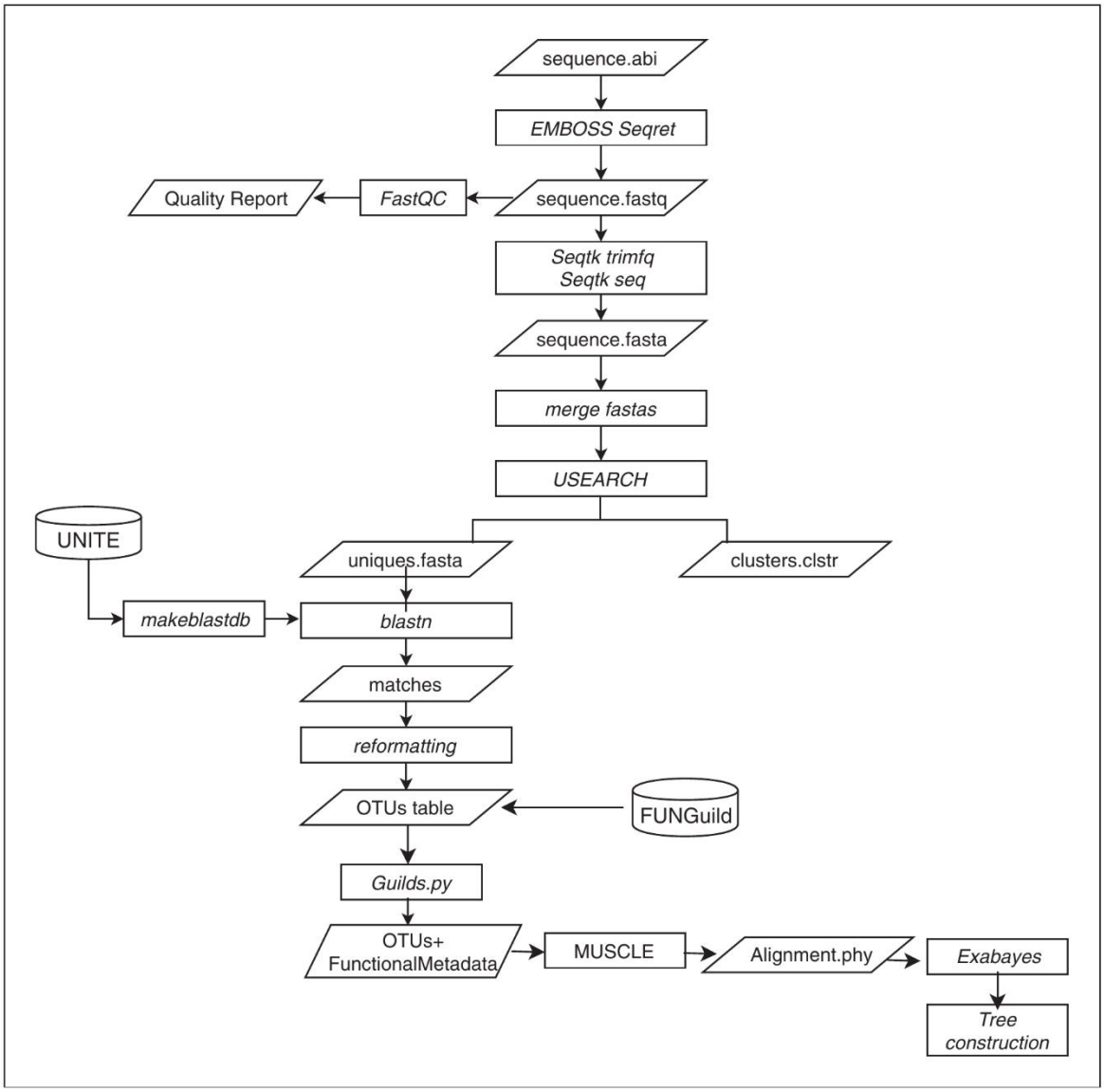


Figure 1 Overview of the pipeline for analysis of ITS data.

classifications based on a curated fungal database: UNITE (Abarenkov et al., 2010). In addition, they provide guidelines for performing Bayesian Inference phylogenetic analysis with a graphic view (Aberer, Kobert, & Stamatakis, 2014) and mapping the ecological roles obtained from FUNGuild into the resulting phylogeny; all are accomplished with bioinformatic tools and several supporting information resources that facilitate the process.

In past decades, employing the fungal nuclear ribosomal operon primers for the large (28S) and the small (18S) subunits was the most used method for the identification of fungi (White, Bruns, Lee, & Taylor, 1990). The species-level identification is mostly done using the ITS region alone, with a high prevalence in fungal taxonomy and systematics studies (Schoch et al., 2012). Systematics studies also use other protein-coding genes (i.e., secondary barcodes); however, they are not as prevalent in databases as ITS data (Schoch et al., 2012). Examples of the regions, including ITS, used in fungal phylogenetic studies have been presented in projects assembling the Fungal Tree of Life (e.g., James et al., 2006; Raja, Miller, Pearce, & Oberlies, 2017).

Even though the NCBI GenBank database is the most complete, curation of the accessions is not regularly done, resulting in many unidentified or misidentified sequences. The UNITE database gives access to developed and high-quality fungal ITS reference sequences that are curated by mycology experts. It also features an annotation management system (PlutoF) that allows users to add pertinent metadata to improve taxonomic designation according to their needs, and has become a powerful tool for mycologists to support the identification of fungal species. Once the taxonomic classification is performed, FUNGuild can be used to parse fungal operational taxonomic units (OTUs) with an ecological guild, independently of sequencing platform or analysis pipeline (Nguyen et al., 2016). Nonetheless, it is important to consider that the ecological guild assignment by FUNGuild is currently based at the genus taxonomic level.

RAW SEQUENCING DATA PROCESSING

BASIC PROTOCOL 1

This protocol describes the preliminary steps before submitting the data to a taxonomic analysis. High-throughput procedures may generate different output sequence formats according to the technology used. For example, ABI machines create `.abi` sequence files. FASTQ is the most common format because it integrates both sequences and associated per-base quality score (Cock et al., 2010). Emboss is a robust and specialized suite of tools for molecular biology, including Seqret, which supports several sequence formats and allows conversion between file formats, e.g., from "ABI" to "FASTQ" files (Rice, Longden, & Bleasby, 2000). FAST QC is an easy and interactive tool to check the quality of a given sequence, providing a summary of quality by position by means of dynamic plots supporting FASTQ files generated by several platforms (Brown, Pirrung, & McCue, 2017). Errors may happen during sequencing; therefore, it is necessary to submit the resulting data to quality control to perform the trimming of bad-quality bases to improve the results of downstream analysis. It is common to find bad-quality regions at both ends of a sequence; they are often related to the sequencing technology used. Seqtk (*https://github.com/lh3/seqtk*) is a fast tool for processing FASTA/FASTQ files that implements the Phred algorithm for quality trimming of sequences. Here, two functions of Seqtk are applied: *trimfq* and *seq*. First, *trimfq* removes the low-quality bases from both ends. Second, *seq* is used to convert the file from FASTQ to FASTA format. Finally, this protocol provides a way of clustering the sequences according to their similarity using USEARCH to generate non-redundant databases and to facilitate the downstream process, avoiding unnecessary use of time and computational efforts (Chen, Wan, Lei, Zobel, & Verspoor, 2016; Fu, Niu, Zhu, Wu, & Li, 2012).

Materials

Hardware

Computer or computational cluster with access to command line environment

Software

ITS sequencing data provided by the sequencing facility; files may have `.abi` or `.fastq` extension
EMBOSS-seqret
FastQC
Seqtk
Usearch

1. Sequence-files format conversion:

   a. Put the raw reads into a new directory.
   b. Write the next command on the terminal inside the directory containing the raw reads.

```
>seqret -sequence file.abi -outseq file.fastq
  -sformat1 abi -osformat2 fastq-sanger
```

Here, the `-sequence` and `-outseq` options correspond to the input and output sequence files, respectively; `-sformat1` is the entry format and `-osformat2` is the desired output format. Detailed documentation on supported formats and further examples are provided at the Emboss web page.

c. Move the recently created `.fastq` files to a new directory. Write the next command in the terminal.

```
>mv *.fastq ../Fastq
```

Once the FASTQ sequence files are created, move them to the `Fastq` directory. This command is specifically for moving multiple sequences that end with the key `.fastq`.

2. Sequence quality control:

a. Execute the tool FASTQC with this command on the directory containing the `.fastq` files.

```
>fastqc ~/Fastq/file.fastq -o FastQC
```

FastQC will run a set of analyses on each sequence contained in a fastq file and create an HTML report summarizing the results; this new output will be located in the `FastQC` directory and can be visualized in a web browser. A green or red check mark accompanying a metric indicates pass or fail conditions; yellow checks correspond to warnings and must be evaluated in each case. When a sequence presents bad quality at the ends, a trimming step is required prior to further analysis (Fig. 2A).

b. Compile all FastQC reports in a single report to visualize and compare the quality of all of your samples.

```
>multiqc ~/FastQC/-o ~/FastQC/multiqc
```

MultiQC uses the log quality files of a given directory as input and specifies an output directory with the flag `-o`, in which an HTML file will be held. The results can be visualized by opening this file (2B).

3. Data refinement:

a. Delete bad-quality ends and produce a FASTA file using Seqtk. Write the next command in the directory containing the fastq files.

```
>seqtk trimfq -q 0.01 file.fastq | seqtk seq -A -L
  500 > file.fasta
```

The function trimfq uses the Phred algorithm to trim both ends of the sequences from the fastq file. If the error rate threshold (`-q`) is not specified, it takes 0.05 as default. Seqtk seq then takes the result of trimfq and uses it as input, so that instead of writing the sequence clipped in fastq format, it is done in FASTA. The option `-L` removes the sequences shorter than a specific length (in this case, 500 bp) and produces an output file(>) called `file.FASTA`.

b. Deletion of duplicated sequences using USEARCH.

```
>usearch -fastx_uniques input.fasta -FASTAout
  uniques.fasta
```

USEARCH uses its dependency `-fastx_uniques` to identify repeated sequences contained in Fasta files. It outputs a file with all the unique sequences concatenated in Fasta format (`-FASTAout`). This program also outputs another file that lists which sequences were duplicated and removed.

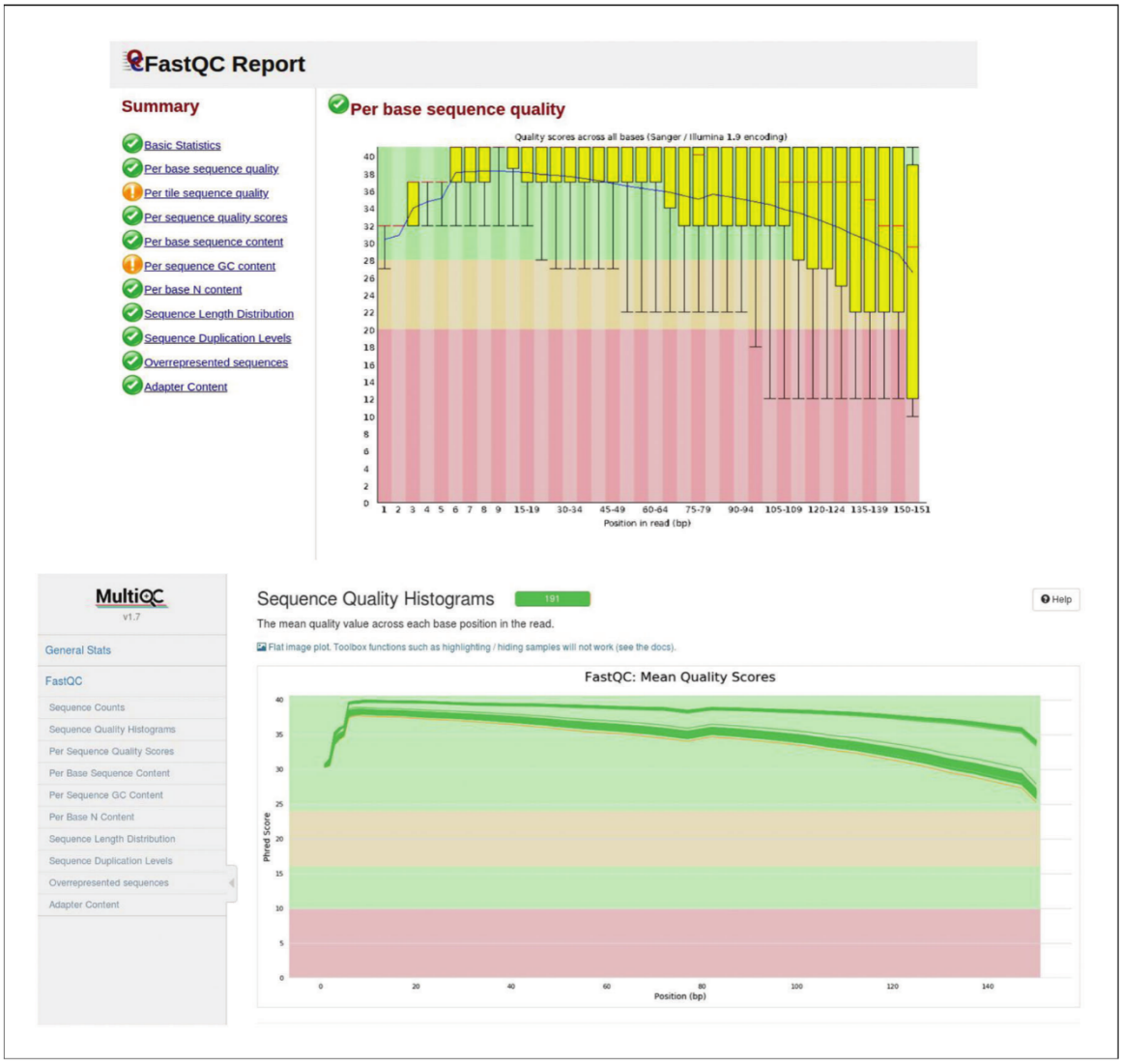


Figure 2 Quality analysis of fastq sequences. (**A**) FastQC quality report for endophytic fungi. (**B**) This graphic shows the quality distribution along all the sequences using MultiQC.

BUILDING A BLAST DATABASE

SUPPORT PROTOCOL

In order to use the tool BLAST with an external database such as UNITE, it must be first installed and configured for BLAST to recognize it. In the case of the UNITE database, the reference files are obtained from the official web page. For this protocol, we used the 8th version, released in November 2018.

Materials

Hardware

Computer or computational cluster with access to command line environment

Software

UNITE (or any other database) reference FASTA file
BLAST

1. Integrate UNITE database to BLAST.

   a. Write the next command in the folder where you place the database reference FASTA file

```
$ makeblastdb -in test.fna -dbtype 'nucl' -hash_
  index
```

This procedure will take as input the Fasta reference from the database (-in) and give, as an output, eight files with the extensions: `.fna.nsq`, `.fna.nsi`, `.fna.nsd`, `.fna.nog`, `.fna.nin`, `.fna.nhr`, `.fna.nhi`, `.fna.nhd`. It is necessary to specify that this will be a nucleotide database (`-dbtype 'nucl'`), since it could be also a protein reference. Here it is also specified that BLAST has to create an index of sequence hash values (`-hash_index`).

BASIC PROTOCOL 2

OBTAINING INFORMATION FROM DATABASES

Programs performing BLAST algorithms compare DNA or protein queries in databases, and have become a crucial step for identifying taxonomic affiliation (Altschul et al., 1997). In this pipeline, we have evaluated the UNITE database (Abarenkov et al., 2010), as it offers highly curated fungal sequences (Nilsson, Abarenkov, Larsson, & Kõljalg, 2011). On the other hand, FUNGuild is a tool that allows the assignment of ecological categories to the organization of the fungal community sequence data, making possible the classification of the different taxa (Nguyen et al., 2016). FUNGuild works with a two-component system, including a database annotated by members of the community and a Python script that assigns functional information to OTUs. It uses the taxonomic strings in the user's table to search against a Guild database to generate an OTU table with functional metadata assigned to them.

Materials

Hardware

Computer or computational cluster with access to command line environment and internet connection

Software

Python package "request" installed

1. BLAST against UNITE:

   a. Write the next command in the folder where you place the database reference FASTA file.

```
>blastn -query input.FASTA -task blastn -db uniteDB
  -out blastout.txt -evalue 10 -outfmt '6 qseqid
  pident sallseqid ' -max_target_seqs 1 -num_threads
  16
```

The nucleotide BLAST uses as input a FASTA file (`-query`) and performs a specific task depending on if the data are nucleotides or proteins (`-task`). Here, we ask to use UNITE as a reference (`-db`) to create an output file (`-out`) called `blastout.txt` and to use an e-value (`-e`) of 10. Finally, we specify an output format (`-outfmt`) to adjust our needs in the former steps of the analysis.

   b. Adjust headers in OTUs table for FUNGuild

```
>echo 'otu_id,sample,taxonomy'> funguild.csv
```

The resulting table obtained from running BLAST must have a header in the first line; any other row of comments must be removed. The header includes a column named "taxonomy," which contains information from a reference database such as UNITE. In Figure 2, the output OTU table is shown as an Excel file.

2. Ecological guild assignment with FUNGuild:

   a. Run FUNGuild with the next command:

```
>python/home/bio_BD/Guilds_v1.1.py -otu./funguild.
  csv -db fungi -m -u
```

The database (`-db`) must be specified between nematode or fungi. Fungi is the default parameter. This command outputs a text file with the respective ecological assignment for the sequences.

PHYLOGENETIC ANALYSIS

BASIC PROTOCOL 3

Once functional metadata has been assigned, a cladogram can be constructed to infer phylogenetic and functional relationships. To accomplish this, a multiple sequence alignment is necessary to provide a measure of the homology between sequences of similar length and to create a model of nucleotide substitution (Edgar, 2004). Once the sequences are aligned, the next step will be using the tool Exabayes to perform a Bayesian Inference phylogenetic analysis using the Markov chain Monte Carlo sampling approach to determine the best evolutionary parameters and construct a phylogram (Aberer et al., 2014). Initially, Exabayes requires two mandatory files: the alignment file from the last step and a configuration file with specifications for different parameters, which allow the user to customize their run. In this case, a tree without partitions is made, since we are using only ITS sequences. As a last step, we use a platform for graphical tree display and visualization. Here, we employ the easy and user-friendly tool, Interactive Tree of Life (ITOL).

Materials

Hardware

Computer or computational cluster with access to command line environment and internet access

Software

Multi-FASTA file with all the sequences
Exabayes
Google account with access to Google spreadsheets

1. Alignment with MUSCLE:

   a. Write the next command in the folder where you place the multi-FASTA file.

   ```
   >muscle -in multiFASTA.fasta -phyiout alignment.phy
   ```

   MUSCLE takes as input the multi-sequence FASTA file (multiFASTA) (`-in`) and can output a series of different formats. In this case, Phylip interleaved format is used to facilitate the downstream analysis (`-phyiout`). Finally, the name of the final alignment is given (`alignment.phy`).

2. Phylogenetic analysis with Exabayes:

   a. Create a configuration file for Exabayes with the next structure. This file must be named with the format `.nex` (i.e., `file.nex`).

   ```
   #NEXUS
   begin run;
   numgen 3e6
   parsimonyStart true
   end;
   ```

   The structure of this file is declared by the line "`begin run;`" followed by all the parameters, which must be indented. The nexus file end is declared by the word "`end`." If any parameter is not specified, the program applies the default values for each parameter.

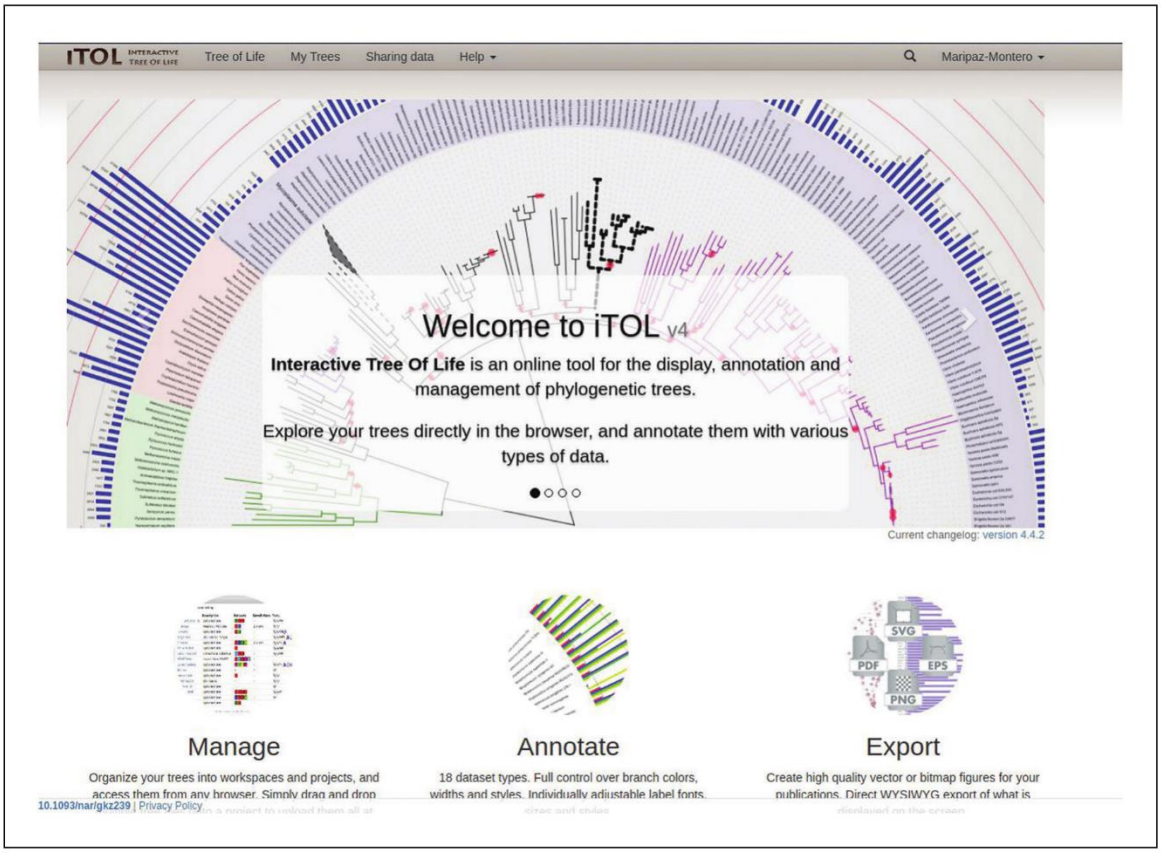


Figure 3 Interactive Tree of Life (ITOL) web platform.

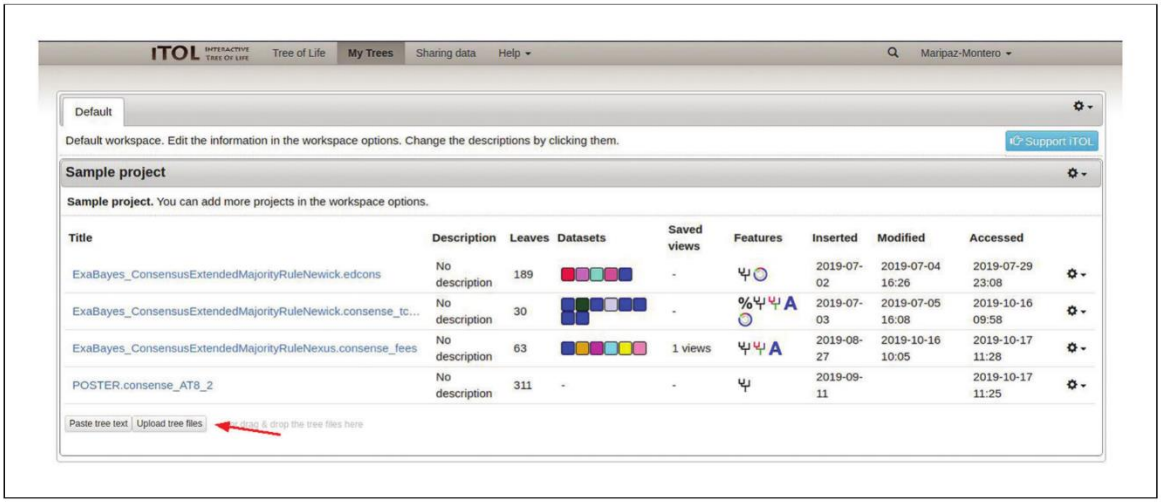


Figure 4 Uploading the Nexus file to ITOL web platform.

For this example, we used the following parameters (the complete list can be found in the Exabayes manual). First, the number of generations (`numgen`) indicates for how many generations Exabayes will run the test. This parameter is important for getting a statistical significance for the resulting tree. The second parameter, `parsimonyStart`, indicates whether or not Exabayes will start sampling from a random topology; if so, this will be set to false.

b. Run Exabayes. Write the next command to execute Exabayes.

```
mpiexec exabayes -f alignment.phy -m DNA -n run_id
  -s 123 -c config.nex
```

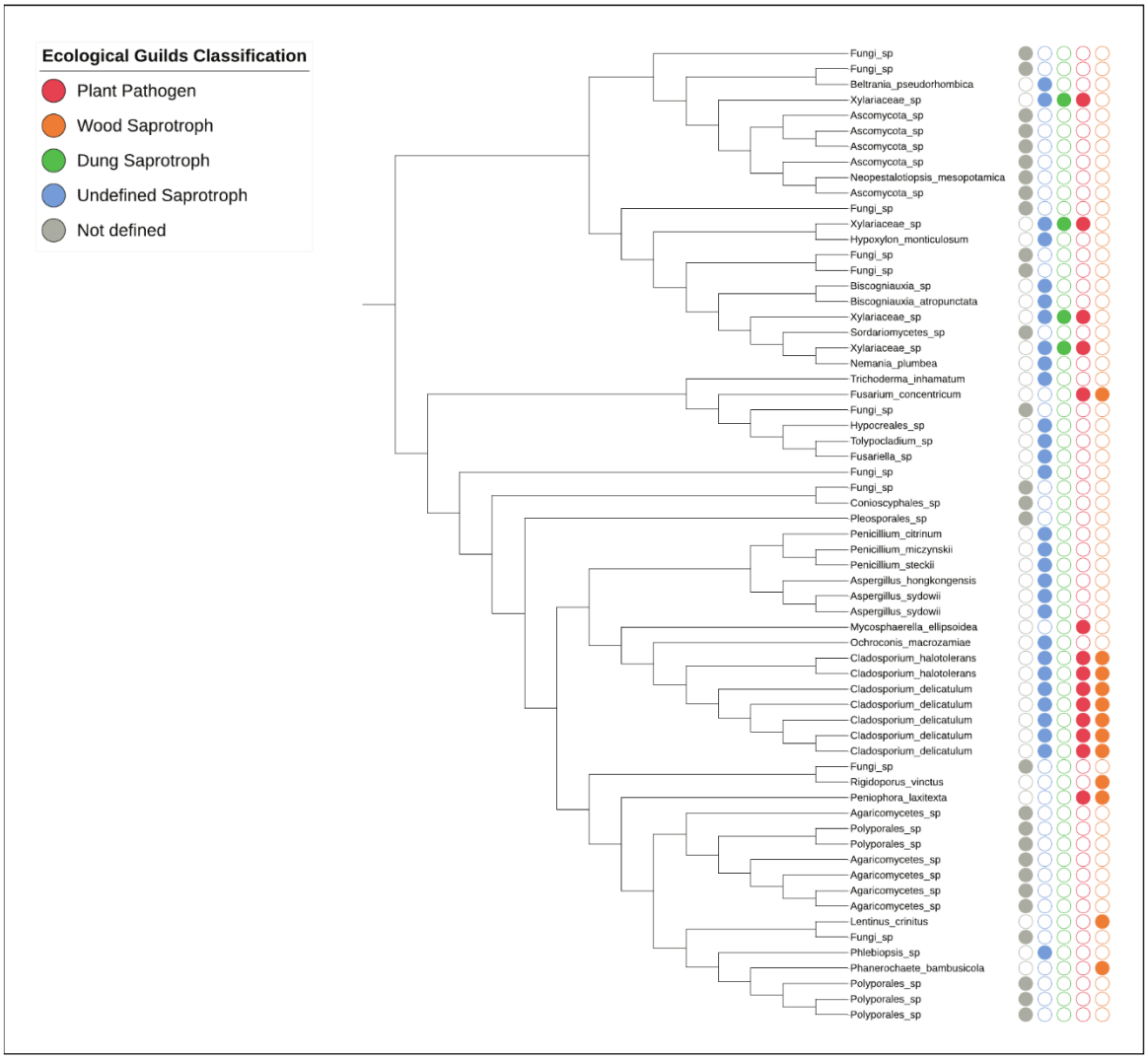


Figure 5 Phylogenetic representation of the endophytic fungi. Each taxon has its respective ecological guilds represented as filled circles.

For this command, the alignment file is specified via `-f`*. Then, the model (*`-m`*) is established depending if the nature of the data is DNA sequences or amino acid sequences. The* `-n` *flag determines the name of the run, and* `-s` *provides a seed so the run can be repeated. Finally, the configuration file is passed via* `-c`*. This will output four types of files:* `ExaBayes_info.myRun`*,* `ExaBayes_topologies.myRun.0`*,* `ExaBayes_parameters.myRun.0`*,* `ExaBayes_diagnostics.myRun`*.*

c. Write the next command to create the consensus tree.

```
mpiexec consense -f ExaBayes_topologies.run* -n
  consense_file
```

This command will sample the topology files (`ExaBayes_topologies.myRun.0`*) generated by the previous step to produce a consensus tree. As a result, this step will create the final tree in two different formats: Newick and Nexus. Here, the flag* `-f` *indicates the input topology files and* `-n` *the name of the output.*

d. Write the next command to generate an overview of the parameters sampled.

```
postProcParam -f ExaBayes_parameters.run* -n Post_
  proc_file
```

This command creates a summary file (`ExaBayes_parameterStatistics`) of the statistics for each parameter. As in the last two steps, `-f` indicates the input file, which in this case will be `ExaBayes_parameters.myRun.0`, generated before, and `-n` is the name of the output file.

3. Create a visual representation of the phylogenetic tree

   There are many available tools for the creation of a graphical representation of the Nexus or Newick files. In this case, the ITOL web platform is used. This tool allows the creation of multiple layers of information, in addition to the taxonomy classification.

   a. Go to ITOL web page. Create an account and subscribe (Fig. 3).
   b. Upload your tree. ITOL accepts either Nexus or Newick formats (Fig. 4).
   c. Modify the settings of the display and add a dataset for taxonomy. In addition, it is possible to display the FUNGuild ecological classification in multiple ways. For this example, a multiple-layer (one layer per ecological guild) binary dataset is generated.
   d. Export and save your tree (Fig. 5).

COMMENTARY

Background Information

A fundamental step in any study where fungi are involved is taxonomic classification, and, certainly, fungal taxonomy is complex. Species were initially defined based on the morphology of asexual and sexual reproductive structures; unfortunately, the number of characters is limited (Grünig, Queloz, & Sieber, 2011). Currently, the easiest and most reliable way to identify fungi is by using molecular tools, which provide rapid and abundant information not only about taxonomy, but also about their biology and ecology (Molina, Horton, Trappe, & Marcot, 2011). Since the recognition of the ITS regions as the primary barcode for fungi (Schoch et al., 2012), a vast amount of sequence data has been generated over recent years. However, finding new species without concordances in databases is frequent, and traditional morphological analyses and descriptions are still needed (Coronado-Ruiz et al., 2018).

In addition, the number of studies involving endophytic fungi through a culturable or metagenomic approach is rapidly increasing, since endophytic fungi have been found in almost all vascular plant species examined to date and are considered important components of fungal biodiversity (Rodríguez, Elissetche, & Valenzuela, 2011). Fungal endophytes can have profound effects on their host's health, biochemistry, and physiology, influencing multitrophic networks and entire ecosystems (Unterseher, 2011); also, they are a very promising source of novel bioactive compounds (Larsen, Smedsgaard, Nielsen, Hansen, & Frisvad, 2005). A useful resource that helps mycologists and ecologists achieve rapid identification of fungi using molecular techniques like ITS sequences with accuracy and highly curated information is the UNITE database (Abarenkov et al., 2010). This database contained 887,395 ITS sequences (as of November 2018). Besides adding new fungal sequences each year, it is possible to associate any number of binary files with well-annotated reference sequences, which go through a strict process prior to inclusion (Nilsson et al., 2011).

To understand the function of an ecosystem or a community, the functional ecology classification to a specific guild is necessary. FUNGuild works with the application of guild assignments to OTUs at the genus level due to the limited ecological lifestyle known for all the fungal species (Nguyen et al., 2016). To understand what FUNGuild does, it is important to first assess the definition of "Guild," a term originally proposed by Root in 1960. "Guild" is defined as a group of species that exploit the same class of environmental resources in a similar way (Simberloff & Dayan, 1991). Those environmental resources can be seen as food, shelter, or habitat, and guild members can be taxonomically distinct (Kornan & Kropil, 2014). For instance, guild categories could be indexed specifically for the study purposes; for example, Höbinger, Schindler, Seaman, Wrbka, and Weissenhofer (2012) defined three categories of guilds for the plant species present in different forest landscapes modified by oil palm plantations. For fungi, we can understand "habitat" as the resources and conditions present in an area that produce occupancy. Therefore, the fungal habitat is not static but dynamic in space and time, especially in growth substrata where nutrient pools can quickly vary, and also shaped by the plant's

genotype (Bonito et al., 2016; Molina et al., 2011). Fungal guilds and their assignments are very important for a better and more complete understanding of ecosystem function. For example, fungal communities are based on their interactions with plants, and may be commensal, mutualistic, or pathogenic, profoundly influencing plants in regard to their growth and general health (Seung-Yoon et al., 2018).

Recommendations for fungal identification have been previously proposed, suggesting the use of barcodes, databases, and phylogenetic approaches (Raja et al., 2017), but detailed procedures are rare or nonexistent. We consider that a complete explanation about methodologies will be useful for new and experienced mycologists. Commonly, fungal identification goes through several steps and is run in various computer packages. We present a pipeline to accelerate and simplify this process using an assembly of informatics tools to process all the data provided by molecular methods going through taxonomic and functional assignments.

To successfully complete the protocols described here, the user needs to confirm that the software listed in the "Necessary Resources" has been already downloaded and installed. Internet Resources includes the download URL for each resource, all of which are freely available.

Critical Parameters

For those researchers managing larger datasets (e.g., metabarcoding), we suggest applying digital normalization to the data, especially during quality control, since it is a critical step for optimization of the protocol. In addition, an adjustment of the number of threads used by the programs according to the capacity of the high-performance computing cluster used is recommended.

Troubleshooting

All the programs used come with a manual to respond to any unexpected errors in coding and management of the results. We suggest following the command line in parallel with user guides for a better understanding of the entire process.

Understanding Results

The procedure described in this article is a process where every step depends on the previous results. While performing the basic processing of data, the user must be attentive to implementing the correct trimming of the sequences once the quality results are ready (Fig. 2). When performing data refinement, the FASTA file produced as output is usually smaller than the initial input, since the bad-quality reads have been removed.

Continuing with the BLAST procedure, once the user does the BLAST against UNITE, checking that the output file contents have the right information is essential. BLAST will output a text file that may be opened with any text editor. This file must have different columns with the identification code of the sequences, the percent identity, and the taxonomic classification. For the next step, FUNGuild must output a similar file that can be checked in the last case. The file must contain columns with the OTU identifications, the sample number (which must be 1, if not specified otherwise), the taxonomy identification, the trophic mode, the guild classification, the growth morphology, the trait confidence ranking, and the citation or source.

After executing the MUSCLE alignment, the program must have created a new file in Phylip format. It is important to check that this file is correctly arranged. The first line describes the dimensions of the alignment; therefore, it must have two numbers separated by a tab. The first number represents the number of samples or sequences, and the second one the number of base pairs of each sequence. Next will be the multiple sequence alignment. This section consists of multiple lines for each sequence, and each one must have an identifier of exactly 10 characters. Including spaces in the name of the sequences could lead to errors, so they must be substituted with other symbols such as underscores or dashes (Fig. 6).

Next, the phylogenetic analysis with Exabayes will output four kinds of files: one with information on the run, including the chosen parameters and the progress of the run (`ExaBayes_info.name`); another one containing the topologies in nexus format (`ExaBayes_topologies.name.0`); one for the values sampled for non-topological parameters (`ExaBayes_parameters.name.0`); and, finally, a file with the diagnostics of acceptance ratios for chain parameters (`ExaBayes_diagnostics.myRun`). This last file is important to keep track of the process and to check the average standard deviation of split frequencies (abbreviated to ASDSF). This parameter is an important criterion for knowing if the iteration over the sampled topologies has been done properly; a value between 0.5% and 5% is considered acceptable.

For the creation of the consensus tree, Exabayes will output two types of topology files: Newick and Nexus. These files can be

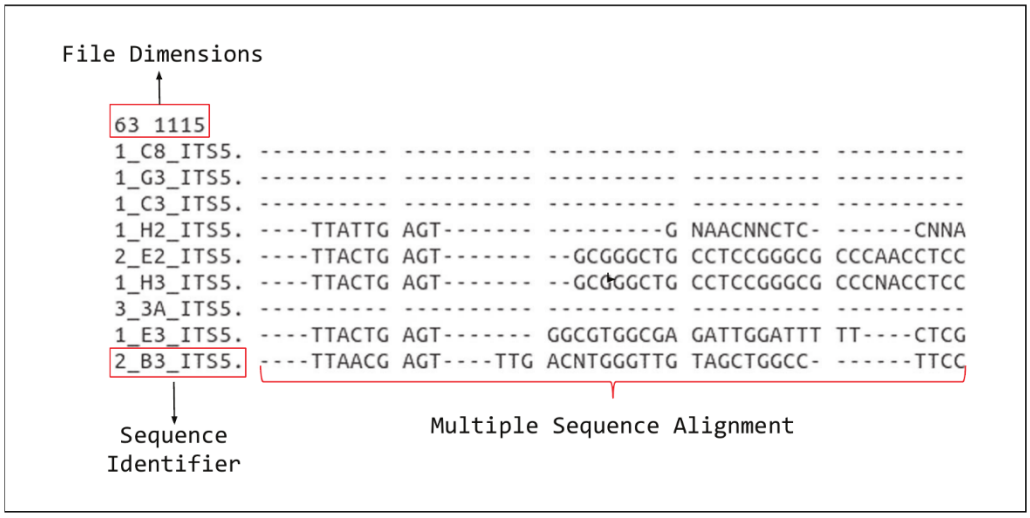


Figure 6 Example for Phylip file structure.

visualized in a variety of platforms such as FigTree, Tree Viewer, and EvolView, among others. In this case, ITOL was used because it allows the user to create multiple layers of annotations for the trees, using multiple formats. In Figure 5, the tree topology is completed with five binary annotation layers—one for each ecological guild—so that, if the taxa fulfill an ecological function, it is represented as a filled circle.

The final step for phylogenetic analysis with Exabayes creates a file that could be visualized in any spreadsheet tool and contains different parameters sampled during the process of topology construction (`ExaBayes_parameterStatistics.params`). The user must pay special attention to the effective sampling size (ESS) of each parameter, because that can tell how well a specific parameter was sampled during the tree reconstruction, and if it is adequate for the number of generations used in the analysis, since a good sampling ESS must be above 100.

Time Considerations

This pipeline, applied to samples with similar sizes to the one presented in this article, may take about a week for completion. Going through the basic processing of the data until the alignment of the samples from the databases may be completed in a couple of days, but the most time-consuming part of this pipeline is the phylogenetic analysis with Exabayes. This tool can be implemented in parallel using servers that allow it, so that the run time decreases substantially.

Acknowledgments

M.M.-V. thanks CeNAT-CONARE, since this research was supported by a machine allocation on the Kabré supercomputer at the Costa Rica National High Technology Center. Funding for this study was provided by *CeNAT-CONARE Scolarships* (2017-2019), University of Costa Rica project B7176, FEES-CONARE project 809-B9-662, and NSF grant DEB-1638976.

Literature Cited

Abarenkov, K., Nilsson, R. H., Larsson, K. H., Alexander, I. J., Eberhardt, U., Erland, S., ... Kõljalg, U. (2010). The UNITE database for molecular identification of fungi–Recent updates and future perspectives. *New Phytologist*, *186*(2), 281–285. doi: 10.1111/j.1469-8137.2009.03160.x.

Aberer, A. J., Kobert, K., & Stamatakis, A. (2014). Exabayes: Massively parallel bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, *31*(10), 2553–2556. doi: 10.1093/molbev/msu236.

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. doi: 10.1093/nar/25.17.3389.

Bonito, G., Hameed, K., Ventura, R., Krishnan, J., Schadt, C. W., & Vilgalys, R. (2016). Isolating a functionally relevant guild of fungi from the root microbiome of Populus. *Fungal Ecology*, *22*, 35–42. doi: 10.1016/j.funeco.2016.04.007.

Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, *33*(19), 3137–3139. doi: 10.1093/bioinformatics/btx373.

Chen, Q., Wan, Y., Lei, Y., Zobel, J., & Verspoor, K. (2016). Evaluation of CD-HIT for constructing non-redundant databases. *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 703–706. doi: 10.1109/BIBM.2016.7822604.

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *38*(6), 1767–1771. doi: 10.1093/nar/gkp1137.

Coronado-Ruiz, C., Avendaño, R., Escudero-Leyva, E., Conejo-Barboza, G., Chaverri, P., & Chavarría, M. (2018). Two new cellulolytic fungal species isolated from a 19th-century art collection. *Scientific Reports*, *8*(1), 7492. doi: 10.1038/s41598-018-24934-7.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*(1), 113. doi: 10.1186/1471-2105-5-113.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. doi: 10.1093/bioinformatics/bts565.

Grünig, C. R., Queloz, V., & Sieber, T. N. (2011). Structure of diversity in dark septate endophytes: From species to genes. In A. M. Pirttilä & A. C. Frank (Eds.), *Endophytes of forest trees* (Vol. *80*, pp. 3–30). Dordrecht, Netherlands: Springer. doi: 10.1007/978-94-007-1599-8_1.

Höbinger, T., Schindler, S., Seaman, B. S., Wrbka, T., & Weissenhofer, A. (2012). Impact of oil palm plantations on the structure of the agroforestry mosaic of La Gamba, southern Costa Rica: Potential implications for biodiversity. *Agroforestry Systems*, *85*(3), 367–381. doi: 10.1007/s10457-011-9425-0.

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., ... Vilgalys, R. (2006). Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature*, *443*(7113), 818–822. doi: 10.1038/nature05110.

Kepler, R., Ban, S., Nakagiri, A., Bischoff, J., Hywel-Jones, N., Owensby, C. A., & Spatafora, J. W. (2013). The phylogenetic placement of hypocrealean insect pathogens in the genus Polycephalomyces: An application of One Fungus One Name. *Fungal Biology*, *117*(9), 611–622. doi: 10.1016/j.funbio.2013.06.002.

Korňan, M., & Kropil, R. (2014). What are ecological guilds? Dilemma of guild concepts. *Russian Journal of Ecology*, *45*(5), 445–447. doi: 10.1134/S1067413614050178.

Larsen, T. O., Smedsgaard, J., Nielsen, K. F., Hansen, M. E., & Frisvad, J. C. (2005). Phenotypic taxonomy and metabolite profiling in microbial drug discovery. *Natural Product Reports*, *22*(6), 672–695. doi: 10.1039/b404943h.

Molina, R., Horton, T. R., Trappe, J. M., & Marcot, B. G. (2011). Addressing uncertainty: How to conserve and manage rare or little-known fungi. *Fungal Ecology*, *4*(2), 134–146. doi: 10.1016/j.funeco.2010.06.003.

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., ... Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248. doi: 10.1016/j.funeco.2015.06.006.

Nilsson, R. H., Abarenkov, K., Larsson, K.-H., & Kõljalg, U. (2011). Molecular identification of fungi: Rationale, philosophical concerns, and the unite database. *The Open Applied Informatics Journal*, *5*(1), 81–86. doi: 10.2174/1874136301105010081.

Raja, H. A., Miller, A. N., Pearce, C. J., & Oberlies, N. H. (2017). Fungal identification using molecular tools: A primer for the natural products research community. *Journal of Natural Products*, *80*(3), 756–770. doi: 10.1021/acs.jnatprod.6b01085.

Rice, P., Longden, I., & Bleasby, A. (2000). Emboss: The European molecular biology open software suite. *Trends in Genetics*, *16*(6), 276–277. doi: 10.1016/S0168-9525(00)02024-2.

Rodríguez, J., Elissetche, J. P., & Valenzuela, S. (2011). Tree endophytes and wood biodegradation. In A. M. Pirttilä & A. C. Frank (Eds.), *Endophytes of forest trees* (Vol. *80*, pp. 81–93). Dordrecht, Netherlands: Springer. doi: 10.1007/978-94-007-1599-8_5.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., André Levesque, C., ... Fungal Barcoding Consortium. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Procedures of the National Academy of Sciences*, *109*(16), 6241–6246. doi: 10.1073/pnas.1117018109.

Seung-Yoon, O., Hae, J. H., Eimes, J. A., Sang-Kuk, H., Chang, S. K., & Young, W. L. (2018). Guild patterns of Basidiomycetes community associated With *Quercus mongolica* in Mt. Jeombong, Republic of Korea. *Mycobiology*, *46*(1), 13–23. doi: 10.1080/12298093.2018.1454009.

Simberloff, D., & Dayan, T. (1991). The guild concept and the structure of ecological communities. *Annual Review of Ecology and Systematics*, *22*(1), 115–143. doi: 10.1146/annurev.es.22.110191.000555.

Unterseher, M. (2011). Diversity of fungal endophytes in temperate forest trees. In A. M. Pirttilä & A. C. Frank (Eds.), *Endophytes of forest trees* (Vol. *80*, pp. 31–46). Dordrecht, Netherlands: Springer. doi: 10.1007/978-94-007-1599-8_2.

White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. L. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications*, *18*(1), 315–322.

Internet Resources

https://unite.ut.ee/
UNITE database (v7.2).

https://github.com/UMNFuN/FUNGuild
FUNGuild.

ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
BLAST+ (v2.7.1).

https://www.drive5.com/usearch/manual/
USEARCH (v.11.6).

http://emboss.sourceforge.net/download/
Emboss (v6.6.0).

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FastQC (v0.11.5).

https://www.ebi.ac.uk/seqdb/confluence/display/JDSAT/MUSCLE+Help+and+Documentation
MUSCLE(3.8.31).

https://github.com/shenwei356/seqkit
SeqKit (v0.7.1).

https://github.com/lh3/seqtk
Seqtk.

https://www.python.org/downloads/
Python.

https://itol.embl.de
ITOL web platform.

Phylogenetic analysis of ITS data from Endophytic fungi using Massive Parallel Bayesian Tree Inference with Exabayes

Análisis Filogenético de Secuencias ITS Provenientes de Hongos Endófitos Utilizando Inferencia Bayesiana Paralela de Árboles con Exabayes

Maripaz Montero-Vargas[1], Jean Carlo Umaña-Jiménez[2], Efraín Escudero-Leiva[3] , Priscila Chaverri-Echandi[4]

Montero-Vargas, M; Umaña-Jiménez, J; Escudero-Leiva, E; Chaverri-Echandi, P. Phylogenetic analysis of ITS data from Endophytic fungi using Massive Parallel Bayesian Tree Inference with Exabayes. *Tecnología en Marcha*. Edición especial 2020. 6th Latin America High Performance Computing Conference (CARLA). Pág 74-79.

https://doi.org/10.18845/tm.v33i5.5079

1 Biologist. Centro Nacional de Computación Avanzada (CNCA), CeNAT-CONARE, Costa Rica. Email: mmontero@cenat.ac.cr .
https://orcid.org/0000-0002-6562-4231
2 Computing Engineer. Centro Nacional de Computación Avanzada (CNCA), CeNAT-CONARE, Costa Rica. Email: jumana@cenat.ac.cr .
https://orcid.org/0000-0003-0857-6007
3 Biologist. Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, Costa Rica & Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, Costa Rica.
https://orcid.org/0000-0003-4440-4296
4 Associate Professor. Universidad de Costa Rica, Escuela de Biología & Centro de Investigaciones en Productos Naturales (CIPRONA), Costa Rica.
https://orcid.org/0000-0002-8486-6033

Keywords

Fungi; ITS; Exabayes; Phylogenetics; Parallelization; Biodiversity.

Abstract

Ecological studies of fungal communities have been favored thanks to the emergence and improvement of independent culture techniques that use the ITS region as a molecular marker. This has allowed a more accurate identification compared to traditional culture-dependent methods.

Next-generation sequencing techniques have increased the amount of data available for the understanding of endophytic fungal communities. An important part of this process is the phylogenetic inference to decipher how the different taxa are related and interact, however, this may become one of the bioinformatic analysis that demands more time.

In response to this, the bioinformatics along with high-performance computing offer solutions to accelerate and make more efficient the tools available for data processing through the implementation of supercomputers and the parallelization of tools

In this study we carried out the processing of ITS sequences to then use the parallelization of Exabayes, software specialized in the analysis and creation of phylogenetic trees.

Thanks to the use of this technique, it was possible to reduce the running time of Exabayes from more than 400 hours to 6 hours, which demonstrates the benefits of the use of high-performance computing platforms.

Palabras clave

Hongos; ITS; Exabayes; Filogenética; Paralelización; Biodiversidad.

Resumen

Los estudios ecológicos de las comunidades fúngicas se han visto favorecidos gracias a la aparición y mejora de técnicas independientes de cultivo que utilizan la región ITS como marcador molecular. Esto ha permitido una identificación más precisa en comparación con los métodos tradicionales dependientes de la cultura.

Las técnicas de secuenciación de próxima generación han aumentado la cantidad de datos disponibles para la comprensión de las comunidades de hongos endofíticos. Una parte importante de este proceso es la inferencia filogenética para descifrar cómo se relacionan e interactúan los diferentes taxones, sin embargo, este puede convertirse en uno de los análisis bioinformáticos que exige más tiempo.

En respuesta a esto, la bioinformática junto con la informática de alto rendimiento ofrecen soluciones para acelerar y hacer más eficientes las herramientas disponibles para el procesamiento de datos a través de la implementación de supercomputadoras y la paralelización de herramientas.

En este estudio llevamos a cabo el procesamiento de secuencias ITS para luego utilizar la paralelización de Exabayes, software especializado en el análisis y creación de árboles filogenéticos.

Gracias al uso de esta técnica, fue posible reducir el tiempo de ejecución de Exabayes de más de 400 horas a 6 horas, lo que demuestra los beneficios del uso de plataformas informáticas de alto rendimiento.

Introduction

Endophyte fungi have been found in almost all vascular plant species examined to date and they are considered important components of fungal biodiversity [1]. They can have profound effects on their host physiology, influence multitrophic networks and entire ecosystems [2]. Fungi and particularly endophytes are a very promising source of novel biologically active compounds [3] so studying them may help to implement new techniques for new biotechnological applications.

An essential part of ecological studies of endophytic fungi is the taxonomic classification and phylogenetic inference [4], [5]. Fungi taxonomy is certainly complex, even now, what was believed to be one species can be in fact be an assemblage of productively isolated lineages. In mycology, species were defined based on the morphology of asexual and sexual reproductive structures, unfortunately, the number of characters is limited generating insecurity in the identification of species [6].

The easiest way to identify endophytic fungi is through molecular methods. ITS region is accepted as a barcode for fungi because of the higher amplification success rate for many fungal groups [7] [8]. Thanks to the next generation sequences the amount of data obtained in the ecological studies has increased exponentially through the years [9], [5], which in turn implies a challenge for the processing and analysis of the information.

Recently is common to join computational efforts, the development of bioinformatics tools and the and specialized databases to make the analysis of the large sets of biological data more fast and efficient [10], [11]. High performance and high throughput computing technologies permit that the processing of such datasets could be automated to accelerate bioinformatics processes [12].

One of the most time-consuming analyses is the Bayesian Tree inference [13]. This is why a critical step in the ecological analysis is to use parallel tools that fulfill this task.

Exabayes is an is a software package that computes Bayesian tree posterior probability using the Markov Chain Monte Carlo sampling approach. This tool applies commonly used evolutionary models and can handle big data sets efficiently. An important aspect of this tool is that it allows the use of Message-passing Interface (MPI) to parallelize the analysis using a computer cluster so that the only limit for the analysis of large data sets is the memory held by the cluster [14].

Exabayes works by performing a calculation of the posterior probability of sampling trees. This tool can divide its analysis on multiple independent runs which in turn analyze multiple chains responsible for sampling the parameters for stochastic simulation to obtain a sample from the posterior distribution of trees [15], [14].

In this paper, we implement the analysis of ITS sequences of endophytic fungi of coffee plants to achieve the reconstruction of their phylogeny through the parallelization of exabayes with the objective of improving the performance of this tool using a computational cluster. These techniques may help ecologist or micologist to process their data in a efficient way, improving the use if computational resources and generating more accurate results.

Methodology and results

For the following experiments, a set of sequence data in Fastq format was used, which were obtained from endophyte fungi extracted from Costa Rica coffee plants. Then a quality control of the sequences was carried out using the FastQC [16] tool for the identification of poor quality bases. All bioinformatic processing was carried out in the Kabre cluster of the National Center of High Technology in San Jose, Costa Rica

The preprocessing of the sequences was made with the Seqtk tool [17], using two of its functions. First, the Trimfq function was used to remove the poor quality bases at both ends of the sequences, identified with a Phred score of less than 20. Then the Seq function was used to change the format of the Fastq sequences to FASTA.

Then duplicate sequences were removed with the USEARCH [18] tool with which a single sequence multifasta file was obtained with 331 sequences of 1331 each. The identification of the Operational Taxonomic Units (OTUs) was then carried out using the UNITE database [10] through the Blast tool of the National Center for Biotechnology Information [19].

For the phylogenetic analysis, an alignment of the sequences was performed using the MUSCLE tool [20]. The resulting file in phylip format is used as input for the Tree inference with Exabayes.

Initially, the analyzes in Exabayes were carried out using nodes, composed of Intel Xeon Phi KNL nodes processors, each one with 64 cores @ 1.3 GHz and 96 GB (architecture A). Historically. A sequential test was performed using 256 ranks and dividing the analysis into four runs and these in turn into four chains using four swaps between chains. This test lasted 447 hours.

To increase efficiency, parallelization of runs and chains was implemented using MPI the parameters described in the documentation: number of runs executed in parallel (-R), number of chains per run executed in parallel (-C) and number of swap attempts between chains per generation.

From these experiments, the best performance was obtained using 64 ranks, and dividing the analysis into 2 runs with 2 strings each. This same distribution was used for parallel runs and parallel chains, with 4 swaps per generation. This test lasted 20:29:01 (hh: mm: ss) reaching statistical significance with = 0.0131916 (table 1).

These tests were also performed with a different architecture (architecture B) using nodes c composed of nodes with Intel (R) Xeon (R) CPUs E3-1225 v5 @ 3.30GHz and 16GB RAM. In the same way, as with the previous tests, the use of the parameters was compared in parallel and better and statistically significant performance (asdsf = 0.0199349) was obtained using 4 ranks and dividing the analysis into 2 runs and 2 chains. It was also specified that the two runs and the two strings will be executed in parallel. With this the best result was obtained with a run time with 6:25:31 (hh: mm: ss) (table 1).

Analysis

As shown in the previous results, Exabayes efficiency improved substantially by balancing the number of runs and chains in which the analysis is divided with the number of runs and chains executed in parallel.

In this way, the better performance of the traditional test tool is observed to test 1. Thus, the number of ranks can be better distributed among the groups and achieve better performance.

With these tests, it was also found that by decreasing swapping among coupled chains, the distribution of resources becomes more efficient by decreasing communication between groups (of runs and chains that are being executed in parallel.

On the other hand, when comparing the performance of Exabayes between the different architectures, there is a significant improvement in the tests carried out on nodes B with respect to nodes A. This indicates that it is more advantageous for Exabayes, to use fewer CPUs but with more clock speed (3.3 GHz), than using a node that has a lot of processors with a very low clock speed (1.3 GHz).

[12] A. Welivita, I. Perera, D. Meedeniya, A. Wickramarachchi and V. Mallawaarachchi, "Managing Complex Workflows in Bioinformatics: An Interactive Toolkit With GPU Acceleration". *IEEE Transactions on NanoBioscience*, vol. 17, no. 3, pp.199-208, 2018.

[13] G. Altekar, S. Dwarkadas, J. Huelsenbeck and F. Ronquist, "Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference". *Bioinformatics*, vol. 20, no. 3, pp.407-415, 2004.

[14] A. Aberer, K. Kobert and A. Stamatakis, (2019). ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era.

[15] B. Rannala and Z. Yang, "Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference". *Journal of Molecular Evolution*, vol. 43, no. 3, pp. 304-311, 1996.

[16] S. Andrews, (2010). FastQC: a quality control tool for high throughput sequence data.

[17] H. Li, (2012). seqtk Toolkit for processing sequences in FASTA/Q formats.

[18] R. Edgar, "Search and clustering orders of magnitude faster than BLAST". *Bioinformatics*, vol. 26, no. 19, pp. 2460-2461, 2010.

[19] G. Boratyn, et al., "BLAST: a more efficient report with usability improvements". *Nucleic Acids Research*, vol. 41, no. W1, pp. W29-W33, 2013.

[20] R. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput". *Nucleic Acids Research*, vol. 32, no. 5, pp.1792-1797, 2004.

4.2 Actividad antagonista contra el patógeno de café *Mycena citricolor.*

La mancha americana de la hoja, causada por *Mycena citricolor*, es una enfermedad importante del café (*Coffea arabica*), principalmente en América Central. Actualmente, existen alternativas limitadas de control de patógenos que son amigables con el medio ambiente y económicamente accesibles. El uso de hongos aislados de la endomicobiota vegetal en sus hábitats nativos está en aumento porque los estudios muestran su gran potencial para el control biológico. Para comenzar a generar una alternativa verde para el control de *M. citricolor*, los objetivos del presente estudio fueron (i) recolectar, identificar, evaluar (in vitro e in planta) y seleccionar hongos endófitos de Rubiaceae silvestres recolectados en bosques maduros. de Costa Rica; (ii) confirmar la colonización endófita en plántulas de café; (iii) evaluar los efectos de los endófitos en el desarrollo de las plántulas; y (iv) corroborar la capacidad antagónica in planta. A través de ensayos de antagonismo *in vitro* e *in planta*, encontramos que de los aislados seleccionados (i.e. *Daldinia eschscholz*ii GU11N, *Nectria pseudotrichia* GUHN1, *Purpureocillium* aff. *lilacinum* CT24, *Sarocladium* aff. *kiliense* CT25, *Trichoderma rifaii* CT5, *T*. aff *crassum* G1C, *T*. aff. *atroviride* G7T, *T*. aff. *strigosellum* GU12 y *Xylaria multiplex* GU14T), *Trichoderma* spp. produjo los mayores porcentajes de inhibición del crecimiento in vitro. Los aislados de Trichoderma CT5 y G1C se probaron luego usando plántulas de *Coffea arabica* cv. caturra. Se verificó la colonización endofítica, seguida de ensayos de promoción y antagonismo del crecimiento *in planta*. Los resultados muestran que los aislados de *Trichoderma* CT5 y G1C tienen potencial para promover el crecimiento de las plantas y antagonizar contra *M. citricolor*, reduciendo la incidencia y la severidad, y previniendo la mortalidad de las plantas.

Publicación:

Escudero-Leyva, E., Granados-Montero, MdM., Orozco-Ortiz, C., Araya-Valverde, E., Alvarado-Picado, E., Chaves-Fallas J.M, Aldrich-Wolfe, L. & Chaverri. P. **2023**. The endophytobiome of wild Rubiaceae as a source of antagonistic fungi against the American Leaf Spot of coffee (*Mycena citricolor*). Journal of Applied Microbiology. 134(5). https://doi.org/10.1093/jambio/lxad090

Journal of Applied Microbiology, 2023, 134, 1–12
https://doi.org/10.1093/jambio/lxad090
Advance access publication date: 27 April 2023
Research Article

JOURNAL OF APPLIED MICROBIOLOGY
OXFORD UNIVERSITY PRESS

The endophytobiome of wild Rubiaceae as a source of antagonistic fungi against the American Leaf Spot of coffee (*Mycena citricolor*)

Efraín Escudero-Leyva[1,2], María del Milagro Granados-Montero[3], Cristofer Orozco-Ortiz[2], Emmanuel Araya-Valverde[2], Eduardo Alvarado-Picado[4], José Miguel Chaves-Fallas[1,5], Laura Aldrich-Wolfe[6], Priscila Chaverri[1,7,*]

[1]Centro de Investigaciones en Productos Naturales (CIPRONA) and Escuela de Biología, Universidad de Costa Rica, San Pedro, San José 11801, Costa Rica
[2]Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, Pavas, San José 10109, Costa Rica
[3]Escuela de Agronomía, Estación Experimental Fabio Baudrit and Centro de Investigaciones en Estructuras Microscópicas, Universidad de Costa Rica, San Pedro, San José 11801, Costa Rica
[4]Centro para el Desarrollo de Alternativas Orgánicas (CeDAO), San Marcos de Tarrazú, San José 10501, Costa Rica
[5]Department of Biology and Whitney R. Harris World Ecology Center, University of Missouri—St. Louis, St. Louis, MO 63121, USA
[6]Department of Biological Sciences, North Dakota State University, Fargo, ND 58102, USA
[7]Department of Natural Sciences, Bowie State University, Bowie, MD 20715, USA

*Corresponding author. Department of Natural Sciences, Bowie State University, 14000 Jericho Park Road, Bowie, MD 20715, USA. E-mail: pchaverri@bowiestate.edu

Abstract

Aims: The American leaf spot, caused by *Mycena citricolor*, is an important disease of coffee (*Coffea arabica*), mostly in Central America. Currently, there are limited pathogen control alternatives that are environment friendly and economically accessible. The use of fungi isolated from the plant endomycobiota in their native habitats is on the rise because studies show their great potential for biological control. To begin to generate a green alternative to control *M. citricolor*, the objectives of the present study were to (i) collect, identify, screen (*in vitro* and *in planta*), and select endophytic fungi from wild Rubiaceae collected in old-growth forests of Costa Rica; (ii) confirm endophytic colonization in coffee plantlets; (iii) evaluate the effects of the endophytes on plantlet development; and (iv) corroborate the antagonistic ability *in planta*.

Methods and results: Through *in vitro* and *in planta* antagonism assays, we found that out of the selected isolates (i.e. *Daldinia eschscholzii* GU11N, *Nectria pseudotrichia* GUHN1, *Purpureocillium* aff. *lilacinum* CT24, *Sarocladium* aff. *kiliense* CT25, *Trichoderma rifaii* CT5, *T.* aff. *crassum* G1C, *T.* aff. *atroviride* G7T, *T.* aff. *strigosellum* GU12, and *Xylaria multiplex* GU14T), *Trichoderma* spp. produced the highest growth inhibition percentages *in vitro*. *Trichoderma* isolates CT5 and G1C were then tested *in planta* using *Coffea arabica* cv. caturra plantlets. Endophytic colonization was verified, followed by *in planta* growth promotion and antagonism assays.

Conclusions: Results show that *Trichoderma* isolates CT5 and G1C have potential for plant growth promotion and antagonism against *Mycena citricolor*, reducing incidence and severity, and preventing plant mortality.

Significance and Impact of Study

The results of this study increase knowledge on the potential use of poorly studied species of endophytic fungi that are not only capable of reducing the impacts of phytopathogens, but also improve plant growth, aiding in the transition to organic agriculture.

Keywords: biofungicide, Costa Rica, endophyte, Hypocreales, *Mycena citricolor*, organic agriculture, phytobiome, *Trichoderma*, Xylariales

Introduction

Global consumer markets are presently demanding more environment-friendly farming practices even when the transition from conventional agriculture may be challenging for the producing countries (Crespo et al. 2020). Coffee (*Coffea arabica*) production is currently threatened, more than ever before, by pests, diseases, and the effects of climate change (Eakin et al. 2014); however, greener alternatives to mitigate those problems are limited (Fuhrimann et al. 2018, Baker et al. 2020). Therefore, the development of innovative strategies that reduce economic losses and environmental impacts of pesticides/fungicides is being encouraged, as this crop is vital for the economies and cultures of many low-income countries (Pineda et al. 2019).

One of the main diseases affecting coffee in Central America is the American leaf spot (*Mycena citricolor*), in addition to coffee leaf rust (*Hemileia vastatrix*) and anthracnose (*Colletotrichum* spp.; Cristóbal-Martínez et al. 2017, Allinne et al. 2018, Granados-Montero et al. 2020). In the last epidemic by *M. citricolor* in 2010, losses amounted to ca. USD 60 million for the coffee industry just in Costa Rica (Granados-Montero et al. 2020). In the previous epidemics of 1995 and 1998, the pathogen affected >3000 hectares of coffee plantations (Borbón 1999). There are cyclic disease outbreaks approximately every 15 years, correlated with the increased rainfall and fungal inoculum abundance (Carvajal 1939, Borbón

Received: February 9, 2023. **Accepted:** April 26, 2023

1999, Granados-Montero et al. 2020). Therefore, preparedness for a possible next outbreak becomes critical.

Unfortunately, chemical-free strategies to manage or reduce the impacts of *M. citricolor*, as well as other pathogens, are very limited. Current approaches include farmers organizing for early detection of diseases, replacing the existing coffee plants with resistant varieties, and transitioning from synthetic fungicides to microorganism-derived compounds or biological control (Avelino et al. 2015). The latter has been favored in part by international certification agencies and international consumer markets, driving the coffee farmers to switch to greener practices (Pinto et al. 2014, Bravo-Monroy et al. 2016).

Various studies demonstrate that plants in their native habitats (i.e. old-growth tropical forests) harbor a high diversity of fungal endophytes with a hypothesized protective effect (Evans et al. 2003, Mulaw et al. 2013, Bussey et al. 2015, Gazis and Chaverri 2015, Aldrich-Wolfe et al. 2020, del Carmen et al. 2021, Rodríguez et al. 2021). Therefore, since the natural distribution of the coffee family (Rubiaceae) is in the tropics (Manns et al. 2012), it is highly probable to find fungal endophytes in those old-growth forests that could have antagonistic abilities against pathogens and be applied in coffee production. In addition, studies show that specifically endophytic species (e.g. some *Trichoderma* spp.) were able to colonize the internal tissues of the plants (e.g. *Hevea brasiliensis* and *Theobroma cacao*) and confer resistance against several pathogens (e.g. *Corynespora cassiicola*, *Moniliophthora roreri*, and *Phytophthora* spp.; Bailey et al. 2008, Mejía et al. 2008, Pujade-Renaud et al. 2019). Based on the above premises, the present study aimed at (i) collecting, identifying, selecting, and screening endophytic fungi from wild Rubiaceae plants for antagonism against *Mycena citricolor*; (ii) confirming endophytic colonization in coffee plantlets; (iii) evaluating the effects of the endophytes in plantlet development; and (iv) corroborating the antagonistic ability *in planta*. The results of this study will provide the groundwork to develop alternatives to synthetic fungicides and aid in the transition to organic or pesticide-free agriculture. These alternatives would consist of endophytic fungi that can fully colonize the internal plant tissues (i.e. roots, stems, and leaves), thus providing a more long-term protection and resistance (Knapp et al. 2012, Xie et al. 2021, Stummer et al. 2022). Additionally, they would provide the crop with a better resilience to other macro- and microclimatic, or biotic and abiotic effects that may appear (Hubbard et al. 2014, Dunlap 2017, Yan et al. 2019, Sadeghi et al. 2020, Hassan et al. 2022, Karimi et al. 2022), because, as the plant grows, the endophytic fungi would continue to grow internally without the need for new applications. The present study reports on the potential of endophytic fungi from wild Rubiaceae that could support this effort.

Materials and methods

Study sites

Plants from old-growth natural forests were sampled in the Miravalles and Rincón de la Vieja National Parks (10.7477 N, 85.1612 W) in the Guanacaste Conservation Area (GCA) and a private forest near Río Claro in the South Pacific Area (SPA) in Costa Rica (8.6389 N, 83.0914 W) (Supplementary Fig. S1). GCA is a 163 000-ha conservation area in Northwestern Costa Rica and an important ecological zone characterized by high species richness derived from the convergence of the dry, wet, and cloud forests (Jiménez-Rodríguezet al. 2015). SPA, just 10 km West of Piedras Blancas National Park, has a typical low-land tropical forest (Höbinger et al. 2012).

Isolation of fungal endophytes

Fungal endophytes were isolated from leaf and sapwood tissues in one individual from each of the 46 wild Rubiaceae species that were found: *Alibertia edulis*, *Allenanthus erythrocarpus*, *Arachnothryx costaricensis*, *Bertiera bracteosa*, *Chimarrhis latifolia*, *Chomelia costaricensis*, *Chomelia microloba*, *Coutarea hexandra*, *Coussarea loftonii*, *Coussarea talamancana*, *Cosmibuena macrocarpa*, *Duroia costaricensis*, *Faramea tamberlikiana*, *Genipa americana*, *Gonzalagunia osaensis*, *Guettarda macrosperma*, *Isertia haenkeana*, *Macrocnemum roseum*, *Notopleura polyphlebia*, *Palicourea eurycarpa*, *Pa. adusta*, *Pentagonia costaricensis*, *Posoqueria grandiflora*, *Po. latifolia*, *Psychotria berteroana*, *Ps. brachiata*, *Ps. guianensis*, *Ps. horizontalis*, *Ps. marginata*, *Ps. mortoniana*, *Ps. panamensis*, *Ps. pubescens*, *Ps. solitodinum*, *Ps. subsessilis*, *Randia grandifolia*, *Ra. lasiantha*, *Ra. aculeata*, *Ronabea emetica*, *Ro. latifolia*, *Rudgea cornifolia*, *Rustia costaricensis*, *Simira maxonii*, *Sommera donnell-smithii*, *Tocoyena pittieri*, and *Warszewiczia coccinea*. Vouchers of all the plants collected were deposited in the Luis A. Fournier Origgi Herbarium (USJ), University of Costa Rica.

Three healthy mature leaves were collected randomly from each plant. For sapwood (vascular cambium + phloem) samples, plants with a diameter-at-breast-height >20 cm were selected. With a knife previously sterilized and at breast height, three 3 × 1 cm rectangles were extracted from the stem (Evans et al. 2003). Samples collected in the forests were transported to the laboratory in plastic bags placed inside a cooler. In the laboratory, 5-mm squares from leaves and bark were cut and surface-sterilized by sequential immersions in 3% sodium hypochlorite (NaOCl) for 1 min, 70% ethanol for 30 s, and finally rinsed repeatedly with sterile water (Abo-Elyousr et al. 2014). Five pieces of the sterilized leaves and bark tissues were placed onto 90-mm Petri plates with potato dextrose agar (PDA) (Difco™ Laboratories, Detroit, MI, USA) and gentamycin as antibiotic. The success of the surface-sterilization was confirmed using the leaf imprint method (e.g. Greenfield et al. 2015). Plates were incubated at room temperature for 5–10 days with a photoperiod of 12/12 h light/dark. Re-isolation and purification of colonies were performed until pure cultures were obtained and then preserved in 1-mL vials with sterile distilled water at room temperature (25 ± 2°C) at the Microbial Biotechnology Laboratory (Natural Products Research Center, University of Costa Rica).

Molecular identification of the fungal isolates

Fungal isolates were grown on PDA for 5–15 days, depending on the isolate and species, or until the colony reached the edge of the Petri dish; then the mycelium was harvested from the surface. DNA extraction was done using PrepMan™ Ultra (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions. Polymerase chain reactions (PCR) used the primers ITS4 and ITS5 (White et al. 1990) of the internal transcriber spacers (ITS) nuclear ribosomal DNA. The 25-µL PCR reaction mixture consisted of 12.5 µL GoTaq® Green Master Mix (Promega Corpora-

tion, Madison, WI, USA), 1-μL forward primer, 1-μL reverse primer, 1 μL of dimethyl sulfoxide, 0.5 μL of bovine serum albumin, 6 μL of UltraPure™ distilled water, and 3 μL genomic DNA template (DNA concentration between 20 and 40 ng μL^{-1}; Herrera et al. 2013). We confirmed amplification through gel electrophoresis. PCR products were sent to Psomagen (MD, USA) for purification and Sanger sequencing. *Trichoderma* (Hypocreales, Sordariomycetes, and Ascomycota) isolates were the focus of the antagonism experiments. Therefore, to obtain a better identification, portions of the translation elongation 1-alpha gene (TEF1) and RNA polymerase II subunit (RPB2) were sequenced using primers 728Mf and 986r (Carbone and Kohn 1999), and RPB2-5f and RPB2-7cr (Liu et al. 1999), respectively (Chaverri et al. 2015, Cai and Druzhinina 2021). Beta-tubulin (TUB) marker (primers T1 and T22) was used to refine the identification of isolates in Xylariales (Sordariomycetes and Ascomycota; Huei-Mei et al. 2005). Edition and alignment of the sequences were done in Geneious v.10.2.3 (https://www.geneious.com). BLASTn algorithm was performed in Geneious with retrieve from UNITE v.2020 database (Abarenkov et al. 2010). The top five hits (>98% identity and >85% coverage) were initially compared, and the top hit with a sequence identity of >99% was used to recognize or name the putative species or Operational Taxonomic Units (OTUs; Gazis et al. 2011).

A cladogram using the ITS sequences was constructed. For this, the closest matches, with "type material" prioritized, were retrieved from the BLAST analyses and aligned with MUSCLE (Edgar 2004). The resulting alignment in Phylip format was submitted to Bayesian Inference analysis with Exabayes v.1.5 (Aberer et al. 2014). Markov chain Monte Carlo (MCMC) was run in parallel, and 15 million generations were done with 25% burn-in. All analyses were run in the Kabré supercomputer (CNCA-CONARE, Costa Rica). A consensus tree was visualized and edited with FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). In the case of *Purpureocillium*, an additional TEF1 cladogram was constructed because its identity with ITS was uncertain.

In vitro dual-plate antagonism tests against *Mycena citricolor*

Even though many endophyte isolates were obtained (Supplementary Table S1), only nine endophytes were selected for *in vitro* tests based on close phylogenetic relationship to taxa with known antagonistic ability (i.e. Hypocreales and Xylariales: *Daldinia eschscholzii* GU11N, *Nectria pseudotrichia* GUHN1, *Purpureocillium* aff. *lilacinum* CT24, *Sarocladium* aff. *kiliense* CT25, *Trichoderma rifaii* CT5, *T.* aff. *crassum* G1C, *T.* aff. *atroviride* G7T, *T.* aff. *strigosellum* GU12, and *Xylaria multiplex* GU14T; Rodriguez et al. 2011, Kim et al. 2018, Guo et al. 2019).

Mycena citricolor was isolated from infected coffee plants (*C. arabica* cv. caturra) in San Marcos de Tarrazú (San José, Costa Rica) plantation "La Laguna." Leaves with well-developed gemmae (pseudopilei) were obtained and the material was transferred to the Center for the Development for Organic Alternatives (CeDAO, CoopeTarrazú) facilities for isolation. Pseudopilei were collected (at least five) and placed in Petri dishes with PDA. Re-isolations were done until pure cultures were obtained; these were preserved with the same methodology as for the endophytes. Out of the isolates obtained, only one (LAMYH1) was selected for subsequent antagonism assay due to the abundant production of pseudopilei in *in vitro* conditions after 4 weeks.

For the dual-plate experiment, the pathogen was grown in a 90-mm-diameter Petri dish with PDA medium, and its radius was measured every 24 h for 5 days at 25 ± 2°C to obtain the control growth rate. After 7 days of growth, a 5-mm-diameter disc was taken from the edge of the colony, transferred to a new Petri plate, and placed 5 mm away from the dish border. After 2 days, identical-size mycelial discs of 7-day-old cultured endophytes were placed in opposition in the same plate and incubated at 25 ± 2°C (Landum et al. 2016). Five replicates were done for each experiment. Growth inhibition (% GI) of the pathogen was measured by the following equation (Abo-Elyousr et al. 2014):

$$\%GI = [(C - T)/C] \times 100,$$

where GI = growth inhibition, C = radial growth of pathogen on control, and T = radial growth of pathogen under treatment.

Direct mycoparasitism was determined by microscopic observation of hyphal coiling (possible necrotrophic contact, Sun et al. 2019) as soon as hyphae of both isolates touched and slightly overlapped to avoid overcrowding of mycelium. Antibiosis is assumed when inhibition halos are formed (Sun et al. 2019).

Positive controls were the pathogen grown onto medium supplemented with various fungicides commonly used in coffee plantations, i.e. azoxystrobin (systemic), chlorothalonil (non-systemic), cyproconazole (systemic), propineb (systemic), tolclofos-methyl (non-systemic), and validamycin-A (non-systemic). The fungicides were added to the media following the protocol and concentrations used in Escudero-Leyva et al. (2022).

In planta antagonism tests

The Supplementary Fig. S3 illustrates the process from *in vitro* plantlet establishment, endophyte inoculation and colonization, plantlet acclimatization, and the *in planta* antagonism assay.

In vitro coffee plantlet establishment

In sterile conditions, 180 embryos were extracted from seeds of *C. arabica* cv. caturra. The embryos were cultivated in Murashige and Skoog medium (MS; Murashige and Skoog 1962) supplemented with Morel vitamins (Sridevi and Giridhar 2014), 30 g L^{-1} sucrose, and 3.5 g L^{-1} Phytagel™ (Sigma–Aldrich Co., St. Louis, MO, USA). The pH of the medium was adjusted to 5.6 before sterilization. Germinated embryos were transferred to the multiplication phase on MS medium supplemented with 30 g L^{-1} sucrose, 1 mg L^{-1} 6-benzyl aminopurine, and 3.5 g L^{-1} Phytagel™. The pH of the medium was adjusted to 5.6. After four multiplication cycles (45 days/cycle), plantlets were transferred to rooting stage on $\frac{1}{2}$ MS medium supplemented with Morel vitamins, 1 mg L^{-1} indole-3-butyric acid, 10 g L^{-1} sucrose, and 3.5 g L^{-1} Phytagel™. The pH was adjusted to 5.7 and growth-room conditions were 26.5 ± 2.5°C and 16 h light/8 h darkness.

Colonization and effect of fungal endophytes on *in vitro* coffee plantlets

A total of 20 rooted plantlets per treatment were transferred to custom-made sterile 11-L modular hermetic storage

containers (42.5 cm length × 25.7 cm height × 16.6 cm wide) filled with 700 g of the substrate Berger™ BM Custom Blend (Quebec, Canada). These previously sterilized containers with lids were kept in a clean growth room to reduce the probability of contamination by other microorganisms that can infect the *in vitro* plantlets and become endophytic. The treatments were: (i) plantlets with CT5 endophyte; (ii) plantlets with G1C endophyte; and (iii) plantlets without endophyte [blank; 20 plants per container × 3 treatments × 3 times (blocks) = 180 plantlets used]. Only *Trichoderma rifaii* (CT5) and *T.* aff. *crassum* (G1C) were used due to their high growth inhibitions *in vitro* against *M. citricolor* and their capacity to produce abundant conidia. Each of the plantlets within treatments (i) and (ii) were inoculated with 15 mL of distilled sterilized water containing 1×10^7 conidia mL^{-1}, which was drained near the base of each plantlet using a pipette (direct spore counts were made using a Neubauer chamber 1/400 mm^2 to measure the conidial concentration). The plants without endophyte were irrigated only with 15 mL of distilled sterile water. This procedure was done three times for all treatments, sterilizing containers and renewing the soil. Plantlet height (only the aerial part), root length, and number of leaves were annotated at the beginning of the experiment and after 4 weeks. To confirm endophyte colonization in different tissues (stems and leaves), three plantlets per container per treatment (27 plantlets total) were haphazardly chosen on each replicate batch following the procedure previously described for the endophytic fungal isolation. Once the fungal colonies emerged from the tissues, their identity was confirmed with morphology and the molecular marker TEF1. We also confirmed that no other fungus emerged from the isolated tissues.

Plantlet establishment in the greenhouse for antagonism assays

After four weeks, all the plantlets (153 total) were transplanted into individual clean plastic bags with 800 g of sterile mix of soil and rice husk to allow for better drainage. Per treatment, only 10 plants were used per block (30 plantlets total per treatment). Afterwards, they were transferred to a greenhouse (conditions were 28 ± 3°C and 85 ± 10% relative humidity) in San Marcos de Tarrazú, Costa Rica. Here, the plants were acclimatized for four weeks following a pesticide/fungicide-free management used by coffee growers of the Tarrazu region: watering every two days; fertilization was performed twice with humic acids with phosphorus (Ecogreen organic fertilizer, Costa Rica; 12% N, P_2O_5 58%, humic and fulvic acids 3%, inert 27%); a total of 20 mL of the product was applied to each plant.

Mycena citricolor inoculation and antagonism assay

From the original 180 *in vitro* plantlets, 27 were sacrificed to confirm endophyte colonization, and others died during the acclimatization period. Therefore, for the antagonism experiment, a total of 90 plants were used and the experiment was designed as follows: The experiment was run using blocks of 10 plants per treatment to favor randomness, for a total of 30 plants per treatment. Height (10–20 cm) and the number of leaves (between 8 and 12) per plant were homogenized for each treatment and trial. Using a dissecting scope, mature *M. citricolor* pseudopilei were carefully removed with fine tweezers and positioned inside synthetic-clay circles previously attached over three healthy leaves per plantlet. Lastly, the plant was carefully atomized with sterile water. A total of 60 mature pseudopilei were applied in each treatment. To maintain the humidity, two 100 L semitransparent boxes were used to keep 10 plants per treatment, adding one inch of water, and closing the lids. Humidity and temperature were measured every week with electronic sensors (model 8541833483, AikTryee, USA). Every week, the plants were monitored for symptoms of infection and signs of the pathogen, and to maintain the humidity >90% by adding water if necessary. After 4 weeks, diameters of lesions caused by *M. citricolor* were measured (in mm). Other parameters, such as number of infected leaves (three leaves/plant were inoculated with the pathogen), defoliation (%), number of pseudopilei, and plant mortality were also recorded.

Statistical analysis

The variables measured in the endophyte colonization process and the antagonism assay were subjected to normality tests and data transformed accordingly: log10 transformation for plant height and square root for root length and number of leaves in the colonization experiment; square root for infected leaves. Then, generalized linear mixed models (GLMM) were used to infer the effect of the treatments using the lmer package v.3.1.3 due to the split plot model design derived from the arrangement of the experiment. The models considered the blocks and containers as random effects and treatments as the fixed effect. Significance between treatments was determined using *post hoc* Tukey HSD test. All tests and graphics were done using R software v.4.1.3 (R Core Team 2021).

Results

Identification and selection of fungal isolates

A preliminary account of 395 culturable Ascomycota and Basidiomycota were isolated from wild Rubiaceae plant species (Supplementary Table S1). After molecular identification and BLAST analysis using the UNITE database, the collection was best represented by the order Glomerellales (232 OTUs), followed by Diaporthales (72 OTUs; Supplementary Table S1). Other fungal orders of interest were less represented (ca. 10%) because they contain species with antagonistic potential, e.g. Xylariales and Hypocreales. The selected isolates were: *Daldinia eschscholzii* (GU11N), *Nectria pseudotrichia* (GUHN1), *Purpureocillium* aff. *lilacinum* (CT24), *Sarocladium* aff. *kiliense* (CT25), *Trichoderma rifaii* (CT5), *T.* aff. *crassum* (G1C), *T.* aff. *atroviride* (G7T), *T.* aff. *strigosellum* (GU12), and *Xylaria multiplex* (GU14T). We did not use Diaporthales or Glomerellales because they are known mostly as pathogens. The selected endophytes, their plant hosts, and their taxonomic assignments are included in Supplementary Table S2, and their position in a Bayesian Inference cladogram are in Supplementary Fig. S2. The resulting sequence data and additional information are available in GitHub (see "Data availability"). Some of the *Trichoderma* isolates (CT5, G1C, GU12, and G7T) were previously identified as *Trichoderma rifaii*, *T.* aff. *crassum*, *T.* aff. *atroviride*, and *T.* aff. *strigosellum*, respectively, in Escudero-Leyva et al. (2022).

In vitro dual plate antagonism tests

From the nine endophytes tested, the four *Trichoderma* species produced the highest pathogen GI after 3 days (Fig. 1a), reaching >50% GI (Fig. 1b); the remaining isolates showed GIs < 30%. Isolate GUHN1 and *Trichoderma* spp. reduced or

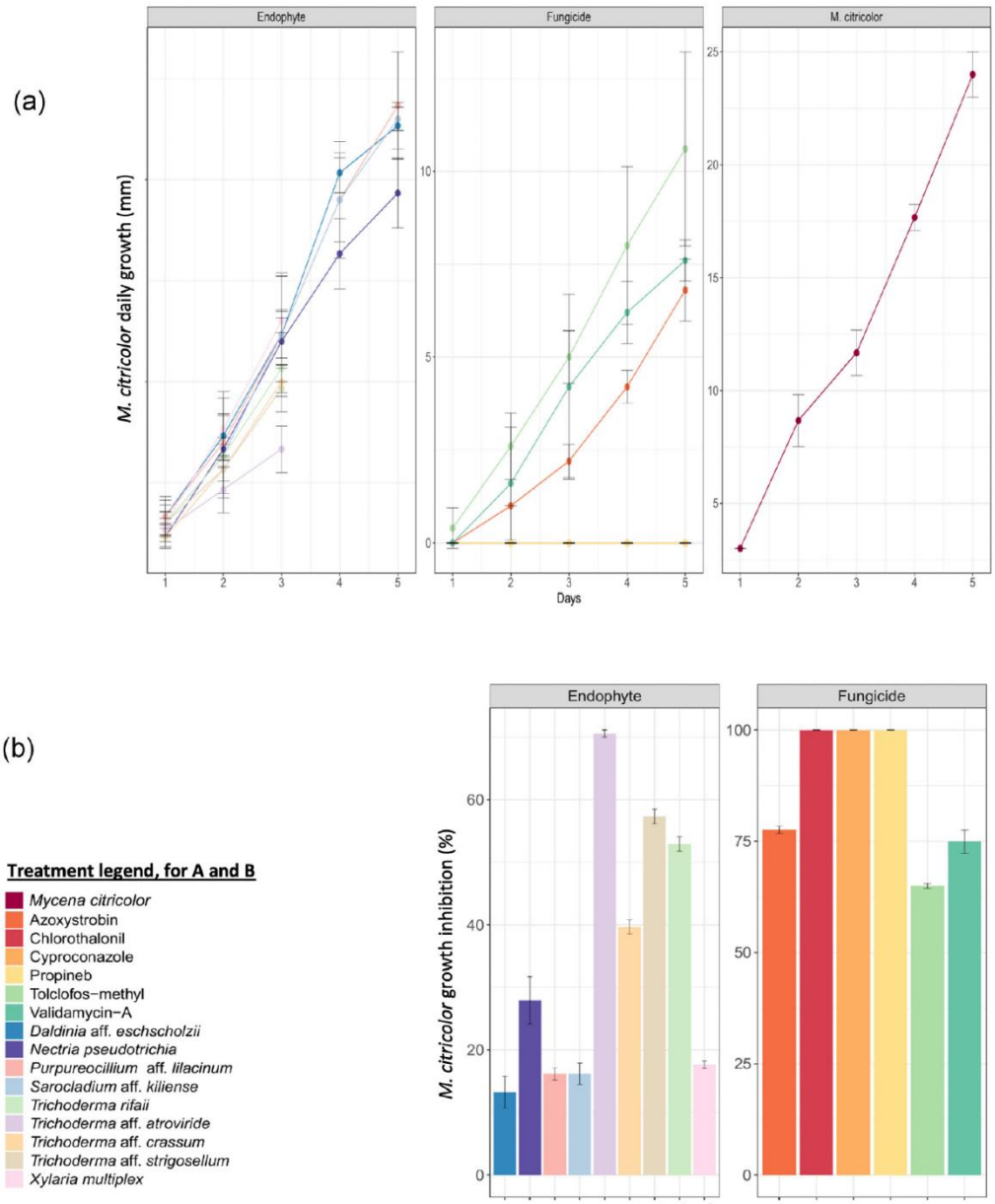


Figure 1. *In vitro* antagonism against *Mycena citricolor* LAMYH1. (a) *Mycena citricolor* mycelial growth after 5 days, alone and in presence of nine selected endophytes and five fungicides. (b) GI percentage against *M. citricolor* produced by endophytes and fungicides.

inhibited the growth of *M. citricolor* within 3–5 days (Fig. 1a). The highest inhibition percentage was registered for GU12 (70.58 ± 0.57%), followed by G7T (57.35 ± 1.15%), CT5 (52.94 ± 1.15%), and G1C (39.70 ± 1.15%). Growth inhibitions by fungicides (positive controls) ranged between 65% and 100% (Fig. 1b). The growth inhibitions produced by *Trichoderma* spp. were 40%–70%, while those exhibited by azoxystrobin, tolclofos-methyl, and validamycin-A were 65%–78%; and chlorothalonil, cyproconazole, and propineb were 100% (Fig. 1b).

Figure 2 shows images of the dual plate antagonism tests. For example, inhibition zones in e.g. *P.* aff. *lilacinum* CT24 (Fig. 2b), *D. eschscholzii* GU11N (Fig. 2c), and *X. multiplex* GU14T (Fig. 2j) indicate antibiosis. On the other hand, all the endophytic *Trichoderma* spp. (Fig. 2f–i) had a direct mycoparasitic activity evidenced by the hyphal coiling (Fig. 2k–m) and growth over *M. citricolor* (Fig. 2h). Figure 2f shows that *T. rifaii* CT5 not only had a direct mycoparasitic effect, but also antibiosis evidenced by an inhibition area. Isolate *Sarocladium* aff. *kiliense* CT25 had an overlapping non-mycoparasitic effect (Fig. 2d). Even though *Trichoderma* isolates GU12 and G7T had higher GIs than the rest, only *T. rifaii* CT5 and *T.* aff. *crassum* G1C were selected for the *in planta* experiments because the former did not produce abundant conidia (Fig. 2h and i).

Effects of endophytes on *in vitro* coffee plantlets

Endophytic colonization was confirmed during the re-isolation process from randomly sectioned plantlets, successfully recovering both inoculated *Trichoderma* isolates: *T. rifaii* CT5 and *T.* aff. *crassum* G1C. No other fungi grew from the sacrificed plantlets. CT5 was recovered from leaves, stems, and roots, while G1C was only re-isolated from stems and roots after four weeks. After one month, only height was statistically significant ($P = 0.0145$) according to the GLMM

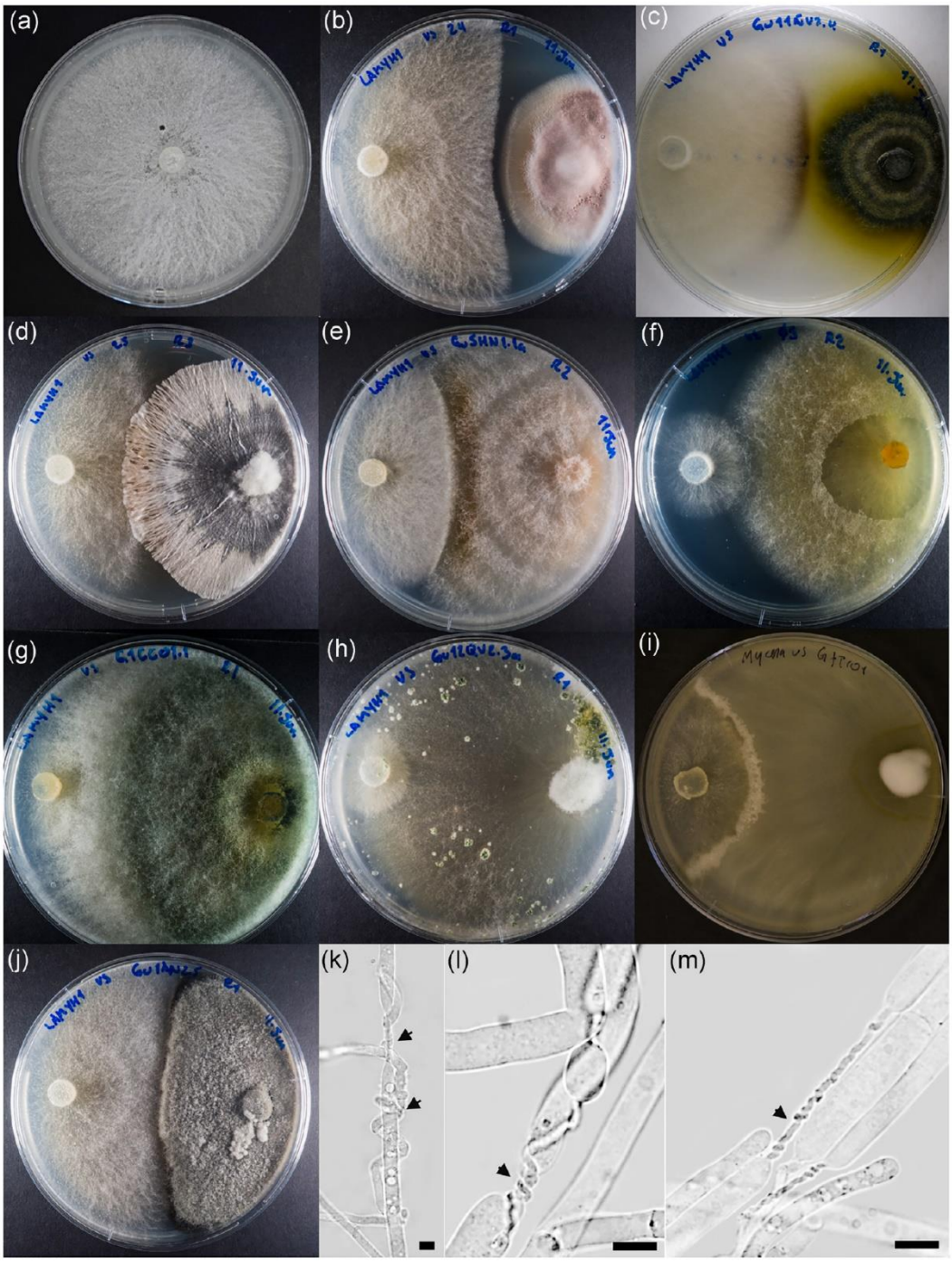


Figure 2. Antagonistic responses produced by the endophytes against *Mycena citricolor* LAMYH1 in PDA plates at 25°C after 5 days. (a) *Mycena citricolor* alone (control). (b)–(j) Dual plates *M. citricolor* (left side of the dual plates) against: (b) *Purpureocillium* aff. *lilacinum* CT24; (c) *Daldinia eschscholzii* GU11N; (d) *Sarocladium* aff. *kiliense* CT25; (e) *Nectria pseudotrichia* GUHN1; (f) *Trichoderma rifaii* CT5; (g) *T.* aff. *crassum* G1C; (h) *T.* aff. *strigosellum* GU12; (i) *T.* aff. *atroviride* G7T; (j) *Xylaria multiplex* GU14T. (k)–(m) Microscopic views of hyphal coiling (arrows) are evident in *T.* aff. *crassum* G1C (k) and *T. rifaii* CT5 (l) and (m). Scale bars: 5 μm.

where the container had no effect over the experiment (Fig. 3). The plants with inoculated endophytes resulted in twice the height compared to the control without endophyte (Fig. 3). The greatest difference was between CT5 and the control (Tukey HSD $P = 0.0359$), followed by G1C and the control (Tukey HSD $P = 0.0218$); there was no difference between CT5 and G1C (Tukey HSD $P = 0.9820$). For the root length and the number of leaves after four weeks, there were no differences among the treatments, although correlation was observed for the height and root length after one month ($P = 0.151$). The effect from replicates was only significant for the height data ($P = 0.001698$), but not for the root length ($P = 0.5333$) or number of leaves ($P = 1$). For the last, only the container presented an effect on the number of leaves ($P = 2.494$ E-06).

In planta antagonism tests

After four weeks, incidence (percent of plants with symptoms) was significantly different between the *Trichoderma*

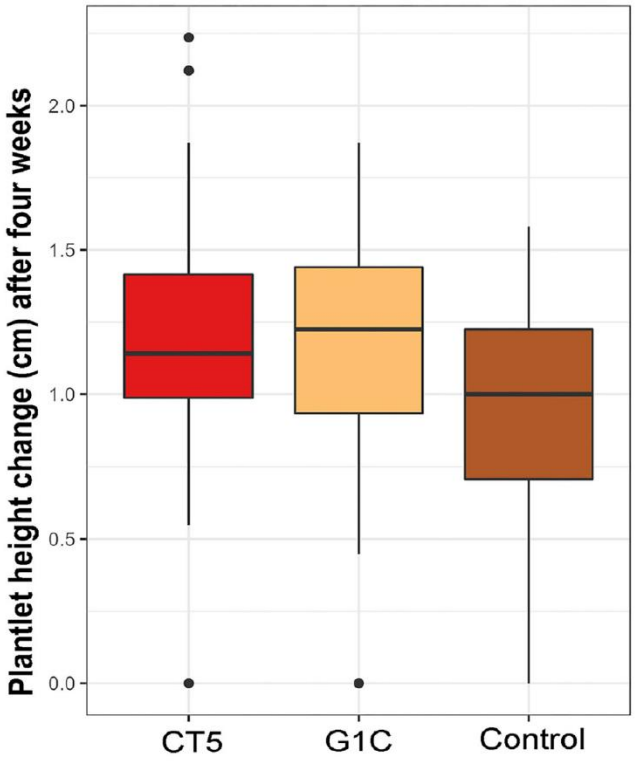


Figure 3. Growth evaluation of *in vitro* plantlet height after four weeks since the inoculation with *Trichoderma* isolates CT5 and G1C. Boxplot shows significative differences in plantlet height ($P < 0.05$).

treatments and the control (93.33 ± 2.54%, $P = 0.00987$), with no block or container effect. The Tukey test shows that plants in the control and CT5 (66.66 ± 4.79%, $P = 0.0228$) treatments were different. The defoliation percentage presented a container effect ($P = 0.00437$) but not a block effect, with differences between the control (50 ± 0.97%) and G1C plantlets (3.40 ± 0.18%, $P = 0.0323$) according to the Tukey test. The percent of infected leaves was significant in the control treatment (45.55 ± 2.05%, $P = 0.00601$), without observing effects from the containers or blocks. The *post hoc* test registered differences between control and CT5 plants (28.88 ± 2.43%, $P = 0.0135$). The leaf spot diameter was different for the control treatment (11.19 ± 5.73 mm, $P = 0.0522$), with absence of container or block effect; the Tukey test registered significance between the control and CT5 treatment (5.66 ± 6.10 mm, $P = 0.0172$). Finally, the percent mortality was significant for the control treatment (13.33 ± 3.46%, $P = 0.0113$), without observing effects derived neither from the container or block; no mortality was observed in the *Trichoderma*-treated plants ($P = 0.0262$). The number of pseudopilei per leaf was the only variable without differences among treatments; there were no container and block effects (Table 1). Images documenting symptoms and signs from this experiment are shown in Supplementary Fig. S4.

Discussion

The present study found that *Nectria pseudotrichia* (GUHN1), *Trichoderma rifaii* (CT5), *T.* aff. *crassum* (G1C), *T.* aff. *atroviride* (G7T), and *T.* aff. *strigosellum* (GU12) isolated from wild Rubiaceae plants were able to inhibit *in vitro* the growth of *Mycena citricolor*. The growth inhibition displayed by *Trichoderma* spp. was comparable to fungicides (*in vitro*). Meanwhile, inhibition was <20% for *Daldinia eschscholzii*, *Purpureocillium* aff. *lilacinum* (CT24), *Sarocladium* aff. *kiliense* (CT25), and *Xylaria multiplex* (GU14T). *In planta*, both *Trichoderma* treatments (CT5 and G1C) were effective in significantly reducing the pathogen incidence and severity, i.e. reduced infection incidence (CT5 = G1C), defoliation (G1C the lowest), number of infected leaves (CT5 = G1C), and size of necrotic areas (CT5 = G1C), and no plant mortality after four weeks (CT5 = G1C). The beneficial effects by these endophytes could represent key approaches in the protection, increased resistance, and *Mycena citicolor* eradication (or inoculum reduction), thus decreasing the impacts of the pathogen since weakening caused by defoliation can also trigger other diseases (Zamora et al. 2017). In a recent study, Escudero-Leyva et al. (2022) showed that CT5 and G1C were able to bioremove several fungicides. However, even though G1C was able to tolerate and bioremove the fungicide cyproconazole, the process resulted in even more toxic compounds. Therefore, caution should be used in the application of this *Trichoderma* isolate as a biological control agent.

As hypothesized, the *Trichoderma* isolates had the highest pathogen inhibition percentages. From the species tested, only *Trichoderma rifaii* has been reported as an endophyte isolated from *Theobroma gileri* in the Amazonian forests of Ecuador (Chaverri et al. 2015). This species belongs to the *T. harzianum* species complex (Chaverri et al. 2015), with many species well known for their antifungal abilities, including direct mycoparasitism through the production of fungal cell-wall degrading enzymes or antibiosis through several types of secondary metabolites (Keswani et al. 2014, Vinale et al. 2014, da Silva et al. 2015, Degenkolb et al. 2015), and induced systemic resistance (Harman et al. 2004, Shoresh et al. 2010, Hermosa et al. 2012, Mendoza-Mendoza et al. 2018). Mechanisms involved in the antagonistic response by the endophytic *T. rifaii* used in this study are hypothesized as direct mycoparasitism (as seen by the hyphal coiling) similar to the *Trichoderma* isolates successfully tested against *Corynespora cassiicola* by De Souza et al. (2008) and Pujade-Renaud et al. (2019), or against *Colletotrichum* spp. in Gazis and Chaverri (2015). On the other hand, even though our results suggest that *Trichoderma* aff. *crassum* is a mycoparasite of *Mycena citricolor*, the species has not been widely studied as a biocontrol agent and thus the mechanisms of antagonism are not yet elucidated (Hoyos-Carvajal et al. 2009, Mendoza et al. 2015, Zhang et al. 2015, Sumida et al. 2018). Regarding the remaining *Trichoderma* species tested *in vitro*, there are reports indicating that *T. viride* has antagonistic abilities against *Fusarium graminearum*, *F. solani*, and *Macrophomina phaseolina*; however, the inhibition percentage was lower compared to other *Trichoderma* species tested in that study (e.g. *T. asperellum*, *T. harzianum*, and *T. longibrachiatum*; Hewedy et al. 2020). In another work, *T. viride* inhibited the growth of *Fusarium verticillioides* through the production of secondary metabolites (Kumar et al. 2021). *Trichoderma strigosellum* belongs to the "*T. koningii* clade" and was isolated from leaf litter from *Pseudomonotes tropenbosii* in the Colombian Amazon Forest (López-Quintero et al. 2013). Interestingly, *T. strigosellum* can colonize *Eucalyptus urophylla* plants when applied in drench, and growth promotion, as well as protection against herbivory, has been

Table 1. Effects produced by *Mycena citricolor* in plants colonized by endophytic *Trichoderma* aff. *crassum* G1C and *T. rifaii* CT5.

Treatment	Leaf spot diameter (mm)	Incidence (%)	Infected leaves (%)	Defoliation (%)	Pseudopilei/leaf	Mortality (%)
CT5	5.7 ± 6.1[a]	66.7 ± 4.8[a]	28.9 ± 2.4[a]	36.7 ± 0.8[a]	2.7 ± 5.6[a]	0[a]
G1C	7.3 ± 6.5[a]	80.0 ± 4.1[a]	36.7 ± 2.4[a]	3.4 ± 0.2[a]	5.4 ± 6.9[a]	0[a]
Control	11.2 ± 5.7[b]	93.3 ± 2.5[b]	45.6 ± 2.1[b]	50 ± 1.0[b]	4.6 ± 4.1[a]	13.3 ± 3.5[b]

Values represent average ± standard deviation. Superscript letters indicate treatments that are significantly different ($P < 0.05$).

observed (Batista et al. 2022). Antagonistic mechanisms against fungal pathogens described for isolates related to *T. koningii* involve the production of chitinolytic enzymes and endochitinases (Barbosa et al. 2001, Metcalf and Wilson 2001).

In contrast to *Trichoderma* spp., the other hypocrealean isolates tested (*Nectria pseudotrichia*, *Purpureocillium* aff. *lilacinum*, and *Sarocladium* aff. *kiliense*) did not show a significant growth inhibition against *Mycena citricolor*. *Purpureocillium lilacinum*, mostly known as an entomopathogen, did not significantly reduce the growth of *M. citricolor* even though antibiosis was evident from the inhibition halo. Despite the limited information on the secondary metabolite diversity produced by *P. lilacinum*, some isolates are reported to have antifungal properties, similar to other entomopathogenic species of *Beauveria* or *Lecanicillium* (Jaber and Ownley 2018). Studies suggest that its antifungal activity could be due to leucinostatin, which has been shown to reduce the growth of *Botrytis cinerea* (Liu et al. 2020). Even less is known about the antifungal properties of *Nectria pseudotrichia* and *Sarocladium kiliense*. The first is a species mostly known as a tropical saprophyte and endophyte (Hirooka et al. 2012), and a few studies show its potential to produce novel and biologically active secondary metabolites (Ariefta et al. 2017, Cota et al. 2018). In our study, *N. pseudotrichia* GUHN1 reduced the growth of *M. citricolor* by ca. 30%, with an evident inhibition halo and diffusing pigment, suggesting antibiosis as its mechanism. *Sarocladium* is a poorly studied genus with only 10 known species, being *S. oryzae* an important pathogen in rice, and the rest possibly endophytes or mycoparasites (Herrera et al. 2013, Giraldo et al. 2015, Potshangbam et al. 2017). Comby et al. (2017) tested *S. kiliense* against *Fusarium graminearum* in wheat spikelets achieving 80% of protection against the disease. In our study, no significant antagonistic activity against *M. citricolor* was produced by *S. kiliense* CT25 isolate; only overlapping colonies were observed.

Xylariales isolates GU11N and GU14T, did not produce an antagonistic effect >30%, even though *Daldinia* spp. have been reported to produce useful antifungal compounds such as polyketides (antifungal), naphtols (nematicide), and phytotoxins (Stadler 2011, Helaly et al. 2018). Sakayaroj et al. (2008) found an endophytic *Xylaria* sp. capable of producing strong antifungal metabolites, such as sordaricin and sordarin. Xylarinic acids from *Xylaria polymorpha* displayed antifungal activity against *Alternaria*, *Aspergillus*, *Fusarium*, *Magnaporthe*, and *Pythium* spp. (Jang et al. 2007). Other genera in the Xylariales, such as *Biscogniauxia* and *Hypoxylon*, synthetize volatile organic compounds with antifungal activity and nodulisporic acids, which have insecticidal properties (Bills et al. 2012, Silva-Hughes et al. 2015). Therefore, Xylariales remains an interesting group to study in the search of secondary metabolites useful in biological control.

Our results support other studies describing growth promotion properties of *Trichoderma* spp. (De Souza et al. 2008, Bae et al. 2009, Tchameni et al. 2011, Alfiky and Weisskopf 2021). The results from our study suggest that plantlets with endophytes had the best performance, i.e. increased plant size. Both groups of plants inoculated with *Trichoderma* were at least twice as tall as the plants without endophytes. Further experiments could elucidate the growth promotion regulators derived from the endophytic fungi or triggered by the infection. For example, the best recognized regulators involved in plant growth promotion are cytokinins, indole acetic acid, and gibberellins (Deng and Cao 2017, Rana et al. 2020). Among known genera of endophytic fungi producing indole acetic acid include *Aspergillus*, *Chaetomium*, *Didymella*, *Curvularia*, *Exophiala*, *Fusarium*, *Paecilomyces*, *Penicillium*, *Phoma*, and *Xylogone* (Priyadharsini and Muthukumar 2017, Turbat et al. 2020). Other fungi, including *Trichoderma* spp., secrete phytohormonal analogues and secondary metabolites, coordinately working to increase growth regulation, in addition to enhancing plant growth through accessing more nutrients by the root system (Mona et al. 2017).

Mycena citricolor infection process starts in the leaves, therefore, the ability of the endophyte to reduce pathogen inoculum (i.e. pseudopilei) is an important characteristic of a biocontrol agent (Busby et al. 2016). The initial amount of inoculum is correlated with pathogen dispersal and disease incidence (Granados-Montero et al. 2020). Notoriously, in our study, the number of *M. citricolor* pseudopilei after four weeks was higher in the treatment with *Trichoderma* aff. *crassum* (G1C); the control and *T. rifaii* (CT5) were statistically lower. Several factors could be affecting those results. First, G1C may not colonize the leaves of the plantlets, as shown by our results. Therefore, if the mechanism of antagonism is only direct parasitism (as suggested by hyphal coiling), then, endophytic hyphae will not be able to reach leaf tissues and hence no antagonistic effect would be observed. Conversely, G1C or the responses of its interaction with the plant and the pathogen could be affecting the regulation of fungal fruit body formation, including secretion of proteins, secondary metabolites, or volatile organic compounds, that then stimulate pseudopilei production (Baars et al. 2020, Wu et al. 2020). On the other hand, CT5 was able to colonize all plantlet tissues, resulting in overall better antagonism (i.e. reduced incidence and severity) and inoculum reduction. However, plantlets with G1C had much less defoliation than with CT5, suggesting that a foliar application of G1C could provide additional benefits (Pertot et al. 2017, El-Deen et al. 2019, Wu et al. 2020); previous studies show that some *Trichoderma* spp. can also penetrate leaves and become endophytic (Bailey et al. 2009).

As has been done with probiotics in humans, manipulation or management of a crop's phytobiome to include plant beneficial microorganisms, will play a key role in revolutionizing modern agriculture (Hyde et al. 2019, Tilocca et al. 2020). Therefore, the ability of a pathogen antagonist to become systemic, which is the colonization of all (or the majority) the

internal plant tissues (e.g. roots, stem, and leaves; Beans 2017), is a desirable trait in an effective and long-term biological control strategy (Argov and Gazit 2008, Leach et al. 2017). In our study, we confirmed the endophytic colonization of *in vitro* coffee plantlets by *Trichoderma rifaii* (CT5) and *T.* aff. *crassum* (G1C). This ability has been shown previously for other *Trichoderma* species obtained from wild *Theobroma* spp., which can become systemic in cacao plants and provide benefits such as mycoparasitism, growth promotion, and drought resistance (Bailey et al. 2008, De Souza et al. 2008, Bae et al. 2009). Another example is *Trichoderma koningiopsis*, which was isolated from wild *Hevea* trees in the Amazon basin and is an effective and systemic endophyte that controls the pathogen *Corynespora cassiicola* (Pujade-Renaud et al. 2019). In *Coffea arabica*, the colonization of roots, stems, and leaves by the entomopathogenic fungus *Beauveria bassiana* has been confirmed in seedlings (Posada and Vega 2006). Additional studies on the route of endophytic colonization and the factors that increase the success of introducing a beneficial phytobiome into plants are still necessary in a strategy to transition from conventional to environment friendly and sustainable agriculture.

In conclusion, *in vitro* and *in planta* screening of wild Rubiaceae endophytes allowed us to select two *Trichoderma* isolates (CT5 and G1C) suitable for plant growth promotion and antagonism against *Mycena citricolor*. The two isolates were also able to colonize internal plantlet tissues, increase plant size, reduce incidence and severity, and prevent plant mortality. However, *Trichoderma rifaii* CT5 would be the primary choice with regards to growth promotion, disease management, fungicide tolerance, and non-toxic bioremoval/bioremediation capacity. Future studies may include evaluating if the endophytes remain in the plants after months or years. We hope that the groundwork presented in this study will foster further studies to elucidate and develop products and solutions based on beneficial microbial consortia.

Acknowledgments

We thank P. Juárez for help in plant identification and H. Castillo for assistance in the field. We also acknowledge J. Porras and other people at Coopetarrazú for their support for this project. E.E.-L. thanks NDSU Graduate Center for Writers for manuscript revision.

Supplementary data

Supplementary data is available at *JAMBIO Journal* online.

Conflict of interest

The authors declare no conflict of interests.

Funding

Funding for this study was provided by CeNAT-CONARE Scholarships (2017–2019); the University of Costa Rica projects B7176 and C0514; FEES-CONARE project 809-B9-662; and US National Science Foundation grant DEB-1638976.

Author contributions

Efraín Escudero-Leyva (Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing), María del Milagro Granados-Montero (Methodology, Writing – original draft), Cristofer Orozco-Ortiz (Investigation, Writing – original draft), Emmanuel Araya-Valverde (Investigation, Methodology, Resources, Writing – original draft), Eduardo Alvarado-Picado (Investigation, Resources, Writing – original draft), José Miguel Chaves-Fallas (Formal analysis, Methodology, Writing – original draft), Laura Aldrich-Wolfe (Formal analysis, Methodology, Writing – review & editing) and Priscila Chaverri (Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing).

Data availability

All newly generated sequences have been deposited in GenBank and supplementary data is publicly available at GitHub https://github.com/EEscuderoL/Endophytes-of-wild-Rubiaceae-as-a-source-of-antagonistic-fungi-against-the-American-Leaf-Spot-of-co.

References

Abarenkov K, Nilsson RH, Larsson KH *et al*. The UNITE database for molecular identification of fungi—recent updates and future perspectives. *New Phytol* 2010;**186**:281–5. https://doi.org/10.1111/j.1469-8137.2009.03160.x.

Aberer AJ, Kobert K, Stamatakis A. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol Biol Evol* 2014;**31**:2553–6. https://doi.org/10.1093/molbev/msu236.

Abo-Elyousr KAM, Abdel-Hafez SII, Abdel-Rahim IR. Isolation of *Trichoderma* and evaluation of their antagonistic potential against *Alternaria porri*. *J Phytopathol* 2014;**162**:567–74. https://doi.org/10.1111/jph.12228.

Aldrich-Wolfe L, Black KL, Hartmann EDL *et al*. Taxonomic shifts in arbuscular mycorrhizal fungal communities with shade and soil nitrogen across conventionally managed and organic coffee agroecosystems. *Mycorrhiza* 2020;**30**:513–27. https://doi.org/10.1007/s00572-020-00967-7.

Alfiky A, Weisskopf L. Deciphering *Trichoderma*–plant–pathogen interactions for better development of biocontrol applications. *J Fungi* 2021;**7**:1–18. https://doi.org/10.3390/jof7010061.

Allinne C, Cerda R, Savary S *et al*. Multiple-disease system in coffee: from crop loss assessment to sustainable management. *Annu Rev Phytopathol* 2018;**56**:611–35. https://doi.org/10.1146/annurev-phyto-080417-050117.

Argov Y, Gazit Y. Biological control of the Mediterranean fruit fly in Israel: introduction and establishment of natural enemies. *Biol Control* 2008;**46**:502–7. https://doi.org/10.1016/j.biocontrol.2008.04.021.

Ariefta NR, Kristiana P, Nurjanto HH *et al*. Nectrianolins A, B, and C, new metabolites produced by endophytic fungus *Nectria pseudotrichia* 120–1NP. *Tetrahedron Lett* 2017;**58**:4082–6. https://doi.org/10.1016/j.tetlet.2017.09.032.

Avelino J, Cristancho M, Georgiou S *et al*. The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Sec* 2015;**7**:303–21. https://doi.org/10.1007/s12571-015-0446-9.

Baars JJP, Scholtmeijer K, Sonnenberg ASM *et al*. Critical factors involved in primordia building in *Agaricus bisporus*: a review. *Molecules* 2020;**25**. https://doi.org/10.3390/molecules25132984.

Bae H, Sicher RC, Kim MS *et al*. The beneficial endophyte *Tricho-*

derma hamatum isolate DIS 219b promotes growth and delays the onset of the drought response in *Theobroma cacao*. *J Exp Bot* 2009;**60**:3279–95. https://doi.org/10.1093/jxb/erp165.

Bailey B, Bae H, Crozier J *et al*. Antibiosis, mycoparasitism, and colonization success for endophytic *Trichoderma* isolates with biological control potential in *Theobroma cacao*. *Biol Control* 2008;**46**:24–35. https://doi.org/10.1016/j.biocontrol.2008.01.003.

Bailey BA, Strem MD, Wood D. *Trichoderma* species form endophytic associations within *Theobroma cacao* trichomes. *Mycol Res* 2009;**113**:1365–76. https://doi.org/10.1016/j.mycres.2009.09.004.

Baker BP, Green TA, Loker AJ. Biological control and integrated pest management in organic and conventional systems. *Biol Control* 2020;**140**:104095. https://doi.org/10.1016/j.biocontrol.2019.104095.

Barbosa MAG, Rehn KG, Menezes M *et al*. Antagonism of *Trichoderma* species on *Cladosporium herbarum* and their enzymatic characterization. *Braz J Microbiol* 2001;**32**:98–104. https://doi.org/10.1590/S1517-83822001000200005.

Batista KOM, Silva DV, Nascimento VL *et al*. Effects of *Trichoderma strigosellum* in *Eucalyptus urophylla* development and leaf-cutting ant behavior. *J Fungi* 2022;**8**:15. https://doi.org/10.3390/jof8010015.

Beans C. Probing the phytobiome to advance agriculture. *Proc Natl Acad Sci USA* 2017;**114**:8900–2. https://doi.org/10.1073/pnas.1710176114.

Bills GF, Gonzá Lez-Mené Ndez V, Martín JS *et al*. *Hypoxylon pulicicidum* sp. nov. (Ascomycota, Xylariales), a Pantropical insecticide-producing endophyte. *PLoS One* 2012;**7**:e46687. https://doi.org/10.1371/journal.pone.0046687.

Borbón O. *Consideraciones sobre la problemática del ojo de gallo (Mycena citricolor) en Costa Rica*. Colegio De Ingenieros Agrónomos, Editores, "Memoria" XI Congreso Agronómico /IV Congreso Nacional De Fitopatología Colegio De Ingenieros Agrónomos, CRI, 1999, 21–5.

Bravo-Monroy L, Potts SG, Tzanopoulos J. Drivers influencing farmer decisions for adopting organic or conventional coffee management practices. *Food Policy* 2016;**58**:49–61. https://doi.org/10.1016/j.foodpol.2015.11.003.

Busby PE, Ridout M, Newcombe G. Fungal endophytes: modifiers of plant disease 2016, , **90**, 645–655

Bussey RO, Kaur A, Todd DA *et al*. Comparison of the chemistry and diversity of endophytes isolated from wild-harvested and greenhouse-cultivated yerba mansa (*Anemopsis californica*). *Phytochem Lett* 2015;**11**:202–8. https://doi.org/10.1016/j.phytol.2014.12.013.

Cai F, Druzhinina IS. In honor of John Bissett: authoritative guidelines on molecular identification of *Trichoderma*. *Fungal Divers* 2021;**107**:1–69. https://doi.org/10.1007/s13225-020-00464-4.

Carbone I, Kohn LM. A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia* 1999;**91**:553–6. https://doi.org/10.1080/00275514.1999.12061051.

Carvajal F. Ojo de gallo (*Omphalia flavida*): segunda parte, la forma perfecta del hongo "*Omphalia flavida*" creciendo libremente en la naturaleza como parásito sobre plantas vivas, en el distrito cafetalero del Zapote (Prov. de San José, CR). *Revista Del Instituto De Defensa Del Café* 1939;**52**:550–65.

Chaverri P, Branco-Rocha F, Jaklitsch W *et al*. Systematics of the *Trichoderma harzianum* species complex and the re-identification of commercial biocontrol strains. *Mycologia* 2015;**107**:558–90. https://doi.org/10.3852/14-147.

Comby M, Gacoin M, Robineau M *et al*. Screening of wheat endophytes as biological control agents against *Fusarium* head blight using two different *in vitro* tests. *Microbiol Res* 2017;**202**:11–20. https://doi.org/10.1016/j.micres.2017.04.014.

Cota BB, Tunes LG, Maia DNB *et al*. Leishmanicidal compounds of *Nectria pseudotrichia*, an endophytic fungus isolated from the plant *Caesalpinia echinata* (Brazilwood). *Mem Inst Oswaldo Cruz* 2018;**113**:102–10. https://doi.org/10.1590/0074-02760170217.

Crespo RJ, Douthat T, Pringle C. Stream friendly coffee: evaluating the impact of coffee farming on high-elevation streams of the Tarrazú coffee region of Costa Rica. *Hydrobiologia* 2020;**847**:1903–23. https://doi.org/10.1007/s10750-020-04221-1.

Cristóbal-Martínez AL, Yáñez-Morales MJ, Solano-Vidal R *et al*. Diversity of *Colletotrichum* species in coffee (*Coffea arabica*) plantations in Mexico. *Eur J Plant Pathol* 2017;**147**:605–14. https://doi.org/10.1007/s10658-016-1029-0.

da Silva LAO, Terrasan CRF, Carmona EC. Purification and characterization of xylanases from *Trichoderma inhamatum*. *Electron J Biotechnol* 2015;**18**:307–13. https://doi.org/10.1016/j.ejbt.2015.06.001.

De Souza JT, Bailey BA, Pomella AWV *et al*. Colonization of cacao seedlings by *Trichoderma stromaticum*, a mycoparasite of the witches' broom pathogen, and its influence on plant growth and resistance. *Biol Control* 2008;**46**:36–45. https://doi.org/10.1016/j.biocontrol.2008.01.010.

Degenkolb T, Fognielsen K, Dieckmann R *et al*. Peptaibol, secondary-metabolite, and hydrophobin pattern of commercial biocontrol agents formulated with species of the *Trichoderma harzianum* complex. *Chem Biodivers* 2015;**12**:662–84. https://doi.org/10.1002/cbdv.201400300.

del Carmen H, Rodríguez M, Evans HC *et al*. New species and records of *Trichoderma* isolated as mycoparasites and endophytes from cultivated and wild coffee in Africa. *Sci Rep* 2021;**11**:5671. https://doi.org/10.1038/s41598-021-84111-1.

Deng Z, Cao L. Fungal endophytes and their interactions with plants in phytoremediation: a review. *Chemosphere* 2017;**168**:1100–6. https://doi.org/10.1016/j.chemosphere.2016.10.097.

Dunlap C. Beneficial microbes and the phytobiome. *Plant Gene* 2017;**11**:1. https://doi.org/10.1016/j.plgene.2017.06.004.

Eakin H, Tucker CM, Castellanos E *et al*. Adaptation in a multi-stressor environment: perceptions and responses to climatic and economic risks by coffee growers in Mesoamerica. *Environ Dev Sustain* 2014;**16**:123–39. https://doi.org/10.1007/s10668-013-9466-9.

Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf* 2004;**5**:1–19. https://doi.org/10.1186/1471-2105-5-113.

El-Deen AN, Alghamdi A, Al-Barty A *et al*. Entomopathogenic bacteria, *Xenorhabdus*: an alternative biocontrol agent for integrated management of root-knot nematode on grapevine. *J Pure Appl Microbiol* 2019;**13**:1499–508. https://doi.org/10.22207/JPAM.13.3.21.

Escudero-Leyva E, Alfaro-Vargas P, Muñoz-Arrieta R *et al*. Tolerance and biological removal of fungicides by *Trichoderma* species isolated from the endosphere of wild Rubiaceae plants. *Front Agron* 2022;**3**:1–14. https://doi.org/10.3389/fagro.2021.772170.

Evans H, Holmes K, Thomas S. Endophytes and mycoparasites associated with an indigenous forest tree, *Theobroma gileri*, in Ecuador and a preliminary assessment of their potential as biocontrol agents of cocoa diseases. *Mycol Progress* 2003;**2**:149–60. https://doi.org/10.1007/s11557-006-0053-4.

Fuhrimann S, Winkler MS, Staudacher P *et al*. Exposure to pesticides and health effects on farm owners and workers from conventional and organic agricultural farms in Costa Rica: protocol for a cross-sectional study. *JMIR Res Protoc* 2018;**8**:e10914. https://doi.org/10.2196/10914.

Gazis R, Chaverri P. Wild trees in the Amazon basin harbor a great diversity of beneficial endosymbiotic fungi: is this evidence of protective mutualism? *Fungal Ecol* 2015;**17**:18–29. https://doi.org/10.1016/j.funeco.2015.04.001.

Gazis R, Rehner S, Chaverri P. Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Mol Ecol* 2011;**20**:3001–13. https://doi.org/10.1111/j.1365-294X.2011.05110.x.

Giraldo A, Gené J, Sutton DA *et al*. Phylogeny of *Sarocladium* (Hypocreales). *Pers Int Mycol J* 2015;**34**:10–24. https://doi.org/10.3767/003158515X685364.

Granados-Montero M, Avelino J, Arauz-Cavallini F *et al*. Leaf litter and *Mycena citricolor* inoculum on the American leaf spot epidemic. *Agron Mesoam* 2020;**31**:77–94. https://doi.org/10.15517/am.v31i1.36614.

Greenfield M, Pareja R, Ortiz V *et al*. A novel method to scale up fungal endophyte isolations. *Biocontrol Sci Technol* 2015;**25**:1208–12.
Guo Y, Ghirardo A, Weber B *et al*. *Trichoderma* species differ in their volatile profiles and in antagonism toward ectomycorrhiza *Laccaria bicolor*. *Front Microbiol* 2019;**10**:1–15. https://doi.org/10.3389/fmicb.2019.00891.
Harman GE, Howell CR, Viterbo A *et al*. *Trichoderma* species—opportunistic, avirulent plant symbionts. *Nat Rev Microbiol* 2004;**2**:43–56. https://doi.org/10.1038/nrmicro797.
Hassan A, Pariatamby A, Ossai IC *et al*. Synergistic association of endophytic fungi enhances tolerance, growth, and heavy metal uptake of *Alocasia calidora* in landfill contaminated soil. *Appl Soil Ecol* 2022;**170**:104307. https://doi.org/10.1016/j.apsoil.2021.104307.
Helaly SE, Thongbai B, Stadler M. Diversity of biologically active secondary metabolites from endophytic and saprotrophic fungi of the ascomycete order Xylariales. *Nat Prod Rep* 2018;**35**:992–1014. https://doi.org/10.1039/C8NP00010G.
Hermosa R, Viterbo A, Chet I *et al*. Plant-beneficial effects of *Trichoderma* and of its genes. *Microbiology* 2012;**158**:17–25. https://doi.org/10.1099/mic.0.052274-0.
Herrera CS, Rossman AY, Samuels GJ *et al*. Revision of the genus *Corallomycetella* with *Corallonectria* gen. nov. for *C. jatrophae* (Nectriaceae, Hypocreales). *Mycosytema* 2013;**32**:518–44.
Hewedy OA, Lateif A, K. S *et al*. Phylogenetic diversity of *Trichoderma* strains and their antagonistic potential against soil-borne pathogens under stress conditions. *Biology* 2020;**9**:1–21. https://doi.org/10.3390/biology9080189.
Hirooka Y, Rossman AY, Samuels GJ *et al*. A monograph of *Allantonectria*, *Nectria*, and *Pleonectria* (Nectriaceae, Hypocreales, Ascomycota) and their pycnidial, sporodochial, and synnematous anamorphs. *Stud Mycol* 2012;**71**:1–210. https://doi.org/10.3114/sim0001.
Höbinger T, Schindler S, Seaman BS *et al*. Impact of oil palm plantations on the structure of the agroforestry mosaic of La Gamba, southern Costa Rica: potential implications for biodiversity. *Agroforest Syst* 2012;**85**:367–81. https://doi.org/10.1007/s10457-011-9425-0.
Hoyos-Carvajal L, Orduz S, Bissett J. Genetic and metabolic biodiversity of *Trichoderma* from Colombia and adjacent neotropic regions. *Fungal Genet Biol* 2009;**46**:615–31. https://doi.org/10.1016/j.fgb.2009.04.006.
Hubbard M, Germida JJ, Vujanovic V. Fungal endophytes enhance wheat heat and drought tolerance in terms of grain yield and second-generation seed viability. *J Appl Microbiol* 2014;**116**:109–22. https://doi.org/10.1111/jam.12311.
Huei-Mei H, Yu-Ming J, Rogers J. Molecular phylogeny of *Hypoxylon* and closely related genera. *Mycologia* 2005;**97**:844–65.
Hyde KD, Xu J, Rapior S *et al*. The amazing potential of fungi: 50 ways we can exploit fungi industrially. *Fungal Divers* 2019;**97**:1–136. https://doi.org/10.1007/s13225-019-00430-9.
Jaber LR, Ownley BH. Can we use entomopathogenic fungi as endophytes for dual biological control of insect pests and plant pathogens? *Biol Control* 2018;**116**:36–45. https://doi.org/10.1016/j.biocontrol.2017.01.018.
Jang YW, Lee IK, Kim YS *et al*. Xylarinic acids A and B, new antifungal polypropionates from the fruiting body of *Xylaria polymorpha*. *J Antibiot* 2007;**60**:696–9. https://doi.org/10.1038/ja.2007.89.
Jiménez-Rodríguez CD, Calvo-Alvarado JC, Jackson JK. Performance of two hydrological models in predicting daily flow under a climate change scenario for mountainous catchments in Northwestern Costa Rica. *Mt Res Dev* 2015;**35**:240–53. https://doi.org/10.1659/MRD-JOURNAL-D-14-00109.1.
Karimi R, Amini H, Ghabooli M. Root endophytic fungus *Piriformospora indica* and zinc attenuate cold stress in grapevine by influencing leaf phytochemicals and minerals content. *Sci Hortic* 2022;**293**:110665. https://doi.org/10.1016/j.scienta.2021.110665.
Keswani C, Mishra S, Sarma BK *et al*. Unraveling the efficient applications of secondary metabolites of various *Trichoderma* spp. *Appl Microbiol Biotechnol* 2014;**98**:533–44. https://doi.org/10.1007/s00253-013-5344-5.
Kim TY, Jang JY, Yu NH *et al*. Nematicidal activity of grammicin produced by *Xylaria grammica* KCTC 13121BP against *Meloidogyne incognita*. *Pest Manag Sci* 2018;**74**:384–91. https://doi.org/10.1002/ps.4717.
Knapp DG, Pintye A, Kovács GM. The dark side is not fastidious—dark septate endophytic fungi of native and invasive plants of semiarid sandy areas. *PLoS One* 2012;**7**:1–8. https://doi.org/10.1371/journal.pone.0032570.
Kumar R, Kumar S, Chaudhary B. Effects of *Trichoderma* species on the growth of *Fusarium verticillioides*. *Bangladesh J Bot* 2021;**50**:423–5. https://doi.org/10.3329/bjb.v50i2.54101.
Landum MC, Félix MR, Alho J *et al*. Antagonistic activity of fungi of *Olea europaea* L. against *Colletotrichum acutatum*. *Microbiol Res* 2016;**183**:100–8. https://doi.org/10.1016/j.micres.2015.12.001.
Leach JE, Triplett LR, Argueso CT *et al*. Communication in the phytobiome. *Cell* 2017;**169**:587–96. https://doi.org/10.1016/j.cell.2017.04.025.
Liu R, Khan RAA, Yue Q *et al*. Discovery of a new antifungal lipopeptaibol from *Purpureocillium lilacinum* using MALDI–TOF–IMS. *Biochem Biophys Res Commun* 2020;**527**:689–95. https://doi.org/10.1016/j.bbrc.2020.05.021.
Liu YJ, Whelsen S, Hall BD. Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit. *Mol Biol Evol* 1999;**16**:1799–808. https://doi.org/10.1093/oxfordjournals.molbev.a026092.
López-Quintero CA, Atanasova L, Franco-Molano AE *et al*. DNA barcoding survey of *Trichoderma* diversity in soil and litter of the Colombian lowland Amazonian rainforest reveals *Trichoderma strigosellum* sp. nov. and other species. *Anton Leeuw* 2013;**104**:657–74. https://doi.org/10.1007/s10482-013-9975-4.
Manns U, Wikstro N, Taylor C *et al*. Historical biogeography of the predominantly neotropical subfamily Cinchonoideae (Rubiaceae): into or out of America? *Int J Plant Sci* 2012;**173**:261–86. https://doi.org/10.1086/663971.
Mejía LC, Rojas EI, Maynard Z *et al*. Endophytic fungi as biocontrol agents of *Theobroma cacao* pathogens. *Biol Control* 2008;**46**:4–14. https://doi.org/10.1016/j.biocontrol.2008.01.012.
Mendoza JLH, Pérez MIS, Prieto JMG *et al*. Antibiosis of *Trichoderma* spp. strains native to northeastern Mexico against the pathogenic fungus *Macrophomina phaseolina*. *Braz J Microbiol* 2015;**46**:1093–101. https://doi.org/10.1590/S1517-838246420120177.
Mendoza-Mendoza A, Zaid R, Lawry R *et al*. Molecular dialogues between *Trichoderma* and roots: role of the fungal secretome. *Fungal Biol Rev* 2018;**32**:62–85. https://doi.org/10.1016/j.fbr.2017.12.001.
Metcalf DA, Wilson CR. The process of antagonism of *Sclerotium cepivorum* in white rot affected onion roots by *Trichoderma koningii*. *Plant Pathol* 2001;**50**:249–57. https://doi.org/10.1046/j.1365-3059.2001.00549.x.
Mona SA, Hashem A, Allah A *et al*. Increased resistance of drought by *Trichoderma harzianum* fungal treatment correlates with increased secondary metabolites and proline content. *J Integr Agric* 2017;**16**:1751–7. https://doi.org/10.1016/S2095-3119(17)61695-2.
Mulaw TB, Druzhinina IS, Kubicek CP *et al*. Novel endophytic *Trichoderma* spp. isolated from healthy *Coffea arabica* roots are capable of controlling coffee tracheomycosis. *Diversity* 2013;**5**:750–66. https://doi.org/10.3390/d5040750.
Murashige T, Skoog F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol Plant* 1962;**15**:473–97. https://doi.org/10.1111/j.1399-3054.1962.tb08052.x.
Pertot I, Giovannini O, Benanchi M *et al*. Combining biocontrol agents with different mechanisms of action in a strategy to control *Botrytis cinerea* on grapevine. *Crop Prot* 2017;**97**:85–93. https://doi.org/10.1016/j.cropro.2017.01.010.
Pineda JA, Piniero M, Ramírez A. Coffee production and women's empowerment in Colombia. *Hum Organ* 2019;**78**:64–74. https://doi.

org/10.17730/0018-7259.78.1.64.
Pinto LFG, Gardner T, McDermott CL *et al*. Group certification supports an increase in the diversity of sustainable agriculture network-rainforest alliance certified coffee producers in Brazil. *Ecol Econ* 2014;**107**:59–64. https://doi.org/10.1016/j.ecolecon.2014.08.006.
Posada F, Vega FE. Inoculation and colonization of coffee seedlings (*Coffea arabica* L.) with the fungal entomopathogen *Beauveria bassiana* (Ascomycota: Hypocreales). 2006, , **47**, 284–289
Potshangbam M, Indira Devi S, Sahoo D *et al*. Functional characterization of endophytic fungal community associated with *Oryza sativa* L. and *Zea mays* L. *Front Microbiol* 2017;**8**:1–15. https://doi.org/10.3389/fmicb.2017.00325.
Priyadharsini P, Muthukumar T. The root endophytic fungus *Curvularia geniculata* from *Parthenium hysterophorus* roots improves plant growth through phosphate solubilization and phytohormone production. *Fungal Ecol* 2017;**27**:69–77. https://doi.org/10.1016/j.funeco.2017.02.007.
Pujade-Renaud V, Déon M, Gazis-Seregina R *et al*. Endophytes from wild rubber trees as antagonists of the pathogen *Corynespora cassiicola*. *Phytopathology* 2019;**109**:1888–99. https://doi.org/10.1094/PHYTO-03-19-0093-R.
R Core Team. *R: a language and environment for statistical computing*.R Foundation for Statistical Computing, Vienna, Austria . 2021. https://www.r-project.org/ (2021-11-01)
Rana KL, Kour D, Kaur T *et al*. Endophytic microbes: biodiversity, plant growth-promoting mechanisms and potential applications for agricultural sustainability. *Anton Leeuw* 2020;**113**:1075–107. https://doi.org/10.1007/s10482-020-01429-y.
Rodriguez MA, Cabrera G, Gozzo FC *et al*. *Clonostachys rosea* BAFC3874 as a *Sclerotinia sclerotiorum* antagonist: mechanisms involved and potential as a biocontrol agent. *J Appl Microbiol* 2011;**110**:1177–86. https://doi.org/10.1111/j.1365-2672.2011.04970.x.
Rodríguez MCH, Evans HC, de Abreu LM *et al*. New species and records of *Trichoderma* isolated as mycoparasites and endophytes from cultivated and wild coffee in Africa. *Sci Rep* 2021;**11**:1–30. https://doi.org/10.1038/s41598-021-84111-1.
Sadeghi F, Samsampour D, Askari Seyahooei M *et al*. Fungal endophytes alleviate drought-induced oxidative stress in mandarin (*Citrus reticulata* L.): toward regulating the ascorbate–glutathione cycle. *Sci Hortic* 2020;**261**:108991. https://doi.org/10.1016/j.scienta.2019.108991.
Sakayaroj J, Pelzing M, Pongcharoen W *et al*. Metabolites from the endophytic fungus *Xylaria* sp. PSU-D14. *Phytochemistry* 2008;**69**:1900–2. https://doi.org/10.1016/j.phytochem.2008.04.003.
Shoresh M, Harman GE, Mastouri F. Induced systemic resistance and plant responses to fungal biocontrol agents. *Annu Rev Phytopathol* 2010;**48**:21–43. https://doi.org/10.1146/annurev-phyto-073009-114450.
Silva-Hughes AF, Wedge DE, Cantrell CL *et al*. Diversity and antifungal activity of the endophytic fungi associated with the native medicinal cactus *Opuntia humifusa* (Cactaceae) from the United States. *Microbiol Res* 2015;**175**:67–77. https://doi.org/10.1016/j.micres.2015.03.007.
Sridevi V, Giridhar P. Establishment of somaclonal variants of Robusta coffee with reduced levels of cafestol and kahweol. *In Vitro Cell Dev Biol Plant* 2014;**50**:618–26. https://doi.org/10.1007/s11627-014-9613-5.
Stadler M. Importance of secondary metabolites in the Xylariaceae as parameters for assessment of their taxonomy, phylogeny, and functional biodiversity. *Curr Res Environ Appl Mycol* 2011;**1**:75–133. https://doi.org/10.5943/cream/1/2/1.
Stummer BE, Zhang X, Yang H *et al*. Co-inoculation of *Trichoderma gamsii* A5MH and *Trichoderma harzianum* Tr906 in wheat suppresses in planta abundance of the crown rot pathogen *Fusarium pseudograminearum* and impacts the rhizosphere soil fungal microbiome. *Biol Control* 2022;**165**:104809. https://doi.org/10.1016/j.biocontrol.2021.104809.
Sumida CH, Daniel JFS, Araujod APCS *et al*. *Trichoderma asperellloides* antagonism to nine *Sclerotinia sclerotiorum* strains and biological control of white mold disease in soybean plants. *Biocontrol Sci Technol* 2018;**28**:142–56. https://doi.org/10.1080/09583157.2018.1430743.
Sun JZ, Liu XZ, McKenzie EHC *et al*. Fungicolous fungi: terminology, diversity, distribution, evolution, and species checklist. *Fungal Divers* 2019;**95**:337–430. https://doi.org/10.1007/s13225-019-00422-9.
Tchameni SN, Ngonkeu MEL, Begoude BAD *et al*. Effect of *Trichoderma asperellum* and arbuscular mycorrhizal fungi on cacao growth and resistance against black pod disease. *Crop Prot* 2011;**30**:1321–7. https://doi.org/10.1016/j.cropro.2011.05.003.
Tilocca B, Cao A, Migheli Q. Scent of a killer: microbial volatilome and its role in the biological control of plant pathogens. *Front Microbiol* 2020;**11**:41. https://doi.org/10.3389/fmicb.2020.00041.
Turbat A, Rakk D, Vigneshwari A *et al*. Characterization of the plant growth-promoting activities of endophytic fungi isolated from *Sophora flavescens*. *Microorganisms* 2020;**8**:1–15. https://doi.org/10.3390/microorganisms8050683.
Vinale F, Sivasithamparam K, Ghisalberti EL *et al*. *Trichoderma* secondary metabolites active on plants and fungal pathogens. *Open Mycol J* 2014;**8**:127–39. https://doi.org/10.2174/18744370014080010127.
White T, Bruns T, Lee S *et al*. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand D, Sninsky J, White T (eds.), *PCR Protocols: A guide to Methods and Applications*. San DiegoCaliforniaAcademic Press, Inc, 1990, 315–22. https://doi.org/10.1016/B978-0-12-372180-8.50042-1
Wu H, Spagnolo A, Marivingt-Mounir C *et al*. Evaluating the combined effect of a systemic phenylpyrrole fungicide and the plant growth-promoting rhizobacteria *Paraburkholderia phytofirmans* (strain PsJN::gfp2x) against the grapevine trunk pathogen *Neofusicoccum parvum*. *Pest Manag Sci* 2020;**76**:3838–48. https://doi.org/10.1002/ps.5935.
Wu T, Zhang Z, Hu C *et al*. A WD40 protein encoding gene Fvcpc2 positively regulates mushroom development and yield in *Flammulina velutipes*. *Front Microbiol* 2020;**11**:498. https://doi.org/10.3389/fmicb.2020.00498.
Xie XG, Zhao YY, Yang Y *et al*. Endophytic fungus alleviates soil sickness in peanut crops by improving the carbon metabolism and rhizosphere bacterial diversity. *Microb Ecol* 2021;**82**:49–61. https://doi.org/10.1007/s00248-020-01555-0.
Yan L, Zhu J, Zhao X *et al*. Beneficial effects of endophytic fungi colonization on plants. *Appl Microbiol Biotechnol* 2019;**103**:3327–40. https://doi.org/10.1007/s00253-019-09713-2.
Zamora K, Castro L, Wang A, Arauz LF, L Uribe. Uso potencial de lixiviados y tés de vermicompost en el control del ojo de gallo del cafeto *Mycena citricolor* 2017, , **41**, 33–55
Zhang X, Harvey PR, Stummer BE *et al*. Antibiosis functions during interactions of *Trichoderma afroharzianum* and *Trichoderma gamsii* with plant pathogenic *Rhizoctonia* and *Pythium*. *Funct Integr Genomics* 2015;**15**:599–610. https://doi.org/10.1007/s10142-015-0456-x.

4.3 Tolerancia a fungicidas.

La transición de la agricultura convencional a la orgánica a menudo se ve desafiada por la adaptación de los agentes de control biológico a entornos muy expuestos a contaminantes agroquímicos. Estudiamos especies de *Trichoderma* aisladas de tejidos de hojas vivas de plantas silvestres de Rubiacaeae para determinar su tolerancia a fungicidas y su potencial de eliminación biológica. Evaluamos la tolerancia *in vitro* a los fungicidas de cuatro aislamientos de *Trichoderma* (T1, T2, T3 y T4) colocando discos de micelio en medios sólidos suplementados con siete fungicidas sistémicos y no sistémicos diferentes. Después de una semana, la mayoría de los fungicidas no inhibieron significativamente el crecimiento de los aislados, excepto en el caso del ciproconazol, donde el único aislado capaz de crecer fue el T1; sin embargo, la morfología de la colonia se vio afectada por la presencia de fungicidas. En segundo lugar, se estableció el potencial de eliminación biológica para aislamientos seleccionados. Para este experimento, los aislamientos T1, T2 y T4 se inocularon de forma independiente en medios líquidos con los fungicidas azoxistrobina, clorotalonil, ciproconazol y trifloxistrobina. Después de 14 días de incubación, se logró una remoción de hasta 89% para clorotalonil, 46,4% para ciproconazol y 33,1% para trifloxistrobina usando biomasa viable. En el caso de la azoxistrobina, la mayor remoción (82,2%) se produjo por adsorción a la biomasa fúngica. Las pruebas ecotoxicológicas en *Daphnia magna* revelaron que T1 tiene el mayor potencial de remoción, logrando una eliminación significativa de todos los fungicidas, al mismo tiempo que detoxifica la matriz acuosa (excepto en el caso del ciproconazol). El aislado T4 también exhibió una eficiencia intermedia, mientras que el aislado T2 no pudo detoxificar la matriz en la mayoría de los casos. La eliminación y desintoxicación de ciproconazol fracasó con todos los aislamientos. Estos hallazgos sugieren que la endosfera de plantas silvestres podría ser un gremio atractivo para encontrar nuevas especies de *Trichoderma* con capacidades prometedoras de biorremediación. Además, los resultados demuestran que se debe prestar atención al combinar ciertos tipos de agroquímicos con hongos antagónicos en estrategias de Manejo Integrado de Plagas y Enfermedades o al hacer la transición a la agricultura orgánica.

Publicación:

Escudero-Leyva, E., Alfaro-Vargas, P., Muñoz-Arrieta, R., Charpentier-Alfaro, C., Granados-Montero, MdM., Valverde-Madrigal, K.S., Pérez-Villanueva, M., Méndez-Rivera, M., Rodríguez-Rodríguez, C.E., Chaverri, P. & Mora-Villalobos, J.A. **2022**. Tolerance and Biological Removal of Fungicides by Trichoderma Species Isolated from the Endosphere of Wild Rubiaceae Plants. Front. Agron. 3:772170. doi: 10.3389/fagro.2021.772170

frontiers in Agronomy

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fagro.2021.772170

Tolerance and Biological Removal of Fungicides by *Trichoderma* Species Isolated From the Endosphere of Wild Rubiaceae Plants

Efraín Escudero-Leyva[1,2†], *Pamela Alfaro-Vargas*[2†], *Rodrigo Muñoz-Arrieta*[2], *Camila Charpentier-Alfaro*[2], *María del Milagro Granados-Montero*[3], *Katherine S. Valverde-Madrigal*[2], *Marta Pérez-Villanueva*[4], *Michael Méndez-Rivera*[4], *Carlos E. Rodríguez-Rodríguez*[4], *Priscila Chaverri*[1,5*] *and J. Aníbal Mora-Villalobos*[2]

[1] Centro de Investigaciones en Productos Naturales (CIPRONA) and Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica, [2] Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), Centro Nacional de Alta Tecnología, Consejo Nacional de Rectores (CeNAT-CONARE)-CONARE, San José, Costa Rica, [3] Escuela de Agronomía and Estación Experimental Fabio Baudrit, Universidad de Costa Rica, San José, Costa Rica, [4] Centro de Investigación en Contaminación Ambiental (CICA), Universidad de Costa Rica, San José, Costa Rica, [5] Department of Plant Sciences and Landscape Architecture, University of Maryland, College Park, MD, United States

The transition from conventional to organic agriculture is often challenged by the adaptation of biological control agents to environments heavily exposed to agrochemical pollutants. We studied *Trichoderma* species isolated from living leaf tissues of wild Rubiacaeae (coffee family) plants to determine their fungicide tolerance and potential for bioremoval. First, we assessed the *in vitro* tolerance to fungicides of four *Trichoderma* isolates (*Trichoderma rifaii* T1, *T.* aff. *crassum* T2, *T.* aff. *atroviride* T3, and *T.* aff. *strigosellum* T4) by placing mycelial plugs onto solid media supplemented with seven different systemic and non-systemic fungicides. After a week, most of the fungicides did not significantly inhibit the growth of the isolates, except in the case of cyproconazole, where the only isolate able to grow was T1; however, the colony morphology was affected by the presence of fungicides. Second, biological removal potential was established for selected isolates. For this experiment, the isolates T1, T2, and T4 were independently inoculated into liquid media with the fungicides azoxystrobin, chlorothalonil, cyproconazole, and trifloxystrobin. After 14 days of incubation, a removal of up to 89% was achieved for chlorothalonil, 46.4% for cyproconazole, and 33.1% for trifloxystrobin using viable biomass. In the case of azoxystrobin, the highest removal (82.2%) occurred by adsorption to fungal biomass. Ecotoxicological tests in *Daphnia magna* revealed that T1 has the highest removal potential, achieving significant elimination of every fungicide, while simultaneously detoxifying the aqueous matrix (except in the case of cyproconazole). Isolate T4 also exhibited an intermediate efficiency, while isolate T2 was unable to detoxify the matrix in most cases. The removal and detoxification of cyproconazole failed with all the isolates. These findings suggest that endosphere of wild plants could be an attractive guild to find new *Trichoderma* species with promising bioremediation capabilities. In addition, the results demonstrate

OPEN ACCESS

Edited by:
Mauricio Schoebitz, University of Concepcion, Chile

Reviewed by:
Ernesto A. Moya-Elizondo, University of Concepcion, Chile
Angela Machuca, Universidad de Concepcion, Chile
Natalia Paulucci, National University of Rio Cuarto, Argentina

***Correspondence:**
Priscila Chaverri
pchaverr@umd.edu

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Pest Management, a section of the journal Frontiers in Agronomy

Received: *09 September 2021*
Accepted: *30 December 2021*
Published: *03 February 2022*

Citation:
Escudero-Leyva E, Alfaro-Vargas P, Muñoz-Arrieta R, Charpentier-Alfaro C, Granados-Montero MdM, Valverde-Madrigal KS, Pérez-Villanueva M, Méndez-Rivera M, Rodríguez-Rodríguez CE, Chaverri P and Mora-Villalobos JA (2022) Tolerance and Biological Removal of Fungicides by Trichoderma Species Isolated From the Endosphere of Wild Rubiaceae Plants. Front. Agron. 3:772170. doi: 10.3389/fagro.2021.772170

that attention should be placed when combining certain types of agrochemicals with antagonistic fungi in Integrated Pest and Disease Management strategies or when transitioning to organic agriculture.

Keywords: biodegradation, bioremediation, ecotoxicity, fungal endophytes, organic agriculture, *Trichoderma*

INTRODUCTION

The rapidly growing demand for food has put high pressure on the environment through the abuse of pesticides, especially fungicides (Duhamel and Vandenkoornhuyse, 2013). This affects production costs of important crops around the world and threatens the livelihoods of millions of people (Jaramillo et al., 2011; Caffarra et al., 2012; Ramirez-Villegas et al., 2012). Coffee (*Coffea arabica*) is currently under great stress because its productivity and quality are affected by pathogenic fungi, causing enormous economic losses. One of the most affected regions is Central America, where coffee farmers are mainly small and low-income (Davis et al., 2012; Paterson et al., 2014; Rikxoort et al., 2014; Avelino et al., 2015; Talhinhas et al., 2017; Verhage et al., 2017). Because the quality of coffee is correlated to adequate management against pathogens, safer environmental practices could help improve coffee cup quality (Feria-Morales, 2002). In addition, organic coffee demand has increased considerably because consumers are preferring food safety and agrochemical-free products (Lee et al., 2015). Therefore, sustainable and environmentally friendly disease management strategies, such as biological control, are urgently needed (Ayalew, 2014; Rice, 2018).

One of the great challenges in the transition from conventional to organic agriculture is the adaptation of introduced beneficial microorganisms (e.g., biological control and bioremediation agents) to a crop field with residual agrochemicals (Gharieb et al., 2004; Shen et al., 2019), or in areas where agrochemicals and antagonistic microorganisms are applied at the same time (Mishra et al., 2014; Nongmaithem, 2015; Palazzini et al., 2018). An alternative and rich guild of microorganisms adapted to these conditions are endophytic fungi, which inhabit the interior tissue of plants without causing apparent disease symptoms, facilitate plant adaptation to the environment, and may act as antagonists against phytopathogens; thus, they can be used as biological control and bioremediation agents (Harman, 2007; Howell, 2007; Stamatiu-Sánchez et al., 2014; Deng and Cao, 2017; Khan and Mohiddin, 2018; Holanda et al., 2019; Pietro-Souza et al., 2020).

Among common fungal endophytes of tropical plants and soil inhabitants, species of the genus *Trichoderma* have been used as successful biocontrollers in important crops (e.g., see review in Ghazanfar et al., 2018). In addition, some *Trichoderma* species can transform xenobiotic agents into non-toxic compounds, e.g., pesticides such as dichlorvos, cyanide pollutants, and even heavy metals (He et al., 2014). There are reports of *Trichoderma atroviride, T. harzianum* sensu lato, *Trichoderma koningii,* and *T. viride* sensu lato, capable of degrading e.g., alachlor, endosulfan, methyl-parathion, monochlorobenzene, neonicotinoids, and pentachlorophenol, among others (He et al., 2014; Vacondio et al., 2015; Cheng et al., 2017; Nunes and Malmlöf, 2018; Nykiel-Szymańska et al., 2018). Therefore, fungi that act as biocontrollers may also be capable of degrading residual pesticides in the fields or within plants, representing an important tool, not only for the transition from conventional to organic agriculture, but to complement different disease management strategies and recovery of "healthy" soil microbiomes.

Considering that: (i) tropical plants in natural forests harbor a high diversity of fungal endophytes (including *Trichoderma*) that generate a protective mutualism and can colonize internal plant tissues and soil (e.g., Evans et al., 2003; Chaverri and Samuels, 2013; Gazis and Chaverri, 2015; Pujade-Renaud et al., 2019); (ii) Neotropics is the Rubiaceae center of biodiversity (Manns et al., 2012); and (iii) *Trichoderma* possess a toolbox of enzymes and secondary metabolites which enable its species to tolerate and degrade toxic chemicals (Morales-Barrera and Cristiani-Urbina, 2008; Asemoloye et al., 2019); we hypothesize that those fungi will have the ability to tolerate and remove systemic and non-systemic fungicides, and that could be potentially applied in coffee fields. In this work, screening of *Trichoderma* spp. from the endosphere of wild Rubiaceae was performed to test their fungicide biodegradation potential. The specific aims of this work were to: (i) determine the ability of the isolated *Trichoderma* spp. to grow in the presence of fungicides in solid phase; and (ii) evaluate their capacity to remove and detoxify different fungicides in liquid medium through analytical and ecotoxicological tests. The results from this study will contribute to the design of disease management strategies that include biocontrol agents that tolerate chemical fungicides and can, at the same time, colonize soil and the plant endosphere. This is particularly critical in Latin American countries, where agricultural practices are characterized by excessive use of agrochemicals, but also an increased demand to transition into more sustainable farming, including organic.

MATERIALS AND METHODS

Identification of *Trichoderma* Isolates

Four isolates of *Trichoderma* were selected for the present study. The isolates were previously obtained from living leaf tissue of Rubiaceae plants from natural forests of Costa Rica and following protocols used in Gazis and Chaverri (2010). Fungi are preserved in 20% glycerol at −80°C and at room temperature at the Natural Products Research Centre (CIPRONA), University of Costa Rica.

Isolates were grown in Petri plates with potato-dextrose-agar (PDA) (Difco™ Laboratories, Detroit, Michigan, U.S.A.) for ca. 7 days. Then, the mycelium was harvested from the surface for DNA extraction using the kit PrepMan™ Ultra

(Applied Biosystems, Foster City, California, U.S.A.). Polymerase Chain Reactions (PCR) were prepared using primers ITS 4 and ITS 5 (Schoch et al., 2012) for internal transcribed spacers nuclear ribosomal DNA (ITS); EF728Mf and EF1R primers (Carbone and Kohn, 1999) for a partial region of translation elongation factor 1-alpha (TEF); and RPB25F and RPB2-7CR primers (Liu et al., 1999) for a partial region of RNA polymerase II subunit (RPB2). The 25-mL PCR reaction consisted of 12.5 mL GoTaq® Green Master Mix (Promega Corporation, Madison, Wisconsin, U.S.A.), 1 mL forward primer, 1 mL reverse primer, 1 mL of dimethyl sulfoxide (DMSO), 0.5 mL of bovine serum albumin (BSA), 6 mL of UltraPure™ distilled water, and 3 mL genomic DNA template (Herrera et al., 2013; Abreu et al., 2014). PCR products were sent to Macrogen (now Psomagen, Maryland, U.S.A.) for purification and sequencing. Sequence edition and alignment were done in Geneious version 10.2.3 (Kearse et al., 2012). To assign taxonomy, BLASTn algorithm was performed using NCBI GenBank database, comparing the queries to type specimens (Robbertse et al., 2017). The recommended barcode for *Trichoderma* identification is TEF (Chaverri et al., 2015). Therefore, TEF sequences with >98% identity were assigned the matching taxon names. Those with 90–97% identity were given the abbreviation "aff." (=*affinis*), which means that it has affinity with but is not identical to the assigned species. Matches using ITS and RPB2 served to support those produced for TEF. Newly generated sequences have been deposited in GenBank (**Supplementary Table 1**).

Fungicide Tolerance Test in Solid Phase

The ability to grow in solid medium containing residual fungicides was assayed for four isolates. Seven fungicides commonly used in coffee plantations were tested: azoxystrobin (systemic), chlorothalonil (non-systemic), cyproconazole (systemic), propineb (systemic), tolclofos-methyl (non-systemic), trifloxystrobin (systemic), and validamycin-A (non-systemic). (Toxicity and accumulation potential for some of those fungicides are discussed in Komárek et al., 2010 and Roman et al., 2021). *Trichoderma* isolates were cultured onto PDA, adding the dose (volume or mass) recommended for coffee: azoxystrobin (1 g/L), chlorothalonil (5 mL/L), cyproconazole (2 mL/L), propineb (5 g/L), tolclofos-methyl (0.5 g/L), trifloxystrobin (0.5 g/L), and validamycin-A (5 mL/L). The concentrations of the active ingredients according to the commercial label are listed in **Supplementary Table 2**. The fungicides were added to the PDA after media sterilization to avoid their degradation. Mycelial plugs (∼5 mm diam.) from 5-day-old cultures were placed 5 mm from the plate's edge. As fungal tolerance was observed with the full dosages, no additional concentrations or dilutions were done. Radial growth (in mm) was measured every 24 h for 5 days with a photoperiod of 12 h at 25°C ± 2. Light was used only in this experiment to induce sporulation; a large number of conidia are needed to produce the desired concentrations/dilutions. Consequently, for the next assay, only isolates T1, T2, and T4 were used because of their ability to produce abundant conidia for biomass production.

Fungicide Removal in Liquid Phase With Fungal Biomass

Biodegradation ability by three *Trichoderma* isolates (T1, T2, and T4) was evaluated against active ingredients of fungicides azoxystrobin, chlorothalonil, cyproconazole, and trifloxystrobin. An independent experiment was performed for each fungicide with three isolates separately (reaction systems, Rxn). Briefly, 1×10^6 spores/mL were inoculated into 250-mL flasks with potato-dextrose-broth (PDB) (Sigma-Aldrich, St. Louis, Missouri, U.S.A.). To obtain the desired volume of spore suspension for each flask, spore counting was performed in a Neubauer chamber in triplicate. With the average values, the number of spores per mL in the stock solution was obtained. After obtaining desired spore concentrations in the stock solution, exact amounts (in mL) were calculated to reach the final concentration of 1×10^6 spores/mL in a final volume of 50 mL. Commercial fungicides were added to the PDB media according to the recommended dose for coffee (**Supplementary Table 2**). The experiments were performed in triplicate, including: (i) abiotic controls (uninoculated) to evaluate degradation by abiotic factors; (ii) heat-killed controls (HKC; containing autoclaved biomass previously grown at identical conditions, but lacking the fungicide) to evaluate adsorption; and (iii) growth control (viable fungal biomass without fungicide) to determine the final dry weight. All cultures were grown for 14 days on a rotary shaker (250 rpm) (Thermo Fisher Scientific, Dubuque, Iowa, U.S.A.) at 21°C in the dark. These experiments were done in the dark to avoid degradation by photolysis.

To determine the fungicide concentration, 1.6 mL of the supernatant was collected from all the replicates at 2, 4, 7, 9, 11, and 14 days. Finally, on day 14, 20 mL samples were taken to determine biomass dry weight by filter-separating the mycelia and subsequently drying at 60°C (overnight) to reach a constant weight. After fungicide quantification (section Quantification of Fungicides), total removal was defined as percentual difference between final fungicide concentration (14 d in the Rxn treatment) and initial concentration (time "zero" in the abiotic control); this includes abiotic, adsorption, and biological removal. Final biodegradation was calculated by subtracting removal due to adsorption (HKC) to the total removal. In the case adsorption was negligible, the highest removal value (adsorption or abiotic control) was subtracted from total removal to estimate final biodegradation.

Analytical Procedures

Sample Processing

Samples were diluted in methanol as follows: 25 μL of chlorothalonil sample was mixed with 1,975 μL methanol (70%), 200 μL cyproconazole with 1,800 μL methanol (50%), 900 μL trifloxystrobin with 900 μL methanol (100%), and 100 μL azoxystrobin with 1,900 μL methanol (100%). All the samples were vortexed and filtered with a PTFE (polytetrafluoroethylene) membrane (0.45 μm). The samples were collected in HPLC vials and stored at 4°C.

Quantification of Fungicides

HPLC analysis was performed using an Agilent 1200 system equipped with a 1200 photodiode array detector (Agilent Technologies, Santa Clara, California, U.S.A.). A volume of 50 μL was injected into a Phenomenex Luna C18 (250 × 4.6 mm, 5 mm) column at a flow rate of 1 mL/min maintained at 35°C. The mobile phase consisted of 0.1% (v/v) formic acid in water (solvent A) and 0.1% (v/v) formic acid in acetonitrile (solvent B). The gradient elution program was as follows: 0–10 min, B increased linearly from 45 to 85%; 10–15 min, isocratic at 85%; 15–20 min, B increased linearly up to 100%, followed by reconditioning of the column. Quantification was performed at a wavelength of 254 nm using calibration curves of analytical standards of each fungicide.

Ecotoxicity Assay: Immobilization Test in *Daphnia magna*

To determine whether the fungal removal process is environmentally relevant, an ecotoxicological assay was employed to confirm detoxification, as some transformation products from pesticide degradation might present higher toxicity than the original compound (Ruíz-Hidalgo et al., 2016). An acute toxicity test was performed with the freshwater microcrustacean *D. magna* using untreated samples (initial work pesticide solutions in liquid culture media) and samples collected after 14 days for each treatment (abiotic and heat-killed controls, and viable fungus). The immobilization test methodology is described in Ramírez-Morales et al. (2021). Briefly, sets of 10 daphnid neonates (<24 h) were placed in 25 mL vials and exposed to 10 mL of each dilution of the sample (prepared in moderately hard reconstituted water); triplicate sets per dilution were incubated in the dark at 23 ± 1°C for 48 h. After incubation, immobility was determined and assumed as mortality. *Daphnia magna* was employed as a benchmark organism due to its sensitivity to xenobiotics, wide distribution, short life cycle, and ease of culturing in the laboratory, which makes it one of the most widely used bioindicators (Tkaczyk et al., 2021). Moreover, the fungicides assayed in this work exhibit high toxicity toward *D. magna*, thus making it an ideal bioindicator to evaluate the detoxification linked to fungicide removal.

Statistics

All the statistics and graphics were performed using R Core Team (2020). The effect of the fungicides over the different *Trichoderma* isolates in solid phase was analyzed using a one-way Analysis of Variance (ANOVA) for each fungicide. Biodegradation data for each fungicide was processed to generate a multiple plot of concentration vs. time and sample variability was calculated as a standard deviation (±SD). The relative sample concentration that resulted in the immobilization of 50% of daphnids (EC_{50}) for the ecotoxicity assay was estimated using the DRC package (Ritz et al., 2015).

RESULTS

Identification of *Trichoderma* Isolates

Results of BLAST searches are summarized in **Supplementary Table 1**. In summary, TEF sequences matched to *T. rifaii* (T1), *T.* aff. *crassum* (T2), *T.* aff. *atroviride* (T3), and *T.* aff. *strigosellum* (T4), with 99, 95, 90, and 96% identity, respectively. Even though ITS and RPB2 are not the recommended barcodes for *Trichoderma*, the BLAST results show support for the close affinities resulting from TEF identifications.

Fungicide Tolerance Test in Solid Phase

The solid phase tolerance test performed for each fungicide showed that azoxystrobin ($P = 0.0001154$), chlorothalonil ($P = 0.0016935$), cyproconazole ($P = 1.558 \times 10^{-19}$), and tolclofos-methyl ($P = 1.72 \times 10^{-5}$) produced differences among *Trichoderma* isolates. Meanwhile, propineb ($P = 0.704234$), validamycin-A ($P = 0.7296269$), and trifloxystrobin ($P = 0.3721598$) had similar effect over the fungi. The result obtained with cyproconazole (**Figure 1A**) was particularly noteworthy because after 5 days of evaluation it produced the highest inhibition on *Trichoderma* isolates T2, T3, and T4, being T1 the only one able to grow (37 mm) in the presence of this compound. Chlorothalonil followed in inhibition effect, allowing <30% of growth for all isolates (**Figure 1B**). Azoxystrobin, propineb, and tolclofos-methyl allowed all *Trichoderma* isolates to grow constantly in PDA supplemented with the fungicide (**Figure 2**). Finally, when exposed to validamycin A, the four isolates were able to reach a growth like that observed in the control (**Figure 1B**).

Even though the four isolates were able to tolerate and grow with most fungicides, some changes in the morphology patterns were observed after 15 days (**Figure 2**). For example, T1, which was able to grow in presence of all fungicides, produced mycelium, but not conidia in some cases. Moreover, occurrence of yellow diffusing pigment in the agar was constant in all T1 treatments. On the other hand, T2 produced concentric rings of mycelium and conidia when exposed to chlorothalonil, while T4 produced abundant green conidia in the medium containing azoxystrobin and validamycin-A, even more notably than the control. The colony surface of T3 was cottony in presence of propineb and validamycin-A, but string-like with tolclofos-methyl. Since conidia were absent from T3 even in the control, this isolate was discarded for the next assay, considering that easy and abundant sporulation for feasible propagation at a large scale is a desirable characteristic of a microorganism intended to be used as a biocontroller or in bioremediation (Hanada et al., 2010).

Fungicide Removal in Liquid Phase With Fungal Biomass

The *Trichoderma* isolates were able to grow in liquid phase with azoxystrobin, chlorothalonil, and trifloxystrobin. After 14 days of incubation, fungicide removals of up to 89% in chlorothalonil, 46% in cyproconazole, and 33% in trifloxystrobin were obtained. For azoxystrobin, higher removal occurred by adsorption onto fungal biomass (**Figure 3**).

Chlorothalonil was the fungicide more efficiently removed. Removal in abiotic (20.4%) and HKC controls (18.4–24.2%) were similar; therefore, biomass adsorption was considered negligible. With a total removal of 89.0% after 14 days, T2 showed the highest biodegradation (68.6%) (**Figure 3B1,B2**). Similar total removal values of 79.6 and 77.7%, with an estimated

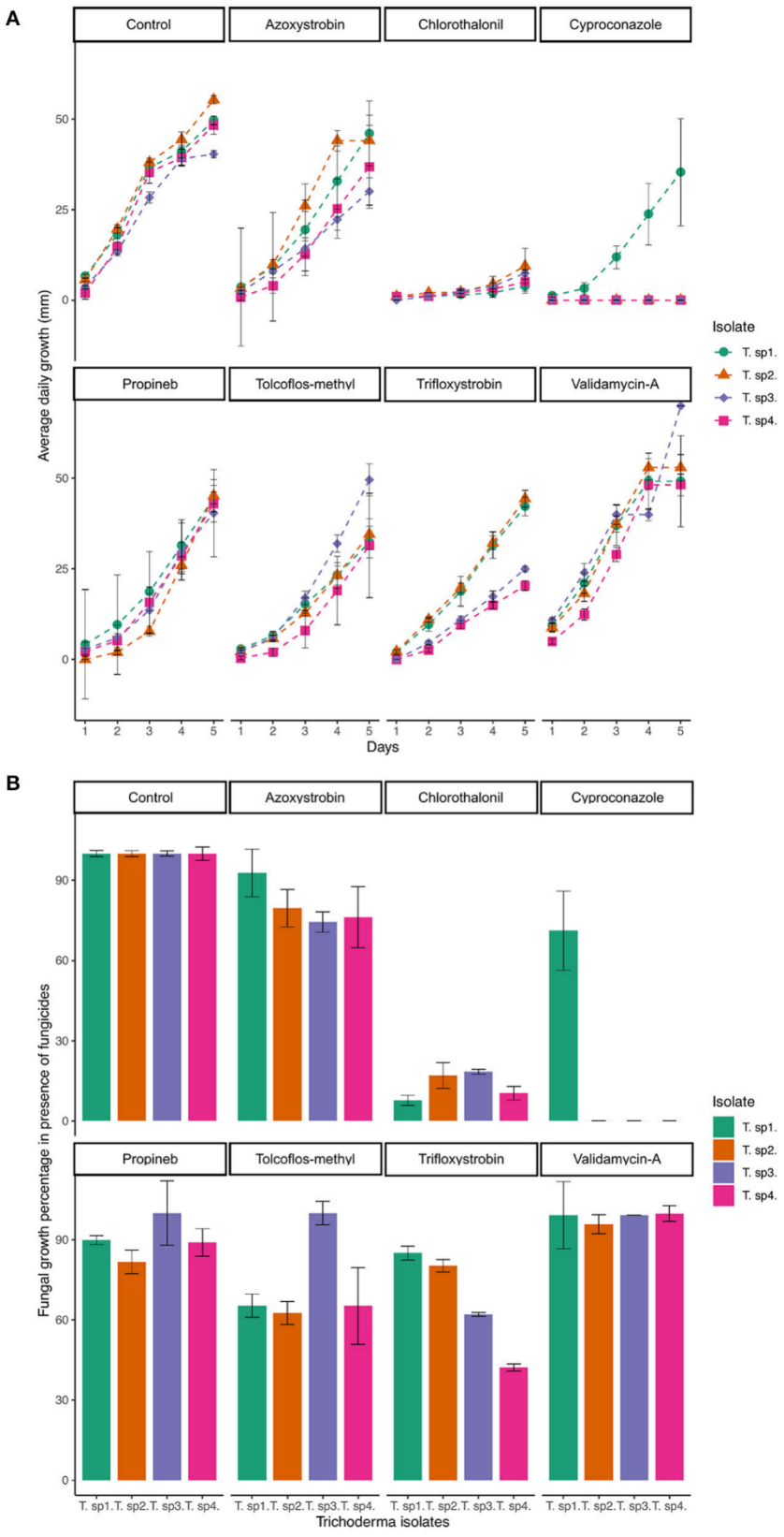


FIGURE 1 | (A) Average daily growth of *Trichoderma* replicates during 5 days. Average was calculated using five replicates per fungicide. **(B)** Fungal growth percentages with respect to the control, in the presence of different fungicides in PDA.

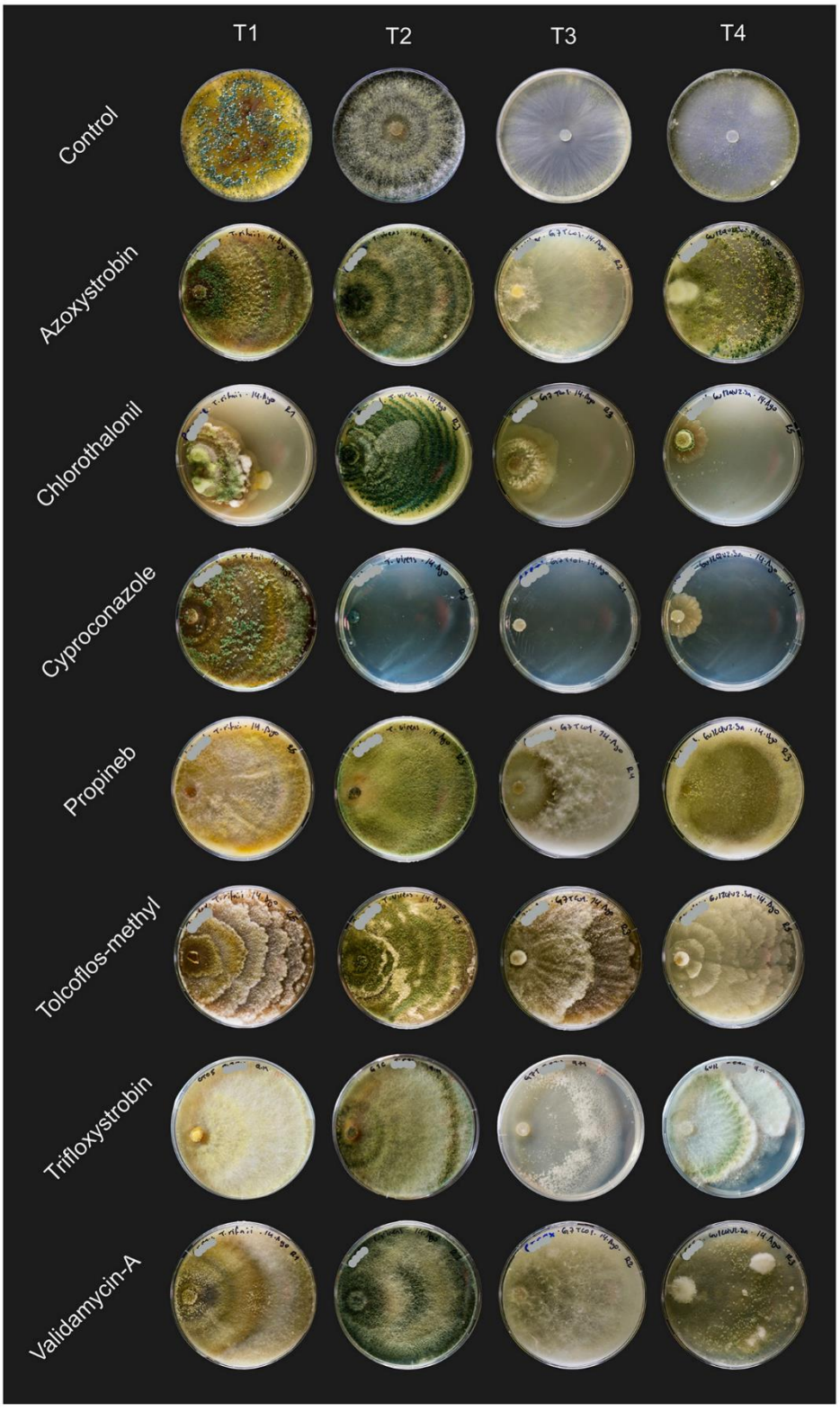


FIGURE 2 | Morphological effects of mycelial growth after 15 days. The isolates in PDA control plates are placed in the upper side of the image. Downwards, different effects on aerial mycelium can be observed in the presence of the seven fungicides.

biodegradation of 59.2 and 53.6% were achieved for T1 and T4, respectively. All tested *Trichoderma* isolates were able to grow in presence of chlorothalonil with a dry weight greater than or equal to the control (**Figure 3B3**).

Isolate T1 was able to grow in presence of cyproconazole, achieving the highest removal of 46.4% (**Figure 3C1,C2**). Considering the abiotic removal of 5.7%, final biodegradation was estimated at 40.8% (**Figure 3C2**). Lower removal was

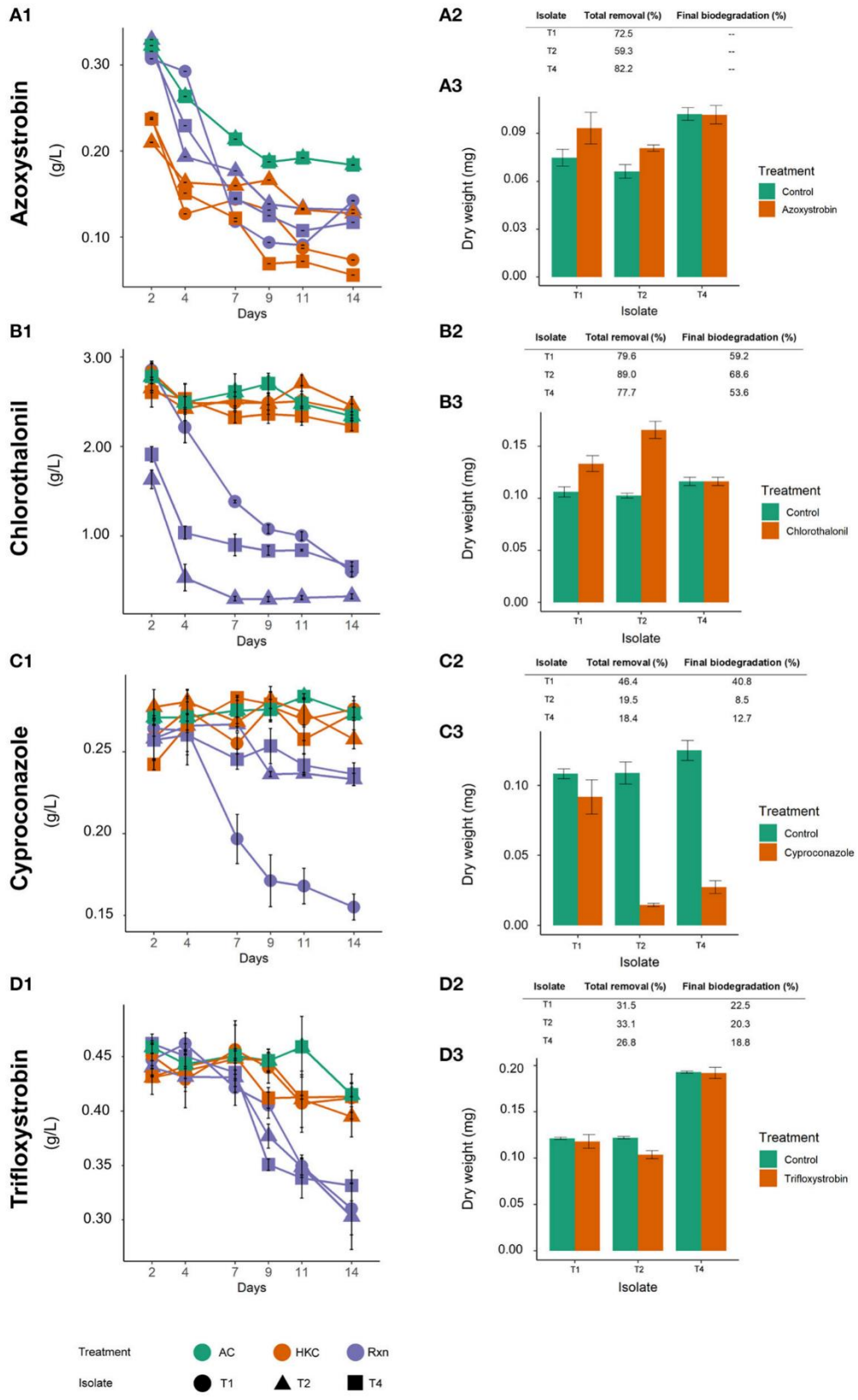


FIGURE 3 | Fungicide removal profiles and final fungal dry weight of *Trichoderma* in the presence of azoxystrobin **(A)**, chlorothalonil **(B)**, cyproconazole **(C)**, and trifloxystrobin **(D)**. AC, Abiotic control; HKC, Heat-killed control; Rxn, viable fungal biomass treatment. In **(A1–D1)**, numbers in Y axis (g/L) refer to concentration of active ingredient.

observed with T2 and T4, whose growth was severely affected by the fungicide, as the dry weight was notably lower than the respective controls (**Figure 3C3**). Final biodegradation was 8.5% for T2 and 12.7% for T4. Adsorption to the fungal biomass was only observed at a low extent for T2 (5.3%) (**Figure 3C1,C2**).

The presence of trifloxystrobin did not hinder fungal growth, as the three isolates were able to produce similar biomass amounts as the control (**Figure 3D3**). Total fungicide removal ranged from 26.8 to 33.1% for all tested strains. Considering abiotic losses of 8.4% and low adsorption (only in T2, 4.5%), final biodegradation was 22.5%, 20.3% and 18.0% for T1, T2, and T4, respectively (**Figure 3D2**).

Based on similar fungicide losses observed in abiotic and HKC (**Figure 3B1,C1,D1**), chlorothalonil, cyproconazole, and trifloxystrobin removal was mainly due to biodegradation, with a negligible role of adsorption. However, in the case of azoxystrobin, high removal occurred by adsorption (**Figure 3A1**). Biomass adsorption in the HKC ranged from 59.3 to 82.2%, while removal in Rxn treatments varied from 57.3 to 66.9% for all isolates (**Figure 3A1,A2**). This compound also exhibited the highest abiotic losses (43.3%), thus revealing estimated adsorption values of 29.2, 16.0, and 38.9% for T1, T2, and T4, respectively. Nonetheless, as suggested below by ecotoxicological assays, biodegradation, and not only adsorption, took place in the Rxn system. As in chlorothalonil and trifloxystrobin, azoxystrobin did not impair fungal growth in any isolate (**Figure 3A3**).

Ecotoxicological Evaluation of the Fungal Treatment With Immobilization Tests in *D. magna*

The results from the detoxification assays are shown in **Table 1**. Initial chlorothalonil working solution showed an EC_{50} value of 0.0050% (corresponding to 0.15 mg/L). After 14 d, both abiotic control and the HKC for every isolate revealed a slightly lower toxicity, similar to initial values (EC_{50} = 0.011–0.015%). However, toxicity clearly decreased in Rxn treatments with isolates T2 (EC_{50} = 0.16%), and especially for T1 (EC_{50} = 6.4%); detoxification was less evident for T4 (EC_{50} = 0.041%).

In cyproconazole, toxicity in the abiotic control decreased after 14 d (EC_{50} = 3.6%) with respect to the initial solution (EC_{50} = 1.09%). Additional detoxification was only slightly observed in HKC (increase in EC_{50} values from 1.1 to 2.4-fold, compared to 14-d abiotic control). However, for every isolate, toxicity was the same or higher in Rxn treatments (EC_{50} = 0.20–3.02%).

A 5-fold decrease was observed in trifloxystrobin abiotic control toxicity, compared to the initial solution (EC_{50} = 0.0067% vs. EC_{50} = 0.0013%). Nonetheless, HKC for every isolate showed similar toxicity to that achieved in the initial solution (EC_{50} = 0.0014–0.0020 %). On the contrary, Rxn treatments for T1 (EC_{50} = 0.077%) and T4 (EC_{50} = 0.058%), but not T2 (EC_{50} = 0.0024%), significantly decreased toxicity with respect to abiotic controls.

Finally, an increase in toxicity was determined for azoxystrobin in abiotic control after 14 d (EC_{50} = 0.015%), with respect to the initial solution (EC_{50} = 0.046%). Detoxification was achieved in HKC for all isolates, revealed as a two-order of magnitude increase in EC_{50} values (1.1–1.7%). Except for T2 (EC_{50} = 0.088%), whose toxicity was lower than the abiotic control, but slightly higher than that of the respective HKC, further detoxification was obtained in Rxn treatments T1 (EC_{50} = 7.4%) and T4 (EC_{50} = 5.1%).

DISCUSSION

Ideally, microorganisms with biocontrol potential should tolerate agrochemicals to successfully prevail in transitional agricultural landscapes and provide benefits to plants against diseases (Sun et al., 2019). In addition, healthy soil microbiomes and ecosystems could be enhanced if biocontrol microorganisms aided in bioremoval or biodegradation of accumulated fungicides (Komárek et al., 2010; Roman et al., 2021). The results of our study show that all *Trichoderma* isolates tested have some level of fungicide tolerance and removal or biodegradation potential. Cyproconazole and chlorothalonil were the most toxic to *Trichoderma*, and propineb and validamycin-A the least. Tolerance to pesticides by *Trichoderma* species related to the *T. harzianum* complex recovered from soil have been previously reported for compounds such as pentachlorophenol (Rigot and Matsumura, 2002) or glyphosate (Levesque and Rahe, 1992). Recently, the tolerance and conidial survival in soil of several isolates of *T. asperellum* and *T. koningiopsis* to azoxystrobin and chlorothalonil, among other eleven fungicides, was demonstrated, resulting in a suggested use of *T. asperellum* in Integrated Pest Management strategies (Widmer, 2019). Moreover, its mycoparasitism mechanism remained even in the presence of azoxystrobin (da Silva et al., 2018).

Other mycoparasitic *Trichoderma* species, e.g., *T. inhamatum*, besides its reported use as a biocontrol agent, is also capable of transforming a highly toxic chromium into a less harmful form (Morales-Barrera and Cristiani-Urbina, 2008). Efforts to provide successful fungal biocontrol species resistant to agrochemicals has led to modification attempts of already-known biocontrol fungal species using different techniques, for example, UV-induced *T. harzianum* and *T. atroviride* mutants resistant to carbendazim and tebuconazole (Hatvani et al., 2006).

Our results indicate a high activity in chlorothalonil biodegradation by the selected *Trichoderma* isolates. In previous studies, chlorothalonil biodegradation was reported by some bacteria such as *Sphingobium* spp., *Stenotrophomonas* spp., and *Ochrobactrum lupini*, among others (Motonaga et al., 1996; Van Eeden et al., 2000; Shi et al., 2011; Man-yun et al., 2014; Hu et al., 2020). This action is possible due to the bacterial enzyme glutathione S-transferase (GST) that can biotransform chlorothalonil (Kim et al., 2004). In addition, this enzyme is responsible for herbicide detoxification (Scarponi et al., 1991; Neuefeind et al., 1997; Scarponi and Del Buono, 2009; Wang et al., 2010). However, in *Trichoderma*, the chlorothalonil biodegradation mechanism through the enzymatic action of GST has not been reported. Nonetheless, this enzyme is present in *T. harzianum* (Shibu et al., 2012; Bernal-Vicente et al., 2015), which is a member of the Harzianum clade; *T. rifaii* (T1) also belongs in

TABLE 1 | Ecotoxicological evaluation of pesticides at initial working concentrations (abiotic control at T = 0 d) and after 14 d of treatment (AC, abiotic control; HKC, heat-killed control; Rxn, viable fungal biomass), according to immobilization tests in *D. magna*.

Pesticide/sample	EC_{50} (%)[a]	Confidence interval (%)[a]	
		Lower	Higher
Azoxystrobin			
AC; T = 0	0.046 (0.14 mg/L)[b]	0.038 (0.12 mg/L)	0.056 (0.23 mg/L)
AC; T = 14 d	0.015	-	-
T1 HKC; T = 14 d	1.7	1.05	2.33
T2 HKC; T = 14 d	1.1	0.11	2.09
T4 HKC; T = 14 d	1.1	0.88	1.29
T1 Rxn; T = 14 d	7.4	6.5	8.3
T2 Rxn T = 14 d	0.088	−0.09	0.10
T4 Rxn T = 14 d	5.1	4.47	5.76
Chlorothalonil			
AC; T= 0	0.0050 (0.15 mg/L)[b]	0.0042 (0.12 mg/L)	0.0056 (0.16 mg/L)
AC; T = 14 d	0.011	0.0061	0.016
T1 HKC; T = 14 d	0.014	0.012	0.016
T2 HKC; T = 14 d	0.015	0.013	0.017
T4 HKC; T = 14 d	0.014	0.012	0.016
T1 Rxn; T = 14 d	6.4	4.34	8.46
T2 Rxn T = 14 d	0.16	0.12	0.21
T4 Rxn T = 14 d	0.041	0.034	0.049
Cyproconazole			
AC; T= 0	1.09 (3.16 mg/L)[b]	0.92 (2.66 mg/L)	1.27 (3.66 mg/L)
AC; T = 14 d	3.60	3.1	4.1
T1 HKC; T = 14 d	8.67	7.43	9.92
T2 HKC; T = 14 d	5.63	-	-
T4 HKC; T = 14 d	4.02	3.38	4.66
T1 Rxn; T = 14 d	3.02	2.44	3.60
T2 Rxn T = 14 d	1.29	1.06	1.52
T4 Rxn T = 14 d	0.20	-	-
Trifloxystrobin			
AC; T= 0	0.0013 (0.0057 mg/L)[b]	0.00098 (0.0044 mg/L)	0.0015 (0.0069 mg/L)
AC; T = 14 d	0.0067	0.0055	0.0081
T1 HKC; T = 14 d	0.0015	0.00095	0.0021
T2 HKC; T = 14 d	0.0020	0.001	0.003
T4 HKC; T = 14 d	0.0014	0.0008	0.002
T1 Rxn; T = 14 d	0.077	0.062	0.09
T2 Rxn T = 14 d	0.0024	0.001	0.0068
T4 Rxn T = 14 d	0.058	0.05	0.07

[a] *EC_{50} values are reported as a percentage of the relative concentration of the sample (considering the pure sample as 100%).*
[b] *For samples corresponding to initial working concentrations, the EC_{50} values are also expressed in parenthesis as pesticide concentrations (mg/L).*

that clade (Chaverri et al., 2015). Therefore, the biodegradation mechanism used by isolate T1 could be related to the expression of GST due to their close phylogenetic affinity (Chen and Zhuang, 2017). In addition, *T. harzianum* has been reported to achieve complete biotransformation of pentachlorophenol (Vacondio et al., 2015), a halogenated substituted monocyclic aromatic fungicide which exhibits a strong similarity to chlorothalonil at the chemical structure level (Gupta, 2017). High removal of this fungicide was observed for three isolates, including biodegradation estimated between 53.6 and 68.6%. Nonetheless, important detoxification was achieved only with isolates T1 (582-fold lower toxicity) and T2 (15-fold lower toxicity); residual toxicity observed in treatment with T4 was only 4-fold lower than the control, suggesting that transformation products of similar toxicity are released by the fungus during chlorothalonil metabolism.

The removal of cyproconazole by the fungi was not that successful. The isolate T1 was able to grow in its presence, while partially degrading it. It is known that species within the *T. harzianum* complex have the capacity to metabolize and tolerate tebuconazole, another triazole (Obanda and Shupe, 2009). For example, the major pathway for tebuconazole biodegradation

consists in triazole ring cleavage and oxidation reactions (Obanda and Shupe, 2009; Wu et al., 2016, 2018). As cyproconazole has the same triazole ring, this pathway might be also used by T1.

Conversely, the same cyproconazole concentration resulted in toxic effects on T2 and T4. The growth of these two isolates was compromised, and consequently, their ability to remove the fungicide was low. In general, triazole fungicides are highly persistent in the environment (Bromilow et al., 1999; Murillo-Zamora et al., 2017), and cyproconazole exhibits one of the longest half-lives, making it very stable (Angioni et al., 2003). Biodegradation of azoles has been demonstrated for fungi such as *Trametes versicolor* and *Fomitopsis palustris* (Woo et al., 2010). Nonetheless, in one study, biodegradation of four triazoles with *T. versicolor* could not be achieved, even after fungal re-inoculation (Murillo-Zamora et al., 2017). This finding agrees with our results on low cyproconazole biodegradation efficiency observed with T2 and T4. Ecotoxicological evaluation revealed that, even though removal took place at some extent with the viable fungi, matrix detoxification did not occur. Moreover, residual toxicity was in every case higher than the control, suggesting formation of metabolites of higher toxicity than the parental compound, as it has been demonstrated in fungal degradation processes (Cruz-Morató et al., 2013; Cambronero-Heinrichs et al., 2018). Interestingly, slight but non-significant detoxification was achieved in HKC, supporting that in this case, the fungal-mediated reactions play a non-desirable role in final ecotoxicological outcome of cyproconazole transformation.

No growth inhibition was observed for strobilurins. *Trichoderma* resistance to trifloxystrobin has been previously reported (Kosanović et al., 2015); however, according to our results, efficiency ascribed to biodegradation was quite low for this fungicide. Despite low removal, detoxification was achieved by isolates T1 (11-fold lower toxicity than the abiotic control after 14 d) and T4 (9-fold lower), exhibiting a higher bioremediation potential than isolate T2, which produced a more toxic residue, at the same level as the HKC. For azoxystrobin, removal due to biodegradation could not be properly estimated, as that in the HKC surpassed the achieved with the viable biomass. At first sight, it could be inferred that all removal takes place by adsorption for all isolates, and that adsorption process is enhanced in thermally-inactivated biomass compared to viable fungi. This has been demonstrated for different kinds of biomass, including fungal, due to surface area modification (Bayramoglu and Arica, 2008; Legorreta-Castañeda et al., 2020). Nonetheless, when ecotoxicological analyses are considered, not only adsorption, but also biodegradation, is possibly taking place in Rxn systems using T1 and T4, in which significant detoxification was achieved (493 and 340-fold lower toxicity than abiotic control for T1 and T4, respectively). The extent of this detoxification was higher than that estimated in HKC by adsorption (and abiotic transformation), which accounted for 73–113-fold lower toxicity values than abiotic control, or for the Rxn isolate T2 (6-fold lower). The occurrence of adsorption does not imply a lack of degradation; on the contrary, degradation may be enhanced by physical proximity after adsorption has taken place, particularly when intracellular enzymes are involved in the process (Lucas et al., 2018; Tormo-Budowski et al., 2021). In general, information regarding strobilurins biodegradation is limited. So far, it is known that the major biological pathway for their degradation is methyl ester group hydrolysis in the toxophore (Clinton et al., 2011), which has been previously described in bacteria for azoxystrobin and trifloxystrobin (Clinton et al., 2011; Howell et al., 2014; Feng et al., 2020). Among fungi, two species of *Aphanoascus* are reported to biotransform azoxystrobin (Baćmaga et al., 2015).

Overall, the highest removal potential among isolates was observed for T1, which achieved considerable elimination of every fungicide, while simultaneously detoxifying the aqueous matrix (except in the case of cyproconazole). Isolate T4 also exhibited an intermediate efficiency, although with lower detoxification capacity than T1. Further research is needed to optimize their application in bioremediation of pesticides. On the other hand, isolate T2, despite being able to remove the fungicides at some extent, in most cases was unable to detoxify the matrix, making this fungus an unsuitable choice for biodegradation.

Lastly, three *Trichoderma* isolates were not confidently identified to species level, i.e., T2, T3, and T4, even though BLAST queries using ITS and RPB2 resulted in high similarity to known species. It has been shown that those two markers alone should not be used for identification (Chaverri et al., 2015; Cai and Druzhinina, 2021). We thus hypothesize that T2, T3, and T4 are new species because TEF sequences match to known species between 90 and 96% identity, below the recommended threshold for that marker (e.g., Chaverri et al., 2015; Cai and Druzhinina, 2021). On the other hand, we cannot confidently conclude that they are new species because additional evidence is needed (i.e., more isolates of the putative new species), as recommended for fungal taxonomy (e.g., see Seifert and Rossman, 2010).

CONCLUSIONS

In this study, endophytic *Trichoderma* isolates were able to tolerate and prevail under the presence of fungicides, generating more questions related to the mechanisms and effects of agrochemicals for the microorganisms living in agroecosystems. Assuming that these *Trichoderma* species can colonize soil and the endosphere of coffee plants (see Bailey et al., 2008; Pujade-Renaud et al., 2019 for examples on *Theobroma* and *Hevea* plants, respectively), and prevail in scenarios of conventional coffee production, then, their combined use with some low-toxicity fungicides (e.g., strobilurins and other compounds derived from microorganisms) could represent an alternative if transitioning to organic agriculture. Our *Trichoderma* isolates were not able to tolerate cyproconazole and chlorothalonil. Therefore, other compounds related to these groups of fungicides should be avoided when developing a strategy to migrate to organic or pesticide-free agriculture, as the same inhibition effect might occur on other beneficial fungi inhabiting the rhizo- or endosphere. Finally, considering its ability to remove fungicides while simultaneously decreasing their toxicity, isolate T1 seems the best candidate for fungicide bioremediation purposes. On the contrary, due to its poor removal and

detoxification capacities, precautions should be taken with the isolate T2, as its use suggests a risk of transforming some fungicides into more hazardous metabolites. The combination of pesticide tolerance, removal, and ecotoxicity assays with biocontrol or bioremediation microorganisms, and interactions between agrochemicals, should be considered in the development of sustainable agriculture strategies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PC, EE-L, and MG-M did the tolerance experiments. PA-V, EE-L, RM-A, CC-A, KV-M, MP-V, MM-R, CR-R, and JM-V conducted the bioremoval experiments. MP-V, MM-R, and CR-R conducted the ecotoxicity assays. EE-L, PA-V, and CR-R analyzed the data. EE-L, PA-V, JM-V, CR-R, and PC wrote the manuscript. PC conceived the study. JM-V and PC lead the study. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided by Universidad de Costa Rica (grants 802-B8-510, 809-C1-604, and B7176), CONARE (grant FEES-809-B9-662), CENAT-CONARE Scholarships 2018 and 2019, U.S. National Science Foundation (DEB-1638976), CENIBiot (grant FP-022-2019), and MICITT (Costa Rica) (grants FI-048B-19 and FI-197B-17).

ACKNOWLEDGMENTS

The authors thank CoopeTarrazú for providing the fungicide formulations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fagro.2021.772170/full#supplementary-material

REFERENCES

Abreu, L. M., Moreira, G. M., Ferreira, D., Rodrigues-Filho, E., and Pfenning, L. H. (2014). Diversity of Clonostachys species assessed by molecular phylogenetics and MALDI-TOF mass spectrometry. *Fungal Biol.* 118, 1004–1012. doi: 10.1016/j.funbio.2014.10.001

Angioni, A., Aguilera Del Real, A., Russo, M., Melis, M., Cabitza, F., and Cabras, P. (2003). Triazole fungicide degradation in peaches in the field and in model systems. *Food Addit. Contam.* 20, 368–374. doi: 10.1080/0265203021000060904

Asemoloye, M. D., Jonathan, S. G., and Ahmad, R. (2019). Degradation of 2, 2-Dichlorovinyl dimethyl phosphate (dichlorvos) through the rhizosphere interaction between Panicum maximum Jacq and some selected fungi. *Chemosphere* 221, 403–411. doi: 10.1016/j.chemosphere.2019.01.058

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., et al. (2015). The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Secur.* 7, 303–321. doi: 10.1007/s12571-015-0446-9

Ayalew, T. (2014). Characterization of organic coffee production, certification and marketing systems: Ethiopia as a main indicator: A Review.. *Asian J. Agric. Res.* 8, 170–180. doi: 10.3923/ajar.2014.170.180

Baćmaga, M., Kucharski, J., and Wyszkowska, J. (2015). Microbial and enzymatic activity of soil contaminated with azoxystrobin. *Environ. Monit. Assess.* 187:4827. doi: 10.1007/s10661-015-4827-5

Bailey, B. A., Bae, H., Strem, M. D., Crozier, J., Thomas, S. E., Samuels, G. J., et al. (2008). Antibiosis, mycoparasitism, and colonization success for endophytic *Trichoderma* isolates with biological control potential in *Theobroma cacao*. *Biol. Control* 46, 24–35. doi: 10.1016/j.biocontrol.2008.01.003

Bayramoglu, G., and Arica, M. Y. (2008). Removal of heavy mercury(II), cadmium(II) and zinc(II) metal ions by live and heat inactivated *Lentinus edodes* pellets. *Chem. Eng. J.* 143, 133–140. doi: 10.1016/j.cej.2008.01.002

Bernal-Vicente, A., Pascual, J. A., Tittarelli, F., Hernández, J. A., and Diaz-Vivancos, P. (2015). *Trichoderma harzianum* T-78 supplementation of compost stimulates the antioxidant defence system in melon plants. *J. Sci. Food Agric.* 95, 2208–2214. doi: 10.1002/jsfa.6936

Bromilow, R. H., Evans, A. A., and Nicholls, P. H. (1999). Factors affecting degradation rates of five triazole fungicides in two soil types: 2 field studies. *Pestic. Sci.* 55, 1135–1142. doi: 10.1002/(SICI)1096-9063(199912)55:12<1135::AID-PS73>3.0.CO;2-1

Caffarra, A., Rinaldi, M., Eccel, E., Rossi, V., and Pertot, I. (2012). Modelling the impact of climate change on the interaction between grapevine and its pests and pathogens: European grapevine moth and powdery mildew. *Agric. Ecosyst. Environ.* 148, 89–101. doi: 10.1016/j.agee.2011.11.017

Cai, F., and Druzhinina, I. S. (2021). In honor of John Bissett: authoritative guidelines on molecular identification of *Trichoderma*. *Fungal Diversity* 107, 1–69. doi: 10.1007/s13225-020-00464-4

Cambronero-Heinrichs, J. C., Masís-Mora, M., Quirós-Fournier, J. P., Lizano-Fallas, V., Mata-Araya, I., and Rodríguez-Rodríguez, C. E. (2018). Removal of herbicides in a biopurification system is not negatively affected by oxytetracycline or fungally pretreated oxytetracycline. *Chemosphere* 198, 198–203. doi: 10.1016/j.chemosphere.2018.01.122

Carbone, I., and Kohn, L. M. (1999). A method for designing primer sets for speciation studies in filamentous Ascomycetes. *Mycologia* 91, 553–556

Chaverri, P., Branco-Rocha, F., Jaklitsch, W., Gazis, R., Degenkolb, T., and Samuels, G. J. (2015). Systematics of the *Trichoderma harzianum* species complex and the re-identification of commercial biocontrol strains. *Mycologia* 107, 558–590. doi: 10.3852/14-147

Chaverri, P., and Samuels, G. J. (2013). Evolution of habitat preference and nutrition mode in a cosmopolitan fungal genus with evidence of interkingdom host jumps and major shifts in ecology. *Evolution* 67, 2823–2837. doi: 10.1111/evo.12169

Chen, K., and Zhuang, W. (2017). Discovery from a large-scaled survey of *Trichoderma* in soil of China. *Sci. Rep.* 2017, 1–37. doi: 10.1038/s41598-017-07807-3

Cheng, Z., Li, C., Kennes, C., Ye, J., Chen, D., Zhang, S., et al. (2017). Improved biodegradation potential of chlorobenzene by a mixed fungal-bacterial consortium. *Int. Biodeterior. Biodegrad.* 123, 276–285. doi: 10.1016/j.ibiod.2017.07.008

Clinton, B., Warden, A. C., Haboury, S., Easton, C. J., Kotsonis, S., Taylor, M. C., et al. (2011). Bacterial degradation of strobilurin fungicides: a role for a promiscuous methyl esterase activity of the subtilisin proteases? *Biocatal. Biotransformation* 29, 119–129. doi: 10.3109/10242422.2011.578740

Cruz-Morató, C., Jelić, A., Perez, S., Petrović, M., Barceló, D., Marco-Urrea, E., et al. (2013). Continuous treatment of clofibric acid by *Trametes versicolor*

in a fluidized bed bioreactor: Identification of transformation products and toxicity assessment. *Biochem. Eng. J.* 75, 79–85. doi: 10.1016/j.bej.2013.03.020

da Silva, M. A. F., de Moura, K. E., de Moura, K. E., Salomão, D., and Patrício, F. R. A. (2018). Compatibility of *Trichoderma* isolates with pesticides used in lettuce crop. *Summa Phytopathol.* 44, 137–142. doi: 10.1590/0100-5405/176873

Davis, A. P., Gole, T. W., Baena, S., and Moat, J. (2012). The impact of climate change on indigenous arabica coffee (*Coffea arabica*): Predicting future trends and identifying priorities. *PLoS ONE* 7:e047981. doi: 10.1371/journal.pone.0047981

Deng, Z., and Cao, L. (2017). Fungal endophytes and their interactions with plants in phytoremediation: A review. *Chemosphere* 168, 1100–1106. doi: 10.1016/j.chemosphere.2016.10.097

Duhamel, M., and Vandenkoornhuyse, P. (2013). Sustainable agriculture: Possible trajectories from mutualistic symbiosis and plant neodomestication. *Trends Plant Sci.* 18, 597–600. doi: 10.1016/j.tplants.2013.08.010

Evans, H. C., Holmes, K. A., and Thomas, S. E. (2003). Endophytes and mycoparasites associated with an indigenous forest tree, *Theobroma gileri*, in Ecuador and a preliminary assessment of their potential as biocontrol agents of cocoa diseases. *Mycol. Progress* 2, 149–160. doi: 10.1007/s11557-006-0053-4

Feng, Y., Zhang, W., Pang, S., Lin, Z., Zhang, Y., Huang, Y., et al. (2020). Kinetics and new mechanism of azoxystrobin biodegradation by an *Ochrobactrum anthropi* strain *SH*14. *Microorganisms* 8:625. doi: 10.3390/microorganisms8050625

Feria-Morales, A. M. (2002). Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. *Food Qual. Prefer.* 13, 355–367. doi: 10.1016/S0950-3293(02)00028-9

Gazis, R., and Chaverri, P. (2010). Diversity of fungal endophytes in leaves and stems of wild rubber trees (*Hevea brasiliensis*) in Peru. *Fungal Ecol.* 3, 240–254. doi: 10.1016/j.funeco.2009.12.001

Gazis, R., and Chaverri, P. (2015). Wild trees in the Amazon basin harbor a great diversity of beneficial endosymbiotic fungi: Is this evidence of protective mutualism? *Fungal Ecol.* 17, 18–29. doi: 10.1016/j.funeco.2015.04.001

Gharieb, M. M., Ali, M. I., and El-Shoura, A. A. (2004). Transformation of copper oxychloride fungicide into copper oxalate by tolerant fungi and the effect of nitrogen source on tolerance. *Biodegradation* 15, 49–57. doi: 10.1023/B:BIOD.0000009962.48723.df

Ghazanfar, M. U., Raza, M., Raza, W., and Qamar, M. I. (2018). *Trichoderma* as potential biocontrol agent, its exploitation in agriculture: a review. *Plant Prot.* 2, 109–135. Available online at: https://esciencepress.net/journals/index.php/PP/article/view/3142/1571

Gupta, P. K. (2017). "Chapter 37 - Herbicides and Fungicides," in *Reproductive and Developmental Toxicology*, ed. R. C. Gupta (New York, NY: Academic Press), 657–679. doi: 10.1016/B978-0-12-804239-7.00037-8

Hanada, R. E., Pomella, A. W. V., Costa, H. S., Bezerra, J. L., Loguercio, L. L., and Pereira, J. O. (2010). Endophytic fungal diversity in *Theobroma cacao* (cacao) and *T. grandiflorum* (cupuaçu) trees and their potential for growth promotion and biocontrol of black-pod disease. *Fungal Biol.* 114, 901–910. doi: 10.1016/j.funbio.2010.08.006

Harman, G. E. (2007). Overview of mechanisms and uses of *Trichoderma* spp.. *Phytopathology* 96, 190–194. doi: 10.1094/phyto-96-0190

Hatvani, L., Manczinger, L., Kredics, L., Szekeres, A., Antal, Z., and Vágvölgyi, C. (2006). Production of *Trichoderma* strains with pesticide-polyresistance by mutagenesis and protoplast fusion. *Int. J. Gen. Mol. Microbiol.* 89, 387–393. doi: 10.1007/s10482-005-9042-x

He, X., Wubie, A. J., Diao, Q., Li, W., Xue, F., Guo, Z., et al. (2014). Biodegradation of neonicotinoid insecticide, imidacloprid by restriction enzyme mediated integration (REMI) generated *Trichoderma* mutants. *Chemosphere* 112, 526–530. doi: 10.1016/j.chemosphere.2014.01.023

Herrera, C., Rossman, A. Y., Samuels, G. J., Lechat, C., and Chaverri, P. (2013). Revision of the genus Corallomycetella with Corallonectria gen. *nov. for C. jatrophae (Nectriaceae, Hypocreales). Mycosytema* 32, 518–544.

Holanda, F. H., Birolli, W. G., Morais, E. S., Sena, I. S., Ferreira, A. M., et al. (2019). Study of biodegradation of chloramphenicol by endophytic fungi isolated from *Bertholletia excelsa* (Brazil nuts). *Biocatal. Agric. Biotechnol.* 20:101200. doi: 10.1016/j.bcab.2019.101200

Howell, C. C., Semple, K. T., and Bending, G. D. (2014). Isolation and characterisation of azoxystrobin degrading bacteria from soil. *Chemosphere* 95, 370–378. doi: 10.1016/j.chemosphere.2013.09.048

Howell, C. R. (2007). Mechanisms employed by *Trichoderma* species in the biological control of plant diseases: the history and evolution of current concepts. *Plant Dis.* 87, 4–10. doi: 10.1094/pdis.2003.87.1.4

Hu, S., Cheng, X., Liu, G., Lu, Y., Qiao, W., Chen, K., et al. (2020). International biodeterioration & biodegradation degradation of chlorothalonil via thiolation and nitrile hydration by marine strains isolated from the surface seawater of the Northwestern Pacific. *Int. Biodeterior. Biodegradation* 154:105049. doi: 10.1016/j.ibiod.2020.105049

Jaramillo, J., Muchugu, E., Vega, F. E., Davis, A., Borgemeister, C., and Chabi-Olaye, A. (2011). Some like it hot: The influence and implications of climate change on coffee berry borer (*Hypothenemus hampei*) and coffee production in East Africa. *PLoS ONE* 6:e024528. doi: 10.1371/journal.pone.0024528

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Khan, M. R., and Mohiddin, F. A. (2018). "Chapter 13 - Trichoderma: Its Multifarious Utility in Crop Improvement," in *Crop Improvement Through Microbial Biotechnology*, eds. R. Prasad, S. S. Gill, and N. Tuteja (Elsevier), 263–291. doi: 10.1016/B978-0-444-63987-5.00013-X

Kim, Y. M., Park, K., Joo, G. J., Jeong, E. M., Kim, J. E., and Rhee, I. K. (2004). Glutathione-dependent biotransformation of the fungicide clorothalonil. *J. Agric. Food Chem.* 52, 4192–4196. doi: 10.1021/jf040047u

Komárek, M. E., Cadková Chrastný, V., Bordas, F., and Bollinger, J. C. (2010). Contamination of vineyard soils with fungicides: A review of environmental and toxicological aspects, *Environ. Int.* 36, 138–151. doi: 10.1016/j.envint.2009.10.005

Kosanović, D., Potočnik, I., Vukojević, J., Stajić, M., Rekanović, E., Stepanović, M., et al. (2015). Fungicide sensitivity of *Trichoderma* spp. from Agaricus bisporus farms in Serbia. *J. Environ. Sci. Heal. - Part B Pestic. Food Contam. Agric. Wastes* 50, 607–613. doi: 10.1080/03601234.2015.1028849

Lee, K. H., Bonn, M. A., and Cho, M. (2015). Consumer motives for purchasing organic coffee: The moderating effects of ethical concern and price sensitivity. *Int. J. Contemp. Hosp. Manag.* 27, 1157–1180. doi: 10.1108/IJCHM-02-2014-0060

Legorreta-Castañeda, A. J., Lucho-Constantino, C. A., Beltrán-Hernández, R. I., Coronel-Olivares, C., and Vázquez-Rodríguez, G. A. (2020). Biosorption of water pollutants by fungal pellets. *Water* 12:1155. doi: 10.3390/W12041155

Levesque, C. A., and Rahe, J. E. (1992). Herbicide interactions with fungal root pathogens, with special reference to glyphosate. *Annu. Rev. Phytopathol.* 30, 579–602.

Liu, Y. J., Whelen, S., and Hall, B. D. (1999). Phylogenetic relationships among Ascomycetes: evidence from an RNA polymerase II subunit. *Mol. Biol. Evol.* 1999, 1799–1808.

Lucas, D., Castellet-Rovira, F., Villagrasa, M., Badia-Fabregat, M., Barceló, D., Vicent, T., et al. (2018). The role of sorption processes in the removal of pharmaceuticals by fungal treatment of wastewater. *Sci. Total Environ.* 610–611, 1147–1153. doi: 10.1016/j.scitotenv.2017.08.118

Manns, U., Wikström, N., Taylor, C.M., and Bremer, B. (2012). Historical biogeography of the predominantly neotropical subfamily cinchonoideae (Rubiaceae): Into or Out of America? *Int. J. Plant Sci.* 173, 261–286. doi: 10.1086/663971

Man-yun, Z., Ying, T., Ye, Z. H. U., Jun, W., Yong-ming, L. U. O., Christie, P., et al. (2014). Isolation and characterization of chlorothalonil-degrading bacterial strain H4 and its potential for remediation of contaminated soil. *Pedosph. An Int. J.* 24, 799–807. doi: 10.1016/S1002-0160(14)60067-9

Mishra, G., Kumar, N., Giri, K., Pandey, S., and Kumar, R. (2014). Effect of fungicides and bioagents on number of microorganisms in

soil and yield of soybean (Glycine max). *Nusant. Biosci.* 6, 45–48. doi: 10.13057/nusbiosci/n060108

Morales-Barrera, L., and Cristiani-Urbina, E. (2008). Hexavalent chromium removal by a *Trichoderma inhamatum* fungal strain isolated from tannery effluent. *Water. Air. Soil Pollut.* 187, 327–336. doi: 10.1007/s11270-007-9520-z

Motonaga, K., Takagi, K., and Matumoto, S. (1996). Biodegradation of chlorothalonil in soil after suppression of degradation. *Biol. Fertil. Soils* 23, 340–345. doi: 10.1007/BF00335964

Murillo-Zamora, S., Castro-Gutiérrez, V., Masís-Mora, M., Lizano-Fallas, V., and Rodríguez-Rodríguez, C. E. (2017). Elimination of fungicides in biopurification systems: Effect of fungal bioaugmentation on removal performance and microbial community structure. *Chemosphere* 186, 625–634. doi: 10.1016/j.chemosphere.2017.07.162

Neuefeind, T., Reinemer, P., and Bieseler, B. (1997). Plant glutathione S-transferases and herbicide detoxification. *Biol. Chem.* 378, 199–205.

Nongmaithem, N. (2015). Compatibility of pesticides with *Trichoderma* spp. and their antagonistic potential against some pathogenic soil borne pathogens. *Indian J. Agric. Res.* 49, 193–196. doi: 10.5958/0976-058X.2015.00030.X

Nunes, S., and Malmlöf, K. (2018). Enzymatic decontamination of antimicrobials, phenols, heavy metals, pesticides, polycyclic aromatic hydrocarbons, dyes, and animal waste. *Enzym. Hum. Anim. Nutr. Princ. Perspect.* 331–359. doi: 10.1016/B978-0-12-805419-2.00017-4

Nykiel-Szymańska, J., Bernat, P., and Słaba, M. (2018). Potential of Trichoderma koningii to eliminate alachlor in the presence of copper ions. *Ecotoxicol. Environ. Saf.* 162, 1–9. doi: 10.1016/j.ecoenv.2018.06.060

Obanda, D. N., and Shupe, T. F. (2009). Biotransformation of tebuconazole by microorganisms: Evidence of a common mechanism. *Wood Fiber Sci.* 41, 157–167.

Palazzini, J. M., Torres, A. M., and Chulze, S. N. (2018). Tolerance of triazole-based fungicides by biocontrol agents used to control *Fusarium* head blight in wheat in Argentina. *Lett. Appl. Microbiol.* 66, 434–438. doi: 10.1111/lam.12869

Paterson, R. R. M., Lima, N., and Taniwaki, M. H. (2014). Coffee, mycotoxins and climate change. *Food Res. Int.* 61, 1–15. doi: 10.1016/j.foodres.2014.03.037

Pietro-Souza, W., de Campos Pereira, F., Mello, I. S., Stachack, F. F. F., and Terezo, A. J., Cunha, C. N., et al. (2020). Mercury resistance and bioremediation mediated by endophytic fungi. *Chemosphere* 240:124874. doi: 10.1016/j.chemosphere.2019.124874

Pujade-Renaud, V., Déon, M., Gazis, R., Ribeiro, S., Dessailly, F., Granet, F., et al. (2019). Endophytes from wild rubber trees as antagonists of the pathogen *Corynespora cassiicola*. *Phytopathology* 109, 1888–1899. doi: 10.1094/PHYTO-03-19-0093-R

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at: https://www.R-project.org/

Ramírez-Morales, D., Masís-Mora, M., Beita-Sandí, W., Montiel-Mora, J. R., Fernández-Fernández, E., Méndez-Rivera, M., et al. (2021). Pharmaceuticals in farms and surrounding surface water bodies: hazard and ecotoxicity in a swine production area in Costa Rica. *Chemosphere* 272:129574. doi: 10.1016/j.chemosphere.2021.129574

Ramirez-Villegas, J., Salazar, M., Jarvis, A., and Navarro-Racines, C. E. (2012). A way forward on adaptation to climate change in Colombian agriculture: Perspectives towards 2050. *Clim. Change* 115, 611–628. doi: 10.1007/s10584-012-0500-y

Rice, R. A. (2018). Coffee in the crosshairs of climate change: agroforestry as abatis. *Agroecol. Sustain. Food Syst.* 42, 1058–1076. doi: 10.1080/21683565.2018.1476428

Rigot, J., and Matsumura, F. (2002). Assessment of the rhizosphere competency and pentachlorophenol-metabolizing activity of a pesticide-degrading strain of *Trichoderma harziamum* introduced into the root zone of corn seedlings. *J. Environ. Sci. Heal. B* 37, 201–210.

Rikxoort, H., Schroth, G., Läderach, P., and Rodríguez-Sánchez, B. (2014). Carbon footprints and carbon stocks reveal climate-friendly coffee production. *Agron. Sustain. Dev. Springer-Verlag/EDP Sci.* 34, 887–897. doi: 10.1007/s13593-014-0223-8

Ritz, C., Baty, F., Streibig, J. C., and Gerhard, D. (2015). Dose-response analysis using R. *PLoS ONE* 10:e146021. doi: 10.1371/journal.pone.0146021

Robbertse, B., Strope, P., Chaverri, P., Gazis, R., and Schoch, C. L. (2017). Taxonomic accuracy for fungi in public sequence databases: applying 'One name one species' in well-defined genera with *Trichoderma/Hypocrea* as a test case. *Database* 2017:bax072. doi: 10.1093/database/bax072

Roman, D. L., Voiculescu, D. I., Filip, M., Ostafe, V., and Isvoran, A. (2021). Effects of triazole fungicides on soil microbiota and on the activities of enzymes found in soil: a review. *Agriculture* 11:893. doi: 10.3390/agriculture11090893

Ruíz-Hidalgo, K., Masís-Mora, M., Barbieri, E., Carazo-Rojas, E., and Rodríguez-Rodríguez, C. E. (2016). Ecotoxicological analysis during the removal of carbofuran in fungal bioaugmented matrices. *Chemosphere* 144, 864–871. doi: 10.1016/j.chemosphere.2015.09.056

Scarponi, L., and Del Buono, D. (2009). Festuca arundinacea, glutathione S-transferase and herbicide safeners: A preliminary case study to reduce herbicidal pollution. *J. Environ. Sci. Heal. - Part B Pestic. Food Contam. Agric. Wastes* 44, 805–809. doi: 10.1080/03601230903238400

Scarponi, L., Perucci, P., and Martinetti, L. (1991). Conjugation of 2-chloroacetanilide herbicides with glutathione: role of molecular structures and of glutathione s-transferase enzymes. *J. Agric. Food Chem.* 39, 2010–2013. doi: 10.1021/jf00011a027

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert,V., Spouge, J. L., Levesque, A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6241–6246. doi: 10.1073/pnas.1117018109

Seifert, K. A., and Rossman, A. Y. (2010). How to describe a new fungal species. *IMA Fungus* 1, 109–116. doi: 10.5598/imafungus.2010.01.02.02

Shen, F. T., Yen, J. H., Liao, C., Chen, W. C., and Chao, Y. T. (2019). Screening of rice endophytic biofertilizers with fungicide tolerance and plant growth-promoting characteristics. *Sustain.* 11:1133. doi: 10.3390/su11041133

Shi, X. Z., Guo, R. J., Takagi, K., Miao, Z. Q., and Li, S. D. (2011). Chlorothalonil degradation by *Ochrobactrum lupini* strain TP-D1 and identification of its metabolites. *World J. Microbiol. Biotechnol.* 27, 1755–1764. doi: 10.1007/s11274-010-0631-0

Shibu, M. A., Hong-Shin, L., Hsueh-Hui, Y., and Kou-Cheng, P. (2012). Trichoderma harzianum ETS 323-mediated resistance in Brassica oleracea var. capitata to Rhizoctonia solani involves the novel expression of a glutathione S-transferase and a deoxycytidine deaminase. *J. Agric. Food Chem.* 60, 10723–10732. doi: 10.1021/jf3025634

Stamatiu-Sánchez, K., Alarcón, A., Ferrera-Cerrato, R., Nava-Díaz, C., Sánchez-Escudero, J., Cruz-Sánchez, J. S., et al. (2014). Tolerancia de hongos filamentosos a endosulfán, clorpirifós y clorotalonil en condiciones *in vitro*. *Rev. Int. Contam. Ambient.* 31, 23–37. Available online at: http://www.scielo.org.mx/pdf/rica/v31n1/v31n1a2.pdf

Sun, J., Zhang, T., Li, Y., Wang, X., and Chen, J. (2019). Functional characterization of the ABC transporter TaPdr2 in the tolerance of biocontrol the fungus *Trichoderma atroviride* T23 to dichlorvos stress. *Biol. Control* 129, 102–108. doi: 10.1016/j.biocontrol.2018.10.004

Talhinhas, P., Batista, D., Diniz, I., Vieira, A., Silva, D. N., Loureiro, A., et al. (2017). The coffee leaf rust pathogen Hemileia vastatrix: one and a half centuries around the tropics. *Mol. Plant Pathol.* 18, 1039–1051. doi: 10.1111/mpp.12512

Tkaczyk, A., Bownik, A., Dudka, J., Kowal, K., and Slaska, B. (2021). *Daphnia magna* model in the toxicity assessment of pharmaceuticals: A review. *Sci. Total Environ.* 763:143038. doi: 10.1016/j.scitotenv.2020.143038

Tormo-Budowski, R., Cambronero-Heinrichs, J. C., Durán, J. E., Masís-Mora, M., Ramírez-Morales, D., Quirós-Fournier, J. P., et al. (2021). Removal of pharmaceuticals and ecotoxicological changes in wastewater using *Trametes versicolor*: A comparison of fungal stirred tank and trickle-bed bioreactors. *Chem. Eng. J.* 410:128210. doi: 10.1016/j.cej.2020.128210

Vacondio, B., Birolli, W. G., Ferreira, I. M., Seleghim, M. H. R., Gonçalves, S., Vasconcellos, S. P., et al. (2015). Biodegradation of pentachlorophenol by marine-derived fungus *Trichoderma harzianum* CBMAI 1677 isolated from ascidian Didemnun ligulum. *Biocatal. Agric. Biotechnol.* 4, 266–275. doi: 10.1016/j.bcab.2015.03.005

Van Eeden, M., Potgieter, H. C., and Van Der Walt, A. M. (2000). Microbial degration of chlorothalonil in agricultural soil: A laboratory investigation. *Environ. Toxicol.* 15, 533–539. doi: 10.1002/1522-7278(2000)15:5<533::AID-TOX25>3.0.CO;2-D

Verhage, F. Y. F., Anten, N. P. R., and Sentelhas, P. C. (2017). Carbon dioxide fertilization offsets negative impacts of climate change on Arabica coffee yield in Brazil. *Clim. Change* 144, 671–685. doi: 10.1007/s10584-017-2068-z

Wang, J., Jiang, Y., Chen, S., Xia, X., Shi, K., Zhou, Y., et al. (2010). The different responses of glutathione-dependent detoxification pathway to fungicide chlorothalonil and carbendazim in tomato leaves. *Chemosphere* 79, 958–965. doi: 10.1016/j.chemosphere.2010.02.020

Widmer, T. L. (2019). Compatibility of *Trichoderma asperellum* isolates to selected soil fungicides. *Crop Prot.* 120, 91–96. doi: 10.1016/j.cropro.2019.02.017

Woo, C., Daniels, B., Stirling, R., and Morris, P. (2010). Tebuconazole and propiconazole tolerance and possible degradation by Basidiomycetes: A wood-based bioassay. *Int. Biodeterior. Biodegrad.* 64, 403–408. doi: 10.1016/j.ibiod.2010.01.009

Wu, H., Shen, J., Jiang, X., Liu, X., Sun, X., Li, J., et al. (2018). Bioaugmentation strategy for the treatment of fungicide wastewater by two triazole-degrading strains. *Chem. Eng. J.* 349, 17–24. doi: 10.1016/j.cej.2018.05.066

Wu, H., Shen, J., Wu, R., Sun, X., Li, J., Han, W., et al. (2016). Biodegradation mechanism of 1H-1,2,4-Triazole by a newly isolated strain *Shinella* sp. NJUST26. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep29675

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

4.4 Diferencias en el volatiloma de un endófito durante su interacción con el patógeno de café *Mycena citricolor*.

Los hongos exhiben una amplia gama de gremios ecológicos, pero aquellos que viven dentro de los tejidos internos de las plantas son particularmente relevantes debido a los beneficios que brindan a sus huéspedes, como disuasión de herbivoría, protección contra enfermedades y promoción del crecimiento. Recientemente, los endófitos han ganado interés como posibles agentes de biocontrol contra patógenos de cultivos, por ejemplo, las plantas de café. Los resultados publicados de la investigación realizada en nuestro laboratorio mostraron que los hongos endófitos aislados de plantas silvestres de Rubiaceae fueron efectivos para reducir los efectos de la mancha americana de la hoja del café (*Mycena citricolor*). Uno de los aislamientos de la planta *Randia grandifolia* fue identificado como *Daldinia eschscholtzii* GU11N. Para investigar más a fondo los mecanismos de antagonismo, los efectos y la química de GU11N contra *M. citricolor*, se compararon los perfiles volátiles de GU11N con y sin la presencia del patógeno mediante ensayos sin contacto y de cultivo dual. El diseño experimental utilizado involucró el muestreo directo de tapones de agar en viales utilizando métodos de espectrometría de masas de cromatografía de gases (GC-MS) y espacio de cabeza (HS) y microextracción en fase sólida de espacio de cabeza (HS-SPME). Además, utilizamos UHPLC-HRMS/MS para identificar compuestos no volátiles de extractos orgánicos de los micelios involucrados en la interacción. Los resultados muestran que se detectaron más compuestos volátiles usando HS-SPME (39 componentes) que con la técnica HS (13 componentes), compartiendo solo 12 compuestos. Las pruebas estadísticas sugieren que *D. eschscholtzii* inhibe el crecimiento de *M. citricolor* a través de la liberación de COV que contienen una combinación de 1,8-dimetoxinaftaleno y compuestos terpénicos que afectan a *M. citricolor* pseudopilei. Se corroboraron los efectos nocivos del 1,8-dimetoxinaftaleno en una prueba *in vitro* contra estructuras reproductivas de *M. citricolor*, confirmando el daño estructural con fotografías SEM. Por otro lado, tras analizar los datos de UHPLC-HRMS/MS, se encontró un predominio de derivados de ácidos grasos entre los compuestos identificados. A pesar de los esfuerzos para anotar la mayoría de las señales utilizando GNPS y SIRIUS/Canopus, una proporción considerable (37,3 %) permaneció sin ningún tipo de

anotación. En conclusión, nuestro estudio sugiere que *Daldinia eschscholtzii* tiene potencial como agente de biocontrol contra *Mycena citricolor* y que el 1,8-dimetoxinaftaleno producido por el hongo endófito podría ser responsable del daño observado en las estructuras reproductivas del patógeno.

Publicación:

Escudero-Leyva, E., Quirós-Guerrero, L., Vásquez-Chaves, V., Pereira-Reyes, R., Chaverri, P. & Tamayo-Castillo, G. **2023**. Differential volatile organic compound expression in the interaction of *Daldinia eschscholtzii* and *Mycena citricolor*. ACS Omega, 8(34), 31373–31388. https://doi.org/10.1021/acsomega.3c03865

ACS Publications
Most Trusted. Most Cited. Most Read.
www.acs.org

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Differential Volatile Organic Compound Expression in the Interaction of *Daldinia eschscholtzii* and *Mycena citricolor*

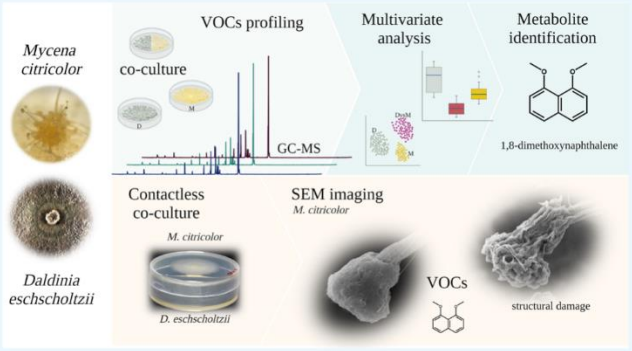
Efraín Escudero-Leyva, Luis Quirós-Guerrero, Víctor Vásquez-Chaves, Reinaldo Pereira-Reyes, Priscila Chaverri, and Giselle Tamayo-Castillo*

Cite This: *ACS Omega* 2023, 8, 31373−31388

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Fungi exhibit a wide range of ecological guilds, but those that live within the inner tissues of plants (also known as endophytes) are particularly relevant due to the benefits they sometimes provide to their hosts, such as herbivory deterrence, disease protection, and growth promotion. Recently, endophytes have gained interest as potential biocontrol agents against crop pathogens, for example, coffee plants (*Coffea arabica*). Published results from research performed in our laboratory showed that endophytic fungi isolated from wild Rubiaceae plants were effective in reducing the effects of the American leaf spot of coffee (*Mycena citricolor*). One of these isolates (GU11N) from the plant *Randia grandifolia* was identified as *Daldinia eschscholtzii* (Xylariales). Its antagonism mechanisms, effects, and chemistry against *M. citricolor* were investigated by analyzing its volatile profile alone and in the presence of the pathogen in contactless and dual culture assays. The experimental design involved direct sampling of agar plugs in vials for headspace (HS) and headspace solid-phase microextraction (HS-SPME) gas chromatography-mass spectrometry (GC-MS) analysis. Additionally, we used ultrahigh-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC-HRMS/MS) to identify nonvolatile compounds from organic extracts of the mycelia involved in the interaction. Results showed that more volatile compounds were identified using HS-SPME (39 components) than those by the HS technique (13 components), sharing only 12 compounds. Statistical tests suggest that *D. eschscholtzii* inhibited the growth of *M. citricolor* through the release of VOCs containing a combination of 1,8-dimethoxynapththalene and terpene compounds affecting *M. citricolor* pseudopilei. The damaging effects of 1,8-dimethoxynaphthalene were corroborated in an in vitro test against *M. citricolor* pseudopilei; scanning electron microscopy (SEM) photographs confirmed structural damage. After analyzing the UHPLC-HRMS/MS data, a predominance of fatty acid derivatives was found among the putatively identified compounds. However, a considerable proportion of features (37.3%) remained unannotated. In conclusion, our study suggests that *D. eschscholtzii* has potential as a biocontrol agent against *M. citricolor* and that 1,8-dimethoxynaphthalene contributes to the observed damage to the pathogen's reproductive structures.



INTRODUCTION

Endophytic fungi are keystone microorganisms in the phytobiome, with a wide variety of symbiotic associations ranging from mutualism to parasitism.[1] Almost every plant on the planet has developed a symbiotic relationship (in the broad sense) with fungal endophytes and, in some cases, their presence translates to benefits.[2,3] Nowadays, several reports highlight the role of volatile organic compounds (VOCs) as a key factor in the fungal ecological interactions with other fungi, bacteria, and plants.[4] These VOCs are known to regulate growth in spores and seeds,[5] modify signaling for herbivore deterrence,[6] pathogen inhibition,[7] among other functions. The production of VOCs by fungi is influenced by numerous factors, including environmental conditions.[8]

Many studies have reported on the fungistatic or fungicidal effects of VOCs from endophytic fungi[9] (see some examples of compounds in Figure 1). Mycofumigation activity has been reported for 5-pentyl-2-furaldehyde (Figure 1.1) from *Irpex* (=*Oxyporus*) *latemarginatus*[10] and 6-pentyl-α-pyrone (6PP) (Figure 1.2) from endophytic *Trichoderma asperellum* IsmT5.[11] The latter induced the accumulation of compounds related to plant defense and increased the production of anthocyanins and trichomes in *Arabidopsis thaliana*. *Induratia kashay* (=*Muscodor kashay*), isolated from the medicinal plant *Aegle marmelos*, has been found to inhibit fungal pathogens such as *Alternaria*

Received: June 1, 2023
Accepted: July 31, 2023
Published: August 18, 2023

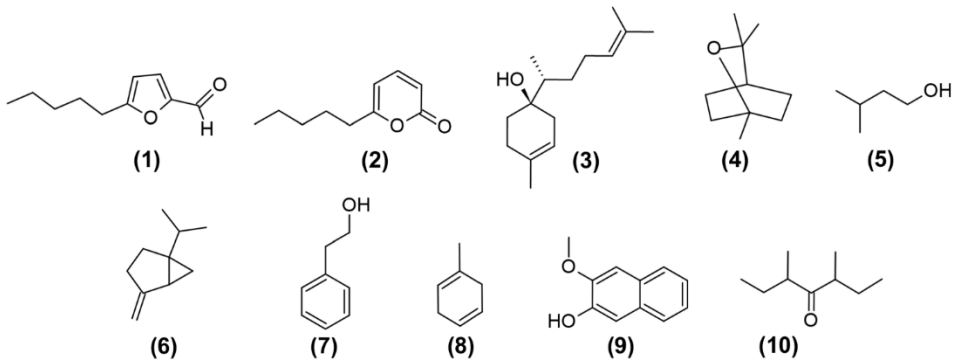


Figure 1. Examples of VOCs reported from several endophytic fungi.

alternata and *Aspergillus flavus* in contactless assays.[12] Among the 23 VOCs reported by Meshram et al.,[12] *β*-bisabolol (Figure 1.3) was identified as the major component. A *Hypoxylon* species recovered from the plant *Persea indica* was found to exhibit antagonistic activity against *Botrytis cinerea*, *Cercospora beticola*, *Phytophtora cinnamomi*, and *Sclerotinia sclerotiorum*[13] according to the authors, 1,8-cineole (Figure 1.4) showed the highest inhibitory effect. Another example is an unidentified *Diaporthe* (=*Phomopsis*) endophyte from *Odontoglossum* sp. producing abundant 3-methylbutanol (Figure 1.5) and sabinene (Figure 1.6) along with other 11 VOCs, which effectively inhibited the growth of *Aspergillus fumigatus*, *Fusarium solani*, *Geotrichum candidum*, *Pythium ultimum*, *P. cinnamomi*, *Phytophthora palmivora*, *Rhizoctonia solani*, and *S. sclerotiorum* by over 40% compared to their controls.[14]

Gas chromatography-mass spectrometry (GC-MS) and solid-phase microextraction (SPME) have facilitated the detection and discovery of VOCs.[15] The constantly increasing wealth of information in databases has helped the annotation and further evaluation of VOCs as long-distance mediators among microorganisms.[16,17] To ensure the accurate detection of VOCs during HS-SPME-GC-MS analysis, it is crucial to consider the configuration and equipment setup. A careful selection of columns and SPME fibers is necessary for detecting a wider range of compound families.[18] Polydimethylsiloxane (PDMS) is the commonly used nonpolar coating for SPME, thanks to its thermal stability and nonpolar affinity.[19] In recent studies on VOCs produced by fungi, the PDMS fiber has been the preferred option for their detection.[20−22]

There is a scarcity of accessible databases for ultrahigh-performance liquid chromatography coupled to mass spectrometry (UHPLC-MS/MS) analysis, which makes identifying compounds challenging. LC-MS/MS metabolomics is subject to variation depending on the equipment and detection technology used, which makes the standardization of data difficult.[23] Nevertheless, LC-MS/MS can provide information on most molecules present in an extract, and while public databases for metabolomic studies are expanding, most of them are currently focused on human metabolites, with databases for metabolites from bacteria and plants being developed at a later stage.[24] This highlights the importance of the microbeMASST initiative, soon to be published by the Dorrestein group, where thousands of annotated and raw MS/MS experiments from microorganisms will be available to the scientific community (https://masst.ucsd.edu/microbemasst/).

The American leaf spot of coffee (*Mycena citricolor*) is currently restricted to Central and South America, causing severe damage to coffee plants, including fruit and leaf drop contributing to economic losses.[25] Despite the lack of information about all of the factors favoring the spread of this pathogen, intense and prolonged rainy seasons have been identified as one of the main ones. For instance, in Costa Rica, between 2010 and 2011, the country estimated a loss of up to 60 million USD.[26−28] Globally, pathogens are responsible for around 40 billion USD in crop losses.[29,30] The use of agrochemicals continues to be the most common strategy for pathogen control despite the well-known negative impacts of their use.[31] Endophytic fungi are emerging as suitable alternatives to hazardous agrochemicals.[32] Moreover, endophytic fungi have shown the ability to promote growth and provide other beneficial effects to their hosts.[33] For example, in a recent study, several species of endophytes from wild Rubiaceae plants from Costa Rica were effective against *M. citricolor* in vitro and in planta, reducing disease incidence and severity and promoting plant growth.[34] However, that study hinted only at potential mechanisms of fungicidal or fungistatic effects. Understanding those mechanisms against pathogens, with specific interest in the release of metabolites possessing protective capabilities for essential crops, could lead to the development of effective alternatives for disease control.[35]

Among the isolates studied by Escudero-Leyva et al.,[34] *Daldinia eschscholtzii* GU11N, obtained from the plant *Randia grandifolia* (Rubiaceae), had an antagonistic effect over *M. citricolor* in an in vitro dual culture experiment. In that study, antibiosis was hypothesized as the mechanism of antagonism. *Daldinia* (Hypoxylaceae, Xylariales, Sordariomycetes, Ascomycota) species are producers of VOCs with fungicidal/fungistatic effects against different pathogens.[36,37] Recent examples include inhibitory effects against pathogenic microorganisms such as *Aspergillus niger*,[38] *Colletotrichum acutatum*,[39] *P. palmivora*,[40] and even nematodes.[41] Benzeneethanol (Figure 1.7), 1-methyl-1,4-cyclohexadiene (Figure 1.8), 3-methoxy-2-naphtol (Figure 1.9), and 3,5-dimethyl-4-heptanone (Figure 1.10) have been identified as common VOCs through the HS-SPME-GC-MS analysis.

To investigate the potential role of VOCs in the antagonistic effect of the endophytic isolate *D. eschscholtzii* GU11N against the coffee pathogen *M. citricolor*, a comprehensive analysis of the volatile chemical profile of *D. eschscholtzii* GU11N was conducted in the presence and absence of the pathogen using HS and HS-SPME-GC-MS analysis. Additionally, we examined the extracted metabolites from the confrontation zone and extracts from both fungi grown independently by using ultrahigh-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC-HRMS/MS) to identify potential metabolites associated with the inhibitory

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

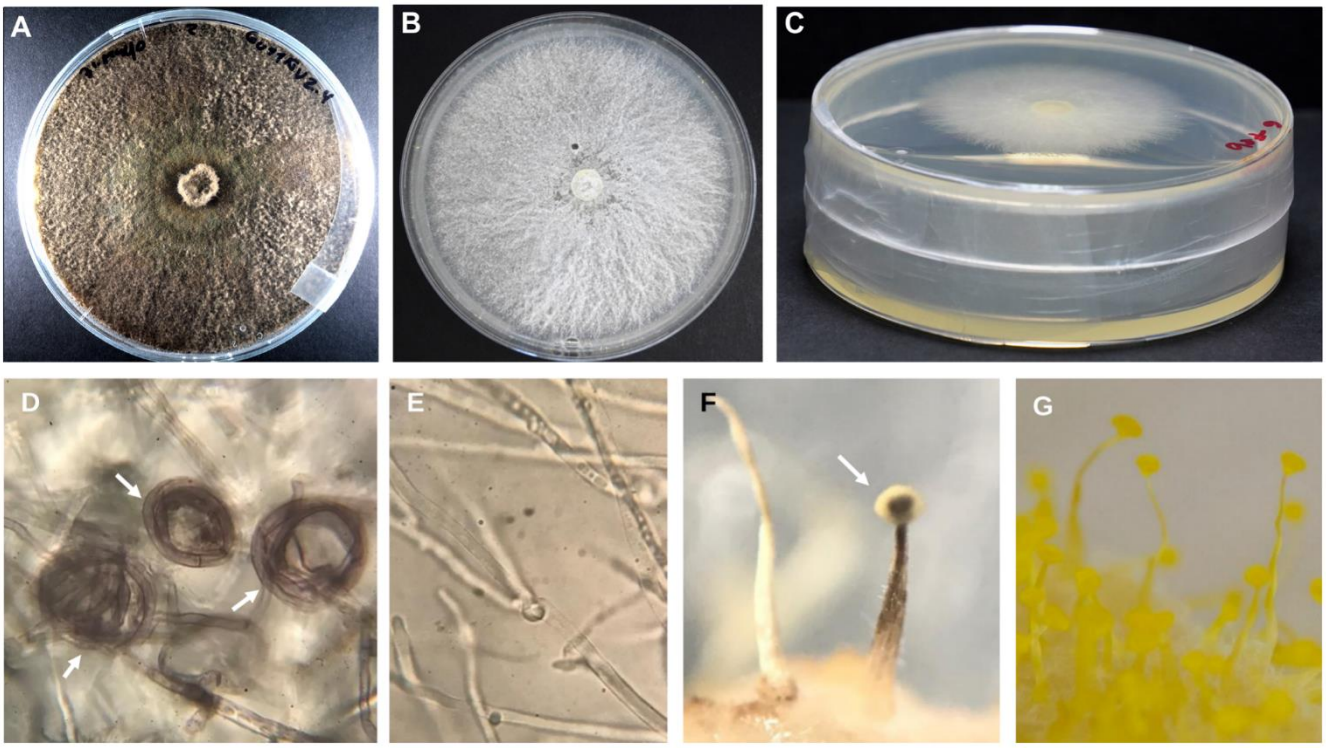


Figure 2. (A) Culture of *D. eschscholtzii* GU11N. (B) Culture of *M. citricolor*. (C) Contactless culture experiment. *M. citricolor* was placed on the top plate. (D) Hyphal damage in *M. citricolor* in the contactless culture experiment; coiled and burned-like hyphae are shown. (E) Normal hyphal growth in *M. citricolor* alone. (F) *M. citricolor* pseudopileus damage in the contactless culture experiment. The abundance of pseudopilei was reduced, and the structures presented a burned-like effect. (G) *M. citricolor* normal pseudopilei; abundant and bright yellow-colored pseudopilei growing without the presence of *D. eschscholtzii* GU11N.

effects. Our primary objectives were (i) to assess the impact of fungal volatile components from *D. eschscholtzii* GU11N on the pathogen within a closed system, (ii) to evaluate the volatile profile of *D. eschscholtzii* GU11N in the presence or absence of the pathogen, and (iii) to determine whether the inhibitory effect could also be attributed to nonvolatile components.

RESULTS AND DISCUSSION

Contactless Experiment. Previous studies have documented the bioactive effects of fungi producing VOCs in enclosed systems, specifically investigating inhibition or growth reduction of specific pathogens (mycofumigation) and even with seeds and fruits placed in proximity to the fungal isolates.[10,38,42] Our contactless experiment revealed that *D. eschscholtzii* GU11N exhibited a fungicidal rather than fungistatic effect against *M. citricolor*. Although there was no significant difference in the mycelial growth of *M. citricolor* between samples in the presence and absence of *D. eschscholtzii* (Figure S1), damaged pseudopilei or mycelial fragments failed to grow even with ample time and suitable media, indicating the potential fungicidal effect of *D. eschscholtzii*. Macroscopic damage occurred in all replicates of *M. citricolor* in the contactless experiment with *D. eschscholtzii* (Figure 2A–C). After 7 days, the damage observed in a 40× light microscope was seen as coiled and burned-like hyphae (Figure 2D). Segments displaying visible signs of damage were excised from hyphae and pseudopilei and transferred to a Petri dish containing fresh PDA media. After a five-day incubation period, no growth was observed in any of the fragments collected from the contactless experiment. Additionally, a reduction in the mycelial density and "burned" pseudopilei was observed (Figure 2F). However, *Mycena* radial growth, both in the presence and absence of *D. eschscholtzii*, exhibited no significant difference (Figure S1). Since *M. citricolor* pseudopilei infect coffee leaves, an assay involving the propagation and evaluation of *D. eschscholtzii* colonizing the leaf tissues and then evaluating its protective effect would be a promising future study. The results of such a study could provide answers to questions related to the mechanisms and responses in the plant–endophyte–pathogen system, where the effect seen under in vitro conditions could be different from that in an exposed environment, where numerous factors happen at the same time (e.g., temperature, humidity, UV radiation). In addition, testing and applying extracts (e.g., as an aerosol or spray application) of *D. eschscholtzii* GU11N's bioactive compounds may provide more evidence on mechanisms or strategies for their delivery.[43]

GC-MS Analysis and Metabolite Identifications. Fungal VOCs are a complex mixture of compounds,[44] making their identification challenging. To address this, the implementation of HS techniques associated with GC-MS is crucial for a more comprehensive analysis. Based on the observations from the contactless experiment, it is evident that an effect occurred in the volatile compartment shared by the coffee pathogen and the fungal endophyte. Direct sampling of this contactless experiment is difficult, which is why we opted to sample from a coculture experiment, segmenting the area near the confrontation zone where the two cultures almost meet. The experimental design involved the use of 10 replicates to improve the reproducibility in the biological samples. From each Petri dish, five agar plugs were sampled and placed in a vial for GC-MS analysis. Similarly, 10 replicates from *D. eschscholtzii* GU11N

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

and 10 replicates from *M. citricolor* were sampled and used as controls.

Data processing of the VOC profiling was performed in two separate ways: uploading the GC-MS raw data to the Data Analysis workflow on the GNPS platform[45] and deconvoluting using MZmine 2 (*v* 2.53).[46] The latter resulted in a more manageable feature table and therefore was used for further analyses (a feature corresponds to a chromatographic peak with a specific mass-to-charge ratio (m/z), retention time (RT), and intensity/area). It contained a total of 40 features derived from the samples of *D. eschscholtzii* GU11N agar plugs (D group) and those from the coculture experiment. The samples from *M. citricolor* plugs (M group) produced no volatiles. HS and HS-SPME experiments were processed independently, but the annotation results were combined in a final consensus annotations list (Table 1). As expected, HS-GC-MS identified a smaller number of VOCs than HS-SPME-GC-MS. Only 12 features were shared between both analysis techniques. A total of 26 features were identified exclusively using SPME. Table 1 presents the retention times, compound names, and chemical classes as well as the experimental and literature Kovats indexes (KIs) for the identified VOCs. The columns containing the up and down arrows were built upon the analysis of variance (ANOVA) analysis, indicating the relative concentration of each feature in the correspondent sampling experiment (HS or SPME). The chemical class of the features was determined using NPClassifier[47] directly from the structural annotation. The KIs shown in Table 1 are usually used for refining the annotations in untargeted metabolomics.[48] Caution is necessary when comparing the indexes due to evolving GC column materials. Nevertheless, KIs remain a powerful piece of information.[49]

The identity of the compounds was then validated with the comparison of the experimental KI and those reported in databases along with the hit rates (HRs) and similar score values, like the cosine provided by the GNPS platform. With all these factors, the identification level (from I to IV)[50] was suggested, being level I when the annotated feature was compared to a standard. Nonetheless, after an exhaustive search in databases, two features remained unknown (**33** and **35**).

It is important to address the differences obtained with the two sampling techniques, HS and HS-SPME (Table 1). The head volume in the first technique usually contains large quantities of carbon dioxide and water, with the latter coming from the sample. The presence of both components makes detection of other analytes difficult at the beginning of the chromatographic run. However, the absence of additional contaminants, such as fiber components, enables the detection of analytes in moderate quantities after the first 2 min.[15] On the other hand, the volatile profile obtained with HS-SPME depends on the polarity of the selected fiber coating. Typically, the PDMS-coated fiber is preferred for detection of low-molecular-weight compounds due to the thermal stability and nonpolar affinity.[19] Other available coatings such as carbowax or polyacrylate may allow the detection of more polar and therefore less volatile components.[15] Here, we selected a 100 μm PDMS-coated fiber when comparing its performance to that of a carbowax fiber. The chemical nature and molecular weights of the determined compounds led to the conclusion that the selection of the fiber was appropriate. Finally, although it is not surprising that with HS-SPME-GC-MS, more VOCs were detected, it is fair to say that the identified features in HS-GC-MS represent more accurately the chemical space involved in the contactless experiment.

Neither set of results (HS or HS-SPME) allowed for robust statistical analysis. Attempts to use internal standards for the quantification of VOCs and comparison of runs were unsuccessful. Consequently, we performed ANOVA using the features that were shared in both techniques for principal component analysis (PCA) and PLS-DA.

To determine if there are statistical differences between VOC levels between the D group and the coculture, an ANOVA was conducted for each sampling method (see Figure S2). In most of the detected features from both sampling methods, a consistent tendency of either increase or decrease was observed when comparing groups, as seen in features **5**, **7**, or **16.** However, features **22** and **27** exhibited contrasting prevalences depending on the specific sampling tool used. Among the listed features, namely, **19**, **21**, **25**, **26**, **37**, and **40**, these were more concentrated in the coculture when sampling with SPME (Table 1). Furthermore, VOC profiles for HS and HS-SPME revealed variations in their concentrations as stated before. The sampling method influences the levels and composition of identified VOCs as stated before. HS sampling provides a broader perspective on the volatile chemical profile, while HS-SPME analysis offers a more targeted and sensitive approach to specific compounds.

The ANOVA revealed significant differences in the VOC compositions between the D group and the coculture group. The D group showed higher levels of 4,4-dimethyl-1,3-cyclopentanedione (**7**), 1,2-dimethyl-4-oxocyclohex-2-enecarbaldehyde (**16**), α-selinene (**28**), and pogostole (**36**), all of which were statistically significant ($p < 0.05$). In contrast, the coculture group showed significantly higher levels of 9-epi-β-caryophyllene (**25**) and 1,8-dimethoxynaphthalene (**37**). These findings suggest that the interaction between *D. eschscholtzii* GU11N and the pathogen in the coculture experiment influences the production of specific VOCs, leading to distinct chemical profiles compared to *D. eschscholtzii* GU11N alone.

When using the complete set of features (40), the best 10 features detected as VIPs in the PLS-DA differ from the loadings obtained from the PCA. However, when performing the ANOVA of the data, it aligned best with the results from the PCA (Supporting Information: PCA and PLS-DA results for HS-SPME data). Therefore, variables or features identified as significant by ANOVA also contribute significantly to the separation between the groups in the PCA plot. The discrepancy between VIPs, PCA, and ANOVA can be a result of variables that may be highly correlated with each other, the presence of outliers, or the sample size. The set of biological replicates used in the experiment was 10 per group, whereas the number of variables was 40. Therefore, in this kind of experiment, where the number of identified VOCs is typically a small number, the number of samples should be increased to obtain data that can be best analyzed using multivariate approaches.[51] One way to circumvent this with the results at hand is to judiciously reduce the number of variables. Along this line, if the set of SPME variables to use ignores those variables that were absent in the HS experiments, the number of variables is reduced to 12. By performing this reduction on data, the PCA considered compound **37** as the only loading associated with the coculture samples (Figure 3A), PLS-DA scores acquired 59% cumulative variance, maintaining the D and coculture clusters, and the VIPs included 9-epi-β-caryophyllene (**25**) and 1,8-dimethoxynaphthalene (**37**), whose concentrations were higher in the coculture (Figure 3B). The last leads us to test the effect of these

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

Table 1. Volatile Organic Compound's Consensus List Identity[a]

features	RT (min)	compound name	chemical class[b]	D HS	coculture HS	D SPME	coculture SPME	KI experimental	KI Adams	KI NIST	annotation level[c]	% HR
1	5.46	2-methyl-1,3-pentanediol	fatty alcohols	nd	nd	↑	↓	930	-	959	III	77
2	5.55	2-ethyl-1-hexanol*	fatty alcohols	↑	↓	↑	↓	935	-	1030	III	83
3	6.30	1-octen-3-ol	fatty alcohols	nd	nd	=	=	977	-	981	III	83
4	6.49	1-nonen-3-ol*	fatty alcohols	↑	↓	nd	nd	987	-	1081	III	84
5	6.59	3-octanone*	oxygenated hydrocarbons	↑	↓	↑	↓	992		989	II	91
6	6.76	3-octanol*	fatty alcohols	nd	nd	↑	↓	1002	-	993	III	88
7	7.14	4,4-dimethyl-1,3-cyclopentanedione[d]	aliphatic ketone[e]	↑	↓	↑	↓	1024	-	1077	III	83
8	7.80	2,4-dimethyl-1-heptanol	fatty alcohols	nd	nd	↑	↓	1062	-	1030	III	77
9	8.91	benzeneethanol*	phenylethanoids	nd	nd	↑	↓	1129	-	1120	II	96
10	9.68	methyl 3-hydroxy-2,4,4-trimethylpentanoate	wax monoesters	nd	nd	↑	↓	1178	-	1097	IV	74
11	10.04	2-((3,3-dimethyloxiran-2-yl)methyl)-3-methylfuran	acyclic monoterpenoids	nd	nd	↑	↓	1202	-	1177	III	75
12	10.17	1-(1-cyclohexen-1-yl)-1-propanone	aliphatic ketone[e]	nd	nd	↑	↓	1210	-	1126	IV	70
13	10.25	heptyl 2,2-dimethylpropanoate	wax monoesters	nd	nd	↑	↓	1216	-	1263	IV	70
14	10.63	2,2,4,6-tetramethyl-3,5-heptanedione	fatty acids	nd	nd	↑	↓	1242		1174	III	79
15	10.83	1-acetyl-2-ethylidenecyclohexanol	aliphatic alcohol[e]	nd	nd	↑	↓	1256	-	1350	IV	70
16	11.05	1,2-dimethyl-4-oxocyclohex-2-enecarbaldehyde	aliphatic aldehyde[e]	↑	↓	↑	↓	1271	-	1285	IV	70
17	11.84	dihydro-5-pentyl-2(3*H*)-furanone	lactone	nd	nd	↑	↓	1327	-	1324	III	77
18	12.30	*α*-cubebene*	cubebane sesquiterpenoids	nd	nd	↓	↑	1361	1345	1351	II	92
19	12.35	2,2-dimethyl-1-(2-hydroxy-1-methylethyl) 2-methylpropionate	fatty acid ester[e]	nd	nd	↓	↑	1366	-	1347	IV	73
20	12.62	2-methylene-4,8,8-trimethyl-4-vinyl-bicyclo[5.2.0]nonane	caryophyllane sesquiterpenoids	nd	nd	↓	↑	1385		1407	III	84
21	12.68	*α*-copaene	copaane sesquiterpenoids	nd	nd	↓	↑	1390	1374	1375	III	87
22	12.88	*β*-elemene	elemane sesquiterpenoids	↓	↑	↑	↓	1404	1434	1390	II	92
23	13.31	*β*-caryophyllene*	caryophyllane sesquiterpenoids	nd	nd	=	=	1439	1417	1408	II	94
24	13.48	*α*-guaiene*	guaiane sesquiterpenoids	↑	↓	↑	↓	1452	1437	1439	II	92
25	13.55	9-epi-*β* -caryophyllene*	caryphyllane sesquiterpenoids	↓	↑	↓	↑	1458	1464	1461	II	96
26	14.00	5-hydroxy-2-methylchroman-4-one	chromones	nd	nd	↓	↑	1495	-	1608	III	76
27	14.16	*β*-selinene	eudesmane sesquiterpenoids	↓	↑	↑	↓	1507	1489	1492	II	90
28	14.26	*α*-selinene	eudesmane sesquiterpenoids	↑	↓	↑	↓	1516	1498	1517	II	90
29	14.36	*δ*-guaiene	guaiane sesquiterpenoids	↓	↑	↑	↓	1524	-	1526	II	90
30	14.40	hexamethylcyclohexane-1,3,5-trione	simple cyclic polyketides	nd	nd	↑	↓	1527	-	1637	IV	72
31	14.65	*β*-vetivenene	eremophilane sesquiterpenoids	nd	nd	↑	↓	1548	-	1540	IV	72
32	15.13	3,4-dimethoxyphenylacetone*	aromatic ketone[e]	nd	nd	↑	↓	1587	-	1507	IV	74
33	15.21	ni	aromatic compound[e]	nd	nd	↑	↓	1596			IV	
34	15.38	*β*-atlantol*	bisabolane sesquiterpenoids	nd	nd	↑	↓	1610	-	1548	IV	87*
35	15.45	ni	aromatic compound[e]	nd	nd	↑	↓	1616	-	-	IV	
36	16.15	pogostole	guaiane sesquiterpenoids	↑	↓	↑	↓	1679	-	1655	II	96

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

Table 1. continued

features	RT (min)	compound name	chemical class[b]	D HS	coculture HS	D SPME	coculture SPME	KI experimental	KI Adams	KI NIST	annotation level[c]	% HR
37	16.20	1,8-dimethoxynaphthalene*	naphthalenes and derivatives	↓	↑	↓	↑	1685	-	1641	I	92
38	17.00	3,5,6,7,8,8a-hexahydro-4,8a-dimethyl-6-(1-methylethenyl)-2(1H) naphthalenone	eudesmane sesquiterpenoids	nd	nd	↑	↓	1760	-	1781	III	78
39	19.23	(*E,E*)-3,7,11,15-tetramethyl-1,6,10,14-hexadecatetraen-3-ol*	acyclic monoterpenoids	nd	nd	↑	↓	1940	-	2020	III	79
40	20.04	(*E*)-15,16-dinorlabda-8(17)-12-dien-14-al	norlabdane diterpenoids	nd	nd	↓	↑	1999	-	1967	III	88

[a]The putative identification for the compounds detected using HS and HS-SPME techniques is shown. Symbology: nd (not detected); ni (not identified); ↑ and ↓ (an increment and reduction); = (no difference of expression). D: Daldinia eschscholtzii GU11N; KI: Kovats index; HR: hit rate; RT: retention time. [b]NPClassifier ontology.[47] [c]Based on Schymanski et al.[50] All features were annotated using NIST 2020 and Wiley 2014 and, if needed, deconvolution with AMDIS before contrasting with the libraries. Consensus RT and area integration were made after deconvolution using MZmine 2.53. Annotation levels are as follows: level I: standard confirmation; level II: all features having a hit rate of ≥90% and ΔRI ≤ 10 IU; level III: all features having a hit rate of >75% and a 100 ≤ ΔRI > 10 iu; and level IV, all of the rest. KI 1176. Pherobase (www.pherobase.com). [d]*Features with >0.8 cosine on a GNPS annotation exercise. [e]Manual annotations for the chemical class.

compounds against the *M. citricolor* reproductive structures, as explained further.

The composition of the chemical classes based on NPClassifier, relative to the number of features, is displayed in Table S1 and Figure S4. The two most represented classes for the D and coculture groups in HS were fatty alcohols (51.2%) and aliphatic ketones (33.9%) and caryophyllene-like sesquiterpenes (46.8%) and fatty alcohols (30.7%), respectively (Figure S5A). In the HS-SPME experiment, aliphatic ketones (16.9%), chromones (16.6%), aliphatic aldehydes (14%), and guaiane sesquiterpenes (11.9%) represented the D group, while the coculture was characterized by chromones (43.8%), naphthalenes and derivatives (31.8%), and caryophyllene-like sesquiterpenes (21.1%, Figure S4B,C).

According to the literature, VOCs previously detected for *D. eschscholtzii* GU11N include alcohols, chromones, and other aromatic compounds, dienes, ketones, aldehydes, and sesquiterpenes.[38,39,52] Among the volatile compounds that have shown statistical significance, we found α-cubebene (**18**), α-bisabolene, and guaizulene. These terpenes can cause damage to hyphal membranes and suppress conidial formation in *A. niger*.[53] Yet, the production and role of sesquiterpenes in fungi is highly variable and might include benefits for plant growth promotion (e.g., β-caryophyllene (**23**))[18] or detrimental effects like the correlation of the incidence of entomopathogenic pests and the emission of α-copaene (**21**).[54,55] The latter has been found to play a role in guiding females toward males and influencing reproductive success.[56] Most of the level-two annotated compounds were identified as sesquiterpenes. Benzochromenes are frequently reported for endophytic fungi. The chromanone 5-hydroxy-2-methylchroman-4-one (**26**) was also found in the endophytic fungus *Cryptosporiopsis sp.* and presented cytotoxic activity against leukemia cells.[57] Other chromenones displayed phytotoxic activity by affecting the mitochondrial membrane of mono- and dicotyledonous plants.[58] The isolate MFLUCC I9-0493, identified as *Daldinia eschscholtzii*, demonstrated significant inhibition of *C. acutatum* in coculture, with elemicin identified as the responsible inhibitor. Similarly, *Daldinia* cf. *concentrica* displayed inhibitory effects against 17 pathogenic fungi, leading to inhibitory ranges spanning from 30 to 100%.[39] In that study, the presence of 4-heptanone and *trans*-2-hexenal was correlated to the observed inhibition. Testing the inhibitory effect using synthetic compounds of these substances demonstrated an enhancement in inhibition performance; however, herbicidal effects were also noted.[38] *Daldinia bambusicola* effectively inhibited the pathogens *Colletotrichum lagenarium* and *P. palmivora*, while *A. alternata*, *B. cinerea*, and *Geotrichum* sp. were less affected (~60% of inhibition).[40] The major components detected in that study were benzeneethanol and 2*H*-1-benzopyran-2-*O*-methoxy-4,7-dihydroxy; however, none of these compounds was individually or collectively tested. In our findings, benzeneethanol was detected in *Daldinia eschscholtzii* GU11N. However, it did not show a difference in its concentration level when compared with the coculture experiment. On the contrary, its expression tended to decrease when *D. eschscholtzii* was exposed to *M. citricolor*. The implementation of sensitive techniques, such as HS-SPME, can enhance our understanding of ecological relations between microorganisms and the environment.[59] On the other hand, complementing these studies with preparative isolation and chemical characterization with tools like nuclear magnetic resonance (NMR) and liquid chromatography-mass spectrometry (LC-MS) is crucial for obtaining a comprehensive chemical

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

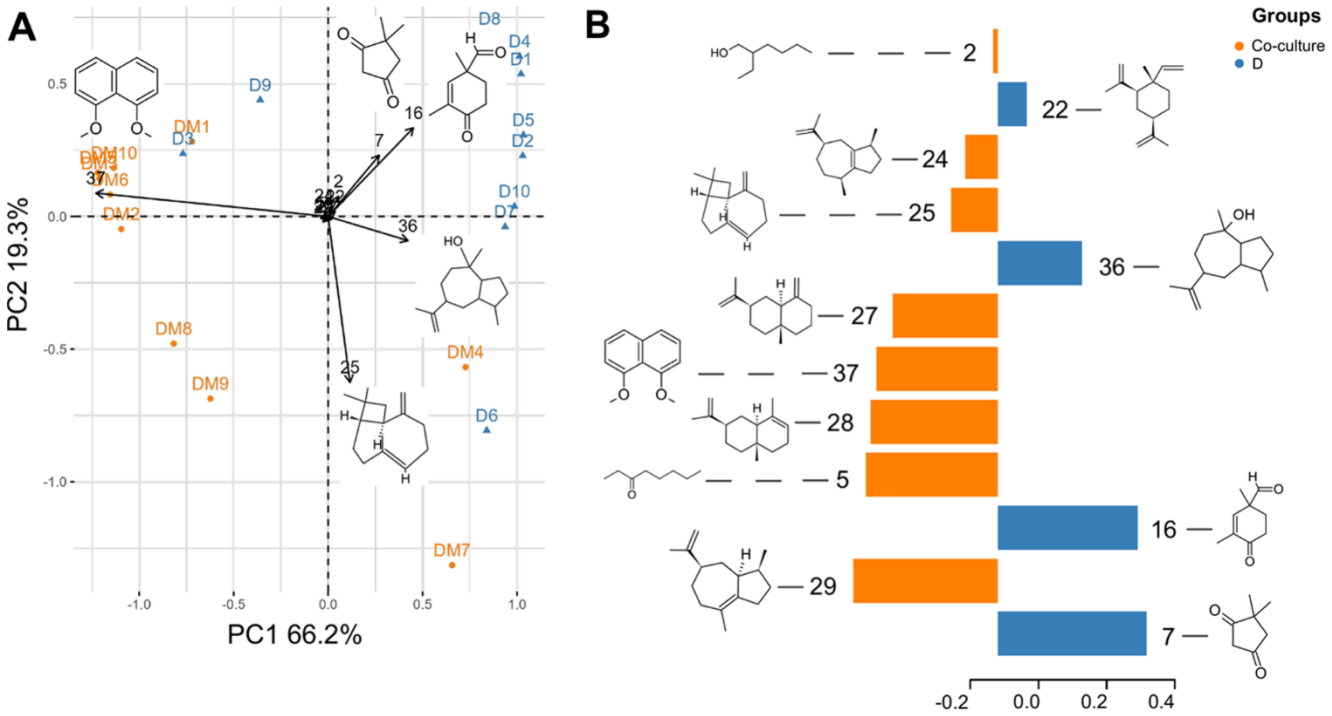


Figure 3. PCA and PLS-DA plots derived from the features present in the HS with HS-SPME data values. (A) PCA biplot with 85.5% cumulative variance. (B) PLS-DA VIPs for each group of the experiment (D and coculture).

annotation and biological testing to gain a better understanding of complex metabolic processes and ecological dynamics within communities.[60,61]

In Vitro Activity Test of Compounds against *M. citricolor* pseudopilei. Compound 1,8-dimethoxynaphthalene (**37**, synthesized from commercial 1,8-dihydroxynaphthalene) and a mixture of terpenes (commercial standard item containing *β*-caryophyllene) were tested at different concentrations to evaluate the effect on *M. citricolor* pseudopilei (Figure 4A,B). The mixture of terpenes (20, 200, and 2000 μg) did not produce any effect over the pseudopilei. On the other hand, the pseudopilei enclosed in the Petri dishes along with the filter discs loaded with 500 and 1000 μg of the compound 1,8-dimethoxynaphthalene (**37**) experienced the "burned-like" effect after 24 h at room temperature, which was similar to the damage observed when the pathogen was confronted against *D. eschscholtzii* GU11N (Figure 4C). None of the negative controls (solvents) produced an effect on pseudopilei. This suggests that **37** produced in the coculture group is responsible for the fungicidal effect. Pseudopilei from the controls and lowest concentrations were able to grow in PDA media and hyphae, as seen after 12 h (Figure 4D). Moderate fungicidal and herbicidal activity has been reported for **37** from an endophytic "*Nodulisporium*" sp. (also a Xylariales) isolate against *Microbotryum violaceum* and *Zymoseptoria* (=*Septoria*) *tritici* in the agar diffusion assay and did not present bactericidal activity.[62] Low inhibitory activity was reported for compound **37** obtained from an endophytic *D. eschscholtzii* isolate from a mangrove in Thailand[63] against *Microsporum gypseum* and *Staphylococcus aureus*. In addition, **37** was also reported from another *D. eschscholtzii* from the same ecosystem, but no activity tests were done.[64] Results from herbicidal tests performed in mono- and dicotyledonous plants have also been reported as low, retarding seed germination.[58] Another biological effect promoted by 1,8-dimethoxynaphthalene (**37**) is the more intense inhibition of the inflammatory process evaluated in murine macrophages when compared to quercetin; in this case, **37** was acquired from endophytic *Hypoxylon investiens*.[65]

Attempts to isolate fair amounts of volatile content and obtain pure compounds were not pursued. Therefore, the actual effect of this type of components on *M. citricolor* pseudopilei was not established, and the effects of **18**, **20**, **21**, **25**, **34**, **36**, and **38** remain unknown. Nonetheless, some works report effective fungicidal effects using terpenes obtained from plants and fungi and some modified derivatives using the disc diffusion method,[36,66,67] as well as by adding the essential oils directly into the culture media where conidia from pathogenic fungi have been previously inoculated.[68]

Comparing the pseudopilei morphology under SEM, notable changes were observed on the surface of the apical portion when they were exposed to 1,8-dimethoxynaphthalene (**37**). Undamaged structures remained turgent, with a consistent tapered shape (Figure 4E,F), whereas for those exposed to synthesized compound **37**, the original shape was obliterated, and the close-up details provided by the SEM photographs (Figure 4G,H) suggest damage related to wall or cell membrane disruption, leading to the fungicidal effect and germination inhibition. This kind of disruption to the fungal membranes is known from the exposure to synthetic (*E*)-2-hexanal at concentrations of 160 and 320 μL/L in a contactless experiment by altering the membrane permeability of *Penicillium aurantiogriseum* (= *Penicillium cyclopium*) conidia, producing a loss of lipids and potassium ions.[69] Similarly, conidia from *A. flavus* experienced membrane alteration and breakdown when exposed to the synthetic hexanal (3.2 and 9 μL/L), directly affecting the mitochondria, augmenting reactive oxygen species production, altering DNA replication, and promoting autophagy.[70] To our knowledge, this is the first time that this kind of effect has been reported, particularly for *M. citricolor* in confrontation against 1,8-dimethoxynaphthalene. Further analysis could include

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

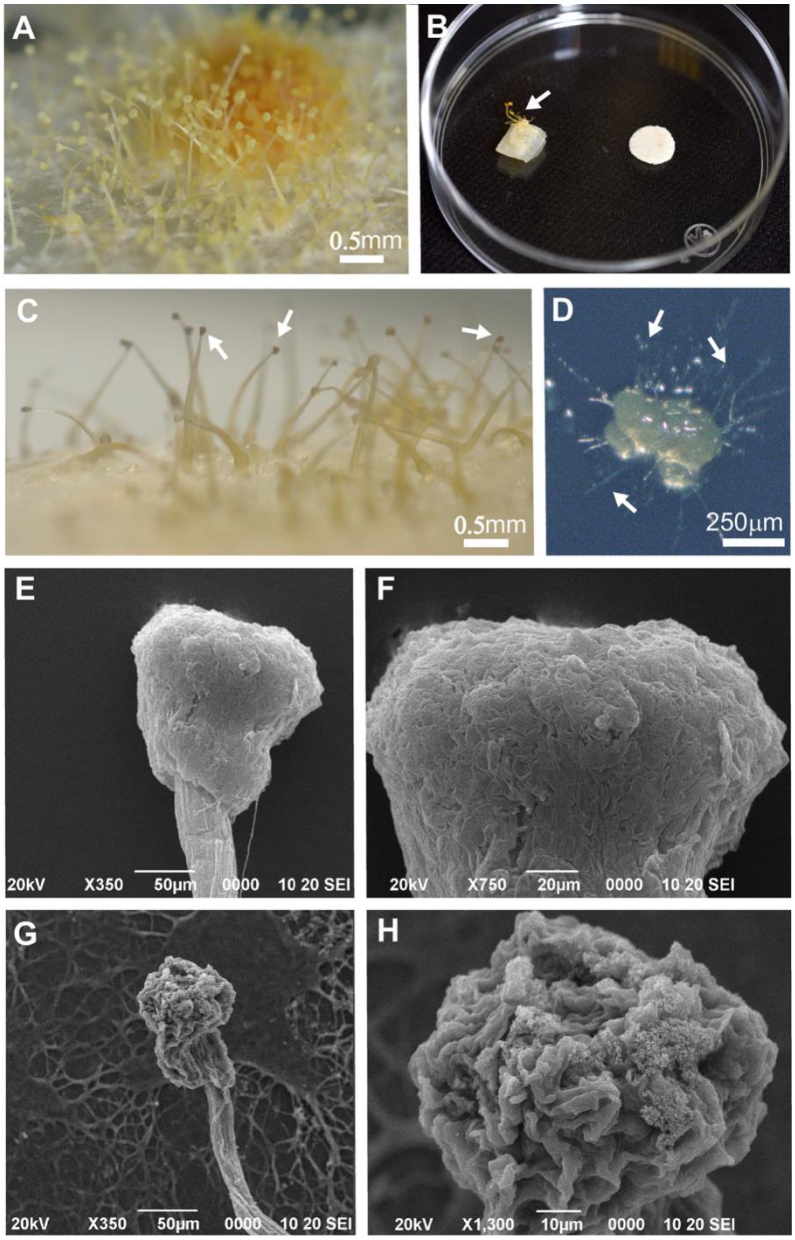


Figure 4. (A) Mature *M. citricolor* pseudopilei grown in PDA media. (B) Experiment with filter paper discs and *M. citricolor* pseudopelei (indicated by the arrow) in Petri dishes (6 cm diam.). (C) "Burning" effect (indicated by the arrow) produced in the pseudopilei in the presence of 1,8-dimethoxynaphthalene (500 and 1000 μg) after 24 h. (D) Pseudopilei from the control without damage and growing in PDA media after 12 h; arrows point to the developing hyphae. (E) SEM photograph from undamaged pseudopilei from the control group. (F) SEM details from the pseusopilei apical surface. (G) SEM image of the pseudopilei group exposed to 1,8-dimethoxynaphthalene (500 μg). (H) Detail of the damaged pseudopilei after exposure to 1,8-dimethoxynaphthalene.

transcriptomics to understand the changes at the metabolomic level and toxicity tests to evaluate the plant and human safety levels of exposure to this compound.

UHPLC-HMRMS/MS Profiling and Metabolite Annotation Overview. The UHPLC-HRMS/MS profiles obtained from the extracts of *Daldinia eschscholtzii* GU11N, *M. citricolor*, and the coculture were processed using MZmine 3.[71] This generated a table of 434 features (Table S2), which were then putatively annotated through the GNPS platform[72] and Sirius 5.[73] Results from both annotation strategies were consolidated and filtered to ensure a minimum of quality before any biological significance. This was done first with a pre-established workflow[74] followed by manual curation. The library search algorithm from GNPS was able to annotate only 5 out of 434 features. In contrast, Sirius provided a 21% annotation rate, while the module CANOPUS[75] was able to annotate 62.7% of the features with a chemical class. The quantitative .csv file was formatted using an in-house Python script to group the features according to the ion identity networking[76] results. The features were grouped based on their annotation network number (ANN), keeping only the highest intensity between the grouped features to perform statistical analysis. This resulted in only one ANN per metabolite, which reduced the number of adducts.

The PCA analysis based on this data showed that the samples from the coculture group formed a unique cluster in the upper left quadrant of the PC1 and PC2 plots, which accounted for

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

55.5% of the total variance (Figure S5). Loadings for features **255** (chemical class of amino acids/peptides) and **293** (chemical class of fatty acid derivatives) were observed in this cluster. Additionally, four coculture samples were grouped together with most of the M group in the right quadrants, with loadings for features **258** (MM 295.2456) and **280** (MM 297.2633) being the main contributors.

The distribution pattern of the *D. eschscholtzii* GU11N samples was evident in the PC2 and PC3 plots, which accounted for 33.1% of the total variance. Specifically, 6 out of the 10 *D. eschscholtzii* GU11N samples showed a correlation with feature **293**, along with four samples from the coculture group (refer to Figure S5A,B). The PLS-DA scores plot provided a clearer separation of samples from each group, allowing for distinct clustering of D, M, and coculture groups in the first two dimensions, explaining 45% of the total variance. From the top 15 VIP features from the first component of the PLS-DA analysis, features **298**, **252**, **247**, **295**, and **284** correlate strongly to the D group. Features **142**, **193**, **224**, **186**, **312** (*N*-acylethanolamine derivative), and **219** (tryptophan alkaloid) are more related to the M group, and features **293**, **62**, **311**, and **239** correlate to the coculture group (Table 2 and Figure S5C).

Table 2. Top 15 VIP Annotation Network Number (ANN) from the PLS-DA Analysis Based on the UHPLC-HRMS/MS Analysis

annotation network number (ANN) relevant in PCA and PLS-DA	group	NPClassifier superclass
62	coculture	purine alkaloids
142	M	-
186	M	pinane monoterpenoids
193	M	-
219	M	unsaturated fatty acids
224	M	-
239	coculture	-
247	D	open chain polyketides
252	D	fatty acyl carnitines
255	coculture	cyclic peptides
258	M	-
280	M	*N*-acyl amines
284	D	cholane steroids
293	D, coculture	fatty acyl carnitines
295	D	cholane steroids
298	D	phenylalanine-derived alkaloids
311	coculture	cholane steroids
312	M	-

Even though one of the current challenges in untargeted metabolomics is the confidence and availability of the spectra in databases, especially those derived from microorganisms such as fungi,[77] we were able to produce a putative feature annotation up to the chemical class level for the extracts obtained from the D, coculture, and M groups. Of the five annotations with statistical importance related to the coculture, only four were putatively annotated. In addition, an extract from an observed matter was released to the medium from *D. eschscholtzii* GU11N when confronted with *M. citricolor* and was submitted to LC-MS analysis. From the UHPLC-HRMS/MS data processing of this sample with MZmine 3, a total of 519 features were obtained and subsequently submitted to GNPS (4.8% annotation) and Sirius 5-CANOPUS (58.4%). As previously mentioned, most of the annotations must be revised and are, at this point, only proposals. The molecular networking derived from these data and supplied with the statistical values from the PLS-DA analysis represented in the size of each node displays some unique and shared annotations among the group samples (Figure 5). Among the clusters I, II, and III, the first one contains more diversity of chemical groups, being the most relevant for the D and the coculture, the node with a molecular weight of 191.6666 and categorized as an alkaloid. Among some of the shared nodes between D and M, many remain without a proposed annotation; nonetheless, nodes such as 332.2209 or 183.0934 are classified as amino acids, peptides, and alkaloids. Cluster II was better represented by the D and coculture groups; however, node 474.3799 was assigned to fatty acids and included a representation of the M group as well. Cluster III contains a node (520.3601) shared between the three sample groups cataloged as an alkaloid, but the node with more statistical significance was that labeled as 520.3618 and signaled as a fatty acid. Finally, cluster IV had an exclusive representation of the D and coculture groups, with fatty acids as the majority of the annotated nodes. Assuming that the coculture could contain compounds derived from both fungi growing in the coculture assay, and since the fungal cell membrane differs in its composition from plants and animals, compounds related to this structure may be pigments, e.g., 1,8-dihydroxynaphthalene, reported as common in *Aspergillus* spp.,[78] or lipids, from which ergosterol is a major component of the fungal cell membrane.[79] Undoubtedly, the coculture represents a rich metabolomic environment, containing fatty acids (45% of the annotations), terpenoids (14%), alkaloids (11.76%), polyketides (11%), amino acids and peptides (7%), carbohydrates (3.30%), and shikimates and phenylpropanoids (2.57%), leaving out 5.51% of the metabolites without annotation (Figure S7). To the best of our knowledge, no metabolite profiling has been done for *M. citricolor*, except for other *Mycena* or related saprophytic species (i.e., *Mycena hematopus* and *Mycena pruinosoviscida*). *M. hematopus* produced bactericidal pyrroloquinoline alkaloids[80] and pigments (e.g., hematopodins or mycenarubins).[81] *Mycena pruinosoviscida* (now *Roridomyces pruinosoviscida*) produced the fungicidal compounds phenylglycols and mycenadiols.[82]

Certainly, evaluation of the role of volatiles and nonvolatile compounds on fungi−fungi interactions poses several challenges from the methodological to the instrumentation capacities; however, this kind of research helps enlighten the metabolomic dynamics of fungal interactions.[83] This knowledge can be applied in the search for applications of fungal endophytes and their natural products to improve crop health, since several variables may shape the multispecies (e.g., plants−pathogens−endophytes) interactions;[84] for example, the chemistry of the plant itself that can affect an antagonistic effect seen in in vitro conditions.[85] In addition, with this knowledge, we can start to understand how multispecies interactions in natural conditions affect the overall phytobiome and, consequently, a plant's immunity and its relationship with other species and the ecosystem.

CONCLUSIONS

The endophytic isolate *Daldinia eschscholtzii* GU11N produced damage to the reproductive structures of the coffee pathogen *M. citricolor* in a contactless assay. Thirteen compounds were detected using HS and 39 by HS-SPME. With HS, the volatile

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

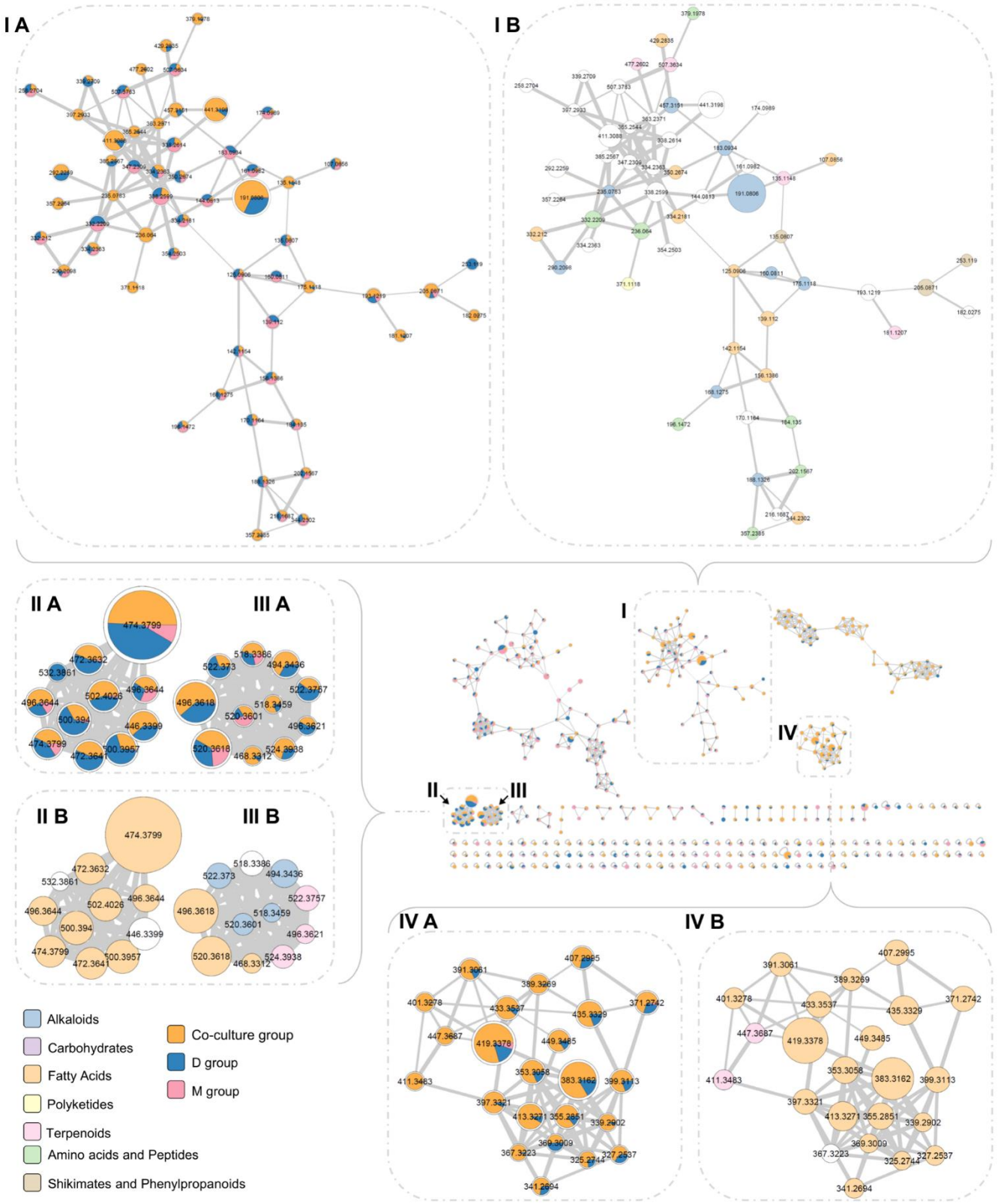


Figure 5. Molecular network derived from UHPLC data analysis. Node sizes represent the VIP value from the PLS-DA analysis. Node labels are molecular weights. Only the clusters with significant statistical values are displayed on a larger scale.

profile for *D. eschscholtzii* was represented by aliphatic ketones and fatty alcohols; sesquiterpenes were also detected. Using HS-SPME, the *D. eschscholtzii* profile was more diverse, detecting the classes found with HS in addition to chromones and polyketides as well as other sesquiterpenoid components. The composition of this profile revealed significant changes when the endophyte was exposed to *M. citricolor* in the coculture experiments. First, the HS profile showed increased detection

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

of caryophyllene sesquiterpenoids and naphthalenes and an evident suppression of aliphatic ketones. Similarly, the HS-SPME analysis added the detection of chromones. 1,8-Dimethoxynaphathalene (**37**) had a statistical correlation with the coculture experiment and produced the "burning" damage over the *M. citricolor* pseudopilei supported by microscopic observations (light and SEM microscopy); these images revealed severe damage to the reproductive structures possibly related to membrane disruption. A mixture of terpenes was tested without effects on the pathogen structures. UHPLC annotations demonstrate that the confrontation area was comprised mostly of fatty acids; followed by terpenoids, alkaloids, and polyketides in similar proportions, amino acids, and peptides; and finally, carbohydrates, shikimates, and phenylpropanoids in the smallest proportions. Understanding the changes derived from stress conditions, such as those produced by endophytic fungi emitting VOCs, could lead to the implementation of alternatives in agriculture for controlling important pests and to a better understanding of multispecies interactions in phytobiomes.

MATERIALS AND METHODS

Fungal Isolates. The endophytic fungus *Daldinia eschscholtzii* (isolate GU11N) was isolated from *R. grandifolia* leaves, while *M. citricolor* (isolate LAMYH1) was recovered from infected coffee plants from San Marcos de Tarrazu, Costa Rica. Both isolates were identified and used in a previous study by our research team.[34] Both fungi are preserved at −80 °C in cryovials with 20% glycerol and in sterile water at room temperature (∼25 °C) in CIPRONA (University of Costa Rica) culture collection.

Contactless Experiment. The "sandwich" technique[38] was used for this experiment. Both fungi were grown in independent 90 mm Petri dishes with potato−dextrose agar (PDA, Difco Laboratories, Detroit, Michigan). The inoculum consisted of mycelial disks (∼5 mm) placed 5 mm from the edge of the Petri dish. *Daldinia eschscholtzii* GU11N mycelial disks were placed on the bottom plate and those of *M. citricolor* on the top plate, facing *D. eschscholtzii* GU11N. Because the growth was faster in *D. eschscholtzii* GU11N, the pathogen was inoculated 2 days before starting the experiment. Radial growth (millimeters) of *M. citricolor* was measured every 24 h for 10 days. Ten replicates were done for the experiment. All plates were incubated at 25 °C with 12 h light/darkness cycles. Mycelia from *M. citricolor* treatments and control were observed under an Olympus BX40 light microscope and photographed with an attached 18 MP digital camera (Omax, China). Colony morphology was documented by using a Sony 6100 camera.

Coculture Assay. For this experiment, 90 mm Petri dishes with 20 mL of PDA were used to grow the *Daldinia eschscholtzii* GU11N (D group) and *M. citricolor* (M group) isolates independently and in a dual experiment with both fungi ("coculture") in the same plate. The coculture was done by placing mycelial disks 5 mm away from the plate's edge; in all treatments, the pathogen was placed 2 days before inoculating the *Daldinia* isolate. After 10 days, five mycelial discs from the edge of each independent culture and three from the confrontation line from the coculture experiment group were collected and placed inside 20 mL screw-top vials with poly(tetrafluoroethylene) (PTFE)/silicone septa caps (Thermo Scientific) for 24 h at 25 °C. Two vials with five PDA disks each were used as blanks along with three empty vials. Half of the mycelium from the independent cultures and the adjacent mycelium from the confrontation line were harvested using a sterile scalpel and stored in preweighted 20 mL vials and immediately lyophilized for 24 h (Labconco Freezone). After this, the vials were weighed to obtain mycelium yield.

HS and HS-SPME-GC-MS Data Acquisition. Vials with discs were pre-equilibrated for 15 min at 45 °C. A 100 μm PDMS SPME fiber (Thermo Scientific) was exposed to the sample headspace for 15 min at the same temperature with stirring in a TriPlus RSH autosampler (Thermo Scientific). Subsequently, the SPME fiber was thermally desorbed at 250 °C for 10 min in a GC-MS injector. Volatile compound measurements were conducted using a Gas Chromatograph Trace 1310 (Thermo Scientific) with a triple quadrupole mass detector TSQ 9000 system (Agilent Technologies), equipped with a TG-5MS capillary column (30 m length × 25 μm i.d. × 1 μm film thickness, Thermo Scientific). The injector of each sample was in split mode with a split ratio of 10, and the flow rate of the carrier gas (helium, 99.999% purity) was 1.2 mL/min. The temperature program was set as follows: initially, 60 °C for 3 min, increased to 200 °C at a rate of 10 °C/min, then increased to 290 °C at a rate of 15 °C/min, and finally maintained for 2 min. The gas chromatograph was set to ionization mode EI, 1 mL/min flow, mass range of 33−550 amu, transfer line at 280 °C, and ion source at 260 °C.

Data Processing with MZmine. The raw files from the HS-SPME experiment were converted from. RAW to an open. mzML format with MS Convert v3.0 Software.[86] Afterward, files were processed with MZmine v.2.5.3.[46] The noise level settled at 1E2 for only MS1. Parameters for the ADAP chromatogram builder were as follows: minimum group size in the number of scans, 5; group intensity threshold, 1E2; minimum highest intensity, 3E2; and m/z tolerance, 0.5. Settings for the ADAP chromatogram deconvolution were as follows: threshold, 8; estimator, "intensity window SN"; minimum feature height, 100; coefficient threshold, 5; peak duration range, 0.00−0.10; and retention time wavelet range, 0.01−0.11 min. Multivariate curve resolution was chosen as the spectral deconvolution option, and settings were adjusted as follows: deconvolution window width (min), 0.048; retention time tolerance (min), 0.08; and the minimum number of peaks, 4. For the deisotoped step, conditions were: m/z tolerance, 0.5; retention time tolerance, 0.08; maximum charge, 1; and representative isotope, "most intense". For the filtering process, the settings were as follows: m/z, 45.0000−250.0000; and retention time, 5.20−20.20 min. The final peak list was exported in .csv and .mgf formats for statistical analysis and spectral organization.

Spectral Organization through the GNPS platform. The online platform GNPS (Global Natural Products Social Network)[87] was used to upload the .mgf file and retrieve a molecular network. The settings were as follows: precursor ion mass tolerance, 0.02 Da; MS/MS fragment ion tolerance, 0.02 Da; and edges were filtered with cosine score >0.7 and >6 peaks. GNPS spectral libraries were used, and the conditions were to restrict a cosine score greater than 0.6 ad at least three matched peaks. Link for the job: https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=2ba96c07ebbb4bb29cca1659229be0f8&view=view_all_annotations_DB

Annotation through LabSolutions Postrun Analysis Software. The HS-SPME raw files were converted to CDF format using the Xcalibur v.4.2 (Thermo Scientific) Tool converter. The files were analyzed with LabSolutions Postrun Analysis software v.4 (Shimadzu Corporation) for compound annotation when the GNPS libraries did not offer a match for spectra. The peaks of each chromatogram were selected

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

following the retention time (RT) indicated by the MZmine feature list (Feature: a peak with an m/z value and area at a given retention time). Each selected feature on the spectrum table was compared against the Wiley 2014 and NIST 2020 databases requesting matches with similarity >80. This procedure was performed for each experiment group (D and M groups, and the coculture) and the blanks. The CDF files from the HS experiment were manually checked, and the RT for the features were corrected using the HS-SPME feature list; then, annotation for the compounds and area calculation was performed using the libraries and configuration described above.

Annotation through AMDIS Software. The NIST 20 mass spectral library v.1.0.8.8 loaded with MS Search v.2.4 and AMDIS 32 was used to corroborate the annotation features from the GNPS and LabSolutions platforms. The parameters used were as follows: the minimum match factor, 80; simple analysis; component width, 8; adjacent peak subtraction, 1; resolution, low; sensitivity, high; and shape requirements, medium.

Statistical Analysis for the GC-MS Data. The quantification table in .csv format obtained from the HS and HS-SPME experiments were normalized by the sum of each sample to perform ANOVA tests in R v.4.1.3.[88] Multivariate tests were done by applying Pareto scaling: principal component analysis (PCA) using the R FactoMineR v.2.4[89] and FactoExtra v.1.0.7[90] packages followed by partial least squares discriminant analysis (PLS-DA) with the mixOmics package v.6.6.2[91] in R v.3.5.1. A secondary analysis of PCA and PLS-DA for the HS-SPME data was done by using only the features detected as well in the HS experiment to fairly compare the detections for both tools.

In Vitro Activity Test of Synthetic Compounds against *M. citricolor* Pseudopilei. Filter paper discs (0.5 mm diameter) were saturated with 1, 10, and 100 μL of the terpene mixture (Supelco, CRM40937) by duplicate. The content of the third fraction (5.09 mg) corresponding to 1,8-dimethoxynaphthalene (Supporting Information: 1,8-dimethoxynaphthalene methylation) was resuspended in 509 μL of ether. Filter discs were saturated by triplicate with 10, 50, and 100 μL using a 100 μL syringe (Hamilton, Switzerland). The discs were air-dried and placed in Petri dishes (6 cm diameter) along with potato−dextrose−agar fragments containing previously grown *M. citricolor* pseudopilei (each fragment contained between 20−40 structures); the dishes were sealed using parafilm and left at room temperature (25 °C) and observed every 8 h with a dissecting microscope. After 24 h, three pseudopilei from each treatment were haphazardly removed and inoculated in Petri dishes with PDA to evaluate the germination and growth. Negative control for 1,8-dimethoxynaphthalene consisted of filter papers saturated with 500 μL of ether and those for the terpene mixture with the same amount of methanol (HPLC grade, J.T. Baker, Trinidad and Tobago); both control groups were performed in triplicate. Photographs in Figure 5A−D were taken using a DSLR camera (Nikon, D7100, Japan) with an 85 mm macro lens (Nikon, micro-Nikkor, Thailand), extension tubes (Vello), and wireless flashes (Nikon, SB-R200).

Scanning Electron Microscopy (SEM). The *M. citricolor* pseudopilei from the control and those with evident damage after exposure to 1,8-dimethoxynaphthalene were prepared and observed in the SEM (Jeol JSM-6390LV, Kyoto, Japan) at 20 kV; the magnification is specified in Figures 5E−H. For sample preparation, *M. citricolor* pseudopilei (around 25 per group) were immersed in 2.5% glutaraldehyde (0.1 M phosphate solution) at 4 °C for 24 h. Afterward, samples were washed with sterile water and dehydrated using sequential immersions of 20 min each in ethanol (from 30 to 95%). Finally, samples were vacuum freeze-dried, and a gold coating was applied (1 min, 1.8 mA, 2.5 kV).

UHPLC-HRMS Sample Preparation. The lyophilized samples were extracted following recommendations by Bertrand et al.[92] with two volumes of 10 mL of a dichloromethane/methanol/water (64:36:8) mix (HPLC grade, Merck), sonicated for 20 min in a water bath sonicator (Cole-Parmer) and filtered through cotton (previously washed with hexane, ethyl acetate, dichloromethane, and methanol all from Merck, LC-MS grade). Extractions were dried under a vacuum using a rotatory evaporator (Buchi, Switzerland). After being weighed, each sample was dissolved with methanol (HPLC quality, Sigma-Aldrich) to a 5 mg/mL concentration. To remove polar compounds and hyphae remnants, the methanol extracts were filtered through a solid-phase extraction cartridge (SPE) C_{18} (Chromabond, 100 mg, Macherey-Nagel, Germany), previously conditioned with 2 mL of methanol, 1:1 methanol/water, and 2 mL of water. Filters with a 0.2 μm nylon membrane (Millex, Merck, Ireland) were attached to the SPE columns, and the system was mounted in a vacuum chamber (SUPELCO Visiprep DL). Extracts were collected and dried under a vacuum using a Speed-Vac system (Savant Instruments). Samples were weighed and dissolved to 5 mg/mL in methanol (HPLC grade, Merck).

UHPLC-HRMS/MS Analysis. Analyses were performed with a Vanquish Duo UHPLC system (Thermo, Scientific, Germany) coupled to an Orbitrap Exploris 120 mass spectrometer (Thermo Scientific, Germany), employing a heated electrospray ionization source (H-ESI) with the following parameters: spray voltage: +3.2 kV; ion transfer tube temperature: 350.00 °C; vaporizer temperature: 400.00 °C; S-lens RF: 45 (arb units); sheath gas flow rate: 45.00 (arb units); and auxiliary gas flow rate: 15.00 (arb. units). The mass analyzer was calibrated using a mixture of caffeine, methionine−arginine−phenylalanine−alanine−acetate (MRFA), sodium dodecyl sulfate, sodium taurocholate, and Ultramark 1621 in an acetonitrile/methanol/water solution containing 1% formic acid by direct injection. Control of the instruments was done using Thermo Scientific Xcalibur 3.1 software v. 4.6.67.17. Full scans were acquired at a resolution of 60,000 fwhm (at m/z 200) and MS2 scans at 17,500 fwhm in the range of 100−1000 m/z, with 1 microscan, time (ms): 200 ms, an RF lens (%): 70; AGC target standard. The centroid data-dependent MS2 (dd-MS^2) scan acquisition events were performed in discovery mode, triggered by Apex detection with a trigger detection (%) of 75. The top 3 abundant precursors (charge states 1 and 2) within an isolation window of 2 m/z were considered for MS/MS analysis. Dynamic exclusion was set at 2 s. A mass tolerance of ±10 ppm was allowed, and the precursor intensity threshold was set at 150×10^3. For precursor fragmentation in the HCD mode, a normalized collision energy of 30% was used. The chromatographic separation was done on a Waters BEH C18 column (50 × 2.1 mm i.d., 1.7 μm, Waters, Milford, MA) using a linear gradient of 5−100% B over 13.50 min, followed by an isocratic step at 100% B until 15.50 min. The mobile phases were (A) water with 0.1% formic acid (Merck, grade HPLC-MS) and (B) acetonitrile with 0.1% formic acid (Merck, grade HPLC-MS). The flow rate was set to 500 μL/min, the injection volume was 2 μL, and the column was kept at 40 °C.

UHPLC-HMRMS/MS Data Processing. The raw data files were converted to. mzML format employing MS Convert software. The mzML files were processed with MZMine 3 software.[46] For mass detection at the MS1 level, the noise level

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

was set to 6.5E3. For MS2 detection, the noise level was set to 7.00. The ADAP chromatogram builder parameters were set as follows: minimum group size in the number of scans: 5; group intensity threshold: 7.0E5; minimum highest intensity: 1.4E6; and scan-to-scan accuracy (m/z): 0.0010 or 5.0 ppm. The local minimum feature resolver algorithm was used for chromatogram deconvolution with the following parameters: MS/MS scan pairing; check (default parameters); chromatographic threshold: 85%; minimum relative height: 0%; minimum absolute height: 7.0E5; minimum ratio of peak/edge: 4.90; peak duration range: 0.01−0.08; and min # of data points: 4. Isotopes were detected using the 13 °C isotope filter with an m/z tolerance of 0.0010 or 5.0 ppm, a retention time tolerance of 0.08 min (absolute), and the maximum charge set at 2. The representative isotope used was the most intense m/z. The files were aligned with the Join aligner algorithm (m/z tolerance: 0.0010 or 5 ppm; weight for m/z: 3; retention time tolerance: 0.08 min (absolute); weight for RT: 1), and the features present in the blanks were removed through the feature list blank subtraction algorithm (minimum number of detection in blanks: 3 and fold change increase: 300%). The resulting list was filtered by retention time: 2.70−7.30; features duration range: 0.02−1.0; and chromatographic fwhm: 0.01−1.0 with the feature list rows filter algorithm. The resulting filtered list was subjected to ion identity networking[76] starting with the metaCorrelate module (RT tolerance, 0.10 min; minimum height, 7.0E5; intensity correlation threshold 7.0E5, and the correlation grouping with the default parameters), followed by ion identity networking (m/z tolerance, 5.0 ppm; check: one feature; minimum height: 1.0E6; annotation library [maximum charge, 2; maximum molecules/cluster, 2; adducts ($[M + H]^+$, $[M + Na]^+$, $[M + K]^+$, $[M + NH4]^+$, $[M + 2H]^{2+}$), modifications ($[M - H_2O]$, $[M - 2H_2O]$, $[M - CO_2]$, [M + HFA], [M + ACN])]; annotation refinement (delete small networks without major ions, yes; delete networks without monomers, yes); add ion identities networks (m/z tolerance, 5 ppm; minimum height, 1.0E6; annotation refinement (minimum size, 1; delete small networks without major ions, yes; delete small networks: link threshold, 4; delete networks without monomers, yes); and check all ion identities by MS/MS (m/z tolerance (MS2), 10 ppm; min-height (in MS2), 1.0E3; check for multimers, yes; check neutral losses (MS1 → MS2), yes) modules. The resulting aligned peak list was exported as an .mgf file for further analysis.

Quantitative Data File Processing. The quantitative .csv file was formatted using an in-house python script to group the features according to the ion identity networking results. The features were grouped according to their annotation network number, keeping only the highest intensity between the grouped features.

Quantitative Data Statistical Analysis. The quantitative table in .csv format was normalized by the sum of each sample, and Pareto scaling was applied to perform principal component analysis (PCA) using the R FactoMineR v.2.4[89] and FactoExtra v.1.0.7[90] packages in R v.4.1.3.[88] Then, partial least squares discriminant analysis (PLS-DA) with the mixOmics package v.6.6.2[91] in R v.3.5.1 was performed.

Spectral Organization and Dereplication Process for Molecular Networking. The quantitative and spectral files obtained from MZmine were uploaded in the GNPS platform to generate feature-based molecular networking. The precursor ion mass tolerance was set to 0.02 Da with an MS/MS fragment ion tolerance of 0.02 Da. A network was created where edges were filtered to have a cosine score above 0.7 and more than six matched peaks. The spectra in the network were then searched against the GNPS' spectral libraries. All matches between network and library spectra were required to have a score above 0.6 and at least three matched peaks. Job link: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=777a3645587a44aea28efc26d043a841

Metabolite Annotation with Sirius. The .mgf file exported from MZmine (using the SIRIUS export module) was processed with SIRIUS (v5.6.2).[73] The parameters were set as follows: possible ionizations: $[M + H]^+$, $[M + NH_4]^+$, $[M - H_2O + H]^+$, $[M + K]^+$, $[M + Na]^+$, and $[M - 2H_2O + H]^+$; instrument profile: Orbitrap; mass accuracy: 5 ppm for MS^1 and 7 ppm for MS2; database for molecular formulas and structures: BIO; and maximum m/z to compute: 1000. ZODIAC was used to improve molecular formula prediction using a threshold filter of 0.99.[93] Metabolite structure prediction was made with CSI: FingerID,[94] and significance was computed with COSMIC.[95] The chemical class prediction was made with CANOPUS[75] using the NPClassifier ontology.[47]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c03865.

> Synthesis of 1,8-dimethoxynaphthalene, including materials, purification, and characterization steps; mycelium growth measurements of *M. citricolor* under both unchallenged and challenged conditions; supplementary statistical analysis for GC-MS experiments and UHPLC-HRMS containing ANOVA (GC-MS only); complementary PCA biplot and PLS-DA with VIPs' figures and representations of the chemical ontology using the NPClassifier nomenclature; supplementary table showing the distribution of features using the chemical ontology for HS and HS-SPME; and supplementary table showing the process in MZmine for UHPLC-HRMS features (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Giselle Tamayo-Castillo − *Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, 11520-2060 San José, Costa Rica; Escuela de Química, Universidad de Costa Rica, 11520-2060 San José, Costa Rica;* orcid.org/0000-0002-4912-8895; Email: giselle.tamayo@ucr.ac.cr

Authors

Efraín Escudero-Leyva − *Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, 11520-2060 San José, Costa Rica; Escuela de Biología, Universidad de Costa Rica, 11520-2060 San José, Costa Rica*

Luis Quirós-Guerrero − *Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1205 Geneva, Switzerland; School of Pharmaceutical Sciences, University of Geneva, 1205 Geneva, Switzerland;* orcid.org/0000-0002-1630-8697

Víctor Vásquez-Chaves − *Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, 11520-2060 San José, Costa Rica*

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

Reinaldo Pereira-Reyes − *Laboratorio Nacional de Nanotecnología (LANOTEC), Centro Nacional de Alta Tecnología, 10109 San Jose, Costa Rica*

Priscila Chaverri − *Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, 11520-2060 San José, Costa Rica; Escuela de Biología, Universidad de Costa Rica, 11520-2060 San José, Costa Rica; Department of Natural Sciences, Bowie State University, Bowie, Maryland 20715, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c03865

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The present manuscript was supported by U.S. National Science Foundation grant DEB-1638976 to P.C., and University of Costa Rica Vice-Rectory for Research projects B7176 and C1604. The isolate GU11N was collected under the SINAC-ACG permit ACG-PI-017-2017. The authors are thankful for the kind help from Lorena Hernández, David Morera, and Karolina Rivera, for their valuable assistantship in the laboratory during sample preparation. The authors acknowledge E.E.-L. for taking and providing the photographs included in this manuscript.

■ REFERENCES

(1) González-Teuber, M.; Palma-Onetto, V.; Aguilera-Sammaritano, J.; Mithöfer, A. Roles of leaf functional traits in fungal endophyte colonization: Potential implications for host−pathogen interactions. *J. Ecol.* **2021**, *109* (12), 3972−3987.

(2) Rodriguez, R. J.; Redman, R. S.; Henson, J. M. The role of fungal symbioses in the adaptation of plants to high stress environments. *Mitig. Adapt. Strateg. Global Change* **2004**, *9* (3), 261−272.

(3) Evans, H. C. The endophyte-enemy release hypothesis: implications for classical biological control and plant invasions. In *Proceedings of the XII International Symposium on Biological Control of Weeds, La Granded Motte, France, April 22−-27, 2007*; Julien, M.; Sforza, R.; Bon, M., Eds.; CABI: United Kingdom, 2008. DOI: 10.1079/9781845935061.0020.

(4) Quintana-Rodriguez, E.; Rivera-Macias, L. E.; Adame-Alvarez, R. M.; Torres, J. M.; Heil, M. Shared weapons in fungus-fungus and fungus-plant interactions? Volatile organic compounds of plant or fungal origin exert direct antifungal activity in vitro. *Fungal Ecol.* **2018**, *33*, 115−121.

(5) Cale, J. A.; Collignon, R. M.; Klutsch, J. G.; Kanekar, S. S.; Hussain, A.; Erbilgin, N. Fungal volatiles can act as carbon sources and semiochemicals to mediate interspecific interactions among bark beetle-associated fungal symbionts. *PLoS One* **2016**, *11* (9), No. e0162197.

(6) Rostás, M.; Cripps, M. G.; Silcock, P. Aboveground endophyte affects root volatile emission and host plant selection of a belowground insect. *Oecologia* **2015**, *177* (2), 487−497.

(7) Thomas, G.; Withall, D.; Birkett, M. Harnessing microbial volatiles to replace pesticides and fertilizers. *Microb. Biotechnol.* **2020**, *13* (5), 1366−1376.

(8) Hung, R.; Lee, S.; Bennett, J. W. Fungal volatile organic compounds and their role in ecosystems. *Appl. Microbiol. Biotechnol.* **2015**, *99* (8), 3395−3405.

(9) Kaddes, A.; Fauconnier, M. L.; Sassi, K.; Nasraoui, B.; Jijakli, M. H. Endophytic fungal volatile compounds as solution for sustainable agriculture. *Molecules* **2019**, *24* (6), 1065.

(10) Lee, S. O.; Kim, H. Y.; Choi, G. J.; Lee, H. B.; Jang, K. S.; Choi, Y. H.; Kim, J.-C. Mycofumigation with *Oxyporus latemarginatus* EF069 for control of postharvest apple decay and *Rhizoctonia* root rot on moth orchid. *J. Appl. Microbiol.* **2009**, *106* (4), 1213−1219.

(11) Kottb, M.; Gigolashvili, T.; Großkinsky, D. K.; Piechulla, B. *Trichoderma* volatiles effecting *Arabidopsis*: From inhibition to protection against phytopathogenic fungi. *Front. Microbiol.* **2015**, *6*, 995.

(12) Meshram, V.; Kapoor, N.; Saxena, S. *Muscodor kashayum* sp. nov. -a new volatile anti-microbial producing endophytic fungus. *Mycology* **2013**, *4* (4), 196−204.

(13) Tomsheck, A. R.; Strobel, G. A.; Booth, E.; Geary, B.; Spakowicz, D.; Knighton, B.; Floerchinger, C.; Sears, J.; Liarzi, O.; Ezra, D. *Hypoxylon* sp., an endophyte of *Persea indica*, producing 1,8-cineole and other bioactive volatiles with fuel potential. *Microb. Ecol.* **2010**, *60* (4), 903−914.

(14) Singh, S. K.; Strobel, G. A.; Knighton, B.; Geary, B.; Sears, J.; Ezra, D. An endophytic *Phomopsis* sp. possessing bioactivity and fuel potential with its volatile organic compounds. *Microb. Ecol.* **2011**, *61* (4), 729−739.

(15) Hübschmann, H.-J. *Handbook of GC-MS: Fundamentals and Applications*, 2nd ed.; McMaster, M., Ed.; Wiley & Sons, Inc.: Germany, 2015; pp 10−12.

(16) Sharifi, R.; Ryu, C. M. Sniffing bacterial volatile compounds for healthier plants. *Curr. Opin. Plant Biol.* **2018**, *44*, 88−97.

(17) Schalchli, H.; Hormazábal, E.; Becerra, J.; Briceño, G.; Hernández, V.; Rubilar, O.; Diez, M. C. Volatiles from white-rot fungi for controlling plant pathogenic fungi. *Chem. Ecol.* **2015**, *31* (8), 754−763.

(18) Guo, Y.; Werner, J.; Ghirardo, A.; Antritter, F.; Benz, J. P.; Schnitzler, J.-P.; Rosenkranz, M. Sniffing fungi − phenotyping of volatile chemical diversity in *Trichoderma* species. *New Phytol.* **2020**, *227* (1), 244−259.

(19) Shirey, R. E. SPME Commercial Devices and Fibre Coatings. In *Handbook of Solid Phase Microextraction*; Elsevier Inc., 2012; pp 99−133 DOI: 10.1016/B978-0-12-416017-0.00004-8.

(20) Xing, S.; Gao, Y.; Li, X.; Ren, H.; Gao, Y.; Yang, H.; Liu, Y.; He, S.; Huang, Q. Antifungal activity of volatile components from *Ceratocystis fimbriata* and its potential biocontrol mechanism on *Alternaria alternata* in postharvest cherry tomato fruit. *Microbiol. Spectrum* **2023**, *11* (1), 1−13.

(21) Rinkel, J.; Babczyk, A.; Wang, T.; Stadler, M.; Dickschat, J. S. Volatiles from the hypoxylaceous fungi *Hypoxylon griseobrunneum* and *Hypoxylon macrocarpum*. *Beilstein J. Org. Chem.* **2018**, *14*, 2974−2990.

(22) Al-Toubi, A. S. S.; Al-Sadi, A. M.; Al-Mahmooli, I. H.; Al-Harrasi, M. M. A.; Al Sabahi, J. N.; Velazhahan, R. Volatile organic compounds emitted by mycoparasitic fungi *Hypomyces perniciosus* and *Cladobotryum mycophilum* suppress the growth of *Agaricus bisporus*. *Czech Mycol.* **2022**, *74* (2), 141−152.

(23) Wolfender, J. L.; Gaudry, A.; Rutz, A.; Quiros-Guerrero, L.-M.; Nothias, L.-F.; Queiroz, E. F.; Defossez, E.; Allard, P.-M. Metabolomics in ecology and bioactive natural products discovery: challenges and prospects for a comprehensive study of the specialised metabolome. *Chimia* **2022**, *76* (11), 954−963.

(24) Ntie-Kang, F.; Svozil, D. An enumeration of natural products from microbial, marine and terrestrial sources. *Phys. Sci. Rev.* **2020**, *5* (8), 1−22.

(25) Koutouleas, A.; Collinge, D. B.; Ræbild, A. Alternative plant protection strategies for tomorrow's coffee. *Plant Pathol.* **2023**, *72* (3), 409−429.

(26) Barquero, M. Cafetaleros perdieron 12% de cosecha por lluvia y plagas. La Nación (San José, Costa Rica), December 17, 2010. https://www.nacion.com/economia/cafetaleros-perdieron-12-de-cosecha-por-lluvia-y-plagas/6NKK6NCILRDDXESMW5PN7PURJY/story/ (accessed 2023−04−25).

(27) Avelino, J.; Cabut, S.; Barboza, B.; Barquero, M.; Alfaro, R.; Esquivel, C.; Durand, J.-F.; Cilas, C. Topography and crop management are key factors for the development of American leaf spot epidemics on coffee in Costa Rica. *Phytopathology* **2007**, *97* (2), 1532−1542.

(28) Granados-Montero, M. D. M.; Avelino, J.; Arauz-Cavallini, F.; Castro-Tanzi, S.; Ureña, N. Leaf litter and *Mycena citricolor* inoculum

on the American leaf spot epidemic. *Agron. Mesoamerican* **2020**, *31* (1), 77−94.
(29) Bolívar-Anillo, H. J.; Garrido, C.; Collado, I. G. Endophytic microorganisms for biocontrol of the phytopathogenic fungus *Botrytis cinerea*. *Phytochem. Rev.* **2020**, *19* (3), 721−740.
(30) Syed Ab Rahman, S. F.; Singh, E.; Pieterse, C. M. J.; Schenk, P. M. Emerging microbial biocontrol strategies for plant pathogens. *Plant Sci.* **2018**, *267*, 102−111.
(31) Durham, T. C.; Mizik, T. Comparative economics of conventional, organic, and alternative agricultural production systems. *Economies* **2021**, *9* (2), 64.
(32) Sirikamonsathien, T.; Kenji, M.; Dethoup, T. Potential of endophytic *Trichoderma* in controlling *Phytophthora* leaf fall disease in rubber (*Hevea brasiliensis*). *Biol. Control* **2023**, *179*, No. 105175.
(33) Yang, Y.; Chen, Y.; Cai, J.; Liu, X.; Huang, G. Antifungal activity of volatile compounds generated by endophytic fungi *Sarocladium brachiariae* HND5 against *Fusarium oxysporum* f. sp. *cubense*. *PLoS One* **2021**, *16* (12), No. e0260747.
(34) Escudero-Leyva, E.; Granados-Montero, M. d. D.; Orozco-Ortiz, C.; Araya-Valverde, E.; Alvarado-Picado, E.; Chaves-Fallas, J. M.; Aldrich-Wolfe, L.; Chaverri, P. The endophytobiome of wild Rubiaceae as a source of antagonistic fungi against the American leaf spot of coffee (*Mycena citricolor*). *J. Appl. Microbiol.* **2023**, *134* (5), No. lxad090.
(35) Pujade-Renaud, V.; Déon, M.; Gazis, R.; Ribeiro, S.; Dessailly, F.; Granet, F.; Chaverri, P. Endophytes from wild rubber trees as antagonists of the pathogen *Corynespora cassiicola*. *Phytopathology* **2019**, *109* (11), 1888−1899.
(36) Helaly, S. E.; Thongbai, B.; Stadler, M. Diversity of biologically active secondary metabolites from endophytic and saprotrophic fungi of the ascomycete order Xylariales. *Nat. Prod. Rep.* **2018**, *35* (9), 992−1014.
(37) Stadler, M.; Læssøe, T.; Fournier, J.; Decock, C.; Schmieschek, B.; Tichy, H.-V.; Peršoh, D. A polyphasic taxonomy of *Daldinia* (Xylariaceae). *Stud. Mycol.* **2014**, *77*, 1−143.
(38) Liarzi, O.; Bar, E.; Lewinsohn, E.; Ezra, D. Use of the endophytic fungus *Daldinia* cf. *concentrica* and its volatiles as bio-control agents. *PLoS One* **2016**, *11* (12), No. e0168242.
(39) Khruengsai, S.; Pripdeevech, P.; Tanapichatsakul, C.; Srisuwannapa, C.; D'Souza, P. E.; Panuwet, P. Antifungal properties of volatile organic compounds produced by *Daldinia eschscholtzii* MFLUCC 19−0493 isolated from *Barleria prionitis* leaves against *Colletotrichum acutatum* and its postharvest infections on strawberry fruits. *PeerJ* **2021**, *9*, No. e11242.
(40) Pandey, A.; Banerjee, D. *Daldinia bambusicola* Ch4/11 an endophytic fungus producing volatile organic compounds having antimicrobial and olio chemical potential. *J. Adv. Microbiol.* **2014**, *1* (6), 330−337.
(41) Sanadhya, P.; Bucki, P.; Liarzi, O.; Ezra, D.; Gamliel, A.; Miyara, S. B. *Caenorhabditis elegans* susceptibility to *Daldinia* cf. *concentrica* bioactive volatiles is coupled with expression activation of the stress-response transcription factor daf-16, a part of distinct nematicidal action. *PLoS One* **2018**, *13* (5), No. e0196870.
(42) Macías-Rubalcava, M. L.; Sánchez-Fernández, R. E.; Roque-Flores, G.; Lappe-Oliveras, P.; Medina-Romero, Y. M. Volatile organic compounds from *Hypoxylon anthochroum* endophytic strains as postharvest mycofumigation alternative for cherry tomatoes. *Food Microbiol.* **2018**, *76*, 363−373.
(43) Bejarano, A.; Puopolo, G. Bioformulation of microbial biocontrol agents for a sustainable agriculture. In *How Research Can Stimulate the Development of Commercial Biological Control Against Plant Diseases*; De Cal, A.; Melgarejo, P.; Magan, N., Eds.; Progress in Biological Control; Springer, Nature: Switzerland, 2020; Vol. *21*, pp 275−294. DOI: 10.1007/978-3-030-53238-3_16.
(44) Inamdar, A. A.; Morath, S.; Bennett, J. W. Fungal Volatile Organic Compounds: More than just a funky smell. *Annu. Rev. Microbiol.* **2020**, *74*, 101−116.
(45) Smirnov, A.; Jia, W.; Walker, D. I.; Jones, D. P.; Du, X. ADAP-GC 3.2: Graphical software tool for efficient spectral deconvolution of Gas Chromatography-High-Resolution mass spectrometry metabolomics Data. *J. Proteome Res.* **2018**, *17* (1), 470−478.
(46) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinf.* **2010**, *11* (1), No. 395.
(47) Kim, H. W.; Wang, M.; Leber, C. A.; Nothias, L.-F.; Reher, R.; Kang, K. B.; van der Hooft, J. J.; Dorrestein, P. C.; Gerwick, W. H.; Cottrell, G. W. NPClassifier: A deep neural network-based structural classification tool for natural products. *J. Nat. Prod.* **2021**, *84* (11), 2795−2807.
(48) Fiehn, O. Metabolomics by Gas Chromatography-Mass Spectrometry: the combination of targeted and untargeted profiling. *Curr. Protoc. Mol. Biol.* **2016**, *114* (1), 30.4.1−30.4.32.
(49) Lucero, M.; Estell, R.; Tellez, M.; Fredrickson, E. A retention index calculator simplifies identification of plant volatile organic compounds. *Phytochem. Anal.* **2009**, *20* (5), 378−384.
(50) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097−2098.
(51) Goodacre, R.; Broadhurst, D.; Smilde, A. K.; Kristal, B. S.; Baker, J. D.; Beger, R.; Bessant, C.; Connor, S.; Capuani, G.; Craig, A.; Ebbels, T.; Kell, D. B.; Manetti, C.; Newton, J.; Paternostro, G.; Somorjai, R.; Sjöström, M.; Trygg, J.; Wulfert, F. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **2007**, *3* (3), 231−241.
(52) Wang, T.; Mohr, K. I.; Stadler, M.; Dickschat, J. S. Volatiles from the tropical ascomycete *Daldinia clavata* (Hypoxylaceae, Xylariales). *Beilstein J. Org. Chem.* **2018**, *14*, 135−147.
(53) Tolouee, M.; Alinezhad, S.; Saberi, R.; Eslamifar, A.; Zad, S. J.; Jaimand, K.; Taeb, J.; Rezaee, M.-B.; Kawachi, M.; Shams-Ghahfarokhi, M.; Razzaghi-Abyaneh, M. Effect of *Matricaria chamomilla* L. flower essential oil on the growth and ultrastructure of *Aspergillus niger* van Tieghem. *Int. J. Food Microbiol.* **2010**, *139* (3), 127−133.
(54) Niogret, J.; Kendra, P. E.; Epsky, N. D.; Heath, R. R. Comparative analysis of terpenoid emissions from Florida host trees of the redbay ambrosia beetle, *Xyleborus glabratus* (Coleoptera: Curculionidae: Scolytinae). *Florida Entomol.* **2011**, *94* (4), 1010−1017.
(55) De Alfonso, I.; Vacas, S.; Primo, J. Role of α-copaene in the susceptibility of olive fruits to *Bactrocera oleae* (Rossi). *J. Agric. Food Chem.* **2014**, *62* (49), 11976−11979.
(56) Nishida, R.; Shelly, T. E.; Whittier, T. S.; Kaneshiro, K. Y. α-Copaene, a potential rendezvous cue for the Mediterranean fruit fly, *Ceratitis capitata*. *J. Chem. Ecol.* **2000**, *26* (1), 87−100.
(57) Pathania, A. S.; Guru, S. K.; Ashraf, N. U.; Riyaz-UI-Hassan, S.; Ali, A.; Tasduq, S. A.; Malik, F.; Bhushan, S. A novel stereo bioactive metabolite isolated from an endophytic fungus induces caspase dependent apoptosis and STAT-3 inhibition in human leukemia cells. *Eur. J. Pharmacol.* **2015**, *765*, 75−85.
(58) Flores-Reséndiz, M.; Lappe-Oliveras, P.; Macías-Rubalcava, M. L. Mitochondrial damage produced by phytotoxic chromenone and chromanone derivatives from endophytic fungus *Daldinia eschscholtzii* strain GsE13. *Appl. Microbiol. Biotechnol.* **2021**, *105* (10), 4225−4239.
(59) Musteata, F. M.; Vuckovic, D. *In Vivo* Sampling with Solid-Phase Microextraction. In *Handbook of Solid Phase Microextraction*; Elsevier, 2012; pp 399−453. DOI: 10.1016/B978-0-12-416017-0.00012-7.
(60) Molinski, T. F. NMR of natural products at the 'nanomole-scale'. *Nat. Prod. Rep.* **2010**, *27* (3), 321−329.
(61) Cuesta-Rubio, O.; Piccinelli, A. L.; Fernandez, M. C.; Hernández, I. M.; Rosado, A.; Rastrelli, L. Chemical characterization of Cuban propolis by HPLC-PDA, HPLC-MS, and NMR: The brown, red, and yellow Cuban varieties of propolis. *J. Agric. Food Chem.* **2007**, *55* (18), 7502−7509.
(62) Dai, J.; Krohn, K.; Flörke, U.; Draeger, S.; Schulz, B.; Kiss-Szikszа, A.; Antus, S.; Kurtán, T.; van Ree, T. Metabolites from the endophytic fungus *Nodulisporium* sp. from *Juniperus cedre*. *Eur. J. Org. Chem.* **2006**, *2006*, 3498−3506.

(63) Kongyen, W.; Rukachaisirikul, V.; Phongpaichit, S.; Sakayaroj, J. A new hydronaphthalenone from the mangrove-derived *Daldinia eschscholtzii* PSU-STD57. *Nat. Prod. Res.* **2015**, *29* (21), 1995−1999.
(64) Yang, L.-J.; Liao, H.-X.; Bai, M.; Huang, G.-L.; Luo, Y.-P.; Niu, Y.-Y.; Zheng, C.-J.; Wang, C.-Y. One new cytochalasin metabolite isolated from a mangrove-derived fungus *Daldinia eschscholtzii* HJ001. *Nat. Prod. Res.* **2018**, *32* (2), 208−213.
(65) Chang, C. W.; Chang, H.-S.; Cheng, M.-J.; Liu, T.-W.; Hsieh, S.-Y.; Yuan, G.-F.; Chen, I.-S. Inhibitory effects of constituents of an endophytic fungus *Hypoxylon investiens* on nitric oxide and interleukin-6 production in RAW264.7 macrophages. *Chem. Biodivers.* **2014**, *11* (6), 949−961.
(66) Bajpai, V. K.; Shukla, S.; Kang, S. C. Chemical composition and antifungal activity of essential oil and various extract of *Silene armeria* L. *Bioresour. Technol.* **2008**, *99* (18), 8903−8908.
(67) Geng, Y.; Gui, K.; Pan, T.; Ye, J.; Li, J.; Feng, J.; Ma, Z.; Lei, P.; Gao, Y. Preparation of terpene-derived fungicidal candidates with a 1,3,4-thiadiazole moiety for natural product-inspired agrochemical discovery. *Ind. Crops Prod.* **2022**, *189*, No. 115889.
(68) Souza, D. P.; Pimentel, R. B. Q.; Santos, A.; Albuquerque, P. M.; Fernandes, A. V.; Junior, S. D.; Oliveira, J. T. A.; Ramos, M. V.; Rathinasabapathi, B.; Gonçalves, J. F. C. Fungicidal properties and insights on the mechanisms of the action of volatile oils from Amazonian Aniba trees. *Ind. Crops Prod.* **2020**, *143*, No. 111914.
(69) Zhang, J.; Tian, H.; Sun, H.; Wang, X. Antifungal activity of trans-2-hexenal against *Penicillium cyclopium* by a membrane damage mechanism. *J. Food Biochem.* **2017**, *41* (2), No. e12289.
(70) Li, S. F.; Zhang, S.-B.; Zhai, H.-C.; Lv, Y.-Y.; Hu, Y.-S.; Cai, J.-P. Hexanal induces early apoptosis of *Aspergillus flavus* conidia by disrupting mitochondrial function and expression of key genes. *Appl. Microbiol. Biotechnol.* **2021**, *105* (18), 6871−6886.
(71) Schmid, R.; Heuckeroth, S.; Korf, A.; et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **2023**, *41* (4), 447−449.
(72) Wang, M.; Carver, J. J.; Phelan, V.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828−837.
(73) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16* (4), 299−302.
(74) Quiros-Guerrero, L. M.; Nothias, L.-F.; Gaudry, A.; Marcourt, L.; Allard, P.-M.; Rutz, A.; David, B.; Queiroz, E. F.; Wolfender, J.-L. Inventa: A computational tool to discover structural novelty in natural extracts libraries. *Front. Mol. Biosci.* **2022**, *9*, No. 1028334.
(75) Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M.; Petras, D.; Gerwick, W. H.; Rousu, J.; Dorrestein, P. C.; Böcker, S. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **2021**, *39* (4), 462−471.
(76) Schmid, R.; Petras, D.; Nothias, L.-F.; et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nat. Commun.* **2021**, *12* (1), No. 3832.
(77) Theodoridis, G.; Gika, H.; Raftery, D.; Goodacre, R.; Plumb, R. S.; Wilson, I. D. Ensuring fact-based metabolite identification in liquid chromatography-mass spectrometry-based metabolomics. *Anal. Chem.* **2023**, *95* (8), 3909−3916.
(78) Hanson, B. J. R. Fungi and the development of microbiological chemistry. In *The Chemistry of Fungi*; Royal Society of Chemistry: Cambridge, 2008; pp 6−15.
(79) Thevissen, K.; Ferket, K. K. A.; François, I. E. J. A.; Cammue, B. P. A. Interactions of antifungal plant defensins with fungal membrane components. *Peptides* **2003**, *24* (11), 1705−1712.
(80) Lohmann, J. S.; Wagner, S.; von Nussbaum, M.; Pulte, A.; Steglich, W.; Spiteller, P. Mycenaflavin A, B, C, and D: Pyrroloquinoline alkaloids from the fruiting bodies of the mushroom *Mycena haematopus*. *Chem. − Eur. J.* **2018**, *24* (34), 8609−8614.
(81) Peters, S.; Spiteller, P. Mycenarubins A and B, red pyrroloquinoline alkaloids from the mushroom *Mycena rosea*. *Eur. J. Org. Chem.* **2007**, *2007*, 1571−1576.
(82) Isaka, M.; Chinthanom, P.; Sappan, M.; Supothina, S.; Boonpratuang, T. Phenylglycol metabolites from cultures of the basidiomycete *Mycena pruinosoviscida* BCC 22723. *Helv. Chim. Acta* **2014**, *97* (7), 909−914.
(83) Azzollini, A.; Boggia, L.; Boccard, J.; Sgorbini, B.; Lecoultre, N.; Allard, P.-M.; Rubiolo, P.; Rudaz, S.; Gindro, K.; Bicchi, C.; Wolfender, J.-C. Dynamics of metabolite induction in fungal co-cultures by metabolomics at both volatile and non-volatile levels. *Front. Microbiol.* **2018**, *9*, 72.
(84) Rai, M.; Agarkar, G. Plant-fungal interactions: What triggers the fungi to switch among lifestyles? *Crit. Rev. Microbiol.* **2016**, *42* (3), 428−438.
(85) Arnold, A. E.; Mejia, L. C.; Kyllo, D.; Rojas, E. I.; Maynard, Z.; Robbins, N.; Herre, E. A. Fungal endophytes limit pathogen damage in a tropical tree. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (26), 15649−15654.
(86) Chambers, M. C.; Maclean, B.; Burke, R.; et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918−920.
(87) Aron, A. T.; Gentry, E.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **2020**, *15* (6), 1954−1991.
(88) R Core Team. *R: A language and environment for statistical computing*, 2020 https://www.r-project.org/ (accessed 2023−01−10).
(89) Lê, S.; Josse, J.; Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Software* **2008**, *25* (1), 1−18.
(90) Kassambara, A.; Mundt, F. *Release 2020: Factoextra: extract and visualize the results of multivariate data analyses*. R package version 1.0.7, 2020. https://CRAN.R-project.org/package=factoextra (accessed 2023−04−15).
(91) Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K. A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **2017**, *13* (11), No. e1005752.
(92) Bertrand, S.; Schumpp, O.; Bohni, N.; Bujard, A.; Azzollini, A.; Monod, M.; Gindro, K.; Wolfender, J.-L. Detection of metabolite induction in fungal co-cultures on solid media by high-throughput differential ultra-high pressure liquid chromatography-time-of-flight mass spectrometry fingerprinting. *J. Chromatogr. A* **2013**, *1292*, 219−228.
(93) Ludwig, M.; Nothias, L.-F.; Dührkop, K.; et al. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nat. Mach. Intell.* **2020**, *2* (10), 629−641.
(94) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (41), 12580−12585.
(95) Hoffmann, M. A.; Nothias, L.-F.; Ludwig, M.; Fleischauer, M.; Gentry, E. C.; Witting, M.; Dorrestein, P. C.; Dührkop, K.; Böcker, S. High-confidence structural annotation of metabolites absent from spectral libraries. *Nat. Biotechnol.* **2022**, *40* (3), 411−421.

https://doi.org/10.1021/acsomega.3c03865
ACS Omega 2023, 8, 31373−31388

4.5 Metabolómica en plantas de café y parientes silvestres de Rubiaceae.

Los productos del metabolismo secundario de las plantas, como los compuestos fenólicos, los flavonoides, los alcaloides y las hormonas, juegan un papel importante en el crecimiento, el desarrollo y la resistencia al estrés de las plantas. La familia de plantas Rubiaceae es extremadamente diversa y abundante en América Central y contiene varios géneros económicamente importantes. Estos son conocidos por la producción de polifenoles bioactivos (por ejemplo, cafeína y quinina), que han tenido un gran impacto en la sociedad humana. El objetivo general de este estudio fue desarrollar un flujo de trabajo de alto rendimiento para identificar y cuantificar los polifenoles vegetales. Primero, se optimizó un método para extraer más de 40 familias de fitoquímicos. Luego, se desarrolló una plataforma metabolómica de alto rendimiento para identificar y cuantificar 184 polifenoles en 15 minutos. El estudio de metabolitos secundarios se realizó en hojas de una variedad comercial de café y dos especies silvestres que también pertenecen a la familia Rubiaceae. El perfilado global se realizó mediante cromatografía líquida de alta resolución y espectrometría de masas de tiempo de vuelo. Las características cuya abundancia fue significativamente diferente entre las especies de café se discriminaron mediante análisis estadístico y se anotaron mediante bases de datos espectrales. La cafeína, la trigonelina y la teobromina fueron muy abundantes en las hojas de café, como se esperaba. Curiosamente, las hojas silvestres de Rubiaceae tenían una mayor diversidad de fitoquímicos en comparación con el café comercial: las moléculas relacionadas con la defensa, como los fenilpropanoides, el ácido giberélico y el monolignol sinaldehído se encontraron más abundantemente en las hojas silvestres de Rubiaceae.

Publicación:

Castro-Moretti, F., Cocuron, J-C., Castillo-Gonzales, H., **Escudero-Leyva, E.**, Chaverri, P., Guerreiro-Filho, O., Slot, C.J. & Alonso, A.P. **2023**. A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography mass spectrometry. Frontiers in Plant Science. 13:1057645. doi: 10.3389/fpls.2022.1057645

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 January 2023
DOI 10.3389/fpls.2022.1057645

OPEN ACCESS

EDITED BY
Marta Sousa Silva,
University of Lisbon, Portugal

REVIEWED BY
Emmanuelle Meudec,
Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France
Melissa Hamner Mageroy,
Norwegian Institute of Bioeconomy Research (NIBIO), Norway

*CORRESPONDENCE
Ana Paula Alonso
Anapaula.Alonso@unt.edu

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 29 September 2022
ACCEPTED 14 December 2022
PUBLISHED 06 January 2023

CITATION
Castro-Moretti FR, Cocuron J-C, Castillo-Gonzalez H, Escudero-Leyva E, Chaverri P, Guerreiro-Filho O, Slot JC and Alonso AP (2023) A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography mass spectrometry. *Front. Plant Sci.* 13:1057645. doi: 10.3389/fpls.2022.1057645

A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography mass spectrometry

Fernanda R. Castro-Moretti[1], Jean-Christophe Cocuron[2], Humberto Castillo-Gonzalez[3], Efrain Escudero-Leyva[4,5], Priscila Chaverri[3,4], Oliveiro Guerreiro-Filho[6], Jason C. Slot[7] and Ana Paula Alonso[1,2]*

[1]BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, TX, United States, [2]BioAnalytical Facility, University of North Texas, Denton, TX, United States, [3]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD, United States, [4]School of Biology and Natural Products Research Center Centro de Investigaciones en Productos Naturales (CIPRONA), University of Costa Rica, San Jose, Costa Rica, [5]Centro Nacional de Alta Technologia-Consejo Nacional de Rectores (CeNAT-CONARE), National Center for Biotechnological Innovations (CENIBiot), San Jose, Costa Rica, [6]Coffee Center Agronomic Institute, Campinas, Sao Paulo, Brazil, [7]Department of Plant Pathology, The Ohio State University, Columbus, OH, United States

Introduction: Products of plant secondary metabolism, such as phenolic compounds, flavonoids, alkaloids, and hormones, play an important role in plant growth, development, stress resistance. The plant family *Rubiaceae* is extremely diverse and abundant in Central America and contains several economically important genera, e.g. *Coffea* and other medicinal plants. These are known for the production of bioactive polyphenols (e.g. caffeine and quinine), which have had major impacts on human society. The overall goal of this study was to develop a high-throughput workflow to identify and quantify plant polyphenols.

Methods: First, a method was optimized to extract over 40 families of phytochemicals. Then, a high-throughput metabolomic platform has been developed to identify and quantify 184 polyphenols in 15 min.

Results: The current metabolomics study of secondary metabolites was conducted on leaves from one commercial coffee variety and two wild species that also belong to the *Rubiaceae* family. Global profiling was performed using liquid chromatography high-resolution time-of-flight mass spectrometry. Features whose abundance was significantly different between coffee species were discriminated using statistical analysis and annotated using spectral databases. The identified features were validated by commercially

available standards using our newly developed liquid chromatography tandem mass spectrometry method.

Discussion: Caffeine, trigonelline and theobromine were highly abundant in coffee leaves, as expected. Interestingly, wild *Rubiaceae* leaves had a higher diversity of phytochemicals in comparison to commercial coffee: defense-related molecules, such as phenylpropanoids (e.g., cinnamic acid), the terpenoid gibberellic acid, and the monolignol sinapaldehyde were found more abundantly in wild Rubiaceae leaves.

KEYWORDS

LC-MS/MS, phenolics, phytochemicals, *Rubiaceae*, secondary metabolism

1 Introduction

Plant secondary metabolites are byproducts of primary metabolism. They play important roles during plant development, reproduction and stress response (Patra et al., 2013; Ma et al., 2016; Böttger et al., 2018; Kessler and Kalske, 2018; Jain et al., 2019). Because plants are sessile organisms, they must endure environmental and biotic pressure. The production of phytochemicals is part of their response to these stresses. Besides their importance for plant adaptation, growth and development, plant secondary compounds are valuable resources for the food, pharmaceutical, and biofuel industries (Boudet, 2007; Korkina, 2007; Cragg and Newman, 2013; Chiocchio et al., 2021). Alkaloids, phenolics, and terpenoids are the three main families that comprise secondary metabolites produced by plants. They are synthesized through malonic acid, mevalonic acid, methylerythritol-phosphate, and shikimate pathways (Jain et al., 2019). Because of their vast structural/chemical diversity, low solubility, and small quantities in plant tissues, the recovery, identification and quantification of phytochemicals are particularly challenging.

Genetics studies in combination with metabolic profile are key for plant breeding and insertion of desired traits, such as specific polyphenols. Recovering lost attributes due to domestication using wild relatives in the breeding program is a promising strategy. However, it has been mostly applied to crops such as rice, wheat, barley and potatoes (McSorley and Phillips, 1992; Peleg et al., 2005; Feuillet et al., 2008; Spooner et al., 2014; Brar and Khush, 2018). Despite the limited number of studies on coffee leaf and other *Rubiaceae* metabolic content, the presence of phenolic compounds has been previously described (Souard et al., 2018; Cangeloni et al., 2022; Montis et al., 2022). Caffeine, chlorogenic acids, mangiferin and trigonelline are the main phytochemicals found in coffee leaves (Cangeloni et al., 2022). Indole alkaloids are the most common secondary metabolite class throughout *Rubiaceae* species, although other classes of alkaloids, terpenes and flavonoids have also been reported (Martins and Nunez, 2015). For example, akuamigine, vincoside, yohimbine and other indole alkaloids have been detected in the *Rubiaceae Uncaria* spp. (Laus and Teppner, 1996; Ndagijimana et al., 2013; Xie et al., 2013). Different species of *Gardenia* sp. produce iridoids, such as genipin and gardenoside, as well as flavonoids and triterpenes (Chen et al., 2009; Kunert et al., 2009; Yang et al., 2013; Wang et al., 2015). Many members the *Rubiaceae* family have been studied for their secondary metabolites with medicinal properties (Chen et al., 2009; Ahmad and Salim, 2015; Martins and Nunez, 2015). For instance, species that are used in traditional medicine from the genera *Borreria* and *Spermacoce* contain alkaloids, flavonoids, iridoids, and terpenoids (Conserva and Ferreira Júnior, 2012). Additionally, medicinal plants with known anti-inflammatory and antioxidant properties, such as species from the genera *Rytignia* and *Canthium multiflorum*, have bioactive compounds like tannins, saponins and flavonoids, coumarins and terpenoids (Chandra Kala, 2015). Therefore, it is important to develop efficient methodologies to monitor plant polyphenols, which will guide breeding programs and boost phytochemical discovery.

In order to fully grasp the diversity of phytochemicals present in leaves of coffee and other *Rubiaceae* species, i) a single extraction procedure allowing to recover the vast diversity phytochemical families is needed, ii) an untargeted metabolomics approach is required to detect unknown/new polyphenols, and iii) high-throughput targeted metabolomics method is necessary to quantify a maximum of secondary compounds within a single run. Developing a fast methodology that isolates most of the secondary metabolites present in leaves is challenging.

Metabolomics is the ideal technique for detecting small quantities of phytochemicals (Jorge et al., 2016; Cocuron et al., 2019; Castro-moretti et al., 2020). Nuclear magnetic resonance (NMR) and mass spectrometry (MS) are the most common analytical tools for performing plant metabolomics. MS coupled

with liquid or gas chromatography (LC and GC, respectively) is a preferred method due to its higher sensitivity and lesser amount of sample requirements (de Falco and Lanzotti, 2018; Liu et al., 2019; Perez de Souza et al., 2019). In this study, two MS instruments were used: a high-resolution quadrupole time-of-flight (HR-Q-TOF) for untargeted metabolomics, and a highly sensitive triple quadrupole for targeted quantification of known metabolites. On one hand, untargeted studies are designed to detect a broad range of molecules in a biological sample (Patti et al., 2012; Perez de Souza et al., 2019). It is common to use spectral libraries to attempt compound identification (Dunn et al., 2012; Perez de Souza et al., 2019; Jez et al., 2021). On the other hand, targeted metabolomics is used to quantify known metabolites using analytical standards (Patti et al., 2012; Roberts et al., 2012; Sawada and Yokota Hirai, 2013). Although complementary, these two approaches have been rarely combined (Montis et al., 2022).

Other approaches have attempted to isolate and quantify plant secondary metabolites; however, a limited number of families (one or two) and compounds (less than 40) were monitored (Sun et al., 2013; Orcic et al., 2014; Bataglion et al., 2015; Lin et al., 2015; Jaini et al., 2017; Cocuron et al., 2019; Gulcin et al., 2019; Marchetti et al., 2019; Quatrin et al., 2019). In this study, a single-extraction method was developed to recover 42 distinct families of phytochemicals. Untargeted and targeted metabolomics were combined to study the secondary metabolites present in coffee and wild *Rubiaceae* leaves. More specifically, a state-of-the-art targeted approach allowing the quantification of 184 phytochemicals was developed and was used to validate the identity of 74 compounds highlighted by the untargeted analysis. Combining both techniques and instruments along with an optimized extraction method resulted in a sensitive and thorough pipeline to detect, classify and quantify secondary metabolites in leaves of coffee and other *Rubiaceae* species. We anticipate that this thorough pipeline will boost the process of detection, classification and quantification of polyphenols in leaves of coffee and other *Rubiaceae*, and will be further applied to other plant organs and species.

2 Materials and methods

2.1 Chemicals

LC-MS-grade acetic acid, acetonitrile, methanol, DMSO, and water were ordered from Thermo Fisher Scientific (Hampton, NH). All non-labeled standards as well as trans-cinnamic acid-β,2,3,4,5,6-d6 were purchased from MilliporeSigma (Burlington, MA). N,N-dimethyltryptamine (N,N-DMT), bufotenin (5-OH-DMT) and psilocybin standards were obtained from Cayman Chemical (Ann Arbor, MI).

2.2 Preparation of standard stocks and working solutions

Stock solutions of dihydrokaempferol, dihydroquercetin, mitragynine, luteolin, luteolin-7-O-glucoside, naringenin, phloretin, piceid, prunetin, pterostilbene, orientin, quercetin, quercitrin, reserpine, rhamnazin, rhamnetin, schaftoside, spiraeoside, swertiajaponin, swertisin, tectochrysin, tricin, vicenin 2 and 3, vincosamide and yohimbine were prepared at 1,000 µM. Acacetin-7-O-rutinoside, afzelin, apigenin-7-glucuronide, calycosin, corynanthine, harmane, hordenine, ipriflavone, tomatidine, xanthohumol, sophoricoside, rauwolscine, idaein, keracyanin and neobavaisoflavone were prepared at 100 µM. All other stock solutions were prepared at 10,000 µM, using methanol or DMSO as solvents. Working solutions were prepared to final concentrations of 100, 50, 10 and 1 µM in methanol/water (40:60; v/v).

2.3 Leaf collection

Untargeted and targeted metabolomic analyses were performed using commercial coffee leaves (*Coffea arabica* cv. Obatã IAC 1669-20 - CC) that were collected in San José, Costa Rica, at the Coopetarrazú plantation. Three mature leaves were harvested from four trees and kept on ice during transportation to the laboratory, where they were flash-frozen in liquid nitrogen, and lyophilized until dryness. Wild *Rubiaceae* leaves were collected from the two species *Isertia hankeana* and *Simira maxonii* (WR1 and WR2, respectively) in a private rainforest of the Golfito (Puntarenas) region of Costa Rica. Three leaves were collected from each tree, one tree per species and kept in ice during their transportation to the laboratory. Then they were flash-frozen in liquid nitrogen and lyophilized (Labconco Freezone, South Kansas City, KS) until dryness. To optimize the extraction and LC-MS/MS method, mature leaves from the wild coffee species *Coffea liberica* var. *dewevrei* and *Coffea salvatrix* were used from plants grown in field conditions, collected at the Agronomic Institute in Campinas (São Paulo, Brazil). Four leaves were collected from each tree, two trees per species, kept in dry ice during harvest and transportation to the laboratory and then they were frozen in liquid nitrogen prior lyophilization until dryness. All collected leaves were lyophilized using a freeze dryer Labconco Freezone 12 plus (South Kansas City, KS). Dried leaves were ground into a fine powder using 15 mL plastic jars with four 20 mm metal beads in a tissue homogenizer Geno/Grinder 2010 from Spex (Metuchen, NJ) for two rounds of 30 sec at 1,750 rpm. Wild *Rubiaceae* leaves were collected from the same tree, grown in very different environmental conditions in comparison to CC; to mitigate variations due to leaf maturity level, light exposure, etc., and to focus on polyphenol differences across species, the ground

powder was pooled, homogenized and divided into five pseudoreplicates per species.

2.4 Intracellular secondary metabolite extraction

The extraction of leaf secondary metabolites was performed after grinding and weighting 10 mg of powdered leaf material. Metabolites were extracted by adding 10 µL of 1 mM trans-cinnamic acid-β,2,3,4,5,6-d6 and 490 µL of 100% methanol followed by grinding with one 5 mm metal bead at 30 Hz for 5 min using a mixer mill MM400 from Retsch (Haan, Germany). Then, the extracts were sonicated at 35-40°C for 20 min and centrifuged at 9,600 *g* for 5 min at room temperature. The supernatants were transferred to 1.5 mL microcentrifuge tubes. The remaining pellets were resuspended in 500 µL of methanol/water (30:70, v/v), sonicated for 20 min at 35-40 °C, and spun down under the same conditions as mentioned before. The supernatants were combined to the first ones, and then, 500 µL of extracts were filtered through 3 kDa Amicon filtering devices (MilliporeSigma, Burlington, MA) at 14,000 *g* for 60 min at room temperature. The resulting eluates were stored at -20°C until LC-MS/MS analysis.

2.5 Untargeted metabolomics

The analysis of the metabolites was carried out using an Exion ultra high-performance liquid chromatography system coupled with a high-resolution mass spectrometer TripleTOF6600+ from AB Sciex (Framingham, MA).

2.5.1 HPLC conditions

The compounds were separated using a C18 Symmetry column (75 x 4.6 mm, 3.5 µm) with a Symmetry C18 pre-column (20 x 3.9 mm; 5µm) from Waters (Milford, MA) as previously described (Cocuron et al., 2019). The temperatures of the column compartment and the autosampler were kept at 30 °C and 15 °C, respectively. The analytes were eluted using a gradient of 0.1% (v/v) acetic acid in acetonitrile (Solvent A) and 0.1% (v/v) acetic acid in water (Solvent B) under a flow rate of 0.8 mL/min. The following gradient was applied: 0-1.0 min, 98% B; 1.0-16.0 min, 98-42% B; 16.0-21.0 min, 42-20% B; 21.0-26.0 min, 20-10% B; 26.0-28.0 min, 10% B; 28.0-28.1, 10-98% B; 28.1-30.0 min, 98% B.

2.5.2 High-resolution discovery using triple TOF

2.5.2.1 Data-dependent acquisition

The mass spectrometer was set to scan metabolites from m/z 100-1500 amu in negative or positive mode. For the negative polarity the ion spray voltage was 4,500 V, the accumulation time was 100 msec, the declustering potential and collision energy were 60 V and 10 V, respectively. MS/MS spectra were acquired over m/z 30-1500 amu with an accumulation time of 25 msec. Parameters such as declustering potential, collision energy, collision energy spread were set to 60 V, 45 V and 15 V, respectively. The parameters for the positive mode were very similar to the ones for the negative mode except for the ion spray voltage, and the declustering potential that were 5,000 V and 35 V, respectively. The total cycling time was 0.65 sec.

The parameters for the electrospray source ionization such as curtain gas (nitrogen), nebulizing gas, heating gas, and the temperature of the source were fixed at 40 psi, 70 psi, 70 psi, and 650 °C, respectively. The source conditions were the same for the negative and positive polarities. An atmospheric-pressure chemical ionization (APCI) negative or positive calibration solution was delivered by a calibrant delivery system every 5 samples to correct for any mass drift that may occur during the run. MS spectra were acquired using Analyst TF 1.8.1 software (AB Sciex, Framingham, MA). It is important to note that the data-dependent acquisition (DDA) was run in negative and positive modes, on a mixture composed of CC, WR1 and WR2 extracts. The precursor ions present in this mixture were then used for the sequential window acquisition of all theoretical mass spectra (SWATH-MS) scan survey.

2.5.2.2 Data-independent acquisition using SWATH-MS

The precursor ion data obtained from the DDA (negative and positive ionizations) were used to generate short overlapping precursor ion windows which are the core of the SWATH-MS mode. Briefly, all precursors ions from the DDA mode (negative or positive polarity) as well as their intensities were exported to an excel file in which a total of 20 variable SWATH-MS windows were created with one amu overlapping mass. These SWATH-MS windows for the negative or positive polarity were saved as a ".txt" file, and uploaded to Analyst to build the mass spectrometry part of the LC-HR-MS/MS acquisition method. It is important to note that the source, MS scan and MS/MS scan parameters for the SWATH-MS mode were the same than the ones used for the DDA scan survey.

For the sequence injection, the total 15 biological samples were placed randomly in the autosampler. Two quality controls (QC) consisting of an equal mixture of each leaf extract, and two blanks containing the internal standard, trans-cinnamic acid-β,2,3,4,5,6-d6 in methanol: water (40:60, v/v), were included; one was injected at the beginning of the sequence, and the other at the end.

2.5.3 Data processing

A non-targeted screening MQ4 workflow was designed using the software SciexOS v.1.6.1 (Sciex, Framingham, MA) with the Smart Confirmation Search algorithm. Results were

sorted by purity, 0.02 Da as precursor mass tolerance, and 0.4 Da as fragment mass tolerance. The intensity threshold was set to 0.05 and minimal purity to 10%. All extract samples, blanks with internal standard, and QCs were considered for constructing the processing method. Quality of the processed data was assessed as follows: i) the difference in the area of the internal standard between the sample extracts and the blanks was less than 5%; and ii) a Principal Component Analysis (PCA) including the QCs was performed for the metabolites monitored in negative and positive modes and showed clustering of the two QCs (Supplemental Figure 1), which is indicative of a low variance. Then, data obtained from untargeted metabolomics were analyzed using the software XCMS (Tautenhahn et al., 2012) with the following parameters: method UPL/UHD Q-TOF matchedFilter; ppm error of 15; minimum peak width of 5; maximum peak width of 20; signal/noise threshold of 6; mzdiff of 0.01; integration method 1; prefilter peaks 3; noise filter 0; and retention time correction method obiwarp. Once peak picking, alignment and integration was performed, a table with mass to charge ratio, retention time, and normalized intensity of each feature (by the intensity of the internal standard, trans-cinnamic acid-β,2,3,4,5,6-d6) was generated. This table was then used for statistical analyses (see 2.7. Statistical Analyses). Further data curation was performed: features with signal intensity lower than 1,000,000.00 count per second were excluded, as well as the features not present on all replicates. Then, the feature with the highest intensity was selected for each peak group. Features from positive and negative ionization were merged and, when the same feature was present in both modes, the one with the highest intensity was selected. Metabolite identification was performed using the software SciexOS v.1.6.1 (Sciex, Framingham, MA) with spectral libraries from the National Institute of Standards and Technology (NIST, Gaithersburg, MA) and a homemade library.

2.6 Targeted metabolomics using LC-MS/MS scheduled multiple reaction monitoring

2.6.1 LC-MS/MS conditions

The detection and quantification of phytochemicals was performed as previously described by (Cocuron et al., 2019) with some slight modifications concerning the LC gradient and the use of scheduled multiple reaction monitoring (sMRM).

The 184 phytochemicals considered in this study were optimized one by one by direct infusion after having been diluted to 1 µM with acetonitrile/water solution (50:50; v/v) containing 0.1% of acetic acid as an additive. The flow for the direct infusion was set to 10 µL/min, and parameters such as declustering potential (DP), collision energy (CE), cell exit potential (CXP) were determined for the five most abundant product ions derived from each precursor ion (see Table 1). The compound optimization was done automatically for the negative and positive polarities using Analyst 1.7 software (AB Sciex, Framingham, MA). The source optimization for the electrospray ionization was conducted using different values for the curtain gas (25, 30, 35, 40 V), the nebulizer gas (GS1; 40, 45, 50, 55, 60, 65 V), the heating gas (GS2; 40, 45, 50, 55, 60, 65 V), the collision activated dissociation (CAD; low, medium, high), the temperature (300, 350, 400, 450, 500, 550, 600, 650°C), and the ionspray voltage (IS; 3000, 3500, 4000, 4500, 5000 V).

The compounds were detected and quantify using an Agilent 1290 Infinity II liquid chromatography system coupled to a hybrid Triple Quadrupole 6500+ from ABSciex (Framingham, MA). The extracts were kept at 10°C in an auto-sampler, and the phytochemicals were separated at 30°C using a reverse phase C18 Symmetry column (4.6 x 75 mm; 3.5 µm) coupled to a Symmetry C18 pre-column (3.9 x 20 mm; 5 µm) from Waters (Milford, MA). The liquid chromatography gradient was made of 0.1% (v/v) acetic acid in acetonitrile (A) and 0.1% (v/v) acetic acid in water (B). The total LC-MS/MS run was 15 min with a flow rate of 800 µL/min. The following gradient was applied to resolve the polyphenols: B= 0-1.0 minute 98%, 1.0-7.0 min 42%, 7.0-9.0 min 20%, 9.0-11.0 min 10%, 11.0-13.0 min 10%, 13.0-13.1 min 98%, and 13.1-15.0 min 98%. The injection needle was rinsed with 50% aqueous methanol. Five µL of external standard mixtures and 5 µL of biological sample were injected onto the column.

Electrospray ionization with polarity switch was applied to the extracts to acquire mass spectra of the different analytes. The settling time between each polarity was 15 msec. Phytochemicals were simultaneously detected as precursor ion/product ion pair using multiple reaction monitoring (MRM) at first to record the retention time for each of the polyphenols considered in this study (see Table 1). The retention times for the compounds were reported in the LC-MS/MS method to create scheduled MRM with MRM detection windows of 60 sec. The cycling time was set to 1.1 sec and the dwell time varied depending on the number of MRMs triggered at a specific point of time during the LC-MS/MS acquisition. The dwell time ranged from 3 to 250 msec. The source parameters for both modes were identical, and they were as followed: 4,500 V for the ionspray voltage, 40 V for the curtain gas, 550°C for the temperature, 50 psi for the nebulizer gas (GS1), 60 psi for the heating gas (GS2), and "Medium" for the collision activated dissociation (CAD).

2.6.2 Data acquisition and processing

Analyst 1.7 software (AB Sciex, Framingham, MA) was used to acquire the data whereas MultiQuant v3.0.3 (AB Sciex, Framingham, MA) was used to integrate the peaks corresponding to the different phytochemicals. Metabolite quantification was performed as previously explained (Arias et al., 2022). Briefly, the total amount of each analyte was calculated using the trans-cinnamic acid-β,2,3,4,5,6-d6 internal standard area, and the known concentration of its corresponding external standard run in parallel to the samples.

10.3389/fpls.2022.1057645

TABLE 1 Compound-dependent parameters for scheduled MRM scan survey, per metabolite and their chemical class (family): retention time (RT) in minutes, precursor mass (Q1) and product mass (Q3), declustering potential (DP), collision energy (CE) and collision cell exit potential (CXP) in volts.

Family	Metabolite	RT	Q1	Q3	DP	CE	CXP
alkaloid	Corynanthine	4.1	355.3	144.0	113	39	16
alkaloid	Dihydrocinchonine	4.2	297.0	279.0	40	31	14
alkaloid	Harmane	3.6	183.0	115.0	50	45	12
alkaloid	Hordenine	1.1	166.0	120.9	24	21	14
alkaloid	Mitragynine	4.9	399.0	174.0	75	41	20
alkaloid	Seneciphylline	3.4	334.0	120.0	105	35	14
alkaloid	Tomatidine	5.6	416.0	161.0	91	49	18
amino acid derivative	5-Hydroxy-tryptophan	3.1	221.0	204.0	20	13	12
amino acid derivative	Tyramine	0.9	138.0	120.9	9	13	14
anthocyanidin	Apigeninidin	4.4	255.0	170.9	127	43	18
anthocyanin	Idaein	3.7	449.1	287.0	60	29	14
anthocyanin glycoside	Keracyanin	3.6	595.2	286.9	75	39	14
benzodioxol	Piperonyloyl	6.4	184.8	142.8	50	13	16
chalcone	Xanthohumol	9.6	355.0	299.0	27	15	16
cinnamaldehyde	p-Coumaraldehyde	5.9	147.1	119.0	-50	-24	-13
cinnamate ester	3,4-Di-O-caffeoylquinic acid	6.6	515.0	353.0	-80	-28	-19
cinnamate ester	3-Caffeoylquinic acid	4.7	353.0	134.0	-25	-62	-15
cinnamate ester	4,5-Di-O-caffeoylquinic acid	6.7	515.0	353.0	-80	-28	-19
cinnamate ester	4-Caffeoylquinic acid	5.3	353.0	173.0	-45	-20	-11
cinnamate ester	5-Caffeoylquinic acid	4.8	353.0	93.0	-25	-56	-11
coumarin	6-Methylcoumarin	7.7	161.0	105.0	80	29	12
coumarin	7,8-Dihydroxy-4-methylcoumarin	5.4	191.0	119.0	-65	-26	-11
coumarin	Scopoletin	5.5	192.9	133.0	100	29	14
coumarin derivative	Esculetin	4.7	178.9	123.0	80	31	14
coumarin derivative	Mellein	8.2	179.0	160.8	11	17	18
cyclic ketone	Isophorone	7.5	138.9	68.9	184	21	8
cyclohexenecarboxylic acid	p-Coumaroyl-shikimate	5.3	321.1	147.0	115	15	10
dihydrochalcone	Phloretin	6.9	274.9	106.9	50	21	12
diterpenoid	Gibberellic acid	5.4	345.0	239.0	-70	-20	-13
diterpenoid	Ginkgolide A	6.6	409.1	345.0	80	27	16
flavanol	Catechin	4.3	291.0	139.0	50	20	16
flavanol	Epicatechin	4.6	291.0	139.0	50	20	16
flavanol	Epigallocatechin	4.1	307.0	138.9	13	19	16
flavanol	Gallocatechin	3.7	307.1	138.9	19	19	16
flavanone	Eriodictyol	6.5	287.0	151.0	-30	-20	-7
flavanone	Hesperetin/Homoeriodictyol	7.3	303.0	177.0	85	25	10

(Continued)

10.3389/fpls.2022.1057645

TABLE 1 Continued

Family	Metabolite	RT	Q1	Q3	DP	CE	CXP
flavanone	Isosakuranetin	8.4	287.0	153.0	80	29	18
flavanone	Naringenin	7.1	272.9	153.0	50	31	16
flavanone	Sakuranetin	8.4	287.0	167.0	90	29	20
flavanone-C-glycoside	Swertiajaponin	4.6	463.1	445.1	60	17	22
flavanone O-glycoside	Naringin	5.3	579.1	271.0	-150	-44	-13
flavanone O-glycoside	Naringenin-7-O-glucoside	5.5	435.1	273.0	70	19	14
flavanonol	Dihydrokaempferol	6.1	287.0	125.0	-60	-28	-5
flavanonol	Dihydroquercetin	5.5	303.0	285.0	-50	-16	-13
flavone	3-Deoxyrobinetin	5.1	284.9	149.0	-110	-36	-9
flavone	Apigenin	7.0	269.0	117.0	-90	-42	-13
flavone	Apigenin-7-glucuronide	6.8	447.1	271.1	120	29	14
flavone	Baicalein	7.3	270.9	122.9	150	43	14
flavone	Chrysin	8.3	253.0	143.0	-110	-36	-9
flavone	Chrysoeriol	7.0	301.0	286.0	80	37	32
flavone	Diosmetin	7.1	300.9	286.0	80	35	32
flavone	Eupatorin	7.9	344.9	283.9	100	41	30
flavone	Flavopiridol	4.8	402.0	341.0	75	33	16
flavone	Genkwanin	8.5	284.8	242.0	80	43	26
flavone	Maysin	5.5	577.0	431.0	50	19	24
flavone	Myricetin	5.8	317.0	151.0	-70	-32	-15
flavone	Scutellarein	6.0	286.9	123.0	110	45	14
flavone	Tectochrysin	10.1	268.9	226.0	80	43	26
flavone	Tricin	7.0	331.0	315.0	100	41	34
flavone C-glycoside	Isoorientin	4.6	449.0	299.0	50	39	14
flavone C-glycoside	Swertisin	5.0	447.2	297.0	60	35	14
flavone-C-glycoside	Isoschaftoside	4.5	565.1	427.1	100	29	20
flavone-C-glycoside	Isovitexin/Vitexin	4.9	433.0	283.0	80	35	14
flavone-C-glycoside	Orientin	4.7	449.0	329.0	120	39	16
flavone-C-glycoside	Rhamnosylisoorientin	4.4	593.0	298.0	-150	-58	-13
flavone-C-glycoside	Vicenin 2	4.2	595.1	577.1	100	21	28
flavone-C-glycoside	Vicenin 3/Schaftoside	4.5	565.1	547.2	80	19	26
flavone-O-glycoside	Acacetin-7-O-rutinoside	5.8	593.1	447.1	60	25	22
flavone-O-glycoside	Apigenin-7-O-glucoside	5.3	432.9	271.0	60	25	14
flavone-O-glycoside	Benzoic acid	7.4	120.9	77.0	-20	-16	-9
flavone-O-glycoside	Luteolin-7-O-glucuronide	6.5	463.0	287.0	130	29	14
flavone-O-glycoside	Myricetin-3-O-Rhamnoside	5.0	465.0	319.0	30	15	16
flavone-O-glycoside	Neodiosmin	5.3	607.0	299.0	-150	-40	-15

(Continued)

10.3389/fpls.2022.1057645

TABLE 1 Continued

Family	Metabolite	RT	Q1	Q3	DP	CE	CXP
flavonol	Fisetin	5.9	286.9	137.0	120	43	16
flavonol	Gossypetin	5.8	319.0	169.1	160	43	18
flavonol	Isorhamnetin	7.2	315.0	300.0	-80	-28	-15
flavonol	Kaempferide	8.5	300.9	229.0	120	53	24
flavonol	Kaempferol	7.1	286.9	153.0	120	43	18
flavonol	Luteolin	6.4	286.9	153.0	120	43	18
flavonol	Morin/Tricetin	6.3	302.9	152.9	140	39	18
flavonol	Quercetin	6.5	301.0	151.0	-80	-28	-15
flavonol	Quercitrin	5.4	449.0	303.0	30	15	16
flavonol	Rhamnazin	8.6	331.0	316.0	140	35	36
flavonol	Rhamnetin	7.8	315.0	165.0	-60	-28	-9
flavonol	Tamarixetin	7.2	316.9	302.0	120	35	14
flavonol-O-glucuronide	Miquelianin	6.7	479.0	303.0	60	21	16
flavonol-O-glycoside	Isoquercetin	5.1	465.1	303.0	30	17	16
flavonol-O-glycoside	Isorhamnetin-3-O-glucoside	5.3	479.1	317.0	30	17	16
flavonol-O-glycoside	Kaempferitrin	4.8	579.1	433.1	50	17	20
flavonol-O-glycoside	Kaempferol-3-O-glucoside	5.4	449.0	287.0	40	17	16
flavonol-O-glycoside	Kaempferol-3-O-glucuronide	6.4	463.0	287.0	50	21	14
flavonol-O-glycoside	Kaempferol-3-O-rutinoside	5.1	595.1	287.1	50	27	14
flavonol-O-glycoside	Kaempferol-7-O-Neohesperidoside	5.1	595.1	287.0	120	31	14
flavonol-O-glycoside	Luteolin-7,3'-Di-O-glucoside/Kaempferol-3-O-sophoroside	4.6	611.1	449.1	150	31	22
flavonol-O-glycoside	Luteolin-7-O-glucoside/Kaempferol-7-O-glucoside	5.4	449.0	287.0	91	27	14
flavonol-O-glycoside	Quercetin-3,4'-O-diglucoside	4.5	627.0	465.0	60	19	24
flavonol-O-glycoside	Quercetin-3-O-galactoside	5.0	465.0	303.0	40	17	16
flavonol-O-glycoside	Rutin	4.8	611.2	303.0	30	29	16
flavonol-O-glycoside	Spiraeoside	5.5	465.0	303.0	120	29	16
flavonol-glycoside	Afzelin	5.7	431.1	285.0	-100	-28	-15
glycosylated hydroquinone	Arbutin	2.9	271.0	161.0	-60	-10	-9
hydroxycinnamic acid	Cynarin	5.4	515.0	190.9	-54	-40	-11
hydroxycinnamic acid	Ferulic acid	5.5	192.9	134.0	-25	-20	-9
hydroxycinnamic acid	p-coumaric acid	5.4	162.9	119.0	-25	-18	-15
hydroxycinnamyl alcohol	p-Coumaryl alcohol	5.1	149.1	131.0	-25	-14	-13
hydroxy monocarboxylic acid	Caffeoyl-shikimate	4.8	337.0	163.0	120	15	10
indole alkaloid	5-Hydroxydimethyltryptamine	1.2	205.0	160.0	40	19	10
indole alkaloid	N,N-dimethyltryptamine	3.3	189.0	144.0	25	23	16
indole alkaloid	Psilocybin	3.1	285.0	205.0	50	23	10
indole alkaloid	Rauwolscine	5.4	355.0	144.0	100	39	14

(Continued)

10.3389/fpls.2022.1057645

TABLE 1 Continued

Family	Metabolite	RT	Q1	Q3	DP	CE	CXP
indole alkaloid	Reserpine	5.5	609.0	195.0	50	47	10
indole alkaloid	Strychnine	3.5	335.0	184.0	120	49	20
indole alkaloid	Theobromine	3.5	181.0	138.0	85	23	14
indole alkaloid	Theophylline	3.9	181.0	124.0	60	25	14
indole alkaloid	Tryptamine	3.1	161.0	144.0	20	13	10
indole alkaloid	Vincosamide	6.2	499.0	337.0	60	23	18
indole alkaloid	Yohimbine	4.0	355.0	144.0	60	37	16
indole alkaloid	Paynantheine	4.9	397.0	174.0	50	37	20
isoflavone	Calycosin	6.5	285.0	269.9	77	31	28
isoflavone	Glycitein	6.3	284.9	270.1	100	35	30
isoflavone	Ipriflavone	10.2	280.8	239.0	114	27	12
isoflavone	Neobavaisoflavone	8.4	323.0	266.9	75	25	14
isoflavone	Prunetin	8.6	284.9	241.9	150	43	26
isoflavone-glycosylated	Sophoricoside	5.4	433.0	271.0	80	17	14
isoflavone-O-glycoside	Genistein-7-O-glucuronide/Baicalin	6.9	447.0	271.0	120	27	14
isoflavone-O-glycoside	Glycitin	4.7	447.0	224.8	42	59	24
lactone	Caffeic acid	6.9	178.9	135.0	-35	-20	-15
lignan	Arctigenin	7.7	371.0	83.0	-71	-24	-15
lignan	Matairesinol	7.1	357.2	82.9	-76	-26	-9
methylxanthine alkaloid	Caffeine	4.3	195.0	138.0	150	27	14
monocarboxilic acid	Cinnamic acid	7.1	148.8	103.0	20	25	12
monohydroxybenzoic acid	4-Hydroxybenzoic acid	4.6	136.9	93.0	-15	-20	-13
monohydroxybenzoic acid	Salicylic acid	9.1	136.9	93.0	-15	-20	-13
monohydroxybenzoic acid	Vanillic acid	4.8	166.9	108.0	-50	-30	-47
mycotoxin	Neosolaniol	5.0	383.2	365.1	117	13	18
mycotoxin	Roridin-L2	7.1	531.3	249.1	29	21	14
octadecanoid	Oxo-phytodienoic acid	9.6	293.0	275.0	35	15	16
O-methylated isoflavone	Brefeldin A	8.6	281.0	245.0	13	9	14
oxopurine alkaloid	1,3,7-Trimethyluric acid	4.0	209.0	194.0	-40	-18	-9
oxylipin	Jasmonic acid	7.3	209.0	59.0	-80	-16	-27
oxylipin	Methyl jasmonate	8.9	225.0	151.0	20	17	10
phenol	Gingerol	8.4	292.8	99.1	-60	-16	-11
phenolic	Rosmarinic acid	6.6	361.0	163.0	12	11	10
phenolic acid	Sinapic acid	5.4	225.0	175.0	20	19	10
phenolic acid	Syringic acid	4.8	199.0	155.0	20	13	10
phenolic alcohol	Caffeyl alcohol	4.5	165.0	147.0	-20	-16	-7
phenolic alcohol	Coniferyl alcohol	5.2	163.0	131.0	45	13	12

(Continued)

10.3389/fpls.2022.1057645

TABLE 1 Continued

Family	Metabolite	RT	Q1	Q3	DP	CE	CXP
phenolic alcohol	Sinapyl-alcohol	5.1	209.0	194.0	-40	-18	-9
phenolic alcohol	Vanillyl alcohol	4.2	137.0	122.0	20	23	14
phenolic aldehyde	2,5-Dihydroxybenzoic acid	6.8	153.0	108.0	-15	-28	-13
phenolic aldehyde	3,4-Dihydroxybenzaldehyde	4.6	139.0	93.0	50	19	10
phenolic aldehyde	3,4-Dihydroxybenzoic acid	6.2	153.0	108.0	-15	-28	-13
phenolic aldehyde	3,4-Dimethoxycinnamic acid	4.1	209.1	163.0	21	27	16
phenolic aldehyde	3,5-Dihydroxybenzaldehyde	6.2	139.0	111.0	50	15	12
phenolic aldehyde	3,5-Dihydroxybenzoic acid	4.1	153.0	109.0	-15	-18	-13
phenolic aldehyde	3,5-Dimethoxybenzaldehyde	7.9	167.0	139.0	20	15	14
phenolic aldehyde	3,5-Dimethoxybenzoic acid	6.9	183.0	124.0	25	21	14
phenolic aldehyde	4-Hydroxybenzaldehyde	5.3	121.0	92.0	-20	-28	-11
phenolic aldehyde	5-Hydroxyconiferaldehyde	5.2	193.0	178.0	-30	-20	-9
phenolic aldehyde	5-Hydroxyconiferyl-alcohol	5.5	195.0	180.0	-135	-18	-9
phenolic aldehyde	Caffeyl aldehyde	5.3	163.0	135.0	-40	-24	-13
phenolic aldehyde	Coniferaldehyde	6.1	179.0	147.0	30	17	10
phenolic aldehyde	Sinapaldehyde	5.1	209.0	177.0	30	15	10
phenolic aldehyde	Vanillin	5.5	153.0	92.9	50	19	10
phenylpropanoid	5-Hydroxy-ferulic acid	4.8	209.0	150.0	-40	-24	-15
phenylpropanoid	Biochanin A	6.3	283.0	267.9	-120	-30	-13
polyketide-derived mycotoxin	Citrinin	12	251.0	233.0	101	23	12
polyphenol	Ellagic acid	5.1	301.1	284.0	-150	-40	-13
pryridine alkaloid	Trigonelline	1.0	138.0	92.0	170	29	10
pyridoisoquinoline	Emetine	5.5	481.0	246.0	180	47	12
pyrrolizine alkaloid	Erucifoline	2.9	350.2	120.0	125	37	14
sesquiterpene	Abscisic acid	6.4	263.0	153.0	-40	-16	-7
sesquiterpene	Alpha-Cyperone	11.1	219.0	111.0	109	29	12
sesquiterpene	Artemisinin	9.4	283.0	265.1	13	11	12
sesquiterpene lactone	Heptelidic Acid	7.6	279.0	205.0	-33	-12	-11
stilbenoid	Piceid	5.1	389.0	227.0	-90	-20	-11
stilbenoid	Pinosylvin	7.9	213.0	135.0	50	19	14
stilbenoid	Pterostilbene	8.9	255.0	240.0	-80	-26	-13
stilbenoid	Resveratrol	6.3	229.0	135.0	50	19	14
stilbenoid	t-trimethoxyresveratrol	10.4	271.0	152.0	50	73	16
stilbenol	3-Hydroxystilbene	9.2	197.0	119.0	45	17	14
terpenoid indole alkaloid	7-Hydroxymitragynine	4.0	415.0	190.0	60	37	10
trihydroxybenzoic acid	Gallic acid	3.3	169.0	125.0	-30	-20	-13
triterpenoid	Enoxolone	11.0	471.0	189.0	230	45	10

2.6.3 Standard curves: Limit of detection, limit of quantification, and linearity range

To determine the limits of detection (LOD) and quantification (LOQ), and the linearity range, standard curves were generated for each metabolite as previously described (Cocuron et al., 2014; Cocuron et al., 2019) with at least six points. Each standard curve was performed in five replicates.

2.6.4 Recovery efficiency, matrix effect, and accuracy intra- and inter- assay

Recovery efficiency (RE) and accuracy intra- and inter- assay were determined using five coffee leaf pseudoreplicates as previously described (Cocuron et al., 2019). To assess the matrix effect (ME), five coffee leaf pseudoreplicates were used according to the procedure previously published (Cocuron et al., 2017) and the following equation:

$$ME = \left(\frac{Analyte\ peak\ area_{Sample\ spiked\ after\ extraction} - Analyte\ peak\ area_{Sample}}{Average\ analyte\ peak\ area_{External\ standard}} \times 100\,\% \right) - 100\,\%$$

In these conditions, a negative value indicates an ion suppression whereas a positive value depicts an ion enhancement due to ME.

2.7 Statistical analyses

Principal component analysis (PCA), partial-least square discriminant analysis (PLS-DA), heatmap and ANOVA-simultaneous component analysis (ASCA), for both untargeted and targeted analyses, were performed after log-transformation and auto-scaling, using MetaboAnalyst 5.0 (Chong et al., 2018).

3 Results

3.1 Untargeted metabolomics

First, the extraction procedure was optimized using a standard mixture containing alkaloids, cinnamate esters, and flavonoids. For that purpose, different solvents/additive and multiple sonication time/temperature combinations were tested: i) defatting beforehand of after extraction with hexanes; ii) methyl ter-butyl ether, ethyl acetate, different percentages of methanol/water as solvents; iii) acetic acid (1%) as additive; iv) under sonication for 10 to 30 min, at temperatures varying from 25 to 40 °C (data not shown). The extracts were injected in the LC-HR-Q-TOF using the same column and solvents described in Cocuron et al. (2019) but a different gradient (see Materials and Methods section). The method which resulted in the best recovery for the most diverse set of metabolites was using two rounds of extraction: the first one 100% methanol, and the second one methanol/water (30:70, v/v) with sonication rounds of 20 min each at 35-40 °C. This procedure was adopted to extract phytochemicals from coffee and wild *Rubiaceae* leaves.

Leaves from three different *Rubiaceae* species were collected in Costa Rica: i) one commercial coffee (CC), *Coffea arabica* cv. Obatã IAC 1669-20 from a plantation, and ii) two wild *Rubiaceae*, *Isertia hankeana* (WR1) and *Simira maxonii* (WR2) from the rainforest. Leaves were freeze-dried and reduced to powder. Secondary metabolites were extracted from leaf powder using the optimized extraction procedure described above, adding a filtering step , and analyzed *via* LC-HR-Q-TOF in positive and negative modes.

Runs in the positive and negative mode resulted in 19,000 and 24,000 features, respectively. Principal component analysis (PCA) of unprocessed data (wiff files) from the positive mode, obtained by the untargeted analysis of the leaf extracts, resulted in three widely separated clusters, grouping the five pseudoreplicates from each plant species together (Figure 1). The principal component (PC 1) explained 53.7% of the variance, separating the CC from WR1 and WR2, whereas the PC 2 explained 38.8% of the variance, separating all species. This can be interpreted as those three plant species having significantly different metabolic profiles.

After chromatogram alignment, data curation was performed on peak intensities, excluding features with signal lower than one million counts per second and not present on all five pseudoreplicates. Features from positive and negative modes were merged, and when they were present on both modes, only the ones with the highest area were selected. This resulted in 324 features in total within the three species extracts and their pseudoreplicates (Supplemental Table 1). National Institute of Standards and Technology (NIST17, Gaithersburg, MD) and in-house library identified 69 and 81 metabolites in the negative mode and the positive mode, respectively. As previously mentioned, for phytochemicals detected in both modes, the highest area was reported (Table 2). Thirty-one families of phytochemicals were identified: flavone, flavone C-glycoside, flavonol, indole alkaloid, monolignol phenylpropanoid and triterpenoid were the most common families of compounds found in leaf extracts (Table 2). Interestingly, flavones and terpenes were among the families of phytochemicals that were not present in CC but were detected in the wild *Rubiaceae* leaves (Table 2). For instance, 6,2'-dihydroxyflavone, afzelin and apigenin were the flavones not detected in CC leaves extracts. Also, robinin, bisdemethoxycurcumin, and madecassic acid were only detected in the wild *Rubiaceae* leaves WR1 and WR2, while epicatechin, vicenin 2, isoquercetrin, 2,3-dehydrosilybin, theobromine, theophylline, trans-3-coumaric acid, and neomangiferin were only present in CC (Table 2).

To validate the findings from the metabolomic profiling, a targeted metabolomic approach was conducted using known quantities of commercially available phytochemical standards.

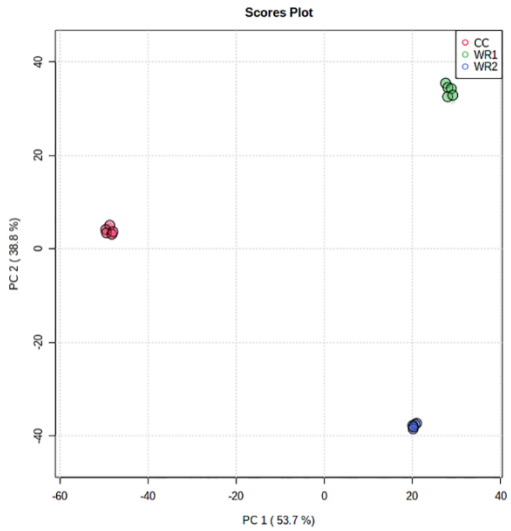


FIGURE 1
PCA analysis performed on features from untargeted metabolomic analysis in the positive mode of leaf extracts of commercial coffee (CC) and two wild *Rubiaceae* species (WR1 and WR2) collected in Costa Rica. The extractions and analyses were performed with five pseudoreplicates per species.

3.2 Targeted metabolomics

3.2.1 Method development

One hundred eighty-four phytochemical standards were individually infused in a triple quadruple to optimize the mass spectrometry parameters (declustering potential - DP, collision energy - CE and collision cell exit potential - CXP) associated with precursor/product ion (Q1/Q3) as shown in Table 1. A multiple reaction monitoring (MRM) scan survey was implemented using these parameters. For the liquid chromatography part, the same column, solvents, and additive than the untargeted metabolomics were used. Mixtures of standards were injected into the LC-MS/MS to obtain the retention time for each phytochemical (Table 1). It is important to note that standards with the same Q1/Q3 transition were separated by their retention times, except for hesperetin and homoeriodictyol, isovitexin and vitexin, vicenin 3 and schaftoside, morin and tricetin, luteolin-7,3'-Di-O-glucoside and kaempferol-3-O-sophoroside, luteolin-7-O-glucoside and Kaempferol-7-O-glucoside, genistein-7-O-glucuronide and baicalin. For the LC-MS/MS analysis of the 184 phytochemicals, polarity switching (between positive and negative modes) was used within the same run, and MRM windows (aka scheduled MRM) were set up for each compound according to its retention time.

A calibration curve was performed for each standard to determine the limits of quantification, detection, and the linearity range (Table 3). The coefficients of correlations were all above 0.98. Acacetin-7-O-rutinoside was the compound with the lowest limits of quantification and detection, at 0.11 and 0.03 fmol, respectively. On the other hand, 2-hydroxyconiferaldehyde was the compound with highest limits of quantification and detection, at 1724.63 and 517.39 fmol, respectively. Tricetin, 2,5-dihydroxybenzoic acid and 4,5-di-O-caffeoylquinic acid were the compounds with the widest range of quantification, from 32.77 to 1,250,000.00 fmol, 81.92 to 1,250,000.00 fmol and 163.84 to 2,500,000.00 fmol, respectively (Table 3).

Recovery efficiency (RE) and matrix effect (ME) were assessed for each metabolite. Ninety percent of the metabolites were recovered with an efficiency of at least 50%. Keracyanin and enoxolone were the compounds with the lowest REs of 23.3 and 28.9%, respectively (Table 4). On the other hand, 5-hydroxytryptophan and 3,4-di-O-caffeoylquinic acid had REs over 100 (Table 4). In parallel, ion suppression or enhancement was measured (i.e. ME > ± 30%) for less than 30% of the phytochemicals. For instance, there was a strong negative ME for 3,4-di-O-caffeoylquinic acid and naringin, underlying ion suppression from the sample matrix. Similarly, 4,5-dicaffeoylquinic acid and gossypetin were the compounds with the highest positive ME, indicating ion enhancement (Table 4).

To verify the method accuracy, at least one representative of each major phytochemical family was added to coffee leaf extracts at three concentrations: 0.25, 0.5, and 1μM. Intra-day and inter-day accuracy percentages were determined by re-injecting the samples on the same day and on three different

10.3389/fpls.2022.1057645

TABLE 2 Families of phytochemical detected by untargeted metabolomics.

Family	Spectral library match	mz/RT (+/-)	CC	WR1	WR2
acyl glycine	Hippuric acid	180.0596/2.1 (-)	3.31E+05 ± 1.24E+04	2.27E+05 ± 2.45E+04	2.06E+05 ± 4.25E+03
amine	Tyramine	138.0913/0.9 (+)	3.37E+06 ± 1.35E+05	5.53E+04 ± 3.27E+03	u.d
benzopyrone	Coumarin	147.0439/10.6 (+)	2.15E+04 ± 3.32E+03	5.58E+05 ± 2.78E+04	5.12E+05 ± 2.93E+04
cinnamate ester	Neochlorogenic acid	353.0877/6.7 (-)	1.08E+07 ± 2.04E+06	1.46E+06 ± 3.03E+05	3.58E+05 ± 7.17E+04
dihyroflavonol	Dihydrokaempferol	287.0563/9.7 (-)	8.64E+04 ± 4.66E+03	1.19E+05 ± 4.98E+03	1.90E+04 ± 1.52E+03
diterpenoid	Gibberellic acid	345.1345/8.4 (-)	u.d	u.d	1.30E+06 ± 9.65E+05
flavan 3-ol	Catechin	289.0715/6.7 (-)	3.99E+05 ± 5.85E+04	1.85E+04 ± 3.20E+03	4.03E+05 ± 1.34E+04
flavan 3-ols	Epicatechin	289.0715/7.0 (-)	1.36E+05 ± 1.57E+04	u.d	u.d
flavanone	Eriodictyol	287.0561/6.8 (-)	9.14E+04 ± 7.19E+03	u.d	2.01E+05 ± 4.65E+03
flavone	6,2'-Dihydroxyflavanone	257.0675/8.3 (-)	u.d	2.16E+06 ± 1.15E+05	9.32E+03 ± 9.54E+02
flavone	Afzelin	431.0978/9.7 (-)	u.d	6.86E+04 ± 7.93E+03	4.07E+04 ± 5.71E+03
flavone	Apigenin	269.0458/11.9 (-)	u.d	u.d	2.69E+05 ± 1.64E+04
flavone	Apigenin 7-glucoside	431.0976/8.8 (-)	2.21E+04 ± 3.10E+03	6.12E+04 ± 4.01E+03	5.06E+03 ± 1.76E+03
flavone	Luteolin	287.0550/8.4 (+)	6.95E+04 ± 9.28E+03	7.96E+05 ± 4.60E+04	2.63E+06 ± 1.19E+05
flavone	Naringenin	273.0755/9.7 (+)	9.78E+04 ± 4.34E+03	4.01E+04 ± 1.16E+04	u.d
flavone C-glycoside	Kaempferol-3-O-rutinoside	593.1525/8.0 (-)	4.36E+04 ± 6.57E+03	3.33E+06 ± 1.38E+05	4.34E+06 ± 2.98E+05
flavone C-glycoside	Kaempferol-7-O-Neohesperidoside	595.1673/8.0 (+)	3.70E+04 ± 5.42E+03	1.98E+06 ± 6.99E+04	3.19E+06 ± 2.32E+05
flavone C-glycoside	Quercetin 3-glucoside	463.0881/8.0 (-)	4.48E+06 ± 3.81E+05	8.91E+04 ± 1.24E+04	1.41E+05 ± 1.16E+04
flavone C-glycoside	Rutin	611.1617/7.7 (+)	1.94E+06 ± 1.45E+05	5.90E+04 ± 5.04E+03	1.54E+04 ± 1.53E+03
flavone C-glycoside	Vicenin 2	595.1518/6.5 (-)	1.64E+06 ± 1.33E+05	u.d	u.d
flavonoid	Tiliroside	593.1311/10.4 (-)	u.d	u.d	6.30E+05 ± 4.19E+04
flavonoid glycoside	Isoquercetrin	465.1034/8.0 (+)	2.31E+06 ± 1.65E+05	u.d	u.d
flavonoid glycoside	Nepetin 7-glucoside	479.1190/9.3 (+)	u.d	u.d	2.77E+05 ± 1.64E+04
flavonol	Kaempferol	285.0401/12.3 (+)	2.72E+04 ± 3.74E+03	1.64E+05 ± 1.01E+04	1.28E+05 ± 5.24E+03
flavonol	Quercitrin	447.0947/8.4 (-)	1.40E+05 ± 4.23E+04	1.95E+06 ± 8.08E+04	5.65E+06 ± 2.06E+05
flavonolignan	2,3-Dehydrosilybin	479.0823/8.0 (-)	1.96E+05 ± 2.74E+04	u.d	u.d
glycosyloxyflavone	Robinin	739.1904/10.2 (-)	u.d	1.18E+06 ± 8.03E+04	5.27E+03 ± 1.06E+03
hydroquinone	3,4-Dihydroxybenzoic acid	153.0196/4.9 (-)	1.05E+06 ± 1.21E+05	2.95E+06 ± 2.65E+05	1.91E+05 ± 7.31E+03
indole alkaloid	Theobromine	181.0715/4.4 (+)	5.73E+06 ± 2.75E+05	u.d	u.d
indole alkaloid	Theophylline	181.0715/5.1 (+)	5.16E+05 ± 2.53E+04	u.d	u.d
indole alkaloid	Yohimbine/Rauwolscine	355.2026/7.1 (+)	1.01E+04 ± 3.91E+03	1.46E+06 ± 6.27E+04	4.89E+04 ± 1.93E+04
indolizine	Isorhyncophylline	385.2123/7.0 (+)	6.02E+04 ± 1.73E+04	3.06E+06 ± 1.15E+05	9.46E+04 ± 3.38E+03
methylxanthine alkaloid	Caffeine	195.0891/6.1 (+)	5.26E+08 ± 1.56E+07	1.50E+05 ± 4.41E+04	1.69E+05 ± 1.89E+04
monohydroxybenzoic acid	Gentisic acid	153.0191/6.0 (-)	1.53E+05 ± 1.10E+04	1.14E+05 ± 6.57E+03	1.03E+05 ± 3.87E+04
monohydroxybenzoic acid	Vanillic acid	167.0344/8.2 (-)	7.58E+04 ± 6.60E+03	2.80E+04 ± 2.22E+03	4.14E+04 ± 4.94E+03
monolignol	Coumaric acid*	163.0403/6.1 (-)	4.51E+06 ± 5.51E+05	1.26E+05 ± 3.99E+03	8.31E+04 ± 2.90E+03
monolignol	Scopoletin	191.0352/8.4 (-)	4.12E+04 ± 4.89E+03	2.81E+05 ± 1.29E+04	2.30E+04 ± 1.40E+03

(Continued)

TABLE 2 Continued

Family	Spectral library match	mz/RT (+/-)	CC	WR1	WR2
monolignol	trans-3-Coumaric acid*	147.0440/7.8 (+)	1.07E+06 ± 1.77E+05	u.d	u.d
oxopurine alkaloid	1,3,7-Trimethyluric acid	209.0684/5.3 (-)	5.81E+05 ± 7.30E+04	3.13E+03 ± 7.31E+02	2.72E+03 ± 2.84E+02
oxylipin	Methyl salicylate	153.0544/7.2 (+)	5.84E+05 ± 2.62E+04	5.59E+04 ± 1.96E+04	1.01E+04 ± 9.88E+02
phenolic aldehyde	2,5-Dihydroxybenzaldehyde	137.0247/6.2 (-)	4.28E+06 ± 1.86E+05	5.52E+05 ± 3.22E+04	1.01E+05 ± 1.23E+04
phenylpropanoid	Benzoic acid	121.0309/7.4 (-)	1.87E+06 ± 8.04E+04	5.18E+06 ± 2.20E+05	5.11E+05 ± 2.90E+04
phenylpropanoid	Esculetin	177.0198/6.6 (-)	1.20E+06 ± 4.69E+05	1.31E+05 ± 1.72E+04	7.18E+04 ± 5.87E+03
phenylpropanoid	Esculin	339.0719/6.4 (-)	3.49E+04 ± 1.28E+04	9.44E+04 ± 2.03E+04	5.19E+04 ± 3.85E+03
phenylpropanoid	p-Methoxycinnamic acid	207.1011/8.1 (+)	7.30E+04 ± 6.47E+03	2.94E+04 ± 1.03E+04	u.d
phenylpropanoid	trans-Cinnamic acid	181.0861/10.7 (+)	u.d	6.00E+06 ± 3.18E+05	u.d
polyphenol	Bisdemethoxycurcumin	309.0964/8.9 (+)	u.d	5.08E+05 ± 3.52E+04	1.08E+04 ± 1.64E+03
polyphenol	Caffeic acid	179.0349/6.7 (-)	4.58E+05 ± 2.24E+04	2.08E+05 ± 1.01E+04	1.25E+05 ± 1.69E+04
pyridine alkaloid	Trigonelline	275.1030/1.0 (+)	2.77E+07 ± 8.40E+05	6.01E+04 ± 9.90E+03	7.92E+03 ± 1.06E+03
sesquiterpene	Abscisic acid	263.1286/10.7 (-)	7.83E+04 ± 4.48E+03	2.96E+04 ± 8.94E+02	1.43E+04 ± 5.80E+02
trihydroxyanthraquinone	Chrysophanol	255.0649/17.6 (+)	u.d	u.d	1.66E+05 ± 1.28E+04
trihydroxybenzoic acid	Gallic acid	169.0140/3.5 (-)	2.23E+05 ± 1.88E+04	4.58E+04 ± 9.22E+03	u.d
triterpene	Sumaresinolic acid	473.3626/11.9 (+)	u.d	u.d	4.91E+05 ± 4.77E+04
triterpenoid	Madecassic acid	503.3371/15.2 (-)	u.d	1.42E+05 ± 2.03E+04	1.39E+06 ± 2.18E+05
triterpenoid	Maslinic acid	471.3469/19.8 (+)	5.32E+04 ± 9.97E+03	4.85E+05 ± 4.70E+04	8.16E+05 ± 4.40E+04
triterpenoid	Soyasaponin I	941.4949/15.3 (-)	u.d	1.01E+05 ± 1.39E+04	u.d
xanthone C-glycoside	Neomangiferin	583.1305/5.3 (-)	4.15E+05 ± 3.03E+04	u.d	u.d

The mass to charge ratio value represents the $[M+H]^+$ and $[M-H]^-$ for the positive and negative modes, respectively, except compounds marked by an asterisk (*) which denotes a loss of water in the source. The areas of secondary metabolites extracted from leaves of commercial coffee (CC) and two wild Rubiaceae species (WR1 and WR2) ± standard deviation (n = 5 pseudoreplicates) are presented. When metabolites were detected in both polarities, only the one with the highest area was reported. Areas labeled as u.d. were under the limit of detection.
For each compound identified using NIST and homemade spectral libraries with > 70% probability, the mass to charge ratio, retention time in min (mz/RT), and mode of detection (positive +, or negative - polarity) are specified.

days, respectively (Table 5). Out of the 25 phytochemicals tested, 18 and 22 had intra-day and inter-day accuracies above 75%, respectively.

3.2.2 Application of the analytical method to quantify phytochemicals in leaf extracts from commercial coffee and wild *Rubiaceae*

Validation of the method was performed using the same biological samples and extracts as the untargeted metabolomics (Supplemental Table 2). From the 184 phytochemicals monitored by LC-MS/MS, 74 were quantifiable in at least one of the *Rubiaceae* species. Interestingly, all the 74 phytochemicals were significantly different in at least one comparison by ANOVA ($p<0.05$). The PCA was consistent with the one from the untargeted metabolomic analysis (Figure 1), resulting in three separate clusters (Figure 2). The principal component 1 (PC 1) explained 70.1% of the variance, separating the CC from the wild *Rubiacaea*, whereas the PC 2 explained 28.9% of the variance, separating all species.

Targeted metabolomic analysis revealed compounds that were highly concentrated in commercial coffee leaf extracts when compared to wild *Rubiaceae* extracts (Figure 3; Supplemental Table 2). Caffeoylquinic acids, like 3, 4 and 5-caffeoylquinic acids, caffeine, trigonelline, vincenin 2, theobromine and others were more abundant in CC (Figure 3). Also, similarly to what was found in the untargeted analysis, some compounds were more abundant and even sometimes found exclusively in wild *Rubiaceae* leaf extracts. For instance, the levels of benzoic acid, 3,4-dimethylcinnamic acid and kaempferol-7-O-neohesperidoside, were the highest in WR1 in comparison to CC and WR2 (Figure 3). Cinnamic acid, yohimbine, corynanthine/rauwolscine, syringic acid, kaempferol were abundant in WR1 but absent in CC (Figure 3). Similarly, the levels of keracyanin, luteolin-7-O-glucoside, idaein,

10.3389/fpls.2022.1057645

TABLE 3 Range, linearity coefficient (R^2), limit of quantification (LOQ) and limit of detection (LOD) of secondary metabolites analyzed in this study.

Metabolite	Range (fmol)	R^2	LOQ (fmol)	LOD (fmol)
1,3,7-Trimethyluric acid	32.77 – 125,000.00	0.9985	4.54	1.36
2,5-Dihydroxybenzoic acid	81.92 – 1250,000.00	0.9995	52.85	15.86
3,4-Dihydroxybenzaldehyde	131.07 – 32,000.00	0.9941	46.81	14.04
3,4-Dihydroxybenzoic acid	327.68 – 32,000.00	0.9935	62.30	18.69
3,4-Dimethoxycinnamic acid	65.54 – 16,000.00	0.9989	22.91	6.87
3,4-di-O-Caffeoylquinic acid	819.20 – 80,000.00	0.9951	364.09	109.23
3,5-Dihydroxybenzaldehyde	327.68 – 1,250,000.00	0.9983	117.87	35.36
3,5-Dimethoxybenzaldehyde	32.77 – 20,000.00	0.9977	9.55	2.87
3,5-Dimethoxybenzoic acid	13.11 – 8,000.00	0.9923	4.68	1.40
3-Caffeoylquinic acid	163.84 – 40,000.00	0.9965	81.92	24.58
3-Deoxyrobinetin	65.54 – 100,000.00	0.9945	26.43	7.93
3-Hydroxystilbene	327.68 – 200,000.00	0.9852	106.74	32.02
4,5-di-O-Caffeoylquinic acid	163.84 – 2,500,000.00	0.9937	148.95	44.68
4-Ccaffeoylquinic acid	163.84 – 100,000.00	0.9988	68.99	20.70
4-Hydroxybenzaldehyde	16.38 – 10,000.00	0.9939	5.43	1.63
4-Hydroxybenzoic acid	32.77 – 8,000.00	0.9994	10.05	3.02
5-Caffeoylquinic acid	65.54 – 250,000.00	0.9928	40.96	12.29
5-Hydroxyconiferaldehyde	4,096.00 – 2,500,000.00	1.0000	1,724.63	517.39
5-Hydroxydimetyltryptamine	131.07 – 5,120.00	0.9920	46.32	13.89
5-Hydroxytryptophan	16.36 – 62,425.00	0.9936	5.20	1.56
6-Methylcoumarin	16.38 – 4,000.00	0.9990	7.99	2.40
7,8-Dihydroxy-4-methylcoumarin	327.68 – 500,000.00	0.9948	114.98	34.49
7-Hydroxymitragynine	163.84 – 100,000.00	0.9970	59.15	17.74
Abscisic acid	16.38 – 10,000.00	0.9978	3.60	1.08
Acacetin-7-O-rutinoside	0.33 – 500.00	0.9957	0.11	0.03
Afzelin	1.31 – 2,000.00	0.9994	0.31	0.09
alpha-Cyperone	16.38 – 4,000.00	0.9979	2.42	0.72
Anthranilic acid	131.07 – 200,000.00	0.9967	80.41	24.12
Apigenin	32.77 – 8,000.00	0.9955	20.61	6.18
Apigenin-7-glucuronide	1.31 – 5,000.00	0.9997	0.45	0.13
Apigenin-7-O-glucoside	3.17 – 1,936.00	0.9971	1.21	0.36
Apigeninidin	64.93 – 39,628.00	0.9984	13.09	3.93
Arbutin	327.00 – 80,000.00	0.9957	52.10	15.63
Arctigenin	327.29 – 79,905.60	0.9928	87.51	26.25
Artemisinin	13.11 – 8,000.00	0.9966	7.62	2.29
Baicalein	3.28 – 5,000.00	0.9993	2.05	0.61

(Continued)

10.3389/fpls.2022.1057645

TABLE 3 Continued

Metabolite	Range (fmol)	R^2	LOQ (fmol)	LOD (fmol)
Baicalin	32.77 – 20,000.00	0.9971	14.83	4.45
Biochanin A	32.77 – 20,000.00	0.9957	20.10	6.03
Brefeldin A	131.07 – 80,000.00	0.9901	42.42	12.73
Caffeic acid	131.07 – 500,000.00	0.9983	48.55	14.56
Caffeine	3.28 – 800.00	0.9984	1.12	0.34
Caffeoyl-shikimate	163.84 – 250,000.00	0.9983	85.11	25.53
Caffeyl alcohol	16.38 – 25,000.00	1.0000	10.40	3.12
Caffeyl aldehyde	6.55 – 25,000.00	0.9961	6.43	1.93
Calycosin	6.55 – 4,000.00	0.9936	1.59	0.48
Catechin	16.38 – 10,000.00	0.9928	2.43	0.73
Chrysin	16.38 – 10,000.00	0.9976	5.67	1.70
Chrysoeriol	3.28 – 800.00	0.9968	1.34	0.40
Cinnamic acid	32.77 – 8,000.00	0.9992	14.63	4.39
Coniferaldehyde	16.38 – 10,000.00	0.9902	4.88	1.46
Coniferyl alcohol	65.54 – 16,000.00	0.9975	24.82	7.45
Corynanthine	1.31 – 320.00	0.9971	0.33	0.10
Coumaric acid	16.38 – 10,000.00	0.9984	7.95	2.39
Cynarin	1,310.72 – 2,000,000.00	0.9910	468.11	140.43
Dihydrokaempferol	13.11 – 3,200.00	0.9902	3.03	0.91
Dihydroquercetin	32.77 – 125,000.00	0.9935	26.21	7.86
Diosmetin	1.31 – 800.00	0.9935	0.56	0.17
Enoxolone	65.54 – 16,000.00	0.9945	10.39	3.12
Epicatechin	3.28 – 5,000.00	0.9952	1.38	0.41
Epigallocatechin	65.54 – 40,000.00	0.9955	11.13	3.34
Eriodictyol	6.55 – 4,000.00	0.9886	1.20	0.36
Erucifoline	16.38 – 25,000.00	0.9970	7.99	2.40
Eupatorin	3.28 – 800.00	0.9930	0.99	0.30
Ferulic acid	65.54 – 16,000.00	0.9914	13.43	4.03
Fisetin	65.54 – 250,000.00	0.9982	20.23	6.07
Flavopiridol	1.31 – 2,000.00	0.9964	0.58	0.17
Gallic acid	163.84 – 625,000.00	0.9944	43.57	13.07
Gallocatechin	65.54 – 40,000.00	0.9957	12.85	3.86
Genistein-7-O-glucuronide	3.28 – 5,000.00	0.9936	1.30	0.39
Genkwanin	6.55 – 4,000.00	0.9933	2.72	0.82
Gibberellic acid	32.77 – 20,000.00	0.9960	11.22	3.37
Gingerol	1,310.72 – 32,000.00	0.9934	379.92	113.98
Ginkgolide A	16.38 – 4,000.00	0.9978	3.19	0.96

(Continued)

10.3389/fpls.2022.1057645

TABLE 3 Continued

Metabolite	Range (fmol)	R^2	LOQ (fmol)	LOD (fmol)
Glycitein	6.55 – 640.00	0.9953	2.32	0.70
Glycitin	64.49 – 15,744.00	0.9869	15.10	4.53
Harmane	1.31 – 800.00	0.9928	0.44	0.13
Hesperetin	3.28 – 2,000.00	0.9975	1.22	0.37
Homoeriodictyol	3.28 – 800.00	0.9998	1.21	0.36
Hordenine	3.28 – 2,000.00	0.9995	0.83	0.25
Idaein	163.84 – 250,000.00	0.9999	159.84	47.95
Ipriflavone	1.31 – 128.00	0.9989	0.29	0.09
Isoorientin	32.77 – 20,000.00	0.9991	12.90	3.87
Isophorone	6.47 – 632.32	0.9812	1.32	0.39
Isoquercetrin	6.55 – 4,000.00	0.9979	2.16	0.65
Isorhamnetin	6.55 – 1,600.00	0.9969	1.50	0.45
Isorhamnetin-3-O-glucoside	1.13 – 5,000.00	0.9964	0.81	0.24
Isosakuranetin	3.28 – 2,000.00	0.9923	1.21	0.36
Isoschaftoside	1.31 – 5,000.00	0.9921	0.52	0.16
Isovitexin	1.26 – 4,800.00	0.9969	0.92	0.28
Jasmonic acid	32.77 – 20,000.00	0.9898	6.39	1.92
Kaempferide	6.55 – 10,000.00	0.9924	2.31	0.69
Kaempferitrin	6.55 – 1,600.00	0.9989	3.58	1.07
Kaempferol	32.77 – 8,000.00	0.9986	13.43	4.03
Kaempferol-3-O-glucoside	6.55 – 640.00	0.9967	3.01	0.90
Kaempferol-3-O-glucuronide	6.55 – 25,000.00	0.9974	2.29	0.69
Kaempferol-3-O-rutinoside	6.55 – 1,600.00	0.9996	2.40	0.72
Kaempferol-3-O-sophoroside	20.48 – 5,000.00	0.9997	8.40	2.52
Kaempferol-7-O-glucoside	3.28 – 800.00	0.9988	0.78	0.24
Kaempferol-7-O-Neohesperidoside	3.28 – 2,000.00	0.9946	1.27	0.38
Keracyanin	32.77 – 20,000.00	0.9992	7.38	2.21
Luteolin	13.11 – 8,000.00	0.9983	4.24	1.27
Luteolin-7,3'-Di-O-glucoside	3.28 – 128.00	0.9996	1.84	0.55
Luteolin-7-O-glucoside	1.31 – 2,000.00	0.9927	0.67	0.20
Luteolin-7-O-glucuronide	3.28 – 2,000.00	0.9923	1.62	0.49
Matairesinol	131.07 – 32,000.00	0.9939	40.96	12.29
Mellein	13.11 – 3,200.00	0.9996	3.96	1.19
Methyl-jasmonate	0.33 – 12.80	0.9878	0.12	0.04
Miquelianin	16.38 – 25,000.00	0.9985	9.36	2.81
Mitragynine	6.55 – 25,000.00	0.9977	4.12	1.24
Morin	32.77 – 20,000.00	0.9963	12.90	3.87

(Continued)

TABLE 3 Continued

Metabolite	Range (fmol)	R^2	LOQ (fmol)	LOD (fmol)
Myricetin	327.68 – 1,250,000.00	0.9900	76.20	22.86
Myricetin-3-O-Rhamnoside	32.77 – 125,000.00	0.9925	9.13	2.74
N,N-Dimethyltryptamine	0.63 – 2,389.50	0.9925	0.19	0.06
Naringenin	6.55 – 4,000.00	0.9950	2.86	0.86
Naringenin-7-O-glucoside	1.31 – 800.00	0.9984	0.30	0.09
Naringin	1.31 – 2,000.00	0.9994	0.78	0.23
Neobavaisoflavone	1.31 – 320.00	0.9983	0.33	0.10
Neodiosmin	1.31 – 2,000.00	0.9999	0.46	0.14
Orientin	16.38 – 10,000.00	0.9866	6.94	2.08
Paynantheine	13.11 – 1,280.00	0.9915	6.43	1.93
p-Coumaraldehyde	131.07 – 12,800.00	0.9928	28.81	8.64
p-Coumaryl-alcohol	131.07 – 32,000.00	0.9915	37.24	11.17
Phloretin	3.28 – 2,000.00	0.9939	1.10	0.33
Piceid	32.77 – 20,000.00	0.9936	7.67	2.30
Pinosylvin	32.77 – 20,000.00	0.9995	14.50	4.35
Prunetin	6.55 – 4,000.00	0.9954	5.46	1.64
Psilocybin	131.07 – 80,000.00	0.9982	43.84	13.15
Pterostilbene	13.11 – 8,000.00	0.9922	3.34	1.00
Quercetin	32.77 – 8,000.00	0.9903	7.20	2.16
Quercetin-3,4'-O-diglucoside	3.28 – 800.00	0.9955	0.74	0.22
Quercetin-3-O-galactoside	16.38 – 10,000.00	0.9921	5.57	1.67
Quercitrin	16.38 – 10,000.00	0.9975	9.93	2.98
Rauwolscine	1.31 – 320.00	0.9954	0.32	0.09
Reserpine	3.28 – 5,000.00	0.9938	1.21	0.36
Resveratrol	65.54 – 40,000.00	0.9935	18.41	5.52
Rhamnazin	6.55 – 4,000.00	0.9988	2.32	0.70
Rhamnetin	13.11 – 8,000.00	0.9971	3.05	0.91
Roridin L2	16.26 – 9,926.40	0.9937	7.49	2.25
Rutin	13.11 – 8,000.00	0.9912	3.23	0.97
Sakuranetin	3.28 – 2,000.00	0.9960	1.17	0.35
Salicylic acid	16.38 – 10,000.00	0.9939	4.45	1.34
Schaftoside	3.28 – 5,000.00	0.9950	1.70	0.51
Scopoletin	65.54 – 250,000.00	0.9953	29.93	8.98
Scutellarein	163.84 – 16,000.00	0.9984	77.10	23.13
Seneciphylline	6.55 – 4,000.00	0.9950	2.23	0.67
Sinapaldehyde	32.77 – 8,000.00	0.9975	18.94	5.68
Sinapic acid	131.07 – 32,000.00	0.9949	67.56	20.27

(Continued)

TABLE 3 Continued

Metabolite	Range (fmol)	R^2	LOQ (fmol)	LOD (fmol)
Sinapyl-alcohol	1,310.00 – 2,000,000.00	0.9939	471.48	141.44
Sophoricoside	3.28 – 2,000.00	0.9958	1.20	0.36
Spiraeoside	3.28 – 5,000.00	0.9962	1.86	0.56
Strychnine	16.38 – 10,000.00	0.9974	4.38	1.31
Swertiajaponin	13.11 – 8,000.00	0.9978	4.52	1.36
Swertisin	6.55 – 1,600.00	0.9973	3.58	1.07
Syringic acid	65.54 – 40,000.00	0.9979	30.06	9.02
Tamarixetin	6.55 – 4,000.00	0.9936	1.46	0.44
Theobromine	6.55 – 4,000.00	0.9990	2.32	0.69
Theophylline	3.28 – 2,000.00	0.9992	0.84	0.25
Tomatidine hydrochloride	6.55 – 4,000.00	0.9906	2.13	0.64
Tricetin	32.77 – 1,250,000.00	0.9994	8.38	2.51
Tricin	1.31 – 2,000.00	0.9972	0.58	0.17
Trigonelline	1.31 – 2,000.00	0.9810	0.42	0.12
Tryptamine	16.38 – 4,000.00	0.9987	5.73	1.72
t-Trimethoxyresveratrol	16.38 – 4,000.00	0.9962	4.72	1.42
Tyramine	13.11 – 3,200.00	0.9997	5.24	1.57
Vanillic acid	327.68 – 32,000.00	0.9942	109.23	32.77
Vanillin	16.38 – 10,000.00	0.9943	4.89	1.47
Vanillyl-alcohol	32.77 – 8,000.00	0.9992	10.30	3.09
Vicenin 3	3.28 – 5,000.00	0.9907	1.45	0.43
Vicenin 2	6.55 – 1,600.00	0.9993	2.02	0.61
Vincosamide	1.31 – 800.00	0.9972	0.38	0.11
Vitexin	1.31 – 2,000.00	0.9948	0.60	0.18
Xanthohumol	3.28 – 800.00	0.9971	0.89	0.27
Yohimbine	1.31 – 320.00	0.9966	0.45	0.14

gibberellic acid, sinalpadehyde and sinapyl-alcohol were found to be the highest in WR2, while harmane and vincosamide, abundant in WR2, were not detected in CC (Figure 3).

To determine the consistency between untargeted and targeted results, an ANOVA-simultaneous comparative analysis (ASCA) was performed on the 74 polyphenols quantified in our samples, and it revealed that both techniques display similar results. Figure 4 are three examples of phytochemicals—caffeine, vicenin and 3-caffeoylquinic acid (3-CQA)—that were found to be significantly more abundant in CC using the untargeted (U) metabolomics. These results were validated by a quantitative targeted (T) approach, which confirmed that the three polyphenols were the highest in CC. These examples illustrate the consistency between the untargeted and targeted metabolomics in our pipeline. Figure 5 depicts the workflow of the metabolomic platform to identify and quantify polyphenols using liquid chromatography mass spectrometry.

4 Discussion

Plant secondary metabolites belong to a wide variety of chemical families and are present at very small concentrations, which makes it challenging to recover with a single extraction procedure and to quantify with a sensitive enough analytical technique. Moreover, most phytochemicals are poorly soluble in water, photosensitive and not thermostable, requiring careful

10.3389/fpls.2022.1057645

TABLE 4 Recovery efficiency (RE) of secondary metabolites and matrix effect (ME) from coffee leaf extracts.

Family	Metabolite	RE	ME
		(%)	(%)
alkaloid	1,3,7-trimethyluric acid	89.90	-24.20
alkaloid	caffeine	76.10	-29.22
alkaloid	corynanthine	56.00	-24.06
alkaloid	erucifoline	61.10	-6.21
alkaloid	harmane	80.30	-23.95
alkaloid	hordenine	72.30	-44.83
alkaloid	mitragynine	46.70	3.28
alkaloid	seneciphylline	61.40	26.53
alkaloid	tomatidine	35.80	-3.46
amino acid derivative	5-hydroxytryptophan	102.70	-44.92
amino acid derivative	tyramine	74.80	-44.75
anthocyanidin	apigeninidin	66.70	13.19
anthocyanin glycoside	keracyanin	23.30	-7.01
cinnamate ester	3,4-di-O-caffeoylquinic acid	194.30	-182.44
cinnamate ester	4,5-dicaffeoylquinic acid	57.20	122.65
cinnamate ester	4-caffeoylquinic acid	82.20	75.55
cinnamic acid derivative	3,4-dimethoxycinnamic acid	76.50	-25.81
coumarin derivative	6-methylcoumarin	78.10	-24.91
coumarin derivative	7,8-dihydroxy-4-methylcoumarin	79.50	-19.74
coumarin derivative	esculetin	82.60	-3.48
coumarin derivative	mellein	81.20	-20.56
coumarin derivative	scopoletin	68.70	-24.87
dihydroxybenzoic acid	2,5-dihydroxybenzoic acid	77.10	-15.47
dihydroxybenzoic acid	3,4-dihydroxybenzoic acid	76.20	0.76
dihydroxybenzoic acid	3,5-dimethoxybenzoic acid	80.50	-24.28
dimethoxybenzene	3,5-dimethoxybenzaldehyde	82.40	-21.11
diterpenoid	gibberellic acid	66.70	-35.04
diterpenoid	ginkgolide A	58.90	-24.42
flavanol	catechin	81.60	-23.50
flavanol	epicatechin	83.10	15.05
flavanol	epigallocatechin	73.60	-85.83
flavanone	eriodictyol	73.10	-29.82
flavanone	hesperetin	80.20	-20.55
flavanone	homoeriodictyol	72.90	-18.14
flavanone	isosakuranetin	80.10	-20.79

(Continued)

10.3389/fpls.2022.1057645

TABLE 4 Continued

Family	Metabolite	RE (%)	ME (%)
flavanone	naringenin	74.30	-22.78
flavanone O-glycoside	naringenin-7-O-glucoside	63.00	-35.42
flavanone-C-glycoside	swertiajaponin	56.20	21.92
flavanone-O-glycoside	naringin	45.80	-90.06
flavanonol	dihydrokaempferol	71.60	-23.12
flavanonol	dihydroquercetin	69.40	-24.51
flavone	3-deoxyrobinetin	60.00	53.64
flavone	apigenin	78.90	-28.58
flavone	apigenin-7-glucuronide	54.20	-16.12
flavone	baicalein	55.20	7.23
flavone	chrysin	87.30	-34.61
flavone	chrysoeriol	78.80	-21.68
flavone	diosmetin	69.80	-16.06
flavone	eupatorin	77.50	-21.60
flavone	flavopiridol	72.20	80.23
flavone	genkwanin	76.90	-22.52
flavone	luteolin	68.80	-12.86
flavone	myricetin	51.00	-18.64
flavone	scutellarein	50.40	51.01
flavone	tectochrysin	67.80	-18.77
flavone	tricetin	53.80	36.58
flavone C-glycoside	isoorientin	53.50	4.16
flavone C-glycoside	swertisin	55.60	-15.14
flavone-C-glycoside	isoschaftoside	66.20	-22.12
flavone-C-glycoside	isovitexin/vitexin	58.80	-16.56
flavone-C-glycoside	orientin	52.10	11.89
flavone-C-glycoside	Vicenin 2	42.00	-10.71
flavone-O-glycoside	acacetin-7-O-rutinoside	61.80	-9.30
flavone-O-glycoside	apigenin-7-O-glucoside	56.90	-25.08
flavone-O-glycoside	baicalin	62.00	-20.13
flavone-O-glycoside	luteolin-7,3'-di-O-glucoside	64.50	-32.82
flavone-O-glycoside	luteolin-7-O-glucoside	42.40	-27.38
flavone-O-glycoside	luteolin-7-O-glucuronide	69.20	-12.69
flavone-O-glycoside	myricetin-3-O-rhamnoside	44.60	47.77
flavone-O-glycoside	neodiosmin	43.60	-88.65
flavonol	fisetin	72.80	17.56

(Continued)

10.3389/fpls.2022.1057645

TABLE 4 Continued

Family	Metabolite	RE (%)	ME (%)
flavonol	gossypetin	38.30	186.65
flavonol	isorhamnetin	57.70	-13.77
flavonol	kaempferide	73.60	-20.18
flavonol	kaempferol	69.30	-21.47
flavonol	morin	54.00	28.56
flavonol	quercetin	67.10	-17.24
flavonol	quercitrin	58.20	-2.56
flavonol	rhamnazin	60.80	-15.60
flavonol	rhamnetin	61.90	-12.18
flavonol	tamarixetin	66.90	-18.37
flavonol-O-glycoside	afzelin	68.60	-26.06
flavonol-O-glycoside	isoquercetrin	68.40	16.38
flavonol-O-glycoside	isorhamnetin-3-O-glucoside	68.50	2.12
flavonol-O-glycoside	kaempferitrin	52.00	0.79
flavonol-O-glycoside	kaempferol-3-O-glucoside	60.20	28.18
flavonol-O-glycoside	kaempferol-3-O-glucuronide	71.20	-18.36
flavonol-O-glycoside	kaempferol-7-O-glucoside	56.30	-29.40
flavonol-O-glycoside	kaempferol-7-O-neohesperidoside	51.30	2.93
flavonol-O-glycoside	quercetin-3,4'-O-diglucoside	49.20	-21.71
flavonol-O-glycoside	quercetin-3-O-galactoside	53.00	-0.89
flavonol-O-glycoside	rutin	33.80	-48.97
flavonol-O-glycoside	spiraeoside	55.40	-20.78
hydroxybenzoic acid	4-hydroxybenzoic acid	83.00	-36.35
hydroxybenzoic acid	gallic acid	76.40	-19.35
hydroxybenzoic acid	salicylic acid	51.40	-74.19
hydroxybenzoic acid	vanillic acid	82.00	-78.18
hydroxycinnamic acid	caffeic acid	65.30	-17.55
hydroxycinnamic acid	coumaric acid	80.20	-25.62
hydroxycinnamic acid	ferulic acid	75.40	-18.89
indole alkaloid	5-hydroxydimethyltryptamine	91.50	-37.98
indole alkaloid	7-hydroxy mitragynine	59.20	-20.88
indole alkaloid	anthranilic acid	72.60	48.89
indole alkaloid	N,N-dimethyltryptemine	79.30	-0.45
indole alkaloid	psilocybin	82.50	45.11
indole alkaloid	rauwolscine	57.20	-22.77
indole alkaloid	reserpine	30.70	0.24

(Continued)

10.3389/fpls.2022.1057645

TABLE 4 Continued

Family	Metabolite	RE (%)	ME (%)
indole alkaloid	theobromine	76.50	-17.88
indole alkaloid	theophylline	77.00	-25.19
indole alkaloid	tryptamine	72.60	-42.00
isoflavone	biochanin A	74.50	-24.03
isoflavone	calycosin	70.90	-21.78
isoflavone	glycitein	81.40	-22.87
isoflavone	ipriflavone	66.40	-24.26
isoflavone	neobavaisoflavone	65.30	-21.12
isoflavone	prunetin	75.60	-17.48
isoflavone-O-glycoside	genistein-7-O-glucuronide	69.30	-19.96
isoflavone-O-glycoside	glycitin	53.50	-12.45
isoflavone-O-glycosylated	sophoricoside	59.00	-43.91
lactone	brefeldin A	68.30	-33.62
lignan	arctigenin	63.50	-21.16
lignan	matairesinol	64.90	-24.19
monocarboxilic acid	cinnamic acid	80.50	-21.71
mycotoxin	neosolaniol	87.90	36.21
mycotoxin	roridin-L2	36.50	-22.76
oxylipin	jasmonic acid	73.30	-25.55
oxylipin	methyl jasmonate	95.30	-30.97
phenol	gingerol	68.80	-20.27
phenolic acid	benzoic acid	90.80	-36.14
phenolic acid	sinapic acid	69.60	-14.47
phenolic acid	syringic acid	80.10	-55.23
phenolic alcohol	caffeyl alcohol	81.80	-33.98
phenolic alcohol	coniferyl alcohol	77.60	-20.35
phenolic alcohol	sinapyl alcohol	67.90	-17.88
phenolic alcohol	vanillyl alcohol	83.70	-22.66
phenolic aldehyde	3,4-dihydroxycbenzaldehyde	93.40	41.47
phenolic aldehyde	4-hydroxybenzaldehyde	78.50	-70.69
phenolic aldehyde	caffeyl aldehyde	77.70	-49.70
phenolic aldehyde	coniferaldehyde	77.40	-20.36
phenolic aldehyde	sinapaldehyde	70.10	-56.06
phenolic aldehyde	vanillin	79.60	-24.83
sesquiterpene	abscisic acid	60.60	-29.79
sesquiterpene	artemisinin	49.30	-22.97

(Continued)

10.3389/fpls.2022.1057645

TABLE 4 Continued

Family	Metabolite	RE (%)	ME (%)
sesquiterpene	heptelidic acid	68.30	-20.29
sesquiterpene	α-cyperone	70.30	-19.83
stilbenoid	3-hydroxystilbene	78.20	-23.62
stilbenoid	piceid	50.90	-7.63
stilbenoid	pinosylvin	77.80	-18.70
stilbenoid	pterostilbene	66.80	-33.96
stilbenoid	resveratrol	64.10	-32.92
stilbenoid	t-trimethoxyresveratrol	58.60	-19.81
triterpenoid	enoxolone	28.90	-25.84

TABLE 5 Intra and inter-day accuracy percentages for secondary metabolites added at three different concentrations (0.25, 0.5 and 1 μM) to coffee leaf extracts.

		Intra-day assay	Inter-day assay
Metabolite	Concentration (μM)	n = 5	n=15
1,3,7-trimethyluric acid	0.25	11.3	16.0
	0.50	25.2	19.9
	1.00	16.3	15.9
3,4-dimethoxycinnamic acid	0.25	3.0	2.1
	0.50	3.0	2.6
	1.00	1.0	0.6
3-Hydroxystilbene	0.25	27.7	12.2
	0.50	6.1	6.8
	1.00	2.2	2.9
4-Hydroxybenzoic acid	0.25	14.5	17.0
	0.50	15.5	18.5
	1.00	14.0	17.1
5-Hydroxytryptophan	0.25	6.5	4.5
	0.50	46.1	28.7
	1.00	43.6	31.8
Apigeninidin	0.25	7.5	5.4
	0.50	9.5	9.2
	1.00	11.4	11.0
Arctigenin	0.25	7.1	7.7
	0.50	3.4	3.7
	1.00	2.6	3.4

(Continued)

TABLE 5 Continued

		Intra-day assay	Inter-day assay
Metabolite	Concentration (μM)	n = 5	n=15
Biochanin A	0.25	6.1	8.5
	0.50	3.6	4.8
	1.00	1.0	1.9
Coniferaldehyde	0.25	7.9	7.0
	0.50	6.4	5.8
	1.00	3.0	3.5
Coniferyl alcohol	0.25	30.8	26.1
	0.50	28.3	23.1
	1.00	18.6	17.1
Cyperone	0.25	7.9	3.2
	0.50	1.8	2.3
	1.00	1.0	1.6
Dihydrokaempferol	0.25	5.9	5.3
	0.50	7.6	7.5
	1.00	6.2	5.7
Epicatechin	0.25	29.3	18.5
	0.50	31.7	23.6
	1.00	4.5	9.2
Ferulic acid	0.25	15.3	11.3
	0.50	17.1	14.6
	1.00	1.6	4.6
Gibberellic acid	0.25	5.5	8.9
	0.50	1.5	7.7
	1.00	4.3	8.2
Genistein-7-O-glucuronide	0.25	3.4	8.2
	0.50	1.8	8.2
	1.00	7.6	7.3
Isorhamnetin-3-O-glucoside	0.25	8.4	5.4
	0.50	4.6	5.3
	1.00	0.5	1.7
Jasmonic acid	0.25	6.5	6.7
	0.50	2.6	4.1
	1.00	2.4	3.0
Luteolin	0.25	1.1	4.9
	0.50	4.0	5.2
	1.00	5.0	3.8

(Continued)

10.3389/fpls.2022.1057645

TABLE 5 Continued

		Intra-day assay	Inter-day assay
Metabolite	Concentration (μM)	n = 5	n=15
Naringenin	0.25	3.1	3.3
	0.50	0.1	1.6
	1.00	2.5	1.6
Orientin	0.25	8.1	15.1
	0.50	8.5	20.3
	1.00	14.0	17.0
Scopoletin	0.25	4.5	4.9
	0.50	6.7	7.5
	1.00	1.6	3.2
Sinapic acid	0.25	30.1	20.4
	0.50	28.7	24.3
	1.00	4.2	6.5
Swertiajaponin	0.25	0.8	25.7
	0.50	6.0	26.9
	1.00	41.4	40.1
Theobromine	0.25	35.4	17.9
	0.50	16.7	12.5
	1.00	5.2	9.1

Intra-day and inter-day accuracy percentages were determined by re-injecting the samples on the same day (n = 5) and on three different days (n = 15), respectively.

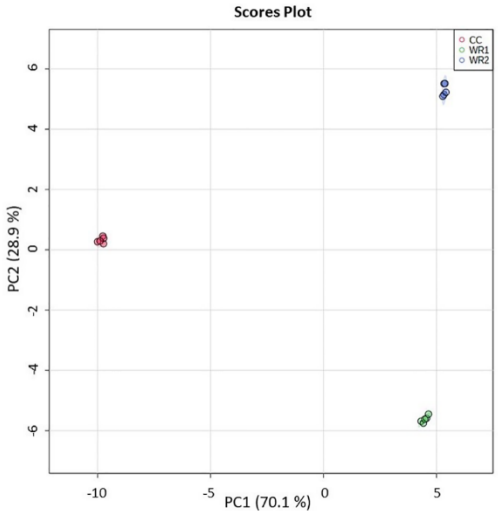


FIGURE 2
PCA analysis performed on targeted metabolomic analysis data from extracts of commercial coffee (CC) and two wild *Rubiaceae* species (WR1 and WR2). The extractions and analyses were performed with five pseudoreplicates per species. The shaded regions represent the 95% confidence intervals.

10.3389/fpls.2022.1057645

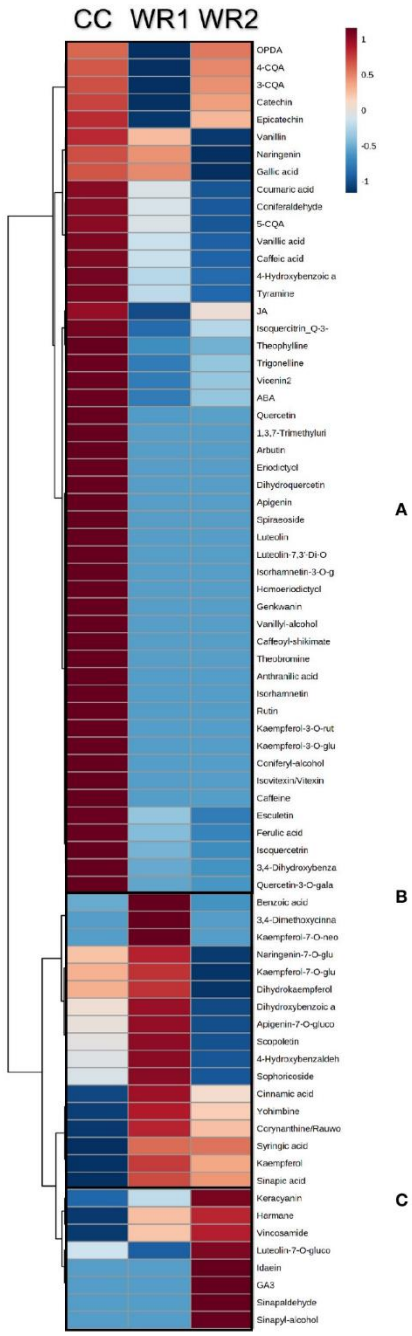


FIGURE 3
Heatmap analysis of the polyphenols from extracts of coffee (CC) and two wild *Rubiaceae* (WR1 and WR2) detected by targeted metabolomics. The metabolomics data were normalized by log transformation, mean-centered, and divided by the standard deviation of each variable. Ward's hierarchical clustering algorithm was used to group metabolites that have the same distribution pattern in the heat map. Color scale represents metabolite relative intensity, with the darkest red and blue symbolizing the highest and the lowest values, respectively. Boxes highlight metabolites that are higher in CC **(A)**, in WR1 **(B)**, and WR2 **(C)**.

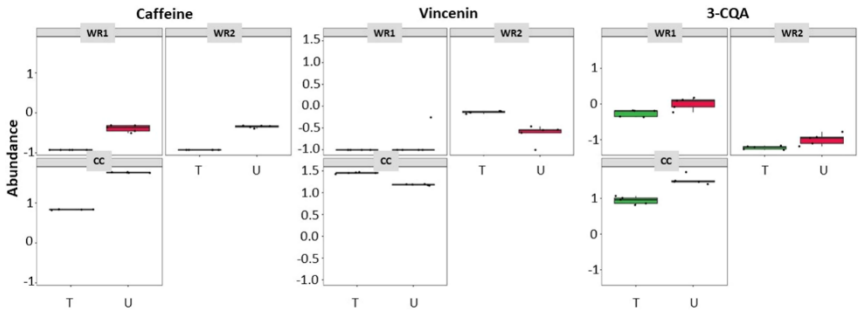


FIGURE 4
ASCA showing similarities of metabolites detected in coffee and wild *Rubiaceae* leaf extracts detected by targeted (T - green boxes) and untargeted (U – red boxes) metabolomics. 3-CQA is 3-caffeoylquinic acid.

procedures for their handling and storage. Choosing the right solvent composition and solid to liquid ratio is critical for isolating plant metabolites (Zhang et al., 2011). In general, the most common solvents are ethanol, methanol, chloroform, and water in different proportions for plant secondary metabolite extraction (Silva et al., 1998; Abubakar and Haque, 2020). Nonetheless, chloroform has a low polarity and is a carcinogen, therefore it is recommended to avoid it (Davidson et al., 2008). In the present study, methanol and water resulted in the most efficient mixture of solvents to recover polyphenols from coffee leaves, combined with sonication at 35-40°C. Indeed, ultrasonication has been widely used to reach higher yields of natural compounds (Annegowda et al., 2010; Masson et al., 2010; Hasan et al., 2017). For instance, ultrasonic-assisted extraction has been used in coffee leaves to improve the extraction of caffeine, trigonelline, rutin, chlorogenic acids, and mangiferin (Chen et al., 2020b). Furthermore, steroidal alkaloids have been successfully recovered from potato peel using ultrasound assisted extraction, obtaining at least 1.5 times more compounds in comparison with other extraction technique (Hossain et al., 2014). A comparative study in *Hibiscus* spp. concluded that using methanol and sonication resulted in better yields of phytosterols (Soares Melecchi et al., 2006). Sonication was also used to successfully extract quinones and flavonoids of six different species of *Dosera* sp. (Marczak et al., 2005). Other extraction techniques cited in literature involve microwave-assisted extraction, pressurized-liquid extraction and supercritical fluid extraction (Zhang et al., 2011; Abubakar and Haque, 2020). However, those are more time consuming and costly techniques (Roopashree and Naik, 2019).

Untargeted metabolomics aims to capture the whole metabolome (Matsuda et al., 2009), while targeted analysis focuses on the use of commercially available standards to detect and measure the quantity of metabolites present in a biological sample (Shimizu et al., 2018). It is therefore essential

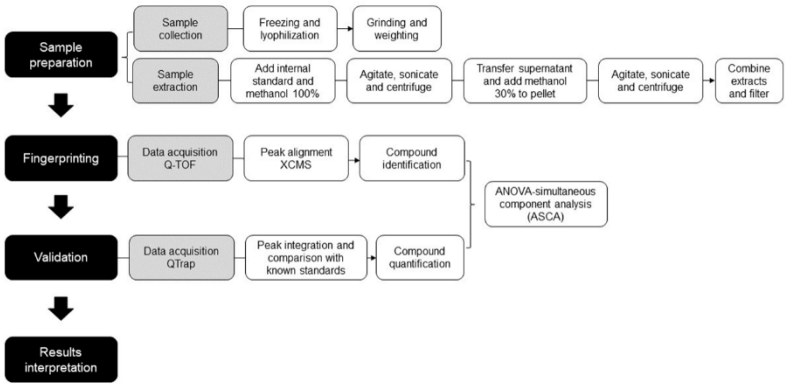


FIGURE 5
Workflow of the metabolomic platform to identify and quantify polyphenols using liquid chromatography mass spectrometry.

to have previous knowledge of the sample composition before performing targeted analysis. Most studies perform untargeted or targeted analysis, but usually do not combine both. However, the combination of both approaches, like the present study, creates a powerful tool to differentiate profiles and then detect/quantify the differences highlighted by metabolite fingerprinting, confirming the results. For instance, untargeted and targeted associated research on three *Coffea* sp. could differentiate the species metabolic profiles of leaf and fruit extracts. Additionally, five phytochemicals (caffeine, mangiferin and three caffeoylquinic acids) were identified, corroborating the identity of the differentiated metabolites between species and tissues (Montis et al., 2022). Also, Zhang et al. characterized cyclopeptides in 20 species of *Rubia* sp. (*Rubiaceae*) using LC-MS/MS (Zhang et al., 2018). Similarly to our study, the authors combined untargeted and targeted metabolomics using LC coupled to a triple TOF and a triple quadrupole, respectively, which provides reliability, precision, and sensitivity, with the additional advantage of requiring very small amounts of plant tissue. Nevertheless, our study reports an extraction that can be performed in less than two hours, and was optimized to recover a wide variety of compounds (over 40 families of phytochemicals). Indeed, 90% of the phytochemicals in our study had a recovery higher than 50%. Finally, 184 phytochemicals can be quantified in a sensitive and specific manner within a single LC-MS/MS run of 15min with polarity switch, which is particularly suitable for high-throughput analyses.

Our study compared the leaf phytochemical profiles from one commercial coffee and two wild *Rubiaceae* species. First, the untargeted metabolomics analysis identified 31 families of phytochemicals, of which flavone, flavone C-glycoside, flavonol, indole alkaloid, monolignol phenylpropanoid and triterpenoid were the most common families (Table 2). Several flavones and terpenes were not present in CC. Indeed, flavones 6,2'-dihydroxyflavone, afzelin and apigenin, the flavonoid tiliroside, flavonol nepetin-7-glucoside and the glycosylflavone robinin were not detected in CC leaves. Interestingly, flavonoids are the most abundant secondary metabolites in human diet (Alara et al., 2021): afzelin has been reported to have anti-inflammatory action (Kim et al., 2019a) while apigenin is an antioxidant and has anticancer properties (Shankar et al., 2017; Kim et al., 2019b). Flavonoids also play important roles in plant development and response to stress (Du Fall and Solomon, 2011; Nakabayashi and Saito, 2015). For example, robinin has been associated with plant drought resilience, as Chrysanthemum plants previously treated with this flavone had enhanced response to water stress and were able to maintain turgor pressure (Elansary et al., 2020). Gibberellic, sumaresinolic and madecassic acids, and soyasaponin were also absent in CC leaf extracts. Gibberellic acid is a plant hormone with multiple functions in growth regulation, flowering and stress response signaling (Bari and Jones, 2009; Schwechheimer and Willige, 2009; Iftikhar et al., 2019; Nagar et al., 2021). Madecassic acid, on the other hand, has been shown to have some medicinal properties, such as anti-inflammatory effects (Won et al., 2010), anti-colitis (Xu et al., 2017), and potential anti-cancer agent (Zhang et al., 2014; Valdeira et al., 2019). Equivalently, soyasaponins have been associated with health promoting properties, such as anti-inflammatory, anti-microbial, and cardiovascular protective activities (Guang et al., 2014; Lee et al., 2020; Wang et al., 2020).

Then, the targeted metabolomic analysis identified several compounds that were highly concentrated in CC or WR leaves (Figure 3; Supplemental Table 2). The most abundant compounds in CC leaves were chlorogenic acids (3,4 and 5-caffeoylquinic acids), caffeine, trigonelline, vicenin 2 and theobromine, which is consistent with previous reports (Campa et al., 2012; Funlayo et al., 2017; Chen, 2019; Cangeloni et al., 2022). Chlorogenic acids are well-known compounds for their antimicrobial activities (Sung and Lee, 2010; Su et al., 2014; Martínez et al., 2017). Indeed, they were shown to have deleterious effects on coffee microbial pathogens: a study showed that coffee plants supplied with silicon–a resistance inducer–had higher levels of chlorogenic acids and were therefore more resistant to *Hemilea vastatrix*, the causal agent of rust (Rodrigues et al., 2011). Caffeine content is one of the most important traits for coffee selection, either to bean processing for beverage consumption or for the pharmaceutical industry (Sawynok, 1995; Leroy et al., 2006; Patay et al., 2017; Carvalho et al., 2019). Having a resourceful method for caffeine detection and quantification, along with other desirable traits, would aid coffee breeders and studies on cultivar development. However, additional research is needed to correlate leaf and berry/bean composition in coffee to perform cultivar selection at earlier stages and speed up the breeding process. The content in kaempferol-7-O-neohesperidoside, kaempferol, yohimbine, corynanthine/rauwolscine, 3,4-dimethylcinnamic, cinnamic, benzoic and syringic acids were more abundant in WR1 leaves. A study in Litchi chinensis seeds revealed that kaempferol-7-O-neohesperidoside had a high cytotoxic activity against lung cancer cells (Xu et al., 2011). Similarly, kaempferol has also shown anti-cancer (Chen et al., 2020a; Kluska et al., 2021; Felice et al., 2022), anti-oxidant (Simunkova et al., 2021) and anti-malarial activities (Somsak et al., 2018), confirming more therapeutic uses of flavonoids. Alkaloids like yohimbine have promising clinical applications (Boğa et al., 2019; Saini et al., 2022), like anti-cancer activity (Jabir et al., 2022), and may be used as chemical markers for botanical selection (Osman et al., 2019). Several polyphenols were more abundant in WR2: harmane, vincosamide, keracyanin, idaein, gibberellic acid, luteolin-7-O-glycoside, synapaldehyde, and sinapyl-alcohol. Anthocyanins play important roles not only in plant reproduction, but also in response to abiotic and biotic stresses

(Liu et al., 2018). Additionally, keracyanin and idaein were proven to have potential anti-inflammatory and anti-cancer activities (Natarajan et al., 2016; Santamarina et al., 2021). Monolignols, key components for lignin biosynthesis, are crucial element for cell wall protection against stresses (Gallego-Giraldo et al., 2018; Xie et al., 2018). Synapaldehyde increases reactive oxygen species in plants and has anti-fungal activity (Millan et al., 2022). Also, this metabolite was more abundant in sugarcane resistant to the causal agent of ratoon stunting (Castro-Moretti et al., 2021), associating its role to plant defense. Overall, WR leaves had a higher diversity of phytochemicals in comparison to the CC. This may be due to the farmer selection of CC for fruit size, caffeine content, and yield at the detriment of other stress resistance and adaptation traits.

5 Conclusions

The present study reports: i) the optimization of an extraction procedure to recover 42 distinct families of phytochemicals from leaves, ii) the development of a robust and sensitive LC-MS/MS method to quantify 184 secondary metabolites, and iii) the complementarity between the untargeted and targeted metabolomics. This approach was applied to characterize the phytochemicals in three different species of *Rubiaceae*, including two wild species and one commercial coffee. The new targeted metabolomics approach was used to validate the identity of 74 compounds highlighted by the untargeted analysis. This work describes a sensitive and thorough pipeline (Figure 5) to detect, classify and quantify secondary metabolites in leaves of coffee and other *Rubiaceae*, and can be further applied to other plant organs and species.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

FC-M, J-CC, PC, JCS, and APA designed and conceptualized the project and method development; HC-G, EE-L, PC, and OG-F collected the leaves and prepared plant materials; FC-M and J-CC tested the different extraction methods, performed the LC-MS/MS method development, and method validation; FC-M. performed data processing and statistical analyses; FC-M, J-CC, and APA drafted the manuscript; APA, JCS and PC supervised, coordinated the project, and acquired funding for the project. All authors have read and agreed to the published version of the manuscript.

Funding

This study was funded in part by National Science Foundation grant DEB-1638999 to JCS, PC and APA, and University of Costa Rica Vice rectory of Research grants B9662 and B7176 to PC. Samples were collected under the Institutional Biodiversity Commission (University of Costa Rica) resolutions VI07723-2018 and VI5152-2016.

Acknowledgments

The authors would like to thank Pedro Juarez for helping with the collection and *Rubiaceae* identification in Costa Rica, Mariana C. Cia and Camila P. Carvalho for helping with plant collection in Brazil, and Mônica Veneziano Labate for access to the freeze dryer. All authors thank the BioAnalytical Facility (BAF) at the University of North Texas for providing the infrastructure and support to perform the metabolomic analyses.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1057645/full#supplementary-material

SUPPLEMENTARY FIGURE 1
PCA scores plot showing clustering of quality control samples for untargeted metabolomics in positive mode **(A)** and negative mode **(B)**. QC, quality control; CC, commercial coffee; WR1 and WR2, wild *Rubiaceae* species.

References

Abubakar, A. R., and Haque, M. (2020). Preparation of medicinal plants: Basic extraction and fractionation procedures for experimental purposes. *J. Pharm. Bioallied Sci.* 12, 1. doi: 10.4103/JPBS.JPBS_175_19

Ahmad, R., and Salim, F. (2015). Oxindole alkaloids of uncaria (Rubiaceae, subfamily cinchonoideae): A review on its structure, properties, and bioactivities. *Stud. Nat. Prod. Chem.* 45, 485–525. doi: 10.1016/B978-0-444-63473-3.00012-5

Alara, O. R., Abdurahman, N. H., and Ukaegbu, C. I. (2021). Extraction of phenolic compounds: A review. *Curr. Res. Food Sci.* 4, 200–214. doi: 10.1016/J.CRFS.2021.03.011

Annegowda, H. V., Anwar, L. N., Mordi, M. N., Ramanathan, S., and Mansor, S. M. (2010). Influence of sonication on the phenolic content and antioxidant activity of terminalia catappa l. leaves. *Pharmacognosy Res.* 2, 368. doi: 10.4103/0974-8490.75457

Arias, C. L., Quach, T., Huynh, T., Nguyen, H., Moretti, A., Shi, Y., et al. (2022). Expression of AtWRI1 and AtDGAT1 during soybean embryo development influences oil and carbohydrate metabolism. *Plant Biotechnol. J.* 20, 1327–1345. doi: 10.1111/PBI.13810

Bari, R., and Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Mol. Biol.* 69, 473–488. doi: 10.1007/s11103-008-9435-0

Bataglion, G. A., Da Silva, F. M. A., Eberlin, M. N., and Koolen, H. H. F. (2015). Determination of the phenolic composition from Brazilian tropical fruits by UHPLC–MS/MS. *Food Chem.* 180, 280–287. doi: 10.1016/J.FOODCHEM.2015.02.059

Boğa, M., Bingül, M., Özkan, E. E., and Şahin, H. (2019). Chemical and biological perspectives of monoterpene indole alkaloids from rauwolfia species. *Stud. Nat. Prod. Chem.* 61, 251–299. doi: 10.1016/B978-0-444-64183-0.00007-5

Böttger, A., Vothknecht, U., Bolle, C., and Wolf, A. (2018). ""Plant secondary metabolites and their general function in plants,"," in *Lessons on caffeine, cannabis & co. learning materials in biosciences* (Cham: Springer), 3–17. doi: 10.1007/978-3-319-99546-5_1

Boudet, A. M. (2007). Evolution and current status of research in phenolic compounds. *Phytochemistry* 68, 2722–2735. doi: 10.1016/J.PHYTOCHEM.2007.06.012

Brar, D. S., and Khush, G. S. (2018). Wild relatives of rice: A valuable genetic resource for genomics and breeding research. In: Mondal, T., and Henry, R. (eds). *The Wild Oryza Genomes. Compendium of Plant Genomes* (Cham: Springer). doi: 10.1007/978-3-319-71997-9_1

Campa, C., Mondolot, L., Rakotondravao, A., Bidel, L. P. R., Gargadennec, A., Couturon, E., et al. (2012). A survey of mangiferin and hydroxycinnamic acid ester accumulation in coffee (Coffea) leaves: biological implications and uses. *Ann. Bot.* 110, 595–613. doi: 10.1093/aob/mcs119

Cangeloni, L., Bonechi, C., Leone, G., Consumi, M., Andreassi, M., Magnani, A., et al. (2022). Characterization of extracts of coffee leaves (Coffea arabica l.) by spectroscopic and chromatographic/spectrometric techniques. *Foods* 11, 2495. doi: 10.3390/FOODS11162495

Carvalho, H. F., da Silva, F. L., de Resende, M. D. V., and Bhering, L. L. (2019). Selection and genetic parameters for interpopulation hybrids between kouilou and robusta coffee. *Bragantia* 78, 52–59. doi: 10.1590/1678-4499.2018124

Castro-Moretti, F. R., Cocuron, J. C., Cia, M. C., Cataldi, T. R., Labate, C. A., Alonso, A. P., et al. (2021). Targeted metabolic profiles of the leaves and xylem sap of two sugarcane genotypes infected with the vascular bacterial pathogen leifsonia xyli subsp. xyli. *Metabolites* 11, 234. doi: 10.3390/metabo11040234

Castro-moretti, F. R., Gentzel, I. N., Mackey, D., and Alonso, A. P. (2020). Metabolomics as an emerging tool for the study of plant–pathogen interactions. *Metabolites* 10 (2), 52. doi: 10.3390/metabo10020052

Chandra Kala, S. (2015). Medicinal attributes of family rubiaceae. *Int. J. Pharm. Biol. Sci.* 5, 179–181. doi: 10.4103/0973-7847.95866

Chen, X. (2019). A review on coffee leaves: Phytochemicals, bioactivities and applications. *Crit. Rev. Food Sci. Nutr.* 59, 1008–1025. doi: 10.1080/10408398.2018.1546667

Chen, X., Ding, J., Ji, D., He, S., and Ma, H. (2020b). Optimization of ultrasonic-assisted extraction conditions for bioactive components from coffee leaves using the taguchi design and response surface methodology. *J. Food Sci.* 85, 1742–1751. doi: 10.1111/1750-3841.15111

Chen, S., Ma, J., Yang, L., Teng, M., Lai, Z. Q., Chen, X., et al. (2020a). Anti-glioblastoma activity of kaempferol *via* programmed cell death induction: involvement of autophagy and pyroptosis. *Front. Bioeng. Biotechnol.* 8. doi: 10.3389/FBIOE.2020.614419/BIBTEX

Chen, Q. C., Zhang, W. Y., Youn, U. J., Kim, H. J., Lee, I. S., Jung, H. J., et al. (2009). Iridoid glycosides from gardeniae fructus for treatment of ankle sprain. *Phytochemistry* 70, 779–784. doi: 10.1016/J.PHYTOCHEM.2009.03.008

Chiocchio, I., Mandrone, M., Tomasi, P., Marincich, L., and Poli, F. (2021). Plant secondary metabolites: An opportunity for circular economy. *Mol* 26, 495. doi: 10.3390/MOLECULES26020495

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., et al. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* 46, W486–W494. doi: 10.1093/nar/gky310

Cocuron, J.-C., Anderson, B., Boyd, A., and Alonso, A. P. (2014). Targeted metabolomics of physaria fendleri, an industrial crop producing hydroxy fatty acids. *Plant Cell Physiol.* 55, 620–633. doi: 10.1093/pcp/pcu011

Cocuron, J.-C., Casas, M. I., Yang, F., Grotewold, E., and Alonso, A. P. (2019). Beyond the wall: high-throughput quantification of plant soluble and cell-wall bound phenolics by liquid chromatography tandem mass spectrometry. *J. Chromatogr. A* 1589, 93–104. doi: 10.1016/J.CHROMA.2018.12.059

Cocuron, J. C., Tsogtbaatar, E., and Alonso, A. P. (2017). High-throughput quantification of the levels and labeling abundance of free amino acids by liquid chromatography tandem mass spectrometry. *J. Chromatogr. A* 1490, 148–155. doi: 10.1016/J.CHROMA.2017.02.028

Conserva, L. M., and Ferreira Júnior, J. C. (2012). Borreria and spermacoce species (Rubiaceae): A review of their ethnomedicinal properties, chemical constituents, and biological activities. *Pharmacogn. Rev.* 6, 46–55. doi: 10.4103/0973-7847.95866

Cragg, G. M., and Newman, D. J. (2013). Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* 1830, 3670–3695. doi: 10.1016/J.BBAGEN.2013.02.008

Davidson, I. W. F., Sumner, D. D., and Parker, J. C. (2008). Chloroform: A review of its metabolism, teratogenic, mutagenic, and carcinogenic potential. *Drug Chem. Toxicol.* 5 (1), 1–87. doi: 10.3109/01480548209017822

de Falco, B., and Lanzotti, V. (2018). NMR spectroscopy and mass spectrometry in metabolomics analysis of salvia. *Phytochem. Rev.* 17, 951–972. doi: 10.1007/S11101-018-9550-8

Du Fall, L. A., and Solomon, P. S. (2011). Role of cereal secondary metabolites involved in mediating the outcome of plant-pathogen interactions. *Metabolites* 1, 64–78. doi: 10.3390/metabo1010064

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., et al. (2012). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* 91 9, 44–66. doi: 10.1007/S11306-012-0434-4

Elansary, H. O., Abdel-Hamid, A. M. E., Yessoufou, K., Al-Mana, F. A., El-Ansary, D. O., Mahmoud, E. A., et al. (2020). Physiological and molecular characterization of water-stressed chrysanthemum under robinin and chitosan treatment. *Acta Physiol. Plant* 42, 1–14. doi: 10.1007/S11738-020-3021-8/FIGURES/12

Felice, M. R., Maugeri, A., De Sarro, G., Navarra, M., and Barreca, D. (2022). Molecular pathways involved in the anti-cancer activity of flavonols: A focus on myricetin and kaempferol. *Int. J. Mol. Sci.* 23, 4411. doi: 10.3390/IJMS23084411

Feuillet, C., Langridge, P., and Waugh, R. (2008). Cereal breeding takes a walk on the wild side. *Trends Genet.* 24, 24–32. doi: 10.1016/J.TIG.2007.11.001

Funlayo, A. A., Adenuga, O. O., Mapayi, E. F., and Olaniyi, O. O. (2017). Coffee: Botany, distribution, diversity, chemical composition and its management. *J. Agric. Vet. Sci.* 10, 57–62. doi: 10.9790/2380-1007035762

Gallego-Giraldo, L., Posé, S., Pattathil, S., Peralta, A. G., Hahn, M. G., Ayre, B. G., et al. (2018). Elicitors and defense gene induction in plants with altered lignin compositions. *New Phytol.* 219, 1235–1251. doi: 10.1111/nph.15258

Guang, C., Chen, J., Sang, S., and Cheng, S. (2014). Biological functionality of soyasaponins and soyasapogenols. *J. Agric. Food Chem.* 62, 8247–8255. doi: 10.1021/JF503047A/ASSET/IMAGES/LARGE/JF-2014-03047A_0001.JPEG

Gulcin, I., Kaya, R., Goren, A. C., Akincioglu, H., Topal, M., Bingol, Z., et al. (2019). Anticholinergic, antidiabetic and antioxidant activities of cinnamon (cinnamomum verum) bark extracts: polyphenol contents analysis by LC-MS/MS. *Int. J. Food Prop.* 22, 1511–1526. doi: 10.1080/10942912.2019.1656232

Hasan, M. M., Bashir, T., and Bae, H. (2017). Use of ultrasonication technology for the increased production of plant secondary metabolites. *Molecules*. 22(7), 1046. doi: 10.3390/molecules22071046.

Hossain, M. B., Tiwari, B. K., Gangopadhyay, N., O'Donnell, C. P., Brunton, N. P., and Rai, D. K. (2014). Ultrasonic extraction of steroidal alkaloids from potato peel waste. *Ultrason. Sonochem.* 21, 1470–1476. doi: 10.1016/J.ULTSONCH.2014.01.023

Iftikhar, A., Ali, S., Yasmeen, T., Arif, M. S., Zubair, M., Rizwan, M., et al. (2019). Effect of gibberellic acid on growth, photosynthesis and antioxidant defense system of wheat under zinc oxide nanoparticle stress. *Environ. pollut.* 254, 113109. doi: 10.1016/J.ENVPOL.2019.113109

Jabir, N. R., Khan, M. S., Alafaleq, N. O., Naz, H., and Ahmed, B. A. (2022). Anticancer potential of yohimbine in drug-resistant oral cancer KB-ChR-8-5 cells. *Moleuclar Biol. Rep.* 49, 9565–9573. doi: 10.1007/s11033-022-07847-7

Jaini, R., Wang, P., Dudareva, N., Chapple, C., and Morgan, J. A. (2017). Targeted metabolomics of the phenylpropanoid pathway in arabidopsis thaliana using reversed phase liquid chromatography coupled with tandem mass spectrometry. *Phytochem. Anal.* 28, 267–276. doi: 10.1002/PCA.2672

Jain, C., Khatana, S., and Vijayvergia, R. (2019). Bioactivity of secondary metabolites of various plants: A review. *Artic. Int. J. Pharm. Sci. Res.* 10, 494–504. doi: 10.13040/IJPSR.0975-8232.10(2).494-04

Jez, J. M., Topp, C. N., Alvarez, S., and Naldrett, M. J. (2021). Mass spectrometry based untargeted metabolomics for plant systems biology. *Emerg. Top. Life Sci.* 5, 189–201. doi: 10.1042/ETLS20200271

Jorge, T. F., Mata, A. T., and António, C. (2016). Mass spectrometry as a quantitative tool in plant metabolomics. *Philos. Trans. R. Soc A Math. Phys. Eng. Sci.* 374 (2079), 20150370. doi: 10.1098/RSTA.2015.0370

Kessler, A., and Kalske, A. (2018). Plant secondary metabolite diversity and species interactions. *Annu. Rev. Ecol. Evol. Syst.* 49, 115–138. doi: 10.1146/annurev-ecolsys-110617-062406

Kim, M., Jung, J., Jeong, N. Y., and Chung, H. J. (2019b). The natural plant flavonoid apigenin is a strong antioxidant that effectively delays peripheral neurodegenerative processes. *Anat. Sci. Int.* 94, 285–294. doi: 10.1007/S12565-019-00486-2/FIGURES/6

Kim, J. H., Kim, M., Kim, J. M., Lee, M. K., Seo, S. J., and Park, K. Y. (2019a). Afzelin suppresses proinflammatory responses in particulate matter-exposed human keratinocytes. *Int. J. Mol. Med.* 43, 2516–2522. doi: 10.3892/IJMM.2019.4162/HTML

Kluska, M., Juszczak, M., Żuchowski, J., Stochmal, A., and Woźniak, K. (2021). Kaempferol and its glycoside derivatives as modulators of etoposide activity in HL-60 cells. *Int. J. Mol. Sci.* 22, 3520. doi: 10.3390/IJMS22073520

Korkina, L. G. (2007). Phenylpropanoids as naturally occurring antioxidants: From plant defense to human health. *Cell. Mol. Biol.* 53, 15–25. doi: 10.1170/T772

Kunert, O., Sreekanth, G., Babu, G. S., Rao, B. V. R. A., Radhakishan, M., Kumar, B. R., et al. (2009). Cycloartane triterpenes from dikamali, the gum resin of gardenia gummifera and gardenia lucida. *Chem. Biodivers.* 6, 1185–1192. doi: 10.1002/CBDV.200800339

Laus, G., and Teppner, H. (1996). The alkaloids of an uncaria rhynchophylla (Rubiaceae-coptosapelteae). *Phyton* (Horn, Austria)36, 185–196.

Lee, K. S., Woo, S. Y., Lee, M. J., Kim, H. Y., Ham, H., Lee, D. J., et al. (2020). Isoflavones and soyasaponins in the germ of Korean soybean [Glycine max (L.) merr.] cultivars and their compound-enhanced BMP-2-induced bone formation. *Appl. Biol. Chem.* 63, 1–8. doi: 10.1186/S13765-020-00508-Y/FIGURES/2

Leroy, T., Ribeyre, F., Bertrand, B., Charmetant, P., Dufour, M., Montagnon, C., et al. (2006). Genetics of coffee quality. *Braz. J. Plant Physiol.* 18, 229–242. doi: 10.1590/S1677-04202006000100016

Lin, Y., Xu, W., Huang, M., Xu, W., Li, H., Ye, M., et al. (2015). Qualitative and quantitative analysis of phenolic acids, flavonoids and iridoid glycosides in yinhua kanggan tablet by UPLC-QqQ-MS/MS. *Mol.* 20, 12209–12228. doi: 10.3390/MOLECULES200712209

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in solanaceous vegetables: A review. *Front. Chem.* 6. doi: 10.3389/FCHEM.2018.00052/BIBTEX

Liu, X., Zhou, L., Shi, X., and Xu, G. (2019). New advances in analytical methods for mass spectrometry-based large-scale metabolomics study. *TrAC Trends Anal. Chem.* 121, 115665. doi: 10.1016/J.TRAC.2019.115665

Ma, D., Li, Y., Zhang, J., Wang, C., Qin, H., Ding, H., et al. (2016). Accumulation of phenolic compounds and expression profiles of phenolic acid biosynthesis-related genes in developing grains of white, purple, and red wheat. *Front. Plant Sci.* 7. doi: 10.3389/FPLS.2016.00528/BIBTEX

Marchetti, L., Pellati, F., Graziosi, R., Brighenti, V., Pinetti, D., and Bertelli, D. (2019). Identification and determination of bioactive phenylpropanoid glycosides of aloysia polystachya (Griseb. et moldenke) by HPLC-MS. *J. Pharm. Biomed. Anal.* 166, 364–370. doi: 10.1016/J.JPBA.2019.01.033

Marczak, L., Kawiak, A., Łojkowska, E., and Stobiecki, M. (2005). Secondary metabolites in *in vitro* cultured plants of the genus drosera. *Phytochem. Anal.* 16, 143–149. doi: 10.1002/PCA.833

Martínez, G., Regente, M., Jacobi, S., Del Rio, M., Pinedo, M., and de la Canal, L. (2017). Chlorogenic acid is a fungicide active against phytopathogenic fungi. pestic. *Biochem. Physiol.* 140, 30–35. doi: 10.1016/J.PESTBP.2017.05.012

Martins, D., and Nunez, C. (2015). Secondary metabolites from rubiaceae species. *Molecules* 20, 13422–13495. doi: 10.3390/molecules200713422

Masson, P., Alves, A. C., Ebbels, T. M. D., Nicholson, J. K., and Want, E. J. (2010). Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS. *Anal. Chem.* 82, 7779–7786. doi: 10.1021/ac101722e

Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K., and Saito, K. (2009). MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.* 57, 555–577. doi: 10.1111/J.1365-313X.2008.03705.X

McSorley, R., and Phillips, M. S. (1992). Origin, evolution, population genetics and resources for breeding of wild barley, hordeum spontaneum, in the fertile crescent. *Barley Genet. Biochem. Mol. Biol. Biotechnol.*, 19–43.

Millan, A. F. S., Gamir, J., Farran, I., Larraya, L., and Veramendi, J. (2022). Identification of new antifungal metabolites produced by the yeast metschnikowia pulcherrima involved in the biocontrol of postharvest plant pathogenic fungi. *Postharvest Biol. Technol.* 192, 111995. doi: 10.1016/J.POSTHARVBIO.2022.111995

Montis, A., Souard, F., Delporte, C., Stoffelen, P., Stévigny, C., and Van Antwerpen, P. (2022). Targeted and untargeted mass spectrometry-based metabolomics for chemical profiling of three coffee species. *Molecules* 27 (10), 3152. doi: 10.3390/MOLECULES27103152

Nagar, S., Singh, V. P., Arora, A., Dhakar, R., Singh, N., Singh, G. P., et al. (2021). Understanding the role of gibberellic acid and paclobutrazol in terminal heat stress tolerance in wheat. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.692252/BIBTEX

Nakabayashi, R., and Saito, K. (2015). Integrated metabolomics for abiotic stress responses in plants. *Curr. Opin. Plant Biol.* 24, 10–16. doi: 10.1016/j.pbi.2015.01.003

Natarajan, T., Anandhi, M., Aiswarya, D., Ramkumar, R., Kumar, S., and Perumal, P. (2016). Idaein chloride induced p53 dependent apoptosis in cervical cancer cells through inhibition of viral oncoproteins. *Biochimie* 121, 13–20. doi: 10.1016/J.BIOCHI.2015.11.008

Ndagijimana, A., Wang, X., Pan, G., Zhang, F., Feng, H., and Olaleye, O. (2013). A review on indole alkaloids isolated from uncaria rhynchophylla and their pharmacological studies. *Fitoterapia* 86, 35–47. doi: 10.1016/J.FITOTE.2013.01.018

Orcic, D., Franciškovic, M., Bekvalac, K., Svircev, E., Beara, I., Lesjak, M., et al. (2014). Quantitative determination of plant phenolics in urtica dioica extracts by high-performance liquid chromatography coupled with tandem mass spectrometric detection. *Food Chem.* 143, 48–53. doi: 10.1016/J.FOODCHEM.2013.07.097

Osman, A. G., Haider, S., Chittiboyina, A. G., and Khan, I. A. (2019). Utility of alkaloids as chemical and biomarkers for quality, efficacy, and safety assessment of botanical ingredients. *Phytomedicine* 54, 347–356. doi: 10.1016/J.PHYMED.2018.03.064

Patay, E. B., Fritea, L., Antonescu, A., Antonescu, A., and Dobjanschi, L. (2017). Coffea arabica: A plant with rich content in caffeine. *The Question of Caffeine*. (London, United Kingdom: IntechOpen) doi: 10.5772/intechopen.68149

Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S., and Yuan, L. (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. biochim. *Biophys. Acta - Gene Regul. Mech.* 1829, 1236–1247. doi: 10.1016/J.BBAGRM.2013.09.006

Patti, G. J., Yanes, O., Siuzadak, G., and Siuzdak, G. (2012). Metabolomics: the apogee of the omics triology. *Nat. Rev. Mol. Cell Biol.* 13, 263–269. doi: 10.1038/nrm3314

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D., et al. (2005). Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell Environ.* 28, 176–191. doi: 10.1111/J.1365-3040.2005.01259.X

Perez de Souza, L., Alseekh, S., Naake, T., and Fernie, A. (2019). Mass spectrometry-based untargeted plant metabolomics. *Curr. Protoc. Plant Biol.* 4, e20100. doi: 10.1002/CPPB.20100

Quatrin, A., Pauletto, R., Maurer, L. H., Minuzzi, N., Nichelle, S. M., Carvalho, J. F. C., et al. (2019). Characterization and quantification of tannins, flavonols, anthocyanins and matrix-bound polyphenols from jaboticaba fruit peel: A comparison between myrciaria trunciflora and m. jaboticaba. *J. Food Compos. Anal.* 78, 59–74. doi: 10.1016/J.JFCA.2019.01.018

Roberts, L. D., Souza, A. L., Gerszten, R. E., and Clish, C. B. (2012). Targeted metabolomics. *Curr. Protoc. Mol. Biol.* 98, 30.2.1–30.2.24. doi: 10.1002/0471142727.MB3002S98

Rodrigues, F. A., Carré-Missio, V., Jham, G. N., Berhow, M., and Schurt, D. A. (2011). Chlorogenic acid levels in leaves of coffee plants supplied with silicon and infected by hemileia vastatrix. *Trop. Plant Pathol.* 36, 404–408. doi: 10.1590/S1982-56762011000600010

Roopashree, K. M., and Naik, D. (2019). Advanced method of secondary metabolite extraction and quality analysis. *J. Pharmacogn. Phytochem.* 8, 1829–1842.

Saini, N., Grewal, A. S., Lather, V., and Gahlawat, S. K. (2022). Natural alkaloids targeting EGFR in non-small cell lung cancer: Molecular docking and ADMET predictions. *Chem. Biol. Interact.* 358, 109901. doi: 10.1016/J.CBI.2022.109901

Santamarina, A. B., Pisani, L. P., Baker, E. J., Marat, A. D., Valenzuela, C. A., Miles, E. A., et al. (2021). Anti-inflammatory effects of oleic acid and the anthocyanin keracyanin alone and in combination: effects on monocyte and macrophage responses and the NF-kappa b pathway. *Food Funct.* 12, 7909–7922. doi: 10.1039/D1FO01304A

Sawada, Y., and Yokota Hirai, M. (2013).). integrated LC-MS/MS system for plant metabolomics. *Comput. Struct. Biotechnol. J.* 4, e201301011. doi: 10.5936/CSBJ.201301011

Sawynok, J. (1995). Pharmacological rationale for the clinical use of caffeine. *Drugs* 49, 37–50. doi: 10.2165/00003495-199549010-00004

Schwechheimer, C., and Willige, B. C. (2009). Shedding light on gibberellic acid signalling. *Curr. Opin. Plant Biol.* 12, 57–62. doi: 10.1016/J.PBI.2008.09.004

Shankar, E., Goel, A., Gupta, K., and Gupta, S. (2017). Plant flavone apigenin: An emerging anticancer agent. *Curr. Pharmacol. Rep.* 3, 423. doi: 10.1007/S40495-017-0113-2

Shimizu, T., Watanabe, M., Fernie, A. R., and Tohge, T. (2018). Targeted LC-MS analysis for plant secondary metabolites. *Methods Mol. Biol.* 1778, 171–181. doi: 10.1007/978-1-4939-7819-9_12/COVER

Silva, G. L., Lee, I.-S., and Kinghorn, A. D. (1998). "Special problems with the extraction of plants," in *Natural products isolation. methods in biotechnology*. Ed. R. J. Cannell (Totowa: Humana Press), 343–363. doi: 10.1007/978-1-59259-256-2_12

Simunkova, M., Barbierikova, Z., Jomova, K., Hudecova, L., Lauro, P., Alwasel, S. H., et al. (2021). Antioxidant vs. prooxidant properties of the flavonoid, kaempferol, in the presence of Cu(II) ions: A ROS-scavenging activity, fenton reaction and DNA damage study. *Int. J. Mol. Sci.* 22 (4), 1619. doi: 10.3390/IJMS22041619

Soares Melecchi, M. I., Péres, V. F., Dariva, C., Zini, C. A., Abad, F. C., Martinez, M. M., et al. (2006). Optimization of the sonication extraction method of hibiscus tiliaceus l. flowers. *Ultrason. Sonochem.* 13, 242–250. doi: 10.1016/J.ULTSONCH.2005.02.003

Somsak, V., Damkaew, A., and Onrak, P. (2018). Antimalarial activity of kaempferol and its combination with chloroquine in plasmodium berghei infection in mice. *J. Pathog.* 2018, 1–7. doi: 10.1155/2018/3912090

Souard, F., Delporte, C., Stoffelen, P., Thévenot, E. A., Noret, N., Dauvergne, B., et al. (2018). Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS. *Food Chem.* 245, 603–612. doi: 10.1016/J.FOODCHEM.2017.10.022

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Bot. Rev.* 80, 283–383. doi: 10.1007/S12229-014-9146-Y/FIGURES/8

Su, Y., Ma, L., Wen, Y., Wang, H., and Zhang, S. (2014). Studies of the *in vitro* antibacterial activities of several polyphenols against clinical isolates of methicillin-resistant staphylococcus aureus. *Molecules* 19, 12630–12639. doi: 10.3390/molecules190812630

Sun, D., Dong, L., Guo, P., Shi, X., Gao, J., Ren, Y., et al. (2013). Simultaneous detection of flavonoids and phenolic acids in herba lysimachiae and herba desmodii styracifolii using liquid chromatography tandem mass spectrometry. *Food Chem.* 138, 139–147. doi: 10.1016/J.FOODCHEM.2012.09.096

Sung, W. S., and Lee, D. G. (2010). Antifungal action of chlorogenic acid against pathogenic fungi, mediated by membrane disruption. *Pure Appl. Chem.* 82, 219–226. doi: 10.1351/PAC-CON-09-01-08/MACHINEREADABLECITATION/RIS

Tautenhahn, R., Patti, G. J., Rinehart, D., and Siuzdak, G. (2012). XCMS online: A web-based platform to process untargeted metabolomic data. *Anal. Chem.* 84, 5035–5039. doi: 10.1021/ac300698c

Valdeira, A. S. C., Darvishi, E., Woldemichael, G. M., Beutler, J. A., Gustafson, K. R., and Salvador, J. A. R. (2019). Madecassic acid derivatives as potential anticancer agents: Synthesis and cytotoxic evaluation. *J. Nat. Prod.* 82, 2094–2105. doi: 10.1021/ACS.JNATPROD.8B00864/ASSET/IMAGES/LARGE/NP-2018-00864R_0003.JPEG

Wang, F., Gong, S., Wang, T., Li, L., Luo, H., Wang, J., et al. (2020). Soyasaponin II protects against acute liver failure through diminishing YB-1 phosphorylation and Nlrp3-inflammasome priming in mice. *Theranostics* 10, 2714–2726. doi: 10.7150/THNO.40128

Wang, Y., Liu, H., Shen, L., Yao, L., Ma, Y., Yu, D., et al. (2015). Isolation and purification of six iridoid glycosides from gardenia jasminoides fruit by medium-pressure liquid chromatography combined with macroporous resin chromatography. *J. Sep. Sci.* 38, 4119–4126. doi: 10.1002/jssc.201500705

Won, J. H., Shin, J. S., Park, H. J., Jung, H. J., Koh, D. J., Jo, B. G., et al. (2010). Anti-inflammatory effects of madecassic acid *via* the suppression of NF-κB pathway in LPS-induced RAW 264.7 macrophage cells. *Planta Med.* 76, 251–257. doi: 10.1055/S-0029-1186142

Xie, S., Shi, Y., Wang, Y., Wu, C., Liu, W., Feng, F., et al. (2013). Systematic identification and quantification of tetracyclic monoterpenoid oxindole alkaloids in uncaria rhynchophylla and their fragmentations in q-TOF-MS spectra. *J. Pharm. Biomed. Anal.* 81–82, 56–64. doi: 10.1016/J.JPBA.2013.03.017

Xie, M., Zhang, J., Tschaplinski, T. J., Tuskan, G. A., Chen, J. G., and Muchero, W. (2018). Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01427

Xu, X., Wang, Y., Wei, Z., Wei, W., Zhao, P., Tong, B., et al. (2017). Madecassic acid, the contributor to the anti-colitis effect of madecassoside, enhances the shift of Th17 toward treg cells *via* the PPARγ/AMPK/ACC1 pathway. *Cell Death Dis.* 838, e2723–e2723. doi: 10.1038/cddis.2017.150

Xu, X., Xie, H., Hao, J., Jiang, Y., and Wei, X. (2011). Flavonoid glycosides from the seeds of litchi chinensis. *J. Agric. Food Chem.* 59, 1205–1209. doi: 10.1021/JF104387Y/ASSET/IMAGES/LARGE/JF-2010-04387Y_0001.JPEG

Yang, L., Peng, K., Zhao, S., Zhao, F., Chen, L., and Qiu, F. (2013). 2-methyl-l-erythritol glycosides from gardenia jasminoides. *Fitoterapia* 89, 126–130. doi: 10.1016/J.FITOTE.2013.05.018

Zhang, X., Bi, Q., Wu, X., Wang, Z., Miao, Y., and Tan, N. (2018). Systematic characterization and quantification of rubiaceae-type cyclopeptides in 20 rubia species by ultra performance liquid chromatography tandem mass spectrometry combined with chemometrics. *J. Chromatogr. A* 1581–1582, 43–54. doi: 10.1016/J.CHROMA.2018.10.049

Zhang, H. F., Yang, X. H., and Wang, Y. (2011). Microwave assisted extraction of secondary metabolites from plants: Current status and future directions. *Trends Food Sci. Technol.* 22, 672–688. doi: 10.1016/J.TIFS.2011.07.003

Zhang, H., Zhang, M., Tao, Y., Wang, G., and Xia, B. (2014). Madecassic acid inhibits the mouse colon cancer growth by inducing apoptosis and immunomodulation. *JBUON* 19, 372–376.

4.6 Herramienta para la anotación de compuestos derivados de microorganismos.

MicrobeMASST, una herramienta de búsqueda de espectrometría de masas (MS) informada taxonómicamente, aborda anotación de metabolitos microbianos en experimentos de metabolómica no dirigidos. Aprovechando una base de datos seleccionada de más de 60 000 monocultivos microbianos, los usuarios pueden buscar espectros de MS/MS conocidos y desconocidos y vincularlos con sus respectivos productores microbianos a través de patrones de fragmentación de MS/MS. La identificación de metabolitos derivados de microbios y productores relativos, sin conocimiento a priori, mejorará enormemente la comprensión del papel de los microorganismos en la ecología y la salud humana.

Publicación:

Zuffa S., Schmid R., Bauermeister A., Gomes PWP., Caraballo-Rodriguez A.M., Abiead Y.E., Aron A.T., Gentry E.C., Zemlin J.., Meehan M.J., Avalon N.E., Cichewicz R.H., Buzun E., Terrazas M.C., Hsu C., Oles R., Ayala A.V., Zhao J., Chu H., Kuijpers M.C.M., Jackrel S.L., Tugizimana F., Nephali L.P., Dubery I.A., Madala N.E., Moreira E.A., Costa-Lotufo L.V., Lopes N.P., Rezende-Teixeira P., Jimenez P.C., Rimal B., Patterson A.D., Traxler M.F., de Cassia Pessotti R., Alvarado-Villalobos D., Tamayo-Castillo G., Chaverri P., **Escudero-Leyva E**, Quiros-Guerrero L., Bory A.J., Joubert J., Rutz A., Wolfender J., Allard P., Sichert A., Pontrelli S., Pullman B.S., Bandeira N., Gerwick W.H., Gindro K., Massana-Codina J., Wagner B.C., Forchhammer K., Petras D., Aiosa N., Garg N., Liebeke M., Bourceau P., Kang K.B., Gadhavi H., de Carvalho L.P.S., dos Santos M.S., Pérez-Lorente A.I., Molina-Santiago C., Romero D., Franke R., Brönstrup M., de León A.V.P., Pope P.B., La Rosa S.L., Barbera G.L., Roager H.M., Laursen M.F., Hammerle F., Siewert B., Peintner U., Licona-Cassani C., Rodriguez-Orduña L., Rampler E., Hildebrand F., Koellensperger G., Schoeny H., Hohenwallner K., Panzenboeck L., Gregor R., O'Neill E.C., Roxborough E.T., Odoi J., Bale N.J., Ding S., Sinninghe Damsté J.S., Guan X.L., Cui J.J., Ju K., Silva D.B., Ribeiro Silva F.M., da Silva G.F., Koolen H.H.F., Grundmann C., Clement J.A., Mohimani H., Broders K., McPhail K.L., Ober-Singleton S.E., Rath C.M., McDonald D., Knight R., Wang M., Dorrestein P.C. A Taxonomically-informed Mass Spectrometry Search Tool for Microbial Metabolomics Data. doi:10.1101/2023.07.20.549584. PPR:PPR695223.

Title - A Taxonomically-informed Mass Spectrometry Search Tool for Microbial Metabolomics Data

Key words (5 words) - metabolomics, microbiome, search tool, bacteria, fungi

Authors list

Simone Zuffa[1,2,*], Robin Schmid[1,2,*], Anelize Bauermeister[1,2,3,*], Paulo Wender P. Gomes[1,2], Andres M. Caraballo-Rodriguez[1,2], Yasin El Abiead[1,2], Allegra T. Aron[4], Emily C. Gentry[5], Jasmine Zemlin[1,6], Michael J. Meehan[1], Nicole E. Avalon[7], Robert H. Cichewicz[8], Ekaterina Buzun[9], Marvic Carrillo Terrazas[9], Chia-Yun Hsu[9], Renee Oles[9], Adriana Vasquez Ayala[9], Jiaqi Zhao[9], Hiutung Chu[9,10], Mirte C. M. Kuijpers[11], Sara L. Jackrel[11], Fidele Tugizimana[12,13], Lerato Pertunia Nephali[12], Ian A. Dubery[12], Ntakadzeni Edwin Madala[14], Eduarda Antunes Moreira[15], Leticia Veras Costa-Lotufo[3], Norberto Peporine Lopes[15], Paula Rezende-Teixeira[3], Paula C. Jimenez[16], Bipin Rimal[17], Andrew D. Patterson[18], Matthew F. Traxler[19], Rita de Cassia Pessotti[19], Daniel Alvarado-Villalobos[20], Giselle Tamayo-Castillo[20,21], Priscila Chaverri[22,23,24], Efrain Escudero-Leyva[25], Luis-Manuel Quiros-Guerrero[26,27], Alexandre Jean Bory[26,27], Juliette Joubert[26,27], Adriano Rutz[26,27,28], Jean-Luc Wolfender[26,27], Pierre-Marie Allard[29,26,27], Andreas Sichert[28], Sammy Pontrelli[28], Benjamin S Pullman[30], Nuno Bandeira[30,1], William H. Gerwick[31,1], Katia Gindro[32], Josep Massana-Codina[32], Berenike C. Wagner[33], Karl Forchhammer[33], Daniel Petras[34], Nicole Aiosa[35], Neha Garg[35,36], Manuel Liebeke[37], Patric Bourceau[37], Kyo Bin Kang[38], Henna Gadhavi[39,40], Luiz Pedro Sorio de Carvalho[39,41], Mariana Silva dos Santos[42], Alicia Isabel Pérez-Lorente[43], Carlos Molina-Santiago[43], Diego Romero[43], Raimo Franke[44], Mark Brönstrup[44,45], Arturo Vera Ponce de León[46], Phillip Byron Pope[47,46], Sabina Leanti La Rosa[46], Giorgia La Barbera[48], Henrik M. Roager[48], Martin Frederik Laursen[49], Fabian Hammerle[50], Bianka Siewert[50], Ursula Peintner[51], Cuauhtemoc Licona-Cassani[52], Lorena Rodriguez-Orduña[52], Evelyn Rampler[53], Felina Hildebrand[53,54], Gunda Koellensperger[53,55], Harald Schoeny[53], Katharina Hohenwallner[53,54], Lisa Panzenboeck[53,54], Rachel Gregor[56], Ellis Charles O'Neill[57], Eve Tallulah Roxborough[57], Jane Odoi[58], Nicole J. Bale[59], Su Ding[59], Jaap S. Sinninghe Damsté[59], Xueli Li Guan[60], Jerry J. Cui[61], Kou-San Ju[61,62,63,64], Denise Brentan Silva[65], Fernanda Motta Ribeiro Silva[65], Gilvan Ferreira da Silva[66], Hector H. F. Koolen[67], Carlismari Grundmann[68], Jason A. Clement[69], Hosein Mohimani[70], Kirk Broders[71], Kerry L. McPhail[72], Sidnee E. Ober-Singleton[73], Christopher M. Rath[74], Daniel McDonald[75], Rob Knight[75,30,76], Mingxun Wang[77], Pieter C. Dorrestein[1,2,^]

* These authors equally contributed to this work

^ Correspondence

Emails and affiliations

[1]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [3]Department of Pharmacology, Institute of Biomedical Sciences, University of São Paulo, Av. Lineu Prestes 1524, São Paulo, SP, 05508-000, Brazil, [4]Department of Chemistry and Biochemistry, University of Denver, Denver, CO, 80210, United States, [5]Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061, United States, [6]Center for Microbiome Innovation, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [7]Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Dr., La Jolla, CA, 92093, United States, [8]Department of Chemistry and Biochemistry, College of Arts and Sciences, University of Oklahoma, 101 Stephenson Parkway, Norman, OK, 73019, United States, [9]Department of Pathology, School of Medicine, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [10]Center for Mucosal Immunology, Allergy, and Vaccines (cMAV), Chiba University-University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [11]Department of Ecology, Behavior and Evolution, School of Biological Sciences, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [12]Department of Biochemistry, Faculty of Science, University of Johannesburg, Auckland Park, Johannesburg, Gauteng, 2006, South Africa, [13]International Research and Development, Omnia Nutriology, Omnia Group (Pty) Ltd, 178 Montecasino Boulevard, Fourways, Johannesburg, Gauteng, 2191, South Africa, [14]Department of Biochemistry and Microbiology, Faculty of Sciences, Agriculture and Engineering, University of Venda, Private Bag X5050, Thohoyandou, Limpopo, 950, South Africa, [15]Department of BioMolecular Sciences, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Avenida do Café, Ribeirão Preto, SP, 14040-903, Brazil, [16]Department of Marine Science, Institute of Marine Science, Federal University of São Paulo, Rua Carvalho de Mendonça, 144, Santos, SP, 11070-100, Brazil, [17]Department of Veterinary and Biomedical Sciences, Pennsylvania State University, 319 Life Sciences Building, University Park, PA, 16802, United States, [18]Department of Veterinary and Biomedical Sciences, Pennsylvania State University, 320 Life Sciences Building, University Park, PA, 16802, United States, [19]Plant and Microbial Biology, College of Natural Resources, University of California Berkeley, 311 Koshland Hall, Berkeley, CA, 94270, United States, [20]Metabolomics & Chemical Profiling, Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, San Pedro de Montes de Oca, San José, 2061, Costa Rica, [21]Escuela de Química, Universidad de Costa Rica, San Pedro de Montes de Oca, San José, 2061, Costa Rica, [22]Microbial Biotechnology, Centro de Investigaciones en Productos Naturales (CIPRONA) & Escuela de Biología, Universidad de Costa Rica, San Pedro de Montes de Oca, San José, 2061, Costa Rica, [23]Escuela de Biología, Universidad de Costa Rica, San Pedro de Montes de Oca, San José, 2061, Costa Rica, [24]Department of Natural Sciences, Bowie State University, Bowie, Maryland, 20715, United States, [25]Microbial Biotechnology, Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, San Pedro de Montes de Oca, San José, 2061, Costa Rica, [26]School of Pharmaceutical Sciences, University of Geneva, Rue Michel-Servet 1, Genève, GE, 1206, Switzerland, [27]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Rue Michel-Servet 1, Genève, GE, 1206, Switzerland, [28]Institute of Molecular Systems Biology, ETH Zurich, Otto-Stern-Weg 3, Zürich, 8093, Switzerland, [29]Department of Biology, University of Fribourg, Chemin du Musée, 10, Fribourg, FR, 1700,

Switzerland, [30]Department of Computer Science and Engineering, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [31]Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [32]Plant Protection, Mycology group, Agroscope, Rte de Duillier, 50, Nyon, VD, 1260, Switzerland, [33]Department of Microbiology and Organismic Interactions, Interfaculty Institute of Microbiology and Infection Medicine, University of Tuebingen, Auf der Morgenstelle 28, Tuebingen, 72076, Germany, [34]Cluster of Excellence "Controlling Microbes to Fight Infections" (CMFI), University of Tuebingen, Auf der Morgenstelle 24, Tuebingen, 72076, Germany, [35]School of Chemistry and Biochemistry, Georgia Institute of Technology, 950 Atlantic Drive, Atlanta, GA, 30332, United States, [36]Center for Microbial Dynamics and Infection, Georgia Institute of Technology, 311 Ferst Drive, Atlanta, GA, 30332, United States, [37]Department of Symbiosis, Metabolic Interactions, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen, 28359, Germany, [38]Research Institute of Pharmaceutical Sciences, College of Pharmacy, Sookmyung Women's University, Cheongpa-ro 47 gil 100, Seoul, 04310, Korea, [39]Mycobacterial Metabolism and Antibiotic Research Laboratory, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK, [40]King's College London, Strand, London, WC2R 2LS, UK, [41]Chemistry Department, The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, 110 Scripps Way, Jupiter, FL, 33458, United States, [42]Metabolomics Science Technology Platform, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK, [43]Department of Microbiology, Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora", Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), Bulevar Louis Pasteur (Campus Universitario de Teatinos), Málaga, Málaga, 29071, Spain, [44]Department of Chemical Biology, Helmholtz Centre for Infection Research, Inhoffenstraße 7, Braunschweig, 38124, Germany, [45]German Center for Infection Research (DZIF), Site Hannover-Braunschweig, Braunschweig, 38124, Germany, [46]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Postboks 5003, Ås, 1433, Norway, [47]Faculty of Biosciences, Norwegian University of Life Sciences, Postboks 5003, Ås, 1433, Norway, [48]Department of Nutrition, Exercise and Sports, University of Copenhagen, Rolighedsvej 26, Frederiksberg, 1958, Denmark, [49]National Food Institute, Technical University of Denmark, Kemitorvet B202, Lyngby, 2800, Denmark, [50]Department of Pharmacognosy, Institute of Pharmacy, University of Innsbruck, Innrain 80-82, Innsbruck, 6020, Austria, [51]Department of Microbiology, University of Innsbruck, Technikerstr. 25, Innsbruck, 6020, Austria, [52]Escuela de Ingeniería y Ciencias, Centro de Biotecnología FEMSA, Tecnologico de Monterrey, Av. Eugenio Garza Sada 2501, Monterrey, Nuevo Leon, 64849, Mexico, [53]Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Waehringer Str. 38, Vienna, 1090, Austria, [54]Vienna Doctoral School in Chemistry (DoSChem), Faculty of Chemistry, University of Vienna, Waehringer Str. 42, Vienna, 1090, Austria, [55]Vienna Metabolomics Center (VIME), University of Vienna, Althanstr. 14,, Vienna, 1090, Austria, [56]Department of Civil and Environmental Engineering, School of Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA, 02142, United States, [57]School of Chemistry, University of Nottingham, University Park, Nottingham, NG72RD, UK, [58]Faculty of Engineering, University of Nottingham, University Park, Nottingham, NG72RD, UK, [59]Department of Marine Microbiology and Biogeochemistry, Netherlands Institute for Sea Research (NIOZ), Landsdiep 4, t Horntje (Texel), 1797 SZ, Netherlands, [60]Lee Kong Chian School of Medicine, Nanyang Technological University, 59 Nanyang Drive, Singapore, Singapore, 636921, Singapore, [61]Department of Microbiology, College of Arts and Sciences, The Ohio State University, 484 W. 12th Ave, Columbus, OH, 43210, United States, [62]Division of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, The Ohio State University, 484 W. 12th Ave, Columbus, OH, 43210, United States, [63]Center for Applied Plant Sciences, The Ohio State University, 484 W. 12th Ave, Columbus, OH, 43210, United States, [64]Infectious Diseases Institute, The Ohio State University, 484 W. 12th Ave, Columbus, OH, 43210, United States, [65]Faculty of Pharmaceutical Sciences, Food and Nutrition, Federal University of Mato Grosso do Sul, Av. Costa e Silva, s/n, Campo Grande, MS, 79070-900, Brazil, [66]Embrapa Amazônia Ocidental, AM-010 Roadway, Manaus, AM, 69010-970, Brazil, [67]Escola Superior de Ciências da Saúde, Universidade do Estado do Amazonas, 1777 Carvalho Leal Avenue, Manaus, AM, 69065-001, Brazil, [68]Department of Pharmaceutical Sciences, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Avenida do Café, Ribeirão Preto, SP, 14040-903, Brazil, [69]Baruch S. Blumberg Institute, 3805 Old Easton Rd., Doylestown, PA, 18902, United States, [70]Computational Biology Department, School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, United States, [71]USDA, Agricultural Research Service, National Center for Agricultural Utilization Research, Mycotoxin Prevention and Applied Microbiology Research Unit, 1815 N. University, Peoria, IL, 61604, United States, [72]Department of Pharmaceutical Sciences, College of Pharmacy, Oregon State University, Weniger Hall, room 341, Corvallis, OR, 97331, United States, [73]Department of Physics, Study of Heavy-Element-Biomaterials, University of Oregon, 1255 E 13th Ave, Basement, Eugene, OR, 97402, United States, [74]1480 64th St #300, Emeryville, CA, 94608, United States, [75]Department of Pediatrics, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [76]Department of Bioengineering, University of California San Diego, 9500 Gilman Dr., San Diego, CA, 92093, United States, [77]Department of Computer Science and Engineering, University of California Riverside, 900 University Ave., Riverside, CA, 92521, United States

Abstract

MicrobeMASST, a taxonomically-informed mass spectrometry (MS) search tool, tackles limited microbial metabolite annotation in untargeted metabolomics experiments. Leveraging a curated database of >60,000 microbial monocultures, users can search known and unknown MS/MS spectra and link them to their respective microbial producers via MS/MS fragmentation patterns. Identification of microbial-derived metabolites and relative producers, without *a priori* knowledge, will vastly enhance the understanding of microorganisms' role in ecology and human health.

Main

Microorganisms drive the global carbon cycle[1] and can establish symbiotic relationships with host organisms, influencing their health, aging, and behavior[2–6]. Microbial populations interact with different ecosystems through the alteration of available metabolite pools and the production of specialized small molecules[7,8]. The vast genetic potential of these communities is exemplified by human-associated microorganisms, which encode approximately 100 times more genes than the human genome[9,10]. However, this metabolic potential remains unreflected in modern untargeted metabolomics experiments, where typically <1% of the annotated molecules can be classified as microbial. This problem particularly affects mass spectrometry (MS)-based untargeted metabolomics, a common technique to investigate molecules produced or modified by microorganisms[11], which famously struggles with spectral annotation of complex biological samples. This is because the majority of spectral reference libraries are biased towards commercially available or otherwise accessible standards of primary metabolites, drugs, or industrial chemicals. Even when metabolites are annotated, extensive literature searches are required to understand whether these molecules have microbial origins and to identify the respective microbial producers. Public databases, such as KEGG[12], MiMeDB[13], NPAtlas[14], and LOTUS[15], can assist in this interpretation, but they are mostly limited to well-established, largely genome-inferred, metabolic models or to fully characterized and published molecular structures. Additionally, while targeted metabolomics efforts aimed at interrogating the gut microbiome mechanistically have been developed[16], these focus only on relatively few commercially-available microbial molecules. Hence, the majority of the microbial chemical space remains unknown, despite the continuous expansion of MS reference libraries. To fill this gap, we have developed microbeMASST (https://masst.gnps2.org/microbemasst/), a search tool that leverages public MS repository data to identify the microbial origin of known and unknown metabolites and map them to their microbial producers.

MicrobeMASST is a community-sourced tool that works within the GNPS[17] ecosystem. Users can search tandem MS (MS/MS) spectra obtained from their experiments against MS/MS spectra previously detected in other extracts of bacterial, fungal, or archaeal monocultures. No other available resource or tool allows linking uncharacterized MS/MS spectra to characterized microorganisms . The microbeMASST reference database of monocultures has been generated through years of community contributions and metadata curation, and it contains microorganisms isolated from plants, soils, oceans, lakes, fish, terrestrial animals, and humans (**Figure 1a**). All available microorganisms are categorized according to the NCBI taxonomy[18] at different taxonomic resolution (*i.e.* species, genus, family, *etc.*) or mapped to the closest taxonomically accurate level, if no NCBI ID was available at the time of database creation. As of June 2023, microbeMASST includes 60,781 liquid chromatography (LC)-MS/MS files, comprising >100 million MS/MS spectra, mapped to 541 strains, 1,336 species, 539 genera, 264 families, 109 orders, 41 classes, and 16 phyla from the three domains of life: Bacteria, Archaea, and Eukaryota (**Figure 1b**). Differently from MASST[19], which uses a precomputed network of ~110 million MS/MS spectra to enable spectral searching, microbeMASST is based on the newly introduced Fast Search Tool (https://fasst.gnps2.org/fastsearch/)[20]. This tool, originally designed for proteomics, drastically improves search speed by several orders of magnitude by indexing all the

MS/MS spectra present in GNPS/MassIVE and restricting the search space to the user input parameters. Because of this, search results are returned within seconds as opposed to 20 min per search or 24-48 hours for modification tolerant searches in the original implementation of MASST. Additionally, microbeMASST leverages the pre-curated file-associated metadata to aggregate results into taxonomic trees. This represents a major enhancement over MASST, where users have to manually inspect result tables and contextualize them, making interpretations tedious.

In microbeMASST, users can search MS/MS spectra using a Universal Spectrum Identifier (USI)[21] or by inputting a precursor ion mass and its spectral fragmentation pattern (**Supplementary Figure 1**). Analogue search can also be enabled to discover molecules related to the MS/MS spectrum of interest across the taxonomic tree[17,19,22]. The microbeMASST web app displays query results in interactive taxonomic trees, which can be downloaded as HTML files. Nodes in the trees represent specific taxa and display rich information, such as taxon scientific name, NCBI taxonomic ID, number of deposited samples, number of found MS/MS matches, and proportion of found matches, which is also visualized through pie charts. Information for an MS/MS match in a particular taxon is propagated upstream through its lineage. The reactive interface of microbeMASST enables filtering of the tree to specific taxonomic levels or to a minimum number of matches observed per taxon. Additionally, three data tables are generated, linking the search job to other resources in the GNPS/MassIVE ecosystem. Each MS/MS query is searched against the public MS/MS reference library of GNPS (587,213 MS/MS spectra, June 2023). Annotations to such reference compounds are listed under the 'Library matches' tab (**Supplementary Figure 2a**). The 'Datasets matches' tab contains information on the matching scans, displaying scientific name, NCBI taxonomic ID and taxonomic rank, number of matching fragment ions, and modified cosine score together with a link to a mirror plot visualization (**Supplementary Figure 2b**). Finally, the 'Taxa matches' tab informs on how many matches were found per taxon and number of samples available for that taxon (**Supplementary Figure 2c**). Quality controls (QCs) and blank samples (n=2,902) present in the reference datasets of microbeMASST have been retained to provide information on possible contaminants and media components. Additionally, data from human cell line cultures (n=1,199) have been included to enable assessment of whether molecules can be produced by both human hosts and microorganisms.

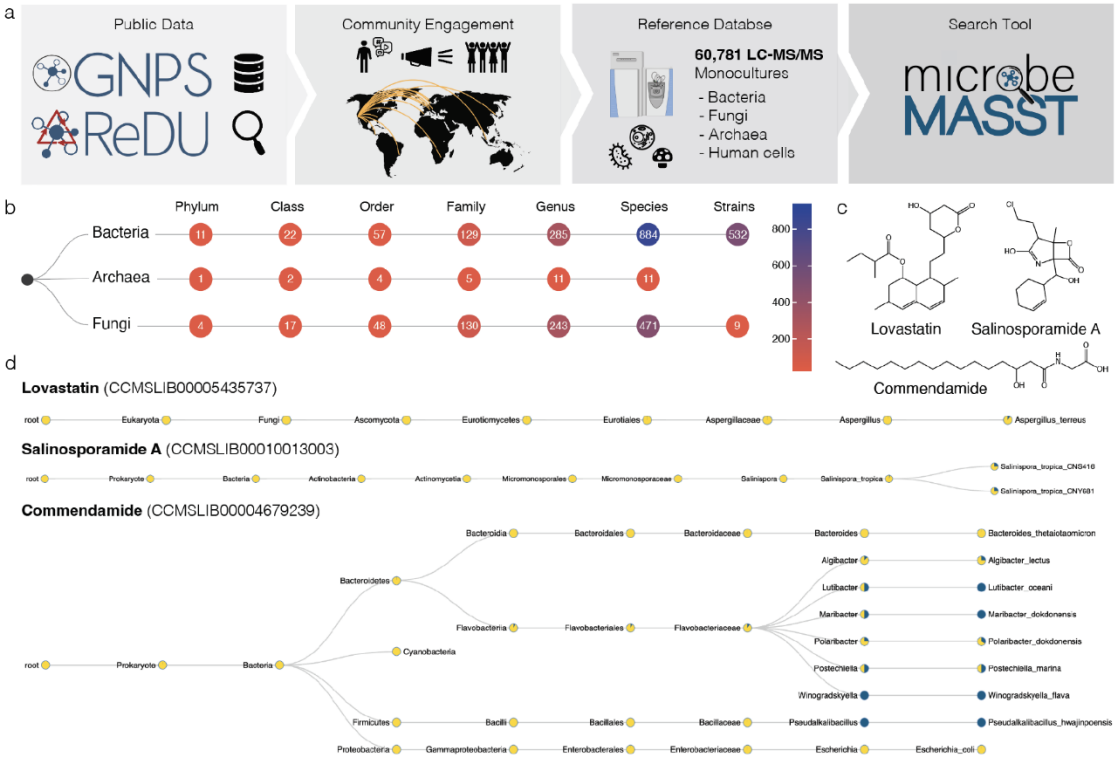


Figure 1. The microbeMASST search tool and reference database.
a) Community contributions of data and knowledge to GNPS[17], ReDU[23], and MassIVE from 2014 to 2022 were used to generate the microbeMASST reference database. Additionally, a public invitation to deposit data in June 2022 resulted in the further deposition of LC-MS/MS files from 25 different laboratories from 15 different countries across the globe, leading to the curation of a total of 60,781 LC-MS/MS files of microbial monoculture extracts. **b**) MicrobeMASST comprises 1,858 unique lineages, across three different domains of life, mapped to 541 unique strains, 1,336 species, 539 genera, 264 families, 109 orders, 41 classes, and 16 phyla. **c**) Examples of medically-relevant small molecules known to be produced by bacteria or fungi. Lovastatin, a cholesterol lowering drug originally isolated from *Aspergillus* genus[24], salinosporamide A, a Phase III candidate to treat glioblastoma produced by *Salinispora tropica*[25], and commendamide, a human G-protein–coupled receptor agonist[26]. **d**) MicrobeMASST search outputs of the three different molecules of interest confirm that they were exclusively found in monocultures of the only known producers. Pie charts display the proportion of MS/MS matches found in the deposited reference database. Blue indicates a match with a monoculture, while yellow represents a nonmatch. Searches were performed using MS/MS spectra deposited in the GNPS reference library: lovastatin (CCMSLIB00005435737), salinosporamide A (CCMSLIB00010013003), and commendamide (CCMSLIB00004679239).

Search results for lovastatin, salinosporamide A, and commendamide MS/MS spectra highlight how microbeMASST can correctly connect microbial molecules to their known producers (**Figure 1c**). In the case of lovastatin, a clinically-used cholesterol-lowering drug originally isolated from *Aspergillus terreus*[24], spectral matches were unique to the genus *Aspergillus* (**Figure 1d**). The MS/MS spectrum for salinosporamide A, a Phase III candidate to treat glioblastoma[27], only matched two strains of *Salinispora tropica* (**Figure 1d**), the only known producer[25]. Commendamide, first observed in cultures of *Bacteroides vulgatus* (recently reclassified as *Phocaeicola vulgatus*), is a G-protein–coupled receptor agonist[26]. Surprisingly it had many matches to several bacterial cultures, including in Flavobacteriaceae (*Algibacter*, *Lutibacter*, *Maribacter*, *Polaribacter, Postechiella*, and *Winogradskyella*) and *Bacteroides*

cultures (**Figure 1d**). Additional examples include searches of mevastatin, arylomycin A4, yersiniabactin, promicroferrioxamine, and the microbial bile acid conjugates[28–30] glutamate-cholic acid (Glu-CA) and glutamate-deoxycholic acid (Glu-DCA) (**Supplementary Figure 3**). Mevastatin, another cholesterol-lowering drug originally isolated from *Penicillium citrinum*[31], was only found in samples classified as fungi. The antibiotic arylomycin A4 was observed in different *Streptomyces* species and it was originally isolated from *Streptomyces* sp. Tue 6075 in 2002[32]. Yersiniabactin, a siderophore originally isolated from *Yersinia pestis*[33], whose monoculture is not yet present in the reference database of microbeMASST, was observed in *Escherichia coli* and *Klebsiella* species, consistent with previous observations[34,35]. Promicroferrioxamine, another siderophore, was observed to match *Micromonospora chokoriensis* and *Streptomyces* species. This molecule was originally isolated from an uncharacterized *Promicromonosporaceae* isolate[36]. The MS/MS spectrum of the gut microbiota-derived Glu-CA, an amidated tri-hydroxylated bile acid, was most frequently observed in cultures of *Bifidobacterium* species, while Glu-DCA was found only in one *Bifidobacterium* strain but also in two *Enterococcus* and *Clostridium* species. None of the aforementioned molecules were found in cultured human cell lines, highlighting the ability of microbeMASST to distinguish MS/MS spectra of molecules that can be exclusively produced by either bacteria or fungi. It is important to acknowledge that MS/MS data generally do not differentiate stereoisomers, but it can nevertheless provide crucial information on molecular families.

MicrobeMASST can be also used to extract microbial information from mass spectrometry-based metabolomics studies without any *a priori* knowledge. To illustrate this, we reprocessed an untargeted metabolomics study comparing germ-free (GF) mice to those harboring microbial communities, also known as specific pathogen-free (SPF) mice[29] (**Figure 2a**). We extracted 10,047 consensus MS/MS spectra uniquely present in SPF mice and queried them with microbeMASST. A total of 3,262 MS/MS spectra were found to have a microbial match. Of these, 837 were also found in human cell lines and for this reason were removed from further analysis. Among the remaining 2,425 MS/MS spectra, 1,673 were exclusively found in bacteria, 95 in fungi, and 657 in both (**Supplementary Figure 4**). These MS/MS spectra were then processed with SIRIUS[37] and CANOPUS[38] to tentatively annotate the metabolites and identify their chemical classes. A file containing all these spectra of interest can be explored and downloaded as .mgf format from GNPS (**see Methods**). To further validate the microbial origin of these MS/MS spectra, we assessed their overlap with data acquired from a different study comparing SPF mice treated with a cocktail of antibiotics to untreated controls[40]. Interestingly, 621 MS/MS spectra were also found in this second dataset and 512 were only present in animals not treated with antibiotics (**Figure 2b**). The distribution of these spectra and their classes across bacterial phyla was visualized using an UpSet plot[39] (**Figure 2c**). Notably, the majority of the spectra classified as terpenoids were commonly observed across phyla while amino acids and peptides appeared to be more phylum specific. Of these 512 spectra, 23% had a level 2 or 3 annotation[41], matching against the GNPS reference libraries (**Supplementary Table 1**). These included the recently described amidated microbial bile acids[19,28–30,42–47], free bile acids originating from the hydrolysis of host derived taurine bile acid conjugates[48], keto bile acids formed via microbial oxidation of alcohols[29], *N*-acyl-lipids belonging to a similar class of metabolites as commendamide[26] (a microbial *N*-acyl lipid), di- and tri-peptides seen in microbial digestion of proteins[49], and soyasapogenol, a byproduct of the microbial digestion of complex saccharides from dietary soyasaponins[29]. Part of the remaining unannotated spectra can be identified as chemical modifications of the above annotated microbial metabolites through spectral similarity obtained from molecular networking (**Supplementary Figure 5**). Based on literature information, the list of annotated MS/MS spectra contained a small number of metabolites traditionally considered to be non-microbial in origin. One interpretation of this finding is that microorganisms are capable of producing metabolites previously described to only be made by mammalian hosts. Notable examples include serotonin[50], γ-aminobutyric acid (GABA)[51], and the glycocholic acid[42,52–54], with microorganisms often being the primary producers of these metabolites in the gut. Additionally, an alternative hypothesis is that microorganisms can also selectively stimulate

the production of host metabolites. Other limitations regarding annotations are discussed in **Methods**. To assess if the observations from the mouse models translate to humans, we searched and found that 455 out of the 512 MS/MS spectra of interest matched to public human data (**Figure 2d**). Interestingly, these spectra were found in both healthy individuals and individuals affected by different health states, including type II diabetes, inflammatory bowel disease (IBD), Alzehimer's diseases and other conditions. These spectra were most commonly found in stool samples (n=110,973 MS/MS matches) followed by blood, breast milk, and the oral cavity as well as other organs including the brain, skin, vagina, and biofluids, such as cerebrospinal fluid and urine (**Figure 2e**). These findings support the concept that a significant number of microbial metabolites reach and influence distant organs in the human body[55].

We anticipate microbeMASST will be a key resource to enhance understanding of the role of microbial metabolites across a wide range of ecosystems, including oceans, plants, soils, insects, animals, and humans. This expanding resource will enable the scientific community to gain valuable taxonomic and functional insights into diverse microbial populations. The mass spectrometry community will play a key role in the evolution of this tool in the future through the continued deposition of data associated with novel microbial monocultures and the expansion of spectral reference libraries. Moreover, microbeMASST holds potential for various applications, ranging from aquaculture and agriculture to biotechnology and the study of microbial-mediated human health conditions. By harnessing the power of public data, we can unlock new opportunities for advancements in multiple fields and deepen our understanding of the intricate relationships between microorganisms and their ecosystems.

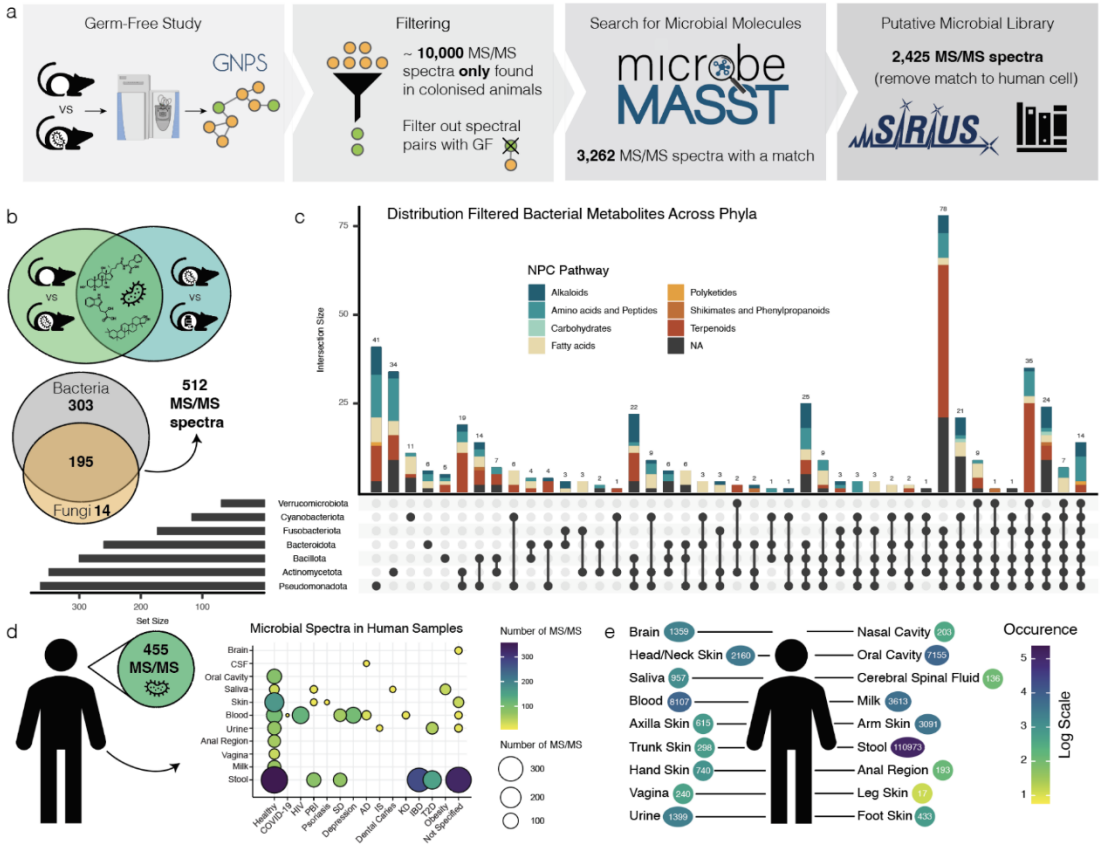


Figure 2. MicrobeMASST can identify microbial MS/MS spectra within mouse and human datasets. **a)** Workflow to extract microbial MS/MS spectra from biochemical profiles of 29 different tissues and biofluids of SPF mice that are not observed in GF mice[29]. The MS/MS spectra unique to SPF mice (10,047) were searched with microbeMASST. A total of 3,262 MS/MS spectra had a match; those MS/MS also matching human cell lines were removed, leaving a total of 2,425 putative microbial MS/MS spectra (**see Methods** to download .mgf file). **b**) The presence of the 2,425 MS/MS spectra was evaluated in an additional animal study looking at antibiotics usage[40]. A total of 512 MS/MS spectra, out of the 621 overlapping, was exclusively found in animals not receiving antibiotics. **c**) UpSet plot of the distribution of the detected MS/MS spectra (512) across bacterial phyla. Terpenoids were more commonly observed across phyla while amino acids and peptides appeared to be more phylum specific **d**) The 512 MS/MS spectra were searched in human datasets and 455 were found to have a match. These MS/MS spectra were present in both healthy individuals and individuals affected by different diseases. **e**) Most of the MS/MS spectra (n=411) matched fecal samples (n=110,973 matches), followed by blood, oral cavity, breast milk, urine, and several other organs. Abbreviations: CSF, cerebral spinal fluid; HIV, human immunodeficiency virus; PBI, primary bacterial infectious disease; SD, sleep disorder; AD, Alzheimer's disease; IS, ischemic stroke; KD, Kawasaki disease; IBD, inflammatory bowel disease; T2D, type II diabetes.

Acknowledgments

This work was carried out through the collaborative microbial metabolite center, which is supported by the National Institutes of Health (NIH) grant U24DK133658, the Alzheimer's gut project U19AG063744 and BBSRC-NSF award 2152526. We thank Travis Adkins and Lisa McCormick from the USDA ARS Culture Collection for their assistance selecting and providing microbial strains used in this research. This project was supported in part by the U.S. Department of Agriculture, Agricultural Research Service. AMCR and HM were supported by the NIH grant 1DP2GM137413. Research reported in this publication was supported in part by the National Center for Complementary and Integrative Health of the NIH under award number F32AT011475 to NEA. KBK is supported by National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (NRF-2020R1C1C1004046, 2022R1A5A2021216, and 2022M3H9A2096191). BS is supported by the Austrian Science Fund (FWF) P31915 and KH is financed by the FWF research group program (grant FG3). DP is supported by the German Research Foundation (DFG) CMFI Cluster of Excellence (EXC 2124) and Collaborative Research Center CellMap (TRR 261). AB, EAM, COG, PCJ, LVCL, NPL are supported by The São Paulo Research Foundation - FAPESP #2018/24865-4, #2019/03008-9, #2020/06430-0, #2022/12654-4, #2015/17177-6, #2020/02207-5, #2021/10603-0) and CNPq. CLC received financial support from StrainBiotech and the FEMSA Biotechnology Centre from Tecnológico de Monterrey. LRO received a scholarship from the Mexican National Council of Science and Technology (CONACYT). PCT and WHG are supported by NIH R01 GM107550. NG is supported by the NSF CAREER Award #2047235. NS, SD and JSSD are supported by ERC Horizon 2020 (grant agreement No 694569) and by a Spinoza award to JSSD from NWO. BW is supported by CMFI Cluster of Excellence (EXC 2124). MS is supported by DZIF (Grant no TTU 09.722). AVPDL, SLLR, and PP are supported by ERA-Net Cofund project BlueBio (grant agreement no. 311913), Research Council of Norway (300846). HMR, MFL and GLB are supported by Novo Nordisk Foundation (grant NNF19OC0056246; PRIMA—toward Personalized dietary Recommendations based on the Interaction between diet, Microbiome and Abiotic conditions in the gut). HMR is supported by the Independent Research Fund Denmark (MOTILITY; grant no. 0171-00006B). ECON is supported by the Nottingham Research Fellowship. AIPL is supported by FPU (FPU19/00289). MFT and RdCP were supported by R35GM12889.CMS is supported by Juan de la Cierva-Incorporación (IJC2018-036923-I) and Proyectos dirigidos por jóvenes investigadores de la Universidad de Málaga (B1-2021_21). DR is supported by Plan Nacional de I+D+i of the Ministerio de Ciencia e Innovación (PID2019-107724GB-I00), and Junta de Andalucía (P20_00479). XLG is supported by Nanyang Assistant Professorship. KSJ is supported by NIH R01 GM137135. AS is supported by EMBO fellowship ALTF 996-2021. SP is supported by ETH Zurich Career Seed Fellowship. RG is supported by the Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology. HC is supported by NIH R01AI167860 and CIFAR. MCT is supported by T32 DK007202 (NIDDK), the National Academies of Sciences, Engineering and Medicine through the Predoctoral Fellowship of the Ford Foundation, and the Howard Hughes Medical Institute (HHMI) Graduate Fellowships grant (GT15123). EEL is supported by VI-Universidad de Costa Rica, Grants numbers C1604 and C0469. PE is supported by VI-Universidad de Costa Rica, Grants numbers C1604 and C0469; U.S. National Science Foundation DEB-1638976. DAV and GTC are supported by VI-Universidad de Costa Rica, Grants numbers C1604 and C0469. KF is supported by CMFI Cluster of Excellence (EXC 2124). HHFK and GFS are supported by Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM). DBS is supported by Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul - FUNDECT (process number: 71/032.390/2022, FUNDECT number: 311/2022). KLM is supported by NIH / 1R01GM132649. PMA is supported by a swissuniversities Open Research Data grant. MCMK and SLJ are supported by NIH R35GM142938. ADP is supported by NIH U01 DK119702 and S10 OD021750. ATA was supported by the Betty and Gordon Moore Foundation. ML and PB are supported by the Max Planck Society. Omnia Group Ltd. is duly acknowledged for microbial cultures. BSP and NB were partially supported by NIH 1R01LM013115 and NSF ABI 1759980. RK is supported by NIH DP1AT010885. LPN is supported by Omnia Group Ltd. Shimadzu South Africa Ltd. is gratefully thanked for the analytical support. We thank James MacRae, head of The Metabolomics STP at the Francis Crick Institute for his guidance.

Disclosures

PCD is an advisor to Cybele, consulted for MSD animal health in 2023, and he is a Co-founder and scientific advisor for Ometa Labs, Arome, and Enveda with prior approval by UC San Diego. MW is a Co-founder of Ometa labs. There are no known conflicts of interest in this work by the USDA, Agricultural Research Service, National Center for Agricultural Utilization Research, Mycotoxin Prevention and Applied Microbiology Research Unit. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Author contributions (names given as initials)

SZ, RS, AB, MW, PCD - Conceptualized the method
RS, SZ, MW - Developed microbeMASST
PCD, SZ, AB, PWPG, AMCR, YEA, ATA, ECG, JZ, MJM, NEA, RHC, EB, MCT, CYH, RO, AVA, JZ, HC, MCMK, SJ, FT, LPN, NEM, IAD, EAM, LVCL, NPL, PRT, PCJ, BR, ADP, MFT, RCP, GTC, PC, EEL, DAV, LMQG, JLW, AS, SP, JJ, KG, JMC, PMA, BCW, KF, DP, NA, NG, ML, PB, KBK, AGG, HG, LPSC, MSS, AIPL, CMS, DR, RF, MB, AVPL, PBP, SLLR, GLB, MFL, HMR, AR, BS, FH, AJB, CLC, LRO, ER, FH, GK, HS, KH, LP, RG, ECON, ETR, JO, NJB, SD, XLG, JJC, KSJ, DBS, FMRS, GFS, HHFK, JAC, HM, KB, KLM, SOS, CMR, and RK Contributed data and curated metadata
SZ - Generated taxonomic tree and performed analyses
RS - Developed the tree visualizer for enriched ontologies and output summaries
BSP, NB - Developed the FASST algorithm
MW - Developed the Fast Search Tool API
SZ, RS, PCD - Tested microbeMASST
SZ, RS, PCD - Wrote the manuscript
All authors reviewed the manuscript

References

1. Jansson, J. K. & Hofmockel, K. S. Soil microbiomes and climate change. *Nat. Rev. Microbiol.* **18**, 35–46 (2020).
2. Malard, F., Dore, J., Gaugler, B. & Mohty, M. Introduction to host microbiome symbiosis in health and disease. *Mucosal Immunol.* **14**, 547–554 (2021).
3. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
4. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278 (2023).
5. Radjabzadeh, D. *et al.* Gut microbiome-wide association study of depressive symptoms. *Nat. Commun.* **13**, 7128 (2022).
6. Morais, L. H., Schreiber, H. L., 4th & Mazmanian, S. K. The gut microbiota-brain axis in behaviour and brain disorders. *Nat. Rev. Microbiol.* **19**, 241–255 (2021).
7. Milshteyn, A., Colosimo, D. A. & Brady, S. F. Accessing Bioactive Natural Products from the Human Microbiome. *Cell Host Microbe* **23**, 725–736 (2018).
8. Paoli, L. *et al.* Biosynthetic potential of the global ocean microbiome. *Nature* **607**, 111–118 (2022).
9. Grice, E. A. & Segre, J. A. The human microbiome: our second genome. *Annu. Rev. Genomics Hum. Genet.* **13**, 151–170 (2012).
10. Tierney, B. T. *et al.* The Landscape of Genetic Content in the Gut and Oral Human Microbiome. *Cell Host Microbe* **26**, 283–295.e8 (2019).
11. Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V., Jarmusch, A. K. & Dorrestein, P. C. Mass spectrometry-based metabolomics in microbiome investigations. *Nat. Rev. Microbiol.* **20**, 143–160 (2022).
12. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
13. Wishart, D. S. *et al.* MiMeDB: the Human Microbial Metabolome Database. *Nucleic Acids Res.* **51**, D611–D620 (2023).
14. van Santen, J. A. *et al.* The Natural Products Atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res.* **50**, D1317–D1323 (2022).
15. Rutz, A. *et al.* The LOTUS initiative for open knowledge management in natural products research. *Elife* **11**, (2022).
16. Han, S. *et al.* A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* **595**, 415–420 (2021).
17. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
18. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, (2020).
19. Wang, M. *et al.* Mass spectrometry searches using MASST. *Nat. Biotechnol.* **38**, 23–26 (2020).
20. Batsoyol, N., Pullman, B., Wang, M., Bandeira, N. & Swanson, S. P-massive: a real-time search engine for a multi-terabyte mass spectrometry database. in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis* 1–15 (IEEE Press, 2022).
21. Deutsch, E. W. *et al.* Universal Spectrum Identifier for mass spectra. *Nat. Methods* **18**, 768–770 (2021).
22. Watrous, J. *et al.* Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1743–52 (2012).
23. Jarmusch, A. K. *et al.* ReDU: a framework to find and reanalyze public mass spectrometry data. *Nat. Methods* **17**, 901–904

(2020).

24. Alberts, A. W. *et al.* Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3957–3961 (1980).
25. Fenical, W. *et al.* Discovery and development of the anticancer agent salinosporamide A (NPI-0052). *Bioorg. Med. Chem.* **17**, 2175–2180 (2009).
26. Cohen, L. J. *et al.* Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4825–34 (2015).
27. CTG Labs - NCBI. https://beta.clinicaltrials.gov/study/NCT03345095?distance=50&intr=NPI-0052&rank=9.
28. Dorrestein, P. *et al.* A synthesis-based reverse metabolomics approach for the discovery of chemical structures from humans and animals. (2021) doi:10.21203/rs.3.rs-820302/v1.
29. Quinn, R. A. *et al.* Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* **579**, 123–129 (2020).
30. Patterson, A. *et al.* Bile Acids Are Substrates for Amine N-Acyl Transferase Activity by Bile Salt Hydrolase. (2022) doi:10.21203/rs.3.rs-2050120/v1.
31. Endo, A. The origin of the statins. 2004. *Atheroscler. Suppl.* **5**, 125–130 (2004).
32. JUDITH SCHIMANA, KLAUS GEBHARDT, ALEXANDRA HÖLTZEL, DIETMAR G. SCHMID, RODERICH SÜSSMUTH, JOHANNES MÜLLER, RÜDIGER PUKALL, HANS-PETER FIEDLER. Arylomycins A and B, New Biaryl-bridged Lipopeptide Antibiotics Produced by Streptomyces sp. Tü 6075. *The Journal of Antibiotics* (2002).
33. Drechsel, H. *et al.* Structure elucidation of yersiniabactin, a siderophore from highly virulent Yersinia strains. *Liebigs Ann.: Org. Bioorg. Chem.* **1995**, 1727–1733 (1995).
34. Schubert, S., Picard, B., Gouriou, S., Heesemann, J. & Denamur, E. Yersinia high-pathogenicity island contributes to virulence in Escherichia coli causing extraintestinal infections. *Infect. Immun.* **70**, 5335–5337 (2002).
35. Lawlor, M. S., O'connor, C. & Miller, V. L. Yersiniabactin is a virulence factor for Klebsiella pneumoniae during pulmonary infection. *Infect. Immun.* **75**, 1463–1472 (2007).
36. Yang, Y.-L. *et al.* Connecting chemotypes and phenotypes of cultured marine microbial assemblages by imaging mass spectrometry. *Angew. Chem. Int. Ed Engl.* **50**, 5839–5842 (2011).
37. Dührkop, K. *et al.* SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **16**, 299–302 (2019).
38. Dührkop, K. *et al.* Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **39**, 462–471 (2021).
39. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
40. Shalapour, S. *et al.* Inflammation-induced IgA+ cells dismantle anti-liver cancer immunity. *Nature* **551**, 340–345 (2017).
41. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
42. Lucas, L. N. *et al.* Dominant Bacterial Phyla from the Human Gut Show Widespread Ability To Transform and Conjugate Bile Acids. *mSystems* e0080521 (2021).

43. Hoffmann, M. A. *et al.* High-confidence structural annotation of metabolites absent from spectral libraries. *Nat. Biotechnol.* **40**, 411–421 (2022).
44. Guzior, D. *et al.* Bile salt hydrolase/aminoacyltransferase shapes the microbiome. (2022) doi:10.21203/rs.3.rs-2050406/v1.
45. Foley, M. H. *et al.* Bile salt hydrolases shape the bile acid landscape and restrict Clostridioides difficile growth in the murine gut. *Nat Microbiol* **8**, 611–628 (2023).
46. Folz, J. *et al.* Human metabolome variation along the upper intestinal tract. *Nat Metab* (2023) doi:10.1038/s42255-023-00777-z.
47. Shalon, D. *et al.* Profiling the human intestinal environment under physiological conditions. *Nature* **617**, 581–591 (2023).
48. Yao, L. *et al.* A selective gut bacterial bile salt hydrolase alters host metabolism. *Elife* **7**, (2018).
49. Bartlett, A. & Kleiner, M. Dietary protein and the intestinal microbiota: An understudied relationship. *iScience* **25**, (2022).
50. Yano, J. M. *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161**, 264–276 (2015).
51. Strandwitz, P. *et al.* GABA-modulating bacteria of the human gut microbiota. *Nat Microbiol* **4**, 396–403 (2019).
52. Maneerat, S., Nitoda, T., Kanzaki, H. & Kawai, F. Bile acids are new products of a marine bacterium, Myroides sp. strain SM1. *Appl. Microbiol. Biotechnol.* **67**, 679–683 (2005).
53. Kim, D. *et al.* Biosynthesis of bile acids in a variety of marine bacterial taxa. *J. Microbiol. Biotechnol.* **17**, 403–407 (2007).
54. Ohashi, K., Miyagawa, Y., Nakamura, Y. & Shibuya, H. Bioproduction of bile acids and the glycine conjugates by Penicillium fungus. *J. Nat. Med.* **62**, 83–86 (2008).
55. Lai, Y. *et al.* High-coverage metabolomics uncovers microbiota-driven biochemical landscape of interorgan transport and gut-brain communication in mice. *Nat. Commun.* **12**, 6000 (2021).
56. Chamberlain, S. A. & Szöcs, E. taxize: taxonomic search and retrieval in R. *F1000Res.* **2**, 191 (2013).
57. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
58. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

Data and code availability

Data used to generate the reference database of microbeMASST are publicly available at GNPS/MassIVE (https://massive.ucsd.edu/). A list with all the accession numbers (MassIVE IDs) of the studies used to generate this tool is available in **Supplementary Table 2**. Interactive examples of the MS/MS queries illustrated in **Figure 1d** and **Supplementary Figure 3** can be generated, visualized, and downloaded from the microbeMASST website (https://masst.gnps2.org/microbemasst/). Known molecules already present in the GNPS library (https://library.gnps2.org/) were used to facilitate interpretation and confirm that specific bacterial and fungal molecules were exclusively observed in the respective monocultures.

- Lovastatin - CCMSLIB00005435737
- Salinosporamide A - CCMSLIB00010013003
- Commendamide - CCMSLIB00004679239
- Mevastatin - CCMSLIB00005435644
- Arylomycin A4 - CCMSLIB00000075066
- Yersiniabactin - CCMSLIB00005435750
- Promicroferrioxamine - CCMSLIB00005716848
- Glutamate-cholic acid (Glu-CA) - CCMSLIB00006582258
- Glutamate-deoxycholic acid (Glu-DCA) - CCMSLIB00006582092

Data used to extract MS/MS spectra exclusively present in colonized (SPF) mice is publicly available in GNPS/MassIVE under the accession number MSV000079949. Data used to validate and assess antibiotics effect on microbial MS/MS spectra of interest is available under the accession number MSV000080918. List of datasets with data acquired from human biosamples that presented matches to the putative microbial MS/MS spectra of interest is available in **Supplementary Table 3**.

The microbeMASST code to query spectra, create interactive trees, and analyze results is available under open source license on GitHub (https://github.com/robinschmid/microbe_masst). Code used to generate the microbeMASST web interface can be accessed on GitHub (https://github.com/mwang87/GNPS_MASST). Code used to perform the analysis and generate the figures present in the manuscript can be downloaded from GitHub (https://github.com/simonezuffa/Manuscript_microbeMASST)

Data collection and harmonization

Data deposited in GNPS/MassIVE was investigated manually and systematically, using ReDU[23] (https://redu.ucsd.edu/), to extract all the publicly available MS/MS files (.mzML or .mzXML formats) acquired from monocultures of bacteria, fungi, archaea, and human cell lines. Only monocultures were included in this search tool to unequivocally associate the production of the detected metabolites to each specific taxon. A total of 60,781 files from 537 different GNPS/MassIVE datasets were selected to be used as a reference database of microbeMASST (**Supplementary Table 2**). These comprise files deposited in response to our call to the scientific community. Between May and July 2022, 25 different research groups deposited 65 distinct datasets in GNPS/MassIVE, comprising a total of 3,142 unique LC-MS/MS files. This represented a 5.45% increase in publicly available MS/MS data acquired from monocultures in just two months. To qualify as a contributor and be credited as one of the authors, researchers had to deposit high resolution LC-MS/MS data acquired either in positive or negative ionization modes from monocultures of either bacteria, fungi, or achaea. Harmonization of the acquired data and metadata represented a challenge. The NCBI taxonomic database is constantly expanding and evolving and ReDU latest updated (December 2021) does not accommodate the latest deposited taxa. For this reason, an additional metadata file (microbeMASST_metadata_*massiveID*) was generated specifically for the microbeMASST project and uploaded to the respective GNPS/MassIVE datasets deposited by the collaborators, if the ReDU workflow failed. All the collected information was finally

aggregate in one single .csv file (microbe_masst_table.csv) that can be found on GitHub, which contains: 1) Full MassIVE path of each sample, 2) Filename of each sample reported as its MassIVE ID/file name to avoid presence of duplicated names, 3) MassIVE ID, 4) Taxonomic name of the isolate as reported by the author submitting the associated metadata, 5) Alternative taxonomic name if the provided taxonomic name was incorrect or not present in NCBI, 6) Associated NCBI ID to the taxonomic name or the alternative taxonomic name, when present, 7) Definition if the taxonomic ID was automatically assigned or manually curated, and information if 8) ReDU metadata is available for that specific file and if the file correspond to a 9) Blank or 10) Quality control (QC) rather than an unique biological sample.

Unique taxonomic names and NCBI IDs were extracted from the metadata associated with the selected samples. When metadata was not available and multiple species of bacteria or fungi were present in the same dataset, samples were generically classified as bacteria or fungi. Concordance between taxonomic names and NCBI IDs was checked by blasting taxonomic names against NCBI (https://www.ncbi.nlm.nih.gov/Taxonomy/TaxIdentifier/tax_identifier.cgi) to obtains respective NCBI IDs and updated taxonomic names. Results were manually investigated and missing IDs were recovered using the NCBI browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi). If the taxonomic name was not found in NCBI, most likely because it was not deposited yet, the NCBI of the closest taxon was retrieved and used. For example, the strain *Staphylococcus aureus* CM05 was unavailable in NCBI and was curated to the species *Staphylococcus aureus* instead.

Taxonomic tree generation

The microbeMASST taxonomic tree was generated using both R 4.2.2 (R Foundation for Statistical Computing) and Python 3.10 (Python Software Foundation). In R, the microbeMASST table was filtered and only unique NCBI IDs were retained (n = 1,834). The classification function from the `taxize` package (v 0.9.100) was used to retrieve the full lineage of each NCBI ID[56]. Main taxonomic ranks (kingdom to strain) plus subgenus, subspecies, and varieties were kept in order to obtain taxonomic trees with a similar number of nodes per lineage. The list of NCBI IDs of all lineages was then imported in Python, where the ETE3 toolkit was used to generate a comprehensive taxonomic tree based on the provided NCBI IDs[57]. The generated Newick tree was then converted into JSON format and information such as taxonomic rank, number of samples available was added. Additionally, children nodes for blanks and QCs were created in order to be visualized in the same tree, if observed.

MASST query

MicrobeMASST web application was built using Dash and Flask open source libraries for Python (https://github.com/mwang87/GNPS_MASST/blob/master/dash_microbemasst.py). The web app can receive as inputs either a USI[21] or an MS/MS spectrum (fragment ions and their intensities). Additionally, batch searches can be performed using a customisable python script that can read either a .tsv file containing a list of USIs or a single .mgf spectra file. Through the manuscript we showcase how we were able to search for more than 10,000 MS/MS spectra contained in a single .mgf file (approximately 2 hours run time). After receiving input information, microbeMASST leverages the Fast Search Tool (https://fasst.gnps2.org/fastsearch/) API and the sample-specific associated metadata to generate taxonomic trees. Fast searches are based on indexing all the MS/MS spectra present in GNPS/MassIVE according to the mass and intensity of their precursor ions and then restricting the search to only a set of relevant spectra that have a greater chance to achieve a high spectral similarity (modified cosine score) to the MS/MS of interest. Searches are customizable and default settings are the following: precursor and fragment ion mass tolerances = 0.05, minimum cosine score threshold = 0.7, minimum number of matching fragment ions = 3, and analog search = False. The JSON file of the microbeMASST taxonomic tree is then filtered according to the results and converted into a D3 JavaScript object that can be visualized as an HTML file.

Applications

We envision microbeMASST having several applications. First, we showcase how researchers can investigate single MS/MS spectra, using the web interface, and obtain matching results if their known or unknown MS/MS spectrum was previously observed in any of the microbial monocultures present in the microbeMASST database. Nine small molecules of interest were investigated using MS/MS spectra already deposited in the GNPS reference library (see **Data and code availability**). Second, we show how microbeMASST can be leveraged to mine for known or unknown microbial metabolites in MS studies. To test this hypothesis, we reprocessed an LC-MS/MS dataset acquired from GF and SPF mice[29]. A comprehensive molecular network was generated (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=893fd89b52dc4c07a292485404f97723). From the obtained job, the qiime2 artifact (qiime2_table.qza), the .mgf file (METABOLOMICS-SNETS-V2-893fd89b-download_clustered_spectra-main.mgf) containing all the captured MS/MS spectra, and the annotation table (METABOLOMICS-SNETS-V2-893fd89b-view_all_annotations_DB-main.tsv) were extracted. The .qza file first converted into a .biom file and then .tsv using QIIME2[58] to extract the feature table. This was then imported in R where only spectra present in tissues and biofluids of SPF animals were retained (n = 10,047). To add an extra layer of filtering, all MS/MS spectra that had an edge (cosine similarity > 0.7) and a delta parent ion mass +/- 0.02 Da with MS/MS spectra present in GF animals were removed (spectral pairs information is contained in networkedges_selfloop). All the MS/MS spectra were then run in batch using a custom python script of microbeMASST (processing time: ~2 hours, Apple M2 Max, 64GB RAM) to obtain microbial matches. Matching and filtered MS/MS spectra (n = 2.425) were aggregated into a single .mgf file that can be downloaded from GNPS (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=aecd30b9febd43bd8f57b88598a05553). Compound class of each MS/MS spectrum, with parent ion mass < 850 Da, was predicted with SIRIUS[37] and CANOPUS[38]. The 2,425 MS/MS spectra were then searched against the MSV000080918 dataset containing animals treated or not with antibiotics[40]. Matching and filtered MS/MS spectra (n = 512) were aggregated into a single .mgf file that can be downloaded from GNPS (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=c33855fc32c948049331e9730189d5c1). A list of the spectra with putative chemical class classification is available in **Supplementary Table 1**. Venn diagrams and UpSet plots were generated in R using `VennDiagram 1.7.3`, `UpSetR 1.4.0`, and `ComplexUpset 1.3.3`. Finally, the 512 MS/MS spectra were searched in batch against the GNPS public repository to observe if they were detected in human datasets (**Supplementary Table 3**)

Technical limitations

Analysis of the results should be considered with these limitations in mind. Molecule detection in microbeMASST is dependent on the availability of specific substrates in the reference monocultures. If the culture lacks the necessary substrates (or any other culture condition) to produce a certain molecule, this molecule will not be detected. Nevertheless, if related substrates are present then a different but related molecule may be produced instead, which can be detected using the analog search. To address this problem, it is crucial for the community to continue to gather data from as many diverse experimental conditions as possible to capture the full range of metabolites produced by different microorganisms. This will help in building the most comprehensive reference database that encompasses diverse microbial metabolic profiles. Isomers and stereoisomers, which are molecules with the same molecular formula but different structural arrangements, often exhibit similar MS/MS spectra. This means that their fragmentation patterns may not provide enough information to distinguish them. Differences in extraction conditions and instrument settings can lead to variations in the obtained MS/MS spectra. For example, the intensity of precursor ions used for fragmentation can impact the resulting spectra. If the precursor ion intensity is low, the fragmented spectrum may lack ions that are present in spectra obtained from high-intensity precursor ions. This may result in "data leakage", as the MS/MS spectrum may be missing ions, and leading to the two molecules not being recognized as the same molecule. To partially overcome this more permissive settings can be created. The majority of the data deposited in public repositories, GNPS included, and used in microbeMASST

were acquired using positive ionization mode, which limits the observation of molecules that cannot be ionized in positive mode. This means that certain metabolites may be underrepresented or not detected at all. The continuous curation of the microbeMASST reference database involves adding more diverse data in terms of ionization modes to improve the coverage of metabolites. Taxonomic tree was generated using associated NCBI IDs provided by the community. Specimen assignment prior to metabolomic analysis can not be checked by microbeMASST. The majority of the deposited data do not have associated genetic information and even if available, it was not used to build taxonomic tree. Thus, specimen mis-identification is possible. By addressing these challenges and continuously curating the reference database with more comprehensive and diverse data, microbeMASST coverage can be expanded to provide valuable insights into the role of microbiota and to facilitate our understanding of microbial metabolism in diverse ecosystems.

CAPÍTULO V. DISCUSION GENERAL

La selección de los hongos endófitos basados en registros conocidos de bases de datos permite el estudio acelerado de especies con actividad antagonista contra patógenos de café. Claramente, la correcta clasificación taxonómica es compleja, resultando el uso de herramientas de biología molecular la forma más rápida para identificación de hongos (Molina et al., 2011). Actualmente, la base de datos UNITE (Abarenkov et al., 2010) resulta ser aquella con información mejor curada para hongos basándose en el marcador ITS. Si esta información se complementa con la encontrada en FUNGuild (Nguyen et al., 2016), en donde, es posible realizar anotaciones sobre la ecología de las especies que se han identificado, entonces, se tiene un panorama más amplio sobre una determinada colección de hongos y así, acelerar las pruebas usando los organismos con un mejor potencial para cumplir el objetivo deseado. En el caso de este trabajo, la aplicación de un flujo de trabajo bioinformático optimizado (Montero-Vargas et al., 2020), permitió seleccionar los mejores organismos para realizar pruebas de antagonismo contra el patógeno de café *Mycena citricolor*. La colección de endófitos de Rubiaceae cuenta con alrededor de 1400 aislamientos correspondientes a 800 especies. A partir de la aplicación del flujo bioinformático, se procedió con la selección de las especies más adecuadas para realizar pruebas de antagonismo.

Como se aprecia en (Escudero-Leyva, Granados-Montero, et al., 2023), *Nectria pseudotrichia* (GUHN1), *Trichoderma rifaii* (CT5), *T.* aff. *crassum* (G1C), *T.* aff. *atroviride* (G7T) y *T.* aff. *strigosellum* (GU12) inhibieron de manera *in vitro* al patógeno *Mycena citricolor* siendo los porcentajes de inhibición obtenidos con las especies de *Trichoderma*, similares a las de ciertos fungicidas. De las especies probadas, solamente *T. rifaii* había sido reportada como endófita de *Thobroma gileri* en Ecuador (Gazis & Chaverri, 2015). Los mecanismos implicados en la respuesta antagónica de *T. rifaii* se proponen como micoparasitismo directo similar a los aislados de *Trichoderma* probados con éxito contra *Corynespora cassiicola* (De Souza et al., 2008b) y Pujade-Renaud et al. (2019), o contra *Colletotrichum* spp. (Gazis y Chaverri, 2015). La otra especie resultante como micoparásita, *Trichoderma* aff. *crassum* no ha sido ampliamente estudiada como

agente de control biológico y, por lo tanto, los mecanismos de antagonismo aún no están dilucidados (Hoyos-Carvajal et al., 2009; Mendoza et al., 2015; Sumida et al., 2018; Zhang et al., 2015). Con respecto a las restantes especies de *Trichoderma* probadas in vitro, existen reportes que indican que *T. viride* tiene habilidades antagónicas contra *Fusarium graminearum*, *F. solani* y *Macrophomina phaseolina*; sin embargo, el porcentaje de inhibición fue menor en comparación con otras especies de *Trichoderma* probadas en ese estudio.

Respecto a los ensayos *in planta*, ambos grupos de plantas inoculadas con *Trichoderma* tenían al menos el doble de altura que las plantas sin endófitos. Futuros experimentos podrían dilucidar los reguladores de la promoción del crecimiento derivados de los hongos endófitos o provocados por la infección. Por ejemplo, los reguladores mejor reconocidos que participan en la promoción del crecimiento de las plantas son las citoquininas, el ácido indolacético y las giberelinas (Deng & Cao, 2017; Rana et al., 2020). De igual manera, ambos grupos de plantas con endófitos resultaron más resistentes a la infección con *M. citricolor*, y cuando esta ocurrió, la severidad de la enfermedad fue reducida en comparación con el grupo de plantas control.

Resalta de manera particular que las especies *T. rifaii* y *T.* aff. *crassum*, las cuales promovieron el crecimiento en plantas de café y proveyeron de defensa contra *M. citricolor*, son también las especies con mejores resultados en las pruebas de tolerancia y remoción de fungicidas. Los microorganismos con potencial de control biológico deben tolerar los agroquímicos para prevalecer con éxito en los sistemas agrícolas en transición y brindar beneficios a las plantas contra las enfermedades (Sun et al., 2019). El ciproconazol y el clorotalonil fueron los más tóxicos para *Trichoderma*, contrario a lo observado con el propineb y la validamicina-A. La tolerancia a pesticidas por parte de especies de *Trichoderma* relacionadas con el complejo *T. harzianum* recuperado del suelo se ha reportado previamente para compuestos como el pentaclorofenol (Rigot & Matsumura, 2002) o el glifosato (Levesque & Rahe, 1992). En general, el potencial de remoción más alto entre los aislamientos fue observado para *T. rifaii*, que logró una eliminación considerable de todos los fungicidas, mientras que simultáneamente detoxificaba la matriz

acuosa (excepto en el caso del ciproconazol). Por otro lado, el aislamieno *T.* aff. *crassum*, a pesar de poder remover los fungicidas en cierta medida, en la mayoría de los casos no pudo detoxificar la matriz, haciendo de este hongo una opción inadecuada para la biodegradación. Como se ha hecho con los probióticos en humanos, la manipulación o la gestión del fitobioma de un cultivo para incluir microorganismos beneficiosos para las plantas, desempeñará un papel clave en la revolución de la agricultura moderna (Hyde et al., 2019; Tilocca et al., 2020).

Dentro de las preguntas que surgen alrededor de los mecanismos de protección conferidos de los endófitos a sus hospederos, la producción de metabolitos secundarios, especialmente los CVOs, llaman la atención por la capacidad de producir efectos fungistáticos y fungicidas. En los resultados obtenidos en el trabajo con *Daldinia eschscholtzii* en confrontación sin contacto contra *M. citricolor* (Escudero-Leyva *et al.*, 2023c), se detectaron 39 compuestos mediante HS-SPME y 13 usando solamente HS. El ANOVA reveló diferencias significativas en las composiciones de CVOs entre el grupo de *Daldinia* (D) y el grupo de cocultivo. El grupo D mostró niveles más altos de 4,4-dimetil-1,3-ciclopentanodiona, 1,2-dimetil-4-oxociclohex-2-enecarbaldehído, α-selineno y pogostole, todos los cuales fueron estadísticamente significativos ($p < 0,05$). Por el contrario, el grupo de cocultivo mostró niveles significativamente más altos de 9-epi-β-cariofileno y 1,8-dimetoxinaftaleno. Estos hallazgos sugieren que la interacción entre *D. eschscholtzii* y el patógeno en el experimento de cocultivo influye en la producción de CVOs específicos, lo que lleva a perfiles químicos distintos en comparación con *D. eschscholtzii* solo. Al utilizar el conjunto completo de variables, las 10 mejores detectadas como VIP en el PLS-DA difieren de las cargas obtenidas del PCA. Sin embargo, al realizar el ANOVA de los datos, se alineó mejor con los resultados de PCA. Por lo tanto, las variables como significativas por ANOVA también contribuyen significativamente a la separación entre los grupos en el gráfico PCA. El PCA consideró al compuesto 1,8-dimetoxinaftaleno como la única carga asociada con las muestras de cocultivo, los VIP del PLS-DA incluyeron 9-epi-β-cariofileno y 1,8-dimetoxinaftaleno. Lo último nos lleva a probar el efecto de estos compuestos contra las estructuras reproductivas de *M. citricolor*, encontrando que los pseudopileos del patógeno colocados en las cajas Petri junto con los discos filtrantes cargados con 500 y

1000 µg del compuesto 1,8-dimetoxinaftaleno experimentaron el efecto de "quemado" después de 24 h a temperatura ambiente, lo que fue similar al daño observado cuando el patógeno se enfrentó a *D. eschscholtzii*. Ninguno de los controles negativos produjo un efecto sobre los pseudopileios. Esto sugiere que el 1,8-dimetoxinaftaleno producido en el grupo de cocultivo es responsable del efecto fungicida. Se ha registrado una actividad inhibitoria baja para el compuesto 1,8-dimetoxinaftaleno obtenido de un endófito de *D. eschscholtzii* de un manglar en Tailandia contra *Microsporum gypseum* y *Staphylococcus aureus* (Kongyen et al., 2015). Hay resultados de pruebas de herbicidas realizadas en plantas mono y dicotiledóneas también registradas como bajas, retrasando la germinación de las semillas (Dai et al., 2006). Aunque uno de los desafíos actuales en la metabolómica no dirigida es la confianza y la disponibilidad de los espectros en las bases de datos, especialmente los derivados de microorganismos como los hongos (Theodoridis et al., 2023), pudimos producir una anotación de características putativas hasta el nivel de clase química para los extractos obtenidos de los grupos *Daldinia*, cocultivo y *Mycena*. Asumiendo que el cocultivo podría contener compuestos derivados de ambos hongos que crecen en el ensayo, y dado que la membrana celular del hongo difiere en su composición de plantas y animales, los compuestos relacionados con esta estructura pueden ser pigmentos, por ejemplo, 1,8-dihidroxinaftaleno, común en *Aspergillus* spp., (Hanson, 2008) o lípidos, de los cuales el ergosterol es un componente principal de la membrana celular fúngica (Thevissen et al., 2003).

Debido a que los endófitos son productores de metabolitos secundarios, poco se sabe sobre cómo estos se expresan según el hospedero al que pertenecen. Por ello, se hace necesario conocer el perfil de metabolitos de sus hospederos para posteriormente hipotetizar y llevar a cabo experimentos que permitan comprender mejor las relaciones entre los endófitos y sus hospederos. La aplicación de metabolómica no dirigida y dirigida sobre *Coffea* sp. podría diferenciar los perfiles metabólicos de especies de extractos de hojas y frutos. Adicionalmente, se identificaron cinco fitoquímicos (cafeína, mangiferina y tres ácidos cafeoilquínicos), corroborando la identidad de los metabolitos diferenciados entre especies y tejidos. Los resultados mostraron que una extracción se puede realizar en menos de dos horas y se optimizó para recuperar una amplia variedad de compuestos (más de 40 familias

de fitoquímicos), de estos, el 90% tuvieron una recuperación superior al 50%, cuantificando 184 fitoquímicos de manera sensible y específica en una sola ejecución de LC-MS/MS de 15 min con cambio de polaridad, lo que es particularmente adecuado para análisis de alto rendimiento (Castro-Moretti et al., 2023). Como se menciona anteriormente, los mayores retos dentro de los experimentos de LC-MS, está la anotación de compuestos provenientes de microorganismos. Por ello, la colaboración con el grupo de Pieter Dorrestein resulta de gran interés. Al integrar los datos de experimentos de LC-MS a la herramienta microbeMASST (Zuffa et al., 2023), integrada dentro de la plataforma GNPS (Communication, 2019), permitirá buscar espectros MS/MS en tándem obtenidos de sus experimentos contra espectros MS/MS detectados previamente en otros extractos de monocultivos bacterianos, fúngicos o arqueales. Ningún otro recurso o herramienta disponible permite vincular espectros MS/MS no caracterizados con microorganismos caracterizados. La base de datos de referencia microbeMASST de monocultivos se ha generado a través de años de contribuciones comunitarias y conservación de metadatos, y contiene microorganismos aislados de plantas, suelos, océanos, lagos, peces, animales terrestres y humanos (Zuffa et al., 2023).

CAPÍTULO VI. CONCLUSIONES GENERALES

En conclusión, se obtuvieron 1400 aislamientos clasificados en 800 especies, de los cuales, posterior a la selección de las especies fúngicas con mejores cualidades basados en su relación filogenética, se procedieron a realizar experimentos de antagonismo contra el patógeno de café, *Mycena citricolor*. De la selección original que incluía géneros como *Sarocladium* o *Fusarium*, se optó por el uso *in vitro* e *in planta* de aislamientos de dos especies de *Trichoderma* provenientes de Rubiaceae silvestre. Estos tuvieron efectos positivos en la promoción del crecimiento de las plantas de café y el antagonismo contra *Mycena citricolor*. Los dos aislamientos también fueron capaces de colonizar los tejidos internos de las plántulas, aumentar el tamaño de las plantas, reducir la incidencia y severidad y prevenir la mortalidad de las plantas. Los aislamientos de *Trichoderma* endófitas fueron capaces de tolerar y prevalecer bajo la presencia de fungicidas. Sin embargo, *Trichoderma rifaii* CT5 sería la opción principal con respecto a la promoción del crecimiento, el manejo de enfermedades, la tolerancia a los fungicidas y la capacidad de biorremoción/biorremediación no tóxica.

El endófito *Daldinia eschscholtzii* GU11N produjo daño en las estructuras reproductivas del patógeno del café *M. citricolor* en un ensayo sin contacto. Usando HS-SPME, el perfil de *D. eschscholtzii* fue más diverso, detectando las clases encontradas con HS además de cromonas y policétidos, así como otros componentes sesquiterpenoides. La composición de este perfil reveló cambios significativos cuando el endófito fue expuesto a *M. citricolo*r en los experimentos de cocultivo. El 1,8-dimetoxinaftaleno tuvo una correlación estadística con el experimento de cocultivo y produjo el daño sobre las estructuras reproductivas de *M. citricolor* respaldado por observaciones con microscopía electrónica.

El estudio de metabolómica dirigida y no dirigida para *Coffea* spp. y especies silvestre de Rubiaceae consiguió la optimización de un procedimiento de extracción para recuperar 42 familias distintas de fitoquímicos de las hojas, un método LC-MS/MS robusto y sensible para cuantificar 184 metabolitos secundarios. Este trabajo describe un proceso sensible y completo para la detección, clasificación y cuantificación de metabolitos secundarios en

hojas de café y otras rubiáceas pudiéndose aplicar a otros órganos y especies de plantas. De manera similar, la colaboración en la plataforma microbeMASST será un recurso clave para mejorar la comprensión del papel de metabolitos microbianos en una amplia gama de ecosistemas, incluidos océanos, plantas, suelos, insectos, animales y humanos. Este recurso en expansión permitirá obtener valiosos conocimientos taxonómicos y funcionales sobre diversas poblaciones microbianas en diversas aplicaciones, que van desde la acuicultura y la agricultura hasta la biotecnología y el estudio de las condiciones de salud humana mediadas por microbios, profundizando la comprensión de las intrincadas relaciones entre los microorganismos y sus ecosistemas.

BIBLIOGRAFIA GENERAL

Abarenkov, K., Nilsson, R. H., Larsson, K. H., Alexander, I. J., Eberhardt, U., Erland, S., Høiland, K., Kjøller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A. F. S., Tedersoo, L., Ursing, B. M., Vrålstad, T., Liimatainen, K., Peintner, U., & Kõljalg, U. (2010). The UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytologist*, *186*(2), 281–285. https://doi.org/10.1111/j.1469-8137.2009.03160.x

Allegra, A., Gentry, E. C., McPhail, K. L., Nothias, L-F., Nothias-Esposito, M., Bouslimani, A., Petras, D., Gauglitz, J. M., Sikora, N., Vargas, F., van der Hooft, J. J. J., Ernst, M., Kang, K. B., Aceves, C. M., Caraballo-Rodríguez, A. M. Koester, I. Weldon, K. C., Bertrand, S., Roulliers C., Sun, K., Tehan, R. M., Boya, C. A., Christian, M. H., Gutiérrez, M., Moreno, A. U., Tejeda, M. J. A., Mojica-Flores, R., Lakey-Beitia, J., Vásquez-Chaves, V., Zhang, Y., Calderón, A. I., Tayler, N., Keyzers R. A., Tugizimana, F., Ndlovu, N., Aksenov, A. A., Jarmusch, A. K., Schmid, R., Truman, A. W., Bandeira N., Wang, M. & Dorrestein, P. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature protocols*, *15*(6), 1954–1991. https:// 10.1038/s41596-020-0317-5

Alpízar, F., Saborío-Rodríguez, M., Martínez-Rodríguez, M. R., Viguera, B., Vignola, R., Capitán, T., & Harvey, C. A. (2020). Determinants of food insecurity among smallholder farmer households in Central America: recurrent versus extreme weather-driven events. *Regional Environmental Change*, *20*(1). https://doi.org/10.1007/s10113-020-01592-y

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., Läderach, P., Anzueto, F., Hruska, A. J., & Morales, C. (2015). The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Security*, *7*(2), 303–321. https://doi.org/10.1007/s12571-015-0446-9

Barberis, L., Michalet, S., Piola, F., & Binet, P. (2021). Root fungal endophytes: identity, phylogeny and roles in plant tolerance to metal stress. *Fungal Biology*, *125*(4), 326–345. https://doi.org/10.1016/j.funbio.2020.11.011

Bernardi-Wenzel, J., García, A., Celso, ;, Filho, J. R., Prioli, A. J., & Pamphile, J. A. (2010). Evaluation of foliar fungal endophyte diversity and colonization of medicinal plant Luehea divaricata (Martius et Zuccarini). *Biol Res*, *43*(43), 375–384.

Busby, P. E., Ridout, M., & Newcombe, G. (2016). Fungal endophytes: modifiers of plant disease. *Plant Molecular Biology*, *90*(6), 645–655. https://doi.org/10.1007/s11103-015-0412-0

Carbone, I., & Kohn, L. M. (1999). A Method for Designing Primer Sets for Speciation Studies in Filamentous Ascomycetes. *Mycologia*, *91*(3), 553–556.

Castro-Moretti, F. R., Cocuron, J. C., Castillo-Gonzalez, H., Escudero-Leyva, E., Chaverri, P., Guerreiro-Filho, O., Slot, J. C., & Alonso, A. P. (2023). A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography mass spectrometry. *Frontiers in Plant Science*, *13*(January), 1–33. https://doi.org/10.3389/fpls.2022.1057645

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D. S., & Xia, J. (2018). MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*, *46*(W1), W486–W494. https://doi.org/10.1093/nar/gky310

Cocuron, J. C., Casas, M. I., Yang, F., Grotewold, E., & Alonso, A. P. (2019). Beyond the wall: High-throughput quantification of plant soluble and cell-wall bound phenolics by liquid chromatography tandem mass spectrometry. *Journal of Chromatography A*, *1589*, 93–104. https://doi.org/10.1016/j.chroma.2018.12.059

Dai, J., Krohn, K., Flörke, U., Draeger, S., Schulz, B., Kiss-Szikszai, A., Antus, S., Kurtán, T., & van Ree, T. (2006). Metabolites from the Endophytic FungusNodulisporium sp. fromJuniperus cedre. *European Journal of Organic Chemistry*, *2006*(15), 3498–3506. https://doi.org/10.1002/ejoc.200600261

De Souza, J. T., Bailey, B. A., Pomella, A. W. V., Erbe, E. F., Murphy, C. A., Bae, H., & Hebbar, P. K. (2008a). Colonization of cacao seedlings by Trichoderma stromaticum, a mycoparasite of the witches' broom pathogen, and its influence on plant growth and resistance. *Biological Control*, *46*(1), 36–45. https://doi.org/10.1016/j.biocontrol.2008.01.010

De Souza, J. T., Bailey, B. A., Pomella, A. W. V., Erbe, E. F., Murphy, C. A., Bae, H., &

Hebbar, P. K. (2008b). Colonization of cacao seedlings by Trichoderma stromaticum, a mycoparasite of the witches' broom pathogen, and its influence on plant growth and resistance. *Biological Control*, *46*(1), 36–45. https://doi.org/10.1016/j.biocontrol.2008.01.010

Deng, Z., & Cao, L. (2017). Fungal endophytes and their interactions with plants in phytoremediation: A review. In *Chemosphere* (Vol. 168, pp. 1100–1106). Elsevier Ltd. https://doi.org/10.1016/j.chemosphere.2016.10.097

Escudero-Leyva, E., Alfaro-Vargas, P., Muñoz-Arrieta, R., Charpentier-Alfaro, C., Granados-Montero, M. del M., Valverde-Madrigal, K. S., Pérez-Villanueva, M., Méndez-Rivera, M., Rodríguez-Rodríguez, C. E., Chaverri, P., & Mora-Villalobos, J. A. (2022). Tolerance and Biological Removal of Fungicides by Trichoderma Species Isolated From the Endosphere of Wild Rubiaceae Plants. *Frontiers in Agronomy*, *3*(February), 1–14. https://doi.org/10.3389/fagro.2021.772170

Escudero-Leyva, E., Granados-Montero, M. del M., Orozco-Ortiz, C., Araya-Valverde, E., Alvarado-Picado, E., Chaves-Fallas, J. M., Aldrich-Wolfe, L., & Chaverri, P. (2023). The endophytobiome of wild Rubiaceae as a source of antagonistic fungi against the American Leaf Spot of coffee (Mycena citricolor). *Journal of Applied Microbiology*, *134*(5), 1–12. https://doi.org/10.1093/jambio/lxad090

Escudero-Leyva, E., Quirós-Guerrero, L., Vásquez-Chaves, V., Pereira-Reyes, R., Chaverri, P., & Tamayo-Castillo, G. (2023). Differential Volatile Organic Compound Expression in the Interaction of Daldinia eschscholtzii and Mycena citricolor. *ACS Omega*, *8*(34), 31373–31388. https://doi.org/10.1021/acsomega.3c03865

Evans, H., Holmes, K., & Thomas, S. (2003). Endophytes and mycoparasites associated with an indigenous forest tree, Theobroma gileri, in Ecuador and a preliminary assessment of their potential as biocontrol agents of cocoa diseases. *Mycological Progress*, *2*(2), 149–160. https://doi.org/10.1007/s11557-006-0053-4

Flores-Reséndiz, M., Lappe-Oliveras, P., & Macías-Rubalcava, M. L. (2021). Mitochondrial damage produced by phytotoxic chromenone and chromanone derivatives from endophytic fungus Daldinia eschscholtzii strain GsE13. *Applied Microbiology and Biotechnology*, *105*(10), 4225–4239. https://doi.org/10.1007/s00253-021-11318-7

Gazis, R., & Chaverri, P. (2010). Diversity of fungal endophytes in leaves and stems of wild rubber trees (Hevea brasiliensis) in Peru. *Fungal Ecology*, *3*(3), 240–254. https://doi.org/10.1016/j.funeco.2009.12.001

Gazis, R., & Chaverri, P. (2015). Wild trees in the Amazon basin harbor a great diversity of beneficial endosymbiotic fungi: Is this evidence of protective mutualism? *Fungal Ecology*, *17*, 18–29. https://doi.org/10.1016/j.funeco.2015.04.001

Gharieb, M. M., Ali, M. I., & El-Shoura, A. A. (2004). Transformation of copper oxychloride fungicide into copper oxalate by tolerant fungi and the effect of nitrogen source on tolerance. *Biodegradation*, *15*(1), 49–57. https://doi.org/10.1023/B:BIOD.0000009962.48723.df

Granados-Montero, M., Avelino, J., Arauz-Cavallini, F., Castro-Tanzi, S., & Ureña, N. (2020). Leaf litter and Mycena citricolor inoculum on the American leaf spot epidemic. *Agronomy Mesoamerican*, *31*(1), 77–94. https://doi.org/10.15517/AM.V31I1.36614

Hanson, J. R. (2008). Pigments and Odours of Fungi. *The Chemistry of Fungi*, 127–146.

Hoyos-Carvajal, L., Orduz, S., & Bissett, J. (2009). Genetic and metabolic biodiversity of Trichoderma from Colombia and adjacent neotropic regions. *Fungal Genetics and Biology*, *46*(9), 615–631. https://doi.org/10.1016/j.fgb.2009.04.006

Hubbard, M., Germida, J. J., & Vujanovic, V. (2014). Fungal endophytes enhance wheat heat and drought tolerance in terms of grain yield and second-generation seed viability. *Journal of Applied Microbiology*, *116*(1), 109–122. https://doi.org/10.1111/jam.12311

Hyde, K. D., Xu, J., Rapior, S., Jeewon, R., Lumyong, S., Niego, A. G. T., Abeywickrama, P. D., Aluthmuhandiram, J. V. S., Brahamanage, R. S., Brooks, S., Chaiyasen, A., Chethana, K. W. T., Chomnunti, P., Chepkirui, C., Chuankid, B., de Silva, N. I., Doilom, M., Faulds, C., Gentekaki, E., … Stadler, M. (2019). The amazing potential of fungi: 50 ways we can exploit fungi industrially. *Fungal Diversity*, *97*(1), 1–136. https://doi.org/10.1007/s13225-019-00430-9

Kongyen, W., Rukachaisirikul, V., Phongpaichit, S., & Sakayaroj, J. (2015). A new hydronaphthalenone from the mangrove-derived Daldinia eschscholtzii PSU-STD57. *Natural Product Research*, *29*(21), 1995–1999.

https://doi.org/10.1080/14786419.2015.1022542

Lee, K. H., Bonn, M. A., & Cho, M. (2015). Consumer motives for purchasing organic coffee: The moderating effects of ethical concern and price sensitivity. *International Journal of Contemporary Hospitality Management*, *27*(6), 1157–1180. https://doi.org/10.1108/IJCHM-02-2014-0060

Levesque, C. A., & Rahe, J. E. (1992). Herbicide interactions with fungal root pathogens, with special reference to glyphosate. *Annual Review of Phytopathology. Vol. 30*, 579–602.

Liarzi, O., Bar, E., Lewinsohn, E., & Ezra, D. (2016). Use of the Endophytic Fungus Daldinia cf. concentrica and Its Volatiles as Bio-Control Agents. *PLOS ONE*, *11*(12), e0168242. https://doi.org/10.1371/journal.pone.0168242

Libert Amico, A., Ituarte-Lima, C., & Elmqvist, T. (2020). Learning from social–ecological crisis for legal resilience building: multi-scale dynamics in the coffee rust epidemic. *Sustainability Science*, *15*(2), 485–501. https://doi.org/10.1007/s11625-019-00703-x

Liu, Y. J., Whelen, S., & Hall, B. D. (1999). Phylogenetic relationships among ascomycetes: Evidence from an RNA polymerse II subunit. *Molecular Biology and Evolution*, *16*(12), 1799–1808. https://doi.org/10.1093/oxfordjournals.molbev.a026092

Manns, U., Wikstro, N., Taylor, C., & Bremer, B. (2012). Historical Biogeography of the Predominantly Neotropical Subfamily Cinchonoideae (Rubiaceae): Into or Out of America? *International Journal of Plant Sciences*, *173*(3), 261–286. https://doi.org/10.1086/663971

Mendoza, J. L. H., Pérez, M. I. S., Prieto, J. M. G., Velásquez, J. D. Q., Olivares, J. G. G., & Langarica, H. R. G. (2015). Antibiosis of trichoderma spp strains native to Northeastern Mexico against the pathogenic fungus Macrophomina phaseolina. *Brazilian Journal of Microbiology*, *46*(4), 1093–1101. https://doi.org/10.1590/S1517-838246420120177

Molina, R., Horton, T. R., Trappe, J. M., & Marcot, B. G. (2011). Addressing uncertainty: How to conserve and manage rare or little-known fungi. *Fungal Ecology*, *4*(2), 134–146. https://doi.org/10.1016/j.funeco.2010.06.003

Montero-Vargas, M., Escudero-Leyva, E., Díaz-Valerio, S., & Chaverri, P. (2020a). Step-

by-Step Pipeline for the Ecological Analysis of Endophytic Fungi using ITS nrDNA Data. *Current Protocols in Microbiology*, *56*(1), 1–14. https://doi.org/10.1002/cpmc.96

Montero Vargas, M., Umaña Jiménez, J. C., Escudero Leyva, E., & Chaverri Echandi, P. (2020b). Phylogenetic analysis of ITS data from Endophytic fungi using Massive Parallel Bayesian Tree Inference with Exabayes. *Revista Tecnología En Marcha, Vol.33(5), Pp.74-79*, 74–79.

Morath, S. U., Hung, R., & Bennett, J. W. (2012). Fungal volatile organic compounds: A review with emphasis on their biotechnological potential. *Fungal Biology Reviews*, *26*(2–3), 73–83. https://doi.org/10.1016/j.fbr.2012.07.001

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., & Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Pinto, L. F. G., Gardner, T., McDermott, C. L., & Ayub, K. O. L. (2014). Group certification supports an increase in the diversity of sustainable agriculture network-rainforest alliance certified coffee producers in Brazil. *Ecological Economics*, *107*, 59–64. https://doi.org/10.1016/j.ecolecon.2014.08.006

Pujade-Renaud, V., Déon, M., Gazis, R., Ribeiro, S., Dessailly, F., Granet, F., & Chaverri, P. (2019a). Endophytes from Wild Rubber Trees as Antagonists of the Pathogen Corynespora cassiicola. *Phytopathology*, *109*(11), 1888–1899. https://doi.org/10.1094/PHYTO-03-19-0093-R

Pujade-Renaud, V., Déon, M., Gazis, R., Ribeiro, S., Dessailly, F., Granet, F., & Chaverri, P. (2019b). Endophytes from Wild Rubber Trees as Antagonists of the Pathogen Corynespora cassiicola. *Phytopathology®*, *109*(11), 1888–1899. https://doi.org/10.1094/PHYTO-03-19-0093-R

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.r-project.org/

Rana, K. L., Kour, D., Kaur, T., Devi, R., Yadav, A. N., Yadav, N., Dhaliwal, H. S., & Saxena, A. K. (2020). Endophytic microbes: biodiversity, plant growth-promoting mechanisms and potential applications for agricultural sustainability. In *Antonie van*

Leeuwenhoek, International Journal of General and Molecular Microbiology (Vol. 113, Issue 8). Springer International Publishing. https://doi.org/10.1007/s10482-020-01429-y

Rigot, J., & Matsumura, F. (2002). Assessment of the rhizosphere competency and pentachlorophenol-metabolizing activity of a pesticide-degrading strain of Trichoderma harziamum introduced into the root zone of corn seedlings. *J. Environ. Sci. Health, B.*, *37*(3), 201–210.

Rodriguez, R. J., White Jr, J. F., Arnold, a E., & Redman, R. S. (2009). Fungal endophytes: diversity and functional roles. *New Phytologist*, *182*(2), 314–330. https://doi.org/10.1111/j.1469-8137.2009.02773.x

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., André Levesque, C., Chen, W., & Consortium, F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Procedures of the National Academy of Sciences*, *109*(16), 6241–6246. https://doi.org/10.1073/pnas.1117018109

Smirnov, A., Jia, W., Walker, D. I., Jones, D. P., & Du, X. (2018). ADAP-GC 3.2: Graphical Software Tool for Efficient Spectral Deconvolution of Gas Chromatography-High-Resolution Mass Spectrometry Metabolomics Data. *Journal of Proteome Research*, *17*(1), 470–478. https://doi.org/10.1021/acs.jproteome.7b00633

Sternhagen, E. C., Black, K. L., Hartmann, E. D. L., Shivega, W. G., Johnson, P. G., McGlynn, R. D., Schmaltz, L. C., Asheim Keller, R. J., Vink, S. N., & Aldrich-Wolfe, L. (2020). Contrasting patterns of functional diversity in coffee root fungal communities associated with organic and conventionally managed fields. *Applied and Environmental Microbiology*, *86*(11). https://doi.org/10.1128/AEM.00052-20

Sumida, C. H., Daniel, J. F. S., Araujod, A. P. C. S., Peitl, D. C., Abreu, L. M., Dekker, R. F. H., & Canteri, M. G. (2018). Trichoderma asperelloides antagonism to nine Sclerotinia sclerotiorum strains and biological control of white mold disease in soybean plants. *Biocontrol Science and Technology*, *28*(2), 142–156. https://doi.org/10.1080/09583157.2018.1430743

Sun, J., Zhang, T., Li, Y., Wang, X., & Chen, J. (2019). Functional characterization of the ABC transporter TaPdr2 in the tolerance of biocontrol the fungus Trichoderma

atroviride T23 to dichlorvos stress. *Biological Control*, *129*(October 2018), 102–108. https://doi.org/10.1016/j.biocontrol.2018.10.004

Talhinhas, P., Batista, D., Diniz, I., Vieira, A., Silva, D. N., Loureiro, A., Tavares, S., Pereira, A. P., Azinheira, H. G., Guerra-Guimaraes, L., Várzea, V., & do Céu, S. M. (2017). The coffee leaf rust pathogen Hemileia vastatrix: one and a half centuries around the tropics. *Molecular Plant Pathology*, 1–13.

Theodoridis, G., Gika, H., Raftery, D., Goodacre, R., Plumb, R. S., & Wilson, I. D. (2023). Ensuring Fact-Based Metabolite Identification in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Analytical Chemistry*, *95*(8), 3909–3916. https://doi.org/10.1021/acs.analchem.2c05192

Thevissen, K., Ferket, K. K. A., François, I. E. J. A., & Cammue, B. P. A. (2003). Interactions of antifungal plant defensins with fungal membrane components. *Peptides*, *24*(11), 1705–1712. https://doi.org/10.1016/j.peptides.2003.09.014

Tilocca, B., Cao, A., & Migheli, Q. (2020). Scent of a Killer: Microbial Volatilome and Its Role in the Biological Control of Plant Pathogens. *Frontiers in Microbiology*, *11*(February). https://doi.org/10.3389/fmicb.2020.00041

Wang, T., Mohr, K. I., Stadler, M., & Dickschat, J. S. (2017). Volatiles from the tropical ascomycete Daldinia clavata (Hypoxylaceae, Xylariales). *Beilstein Journal of Organic Chemistry*, *14*, 135–147. https://doi.org/10.3762/bjoc.14.9

Zhang, X., Harvey, P. R., Stummer, B. E., Warren, R. A., Zhang, G., Guo, K., Li, J., & Yang, H. (2015). Antibiosis functions during interactions of Trichoderma afroharzianum and Trichoderma gamsii with plant pathogenic Rhizoctonia and Pythium. *Functional and Integrative Genomics*, *15*(5), 599–610. https://doi.org/10.1007/s10142-015-0456-x

Zuffa, S., Schmid, R., Bauermeister, A., Gomes, P. W. P., Caraballo-rodriguez, A. M., Buzun, E., Terrazas, M. C., Hsu, C., Oles, R., Ayala, A. V., Zhao, J., & Chu, H. (2023). *A Taxonomically-informed Mass Spectrometry Search Tool for Microbial Metabolomics Data Key words (5 words) - metabolomics , microbiome , search tool , bacteria , fungi Authors list * These authors equally contributed to this work ^ Correspondenc*. https://doi.org/https://doi.org/10.1101/2023.07.20.549584