GBE

High-Quality Draft Genome for *Armillaria mexicana*, a Recently Described Species in North America

Jane E. Stewart [1,*], Mee-Sook Kim [2,*], Jorge R. Ibarra Caballero [1], Patrick I. Bennett [3], John W. Hanna [3], Rubén D. Elías-Román [4], Ned B. Klopfenstein [3,*]

[1]Department of Agricultural Biology, Colorado State University, Ft. Collins, CO 80523, USA

[2]United States Department of Agriculture, Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331, USA

[3]United States Department of Agriculture, Forest Service, Rocky Mountain Research Station, Moscow, ID 83843, USA

[4]Departamento de Agronomía, División de Ciencias de la Vida, Campus Irapuato-Salamanca, Universidad de Guanajuato, Irapuato, Guanajuato, México

*Corresponding authors: E-mails: jane.stewart@colostate.edu; meesook.kim@usda.gov; ned.klopfenstein@usda.gov.

Accepted: August 29, 2025

Abstract

The genus *Armillaria* is a broad group of basidiomycete (*Agaricales*, *Physalacriaceae*) plant pathogens that can have detrimental effects on woody hosts in forested, urban, and horticultural landscapes. Several species are known as aggressive root pathogens on both conifers and deciduous woody plants. *Armillaria* species are considered white rot fungi due to their capacity to degrade both lignin and cellulose in woody tissues. *Armillaria mexicana* was recently described as a new species found in peach (*Prunus persica*) orchards of Coatepec Harinas, State of Mexico. However, a subsequent study identified *A. mexicana* on avocado (*Persea americana*) and pine (*Pinus* sp.), suggesting that *A. mexicana* may be able to infect diverse hosts that are planted in disturbed areas following deforestation. We assembled a reference genome for *A. mexicana* consisting of 38 contigs constructed using PacBio and Illumina sequencing reads. Genome annotation and comparison with *A. mellea*, a phylogenetic sister taxon, revealed notable differences, including a larger number of interspersed repeats in *A. mexicana* and the absence of small RNAs, which were detected in *A. mellea*. While *A. mexicana* (49 Mb) has a roughly 30% smaller genome than *A. mellea* (70 Mb), it has a similar number of genes encoding pectinases and nonribosomal peptide synthetase (NRPS) and NPRS-like secondary metabolites, which may influence the pathogenicity of *A. mexicana*. This reference genome of *A. mexicana* allows future genomic comparisons that can help characterize the evolutionary history and enhance our understanding of the molecular mechanisms involved in pathogenicity and wood decomposition of *Armillaria* species.

Key words: *Armillaria mexicana*, fungal root pathogen, pectinases.

Significance

Armillaria root disease of trees and shrubs is caused by fungi in the genus *Armillaria*, which are found globally in diverse forest ecosystems, urban, and horticultural settings. Recent genomic studies of *Armillaria* species have identified genes associated with host specialization, lifestyle, and other enzymatic activities that vary across ecosystems and continents. *Armillaria mexicana*, recently described as a destructive pathogen of fruit trees in Mexico, has been found on diverse hosts including peach, olive, avocado, oak, and pine. Phylogenetic analyses place *A. mexicana* as ancestral to other Northern Hemisphere *Armillaria* species, with *A. mellea*, a well-known, broad-host pathogen of mostly deciduous tree species, as its closest known relative. This high-quality genome presented here represents the first for *A. mexicana* and enables robust genomic comparisons with other species. This resource will provide opportunities to examine the evolution of pathogenicity, mechanisms of host interaction, and genes involved in wood decay across the genus.

Introduction

Armillaria root disease of trees and woody shrubs is caused by *Armillaria* spp., a broad group of basidiomycete fungi that are found globally across temperate and tropical forest ecosystems, where they play diverse ecological roles that range from beneficial ectomycorrhizal symbionts to saprophytic decomposers or virulent pathogens (Kim et al. 2022). More than 40 *Armillaria* species are found globally, with 10 species known in North America, and pathogenic species are known to infect coniferous and hardwood tree species (Kim et al. 2022). Large-scale damage can occur in affected areas because a single *Armillaria* genet (individual) can spread and grow up to 965 hectares (Ferguson et al. 2003). Two species that cause significant damage in tree orchards include *A. mellea*, which is found broadly across host species and geographic areas in the Northern Hemisphere and has been introduced to areas of Africa (Coetzee et al. 2001), and *A. mexicana*, which is a recently described species and was identified causing significant damage in peach (*Prunus persica*), plum (*Prunus salicina*), and avocado (*Persea americana*) orchards in Mexico but has also been found in natural forests on oaks (*Quercus* spp.) and pines (*Pinus* spp.) (Elías-Román et al. 2018, 2019; Alvarado-Rosales et al. 2023).

Armillaria belongs to a group known as white rot fungi that can break down both lignin and cellulose to degrade their woody hosts, and *Armillaria* species have other features not yet well understood in fungi, including bioluminescence and longevity (Sahu et al. 2023). Several genomic studies of *Armillaria* species have identified genes encoding many plant cell wall-degrading enzymes (PCWDEs), including several pectinases, which appear to be particularly abundant in *Armillaria* species compared to other white rot fungi. These comparative studies have also revealed a rich repertoire of PCWDEs regardless of ecological lifestyle. In combination, genome evolution studies of *Armillaria* species suggest that they have developed novel genetic innovations to become unique and aggressive pathogens (Sipos et al. 2017; Ibarra Caballero et al. 2023; Sahu et al. 2023).

Recent phylogenetic analyses place *A. mexicana* as ancestral to other Northern Hemisphere *Armillaria* spp., while exannulate species now belong to *Desarmillaria* forming a group basal to *Armillaria* spp.; furthermore, *A. mexicana* is the only species known to exist in its phylogenetic lineage, but it is sister to the clade containing *A. mellea* (Klopfenstein et al. 2017; Coetzee et al. 2018). However, few morphological similarities exist between *A. mexicana* and *A. mellea*, such as the lack of clamp connections, a feature common among other *Armillaria* species (Elías-Román et al. 2018). As *A. mexicana* and *A. mellea* are sister taxa, yet are morphologically distinct, our goal was to sequence *A. mexicana* and compare its genome to its closest known relative, *A. mellea*. This is the first genome sequence of *A. mexicana*.

Results

Sequencing yielded 69,319,592 short reads (from Illumina) and 4,239,547 long reads (from PacBio) for *A. mexicana* (NCBI accession number: IZTA4595; BioSample #SAMN46478975). The PacBio coverage was 1,041×. There was an overall mapping rate of 95.6% of short reads to the PacBio reads-assembled genome. The final genome assembly consisted of 38 contigs containing 49,049,582 bp, with an average contig length of 1,290,779 bp and N50 of 3,015,657 bp. The GC content was 47.98%. A 95.6% completeness and 65 (3.7%) complete and duplicate BUSCOs were obtained using BUSCO (Manni et al. 2021) against the basidiomycota_odb10 lineage dataset, and 94.2% completeness and 114 (2.9%) completed and duplicated BUSCOs were identified against the agaricales_odb10 lineage dataset, highlighting the quality of the genome assembly. The completeness of the predicted transcriptome was estimated against the Basidiomycota and Agaricales datasets, and scores of 95.9% and 93.8% completeness were obtained, respectively. A search for telomeres identified one contig with telomeres on both ends, indicating that at least one putative chromosome was assembled. Telomeres were identified on a single end of eight other contigs, suggesting chromosome fragmentation in the assembly. A synteny analysis with *A. ostoyae* (C18/9) highlighted that 2 of the 11 chromosomes of *A. ostoyae* were present in *A. mexicana*, though it is not clear how many total chromosomes *A. mexicana* has (Fig. S1). A total of 11,939 genes and 12,200 transcripts made up of 78,930 exons and corresponding proteins were predicted by Braker3 (Table 1) (Gabriel et al. 2024). A total of 12,285,276 bp (25.05%) of the genome sequence were detected and masked using RepeatMasker (Smit et al. 2015), and these masked regions were mainly long terminal repeat (LTR) elements (18.14%). Notably, no small RNAs were detected in the *A. mexicana* genome (Table S1). Close to 550 putative protein coding genes were annotated as carbohydrate-active enzymes (CAZymes), including auxiliary activities (AA), glycoside

Table 1 *Armillaria mexicana* genome[a] assembly metrics

Description	Value
Assembled genome size	49,049,582
Number of contigs	38
Contig N50, bp	3,015,657
L50	7
GC content, %	47.98
BUSCO completeness, % (Basidiomycota)	95.6
Completed and duplicated genome BUSCOs	3.7%
Number of predicted genes	11,939
Completed and duplicated transcriptome BUSCOS (Basidiomycota)	95.9 & 5.9%

[a]NCBI accession number: IZTA4595; BioSample #SAMN46478975.

hydrolases (GH), polysaccharide lyases (PL), glycosyltransferases (GT), carbohydrate esterases (CE), and non-catalytic carbohydrate-binding modules (CBM) (Table S1), and these were compared directly to those in the *A. mellea* genome (Table S1). In addition, 44 genomic regions were annotated as part of secondary metabolite clusters, with the highest number annotated as being associated with terpenes in both *A. mexicana* (ca. 16) and *A. mellea* (ca. 23) (Table S1). A synteny plot of the *A. mexicana* and *A. mellea* genomes highlighted many contigs in *A. mellea* that did not map to *A. mexicana*, and nine contigs were found in *A. mexicana* that did not align to *A. mellea*, suggesting that more unique genomic regions may be present in *A. mellea*, which may confer novel biological features (Fig. 1).

Discussion

In a previous study, *A. mexicana* was found to have the largest internal transcribed spacer 1 and 2 (1,299 and 582 bp, respectively) of any fungus known to date, and this was confirmed in this reference genome (Elías-Román et al. 2018). Furthermore, *A. mexicana* was found to be morphologically unique compared to other North American *Armillaria* species (Elías-Román et al. 2018). In phylogenetic analyses, it was placed basal to other *Armillaria* species, suggesting that it may be ancestral to other species, with *A. mellea* as its closest relative (Klopfenstein et al. 2017). The genome size of *A. mexicana* (49 Mb) was smaller than that of other *Armillaria* spp. including *A. mellea* (70 Mb), *A. altimontana* (73 Mb), and *A. solidipes* (55 Mb). In a recent study comparing carbohydrate enzymes of *Armillaria* species across a broad suite of other basidiomycetes, some cellulase gene families were expanded in the *Armillaria* genus, including AA16, AA3_1, GH1, and GH45. We observed that *A. mexicana* has a similar number of these genes but with reduced numbers in the GH45 cellulase family. Pectinase gene families are also expanded within the *Armillaria* genus, including carbohydrate esterase CE8 and polysaccharide lyase PL1. We found that *A. mexicana* has higher numbers of CE8 (8), and PL1 (6), compared to other *Armillaria* species, except for *A. mellea*, which has similar numbers at 10 and 6, respectively.

More direct comparisons between *A. mexicana* and *A. mellea* highlighted similarities and differences across these two sister taxa. For example, *A. mellea* has an expanded genome compared to *A. mexicana*. This is apparent when comparing the synteny of the two genomes; multiple contigs in *A. mellea* do not map to the genome of *A. mexicana*. To further investigate, we examined the 409 *A. mellea* contigs that did map to *A. mexicana* genome. We found that putative genes were identified on only 12 of these contigs, while large regions of the contigs were identified as repetitive sequences. These repetitive elements may partially account for the lack of synteny observed between two genomes. Also, when comparing the number of specific carbohydrate-degrading enzymes and secondary metabolites, *A. mellea* has more in almost all categories. However, *A. mellea* and *A. mexicana* have similar numbers of genes encoding pectinases, even though the genome of *A. mellea* is 1.42 times larger. These pectinase genes are likely associated with the pathogenicity of *A. mexicana* and may influence its host range. In a recent study, *A. mexicana* was found to be more virulent to peach and plum rootstocks than *A. mellea* (Elías-Román et al. 2019), indicating that gene content alone may not fully explain pathogenic potential. In contrast, *A. mellea* has a larger repertoire of diverse carbohydrate-degrading enzymes and also a higher number of secondary metabolite clusters. This perhaps allows *A. mellea* to cause disease on an expanded range of hosts, as has been documented in *Collectotrichum* species (Baroncelli et al. 2016). The new genome of *A. mexicana* enables robust genomic comparison across the genus and provides a crucial foundation to enhance our understanding of the pathogenicity and wood-degrading abilities of *Armillaria* spp. and their host interactions.

Methods

In this study, we performed whole-genome Illumina and PacBio sequencing of the type specimen (extype MEX87 derived from the holotype IZTA4595). This isolate was originally found on a dead peach tree in Coatepec Harinas, State of Mexico. Mycelia of *A. mexicana* isolate MEX87 were grown for 2 weeks in a liquid broth medium consisting of 1% malt extract, 0.5% yeast extract, and 1% glucose. The liquid cultures were filtered using sterilized MiraCloth (Millipore Sigma, St. Louis, MO, United States), rinsed thoroughly with sterile, deionized water, and dried using sterilized paper towels. The filtered mycelium was then briefly frozen at −80 °C before lyophilizing at −50 °C overnight in a Labconco freeze dryer (Labconco Corporation, Kansas City, MO, United States). The lyophilized tissue was then weighed, and 15 to 20 mg was added to each of 12 sterilized 1.5 ml microcentrifuge tubes. The tissue was ground gently in lysis buffer containing polyvinylpyrrolidone (PVP) (2% w/v) using a sterilized disposable plastic tissue grinder (pellet pestle). DNA was extracted using the Omega Bio-tek Mag-Bind plant DNA kit (Omega Bio-tek, Inc., Norcross, GA, United States) following the manufacturer's protocol for use of the kit with the KingFisher Duo Prime purification system (Thermo Fisher Scientific, Waltham, MA, United States). The DNA was eluted in 75 µl of elution buffer. DNA quantity and quality were assessed using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), Qubit 2.0 fluorometer with the Qubit dsDNA BR assay kit (Thermo Fisher Scientific), and the Agilent 4200 TapeStation system with a genomic DNA ScreenTape (Agilent Technologies, Santa Clara, CA,

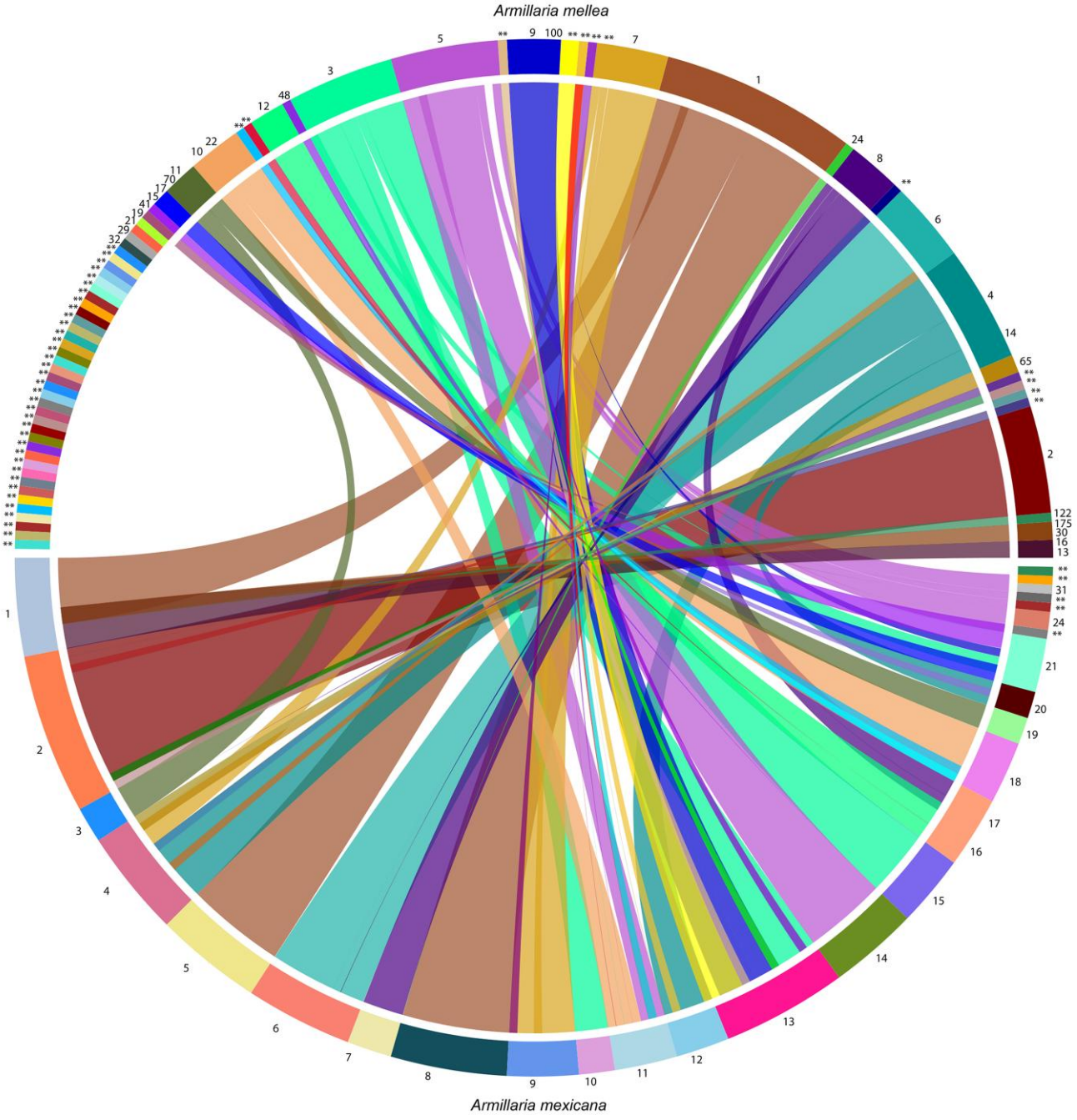

Fig. 1. Blocks of synteny comparing the 38 contigs of *Armillaria mexicana* (MEX87) against 474 contigs of *Armillaria mellea* (ELDO17). The ** symbols represent synteny blocks that contain multiple contigs.

United States), following the manufacturer's protocols. The highest quality samples were pooled prior to submitting to Psomagen (Rockville, MD) for sequencing.

Genomic sequencing was completed using both the NovaSeq 6000 (Illumina, San Diego, CA, United States) with 150 bp paired-end reads and the PacBio Revio system (Pacific Biosciences, Menlo Park, CA, United States). The data were submitted to GenBank, and during submission, a contamination check was performed using the Foreign Contamination Screen (FCS), and no contamination was found. Genome assembly was performed using hifiasm 0.19.6-r597 (Cheng et al. 2021). Haplotigs and heterozygous overlaps were removed using purge_dups 1.2.6 (Guan et al. 2020). Illumina reads were then mapped to the genome assembly using bowtie 2-2.5.2 (Langmead and Salzberg 2012) with an overall alignment rate of

95.64%. Obtained bam files were used to perform a final polishing of the assembly using Pilon 1.24 (Walker et al. 2014). To access the completeness of the genome assembly and the predicted transcriptome, BUSCO v5.5.0 was used (Simão et al. 2015). The telomere regions were search using "Telomere Search" (McGowan 2021) with default settings, and to confirm these results, we downloaded the genome of *A. ostoyae* C18/9 which has been assembled into 11 chromosomes (Heinzelmann et al. 2020) and performed a synteny analysis using SyMap 5 (data not shown; Soderlund et al. 2011).

Interspersed repeats and low complexity DNA sequences present in the genome assembly were masked using RepeatModeler 2.0.5 and RepeatMasker 4.1.7-p1. The masked genome was structurally annotated using Braker 3.0.8 (Gabriel et al. 2024). RNA-Seq data from a closed related species (*A. mellea*, NCBI SRR2545913) were used for training, and the "fungus" parameter was included to run the algorithm with the fungal branch point model. Functional annotations were made searching for extracellular proteins using DeepLoc 2.0 (Thumuluri et al. 2022), CAZymes using the dbcan3 server (Zheng et al. 2023), and secondary metabolites using antiSMASH (Blin et al. 2023).

The genome of *A. mellea*, isolate ELDO17, was downloaded from the genome portal of JGI (Armmel1) and run through the same databases (Deeploc 2.0, dbcan3, and antiSMASH) as *A. mexicana*. To compare the synteny of *A. mexicana* and *A. mellea*, we performed a synteny analysis using SyMap 5 (Fig. 1).

Supplementary Material

Supplementary material is available at Genome Biology and Evolution online.

Acknowledgments

This project has been supported by USDA Forest Service (FS), State, Private, and Tribal Forestry, Forest Health Protection—Special Technology and Development Program (STDP R5-2023-01 and STDP R2-2023-01), and FS Research & Development Infrastructure Investment and Jobs Act Research Project- ERR09. This research was partially supported by the employees of the USDA Forest Service, and the findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA determination or policy.

Data Availability

The data underlying this article are available in the GenBank BioSample database at https://www.ncbi.nlm.nih.gov/biosample/and can be accessed with SAMN46478975.

Literature Cited

Alvarado-Rosales D, et al. New hosts for *Desarmillaria caespitosa* and *Armillaria mexicana* in Veracruz and Michoacán, Mexico. Revista Chapingo Serie Ciencias Forestales y del Ambiente. 2023:29:51–59. http://doi.org/10.5154/r.rchscfa.2022.05.034.

Baroncelli R, et al. Gene family expansions and contractions are associated with host range in plant pathogens of the genus Colletotrichum. BMC Genom. 2016:17:555. https://doi.org/10.1186/s12864-016-2917-6.

Blin K, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures, and visualization. Nucleic Acids Res. 2023:51:W46–W50. https://doi.org/10.1093/nar/gkad344.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Coetzee MPA, et al. The root rot fungus *Armillaria mellea* introduced into South Africa by early Dutch settlers. Mol Ecol. 2001:10:387–396. https://doi.org/10.1046/j.1365-294X.2001.01187.x.

Coetzee MPA, Wingfield BD, Wingfield MJ. *Armillaria* root-rot pathogens: species boundaries and global distribution. Pathogens. 2018:7:83. https://doi.org/10.3390/pathogens7040083.

Elías-Román RD, et al. *Armillaria mexicana*, a newly described species from Mexico. Mycologia. 2018:110:347–360. https://doi.org/10.1080/00275514.2017.1419031.

Elías-Román RD, et al. 'Mondragon': a clonal plum rootstock to enhance management of *Armillaria* root disease in peach orchards of Mexico. Crop Protection. 2019:121:89–95. https://doi.org/10.1016/j.cropro.2019.03.011.

Ferguson BA, Dreisbach TA, Parks CG, Filip GM, Schmitt CL. Coarse-scale population structure of pathogenic *Armillaria* species in a mixed-conifer forest in the Blue Mountains of Northeast Oregon. Can J For Res. 2003:33:612–623. https://doi.org/10.1139/x03-065.

Gabriel L, et al. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 2024:34:769–777. https://doi.org/10.1101/gr.278090.123.

Guan D, et al. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020:36:2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

Heinzelmann R, Rigling D, Sipos G, Münsterkötter M, Croll D. Chromosomal assembly and analyses of genome-wide recombination rates in the forest pathogenic fungus *Armillaria ostoyae*. Heredity (Edinb). 2020:124:699–713. https://doi.org/10.1038/s41437-020-0306-z.

Ibarra Caballero JR, et al. Genomic comparisons of two *Armillaria* species with different ecological behaviors and their associated soil microbial communities. Microb Ecol. 2023:85:708–729. https://doi.org/10.1007/s00248-022-01989-8.

Kim M-S, et al. Armillaria root diseases of diverse trees in wide-spread global regions. Forest Microbiology. 2022:2:361–378. https://doi.org/10.1016/B978-0-323-85042-1.00004-5.

Klopfenstein NB, et al. Insights into the phylogeny of Northern Hemisphere *Armillaria*: neighbor-net and Bayesian analyses of translation elongation factor 1-α gene sequences. Mycologia. 2017:109:75–91. https://doi.org/10.1080/00275514.2017.1286572.

Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012:9:357–359. https://doi.org/10.1038/nmeth.1923.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EV. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021:38:4647–4654. https://doi.org/10.1093/molbev/msab199.

McGowan. 2021. Telomere search. *Github*. https://github.com/jamiemcg/TelomereSearch

Sahu N, et al. Vertical and horizontal gene transfer shaped plant colonization and biomass degradation in the fungal genus *Armillaria*. Nat Microbiol. 2023:8:1668–1681. https://doi.org/10.1038/s41564-023-01448-1.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015: 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Sipos G, et al. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. Nat Ecol Evol. 2017:1:1931–1941. https://doi.org/10.1038/s41559-017-0347-8.

Smit AFA, Hubley R, Green P. *RepeatMasker Open-4.0*. 2013-2015. http://www.repeatmasker.org

Soderlund C, Bomhoff M, Nelson W. SyMAP: a turnkey synteny system with application to multiple large duplicated plant sequenced genomes. Nucleic Acids Res. 2011:39:e68. https://doi.org/10.1093/nar/gkr123.

Thumuluri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. Nucleic Acids Res. 2022:50: W228–W234. https://doi.org/10.1093/nar/gkac278.

Walker BJ, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014:9:e112963. https://doi.org/10.1371/journal.pone.0112963.

Zheng J, et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023:51:W115–W121. https://doi.org/10.1093/nar/gkad328.

Associate editor: Qi Zhou