AMERICAN SOCIETY FOR MICROBIOLOGY
Microbiology Resource Announcements
GENOME SEQUENCES

Genome Sequence and Assembly of 18 *Fusarium* Isolates from Florida Citrus under High Huanglongbing Disease Pressure and California Citrus under Low Huanglongbing Disease Pressure

Tania Kurbessoian,[a,c] Gretchen Heimlich-Villalta,[a,c] Nichole Ginnan,[a,c] Flavia Campos Vieira,[a,c] Philippe E. Rolshausen,[b] M. Caroline Roper,[a,c] Jason E. Stajich[a,c]

[a]Department of Microbiology and Plant Pathology, University of California, Riverside, California, USA
[b]Department of Botany and Plant Sciences, University of California, Riverside, California, USA
[c]Institute for Integrative Genome Biology, University of California, Riverside, California, USA

ABSTRACT The genomes of eighteen *Fusarium* isolates cultured from diseased and healthy citrus trees were sequenced, assembled, and annotated. Isolate species identification was confirmed using single marker (TEF1-alpha) phylogenetic assessment. Studies of the traits and genotypes of plant-associated isolates are important to understanding the fungal contribution to phytobiomes of citrus.

Fungal isolates were cultured from leaf, stem, and root tissue of California (Riverside County) and Florida (Marion, Lake, and Martintown Counties) citrus trees. Samples were taken from *Citrus* sp. scions grafted onto *Citrus aurantium* and × *Citroncirus* species rootstocks. California trees had low Huanglongbing (HLB) disease pressure; Florida trees had high HLB pressure. Fungal taxonomy was initially assigned by internal transcribed spacer (ITS) sequencing (1) and confirmed using single marker phylogenetic analysis.

Strains were grown on potato dextrose agar (PDA) for 1 week, collected by scraping fungal mycelial tissue from the medium, and frozen in liquid nitrogen. High-molecular-weight DNA was extracted from fungal tissue based on reference 2. Genomic libraries for the 18 isolates were constructed with the Illumina DNA Prep kit with 10-bp IDT UDI indices and sequenced on an Illumina NextSeq 2000 sequencer in 2- by 151-bp paired-end format at the MiGS sequencing center (Pittsburgh, PA). Reads were trimmed and demultiplexed by the bcl-convert workflow to produce Fastq files. Two strains (*Fusarium oxysporum* CF00159 and *Fusarium falciforme* CF00175) were additionally sequenced with Oxford Nanopore Technologies (ONT) (3) at MiGS. An average of 5.1 million Illumina reads and 0.5 million ONT reads were produced (Table 1).

Genome assembly of Illumina reads was performed using AAFTF (4–8) for performing filtering and trimming steps for data quality and SPAdes (3.15.4) (9) for assembly. Default parameters for the underlying tools were applied throughout. Assembly of the two ONT-sequenced isolates was performed using Canu (v.2.2) (10) and Flye (v.2.9-b1774) (11), followed by assembly polishing with Medaka (v.1.6) (12). Both Nanopore assemblies were processed with five rounds of polishing by Pilon (v1.24) (13) Canu (CF159) and Flye (CF175) assemblies with Illumina sequencing reads via the AAFTF 'pilon' step. Contigs were reordered and renamed from largest to smallest with the 'sort' command. Assembly summary statistics were calculated with the 'assess' tool in AAFTF and genome completeness by BUSCO (v5.2.2) (14) with the sordariomycetes_odb10 database of 3,817 marker genes. Genome annotation was performed with Funannotate (v.1.8.10) (15–32) using default parameters for the underlying tools applied throughout. Genome sequencing, assembly, and protein coding gene annotation statistics of the 18 genomes are summarized in Table 1.

BLASTN was used to capture translation elongation factor 1 (TEF1) (MG183712) sequences of each genome assembly for species identification (33–35). A multiple sequence alignment

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Jason E. Stajich, jason.stajich@ucr.edu.

The authors declare no conflict of interest.

Received 10 February 2023
Accepted 30 March 2023
Published 12 April 2023

TABLE 1 Strain and species designation, isolation source, sequencing read, assembly, and annotation statistics, and accession numbers[d]

Species	Strain ID[e]	Host/tissue	Location	GenBank accession no.	SRA accession no.	No. of read pairs	Coverage	No. of contigs	Genome size (Mbp)	Contig L_{50}	Contig N_{50} (kbp)	G+C content (%)	Genome completion (BUSCO %)	No. of genes	Telomeres found[f]
F. solani	CF00177	*Citrus jambhiri* Lush. (Schaub rough lemon),	Riverside County, CA	JAOQAT000000000	SRR21444562	5,208,997	28.56	1,595	54.7	29	560	50.61	99.8	17,742	8F, 7R
F. falciforme	CF00178	*Citrus jambhiri* Lush. (Schaub rough lemon),	Riverside County, CA	JAOQAU000000000	SRR21444563	4,824,837	24.65	1,010	56.4	32	504	49.12	99.8	15,807	3F, 4R
F. falciforme	CF00179	*Citrus jambhiri* Lush. (Schaub rough lemon),	Riverside County, CA	JAOQAV000000000	SRR21444564	6,299,623	33.75	866	54.3	23	663	49.35	99.9	14,964	1F, 1R
F. falciforme	CF00180	*Citrus aurantifolia* Christm. (Mexican lime)/x *Cintroncirus* spp. (Swingle),	Riverside County, CA	JAOQAW000000000	SRR21444565	6,641,869	34.75	1,035	55.3	27	595	49.49	99.9	14,885	4F, 3R
F. falciforme	CF00175	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), root	Marion County, FL	JAOQBJ000000000	SRR21444571	5,036,413	39.66 (24.4)[a]	296	65.5	9	2,609	47.69	98.2	16,794	1F, 3R
F. falciforme (ONT)	CF00175	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), root	Marion County, FL		SRR21444561	674,410[b]	15.26				6.831[c]				
F. equiseti	CF00095	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), stem	Lake County, FL	JAOQBH000000000	SRR21444559	4,858,320	37.09	89	37.1	8	1,904	47.76	99.9	12,206	4F, 8R
F. irregulare	CF00137	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), stem	Marion, County, FL	JAOQBA000000000	SRR21444569	6,189,289	46.7	41	37.9	8	1,635	48.05	99.8	12,502	9F, 6R
F. irregulare	CF00143	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), stem	Marion County, FL	JAPDHF000000000	SRR21444553	4,877,712	36.52	42	38.1	9	1,635	48.03	99.8	12,865	10F, 5R
F. oxysporum	CF00115	*Citrus sinensis* L. Osbeck (Hamlin)/× *Citroncirus* spp. (Swingle), leaf	Lake County, FL	JAOQAX000000000	SRR21444566	4,443,693	27.44	208	45.0	13	1,348	47.41	94.7	15,474	0F, 2R
F. oxysporum	CF00132	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), root	Marion County, FL	JAOQAY000000000	SRR21444567	5,666,638	33.92	383	48.6	11	1,590	47.63	99.9	16,669	3F, 3R
F. oxysporum	CF00141	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), leaf	Marion County, FL	JAOQBB000000000	SRR21444552	4,806,668	29.53	239	45.0	14	988	47.44	94.6	15,308	12F, 8R
F. oxysporum	CF00144	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), leaf	Marion County, FL	JAOQBC000000000	SRR21444554	3,961,972	25.02	107	42.2	8	1,593	47.52	92.2	14,579	13F, 12R
F. oxysporum	CF00145	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), leaf	Marion County, FL	JAOQBD000000000	SRR21444555	4,867,827	30.81	60	42.9	7	2,090	47.5	92.3	14,883	13F, 9R
F. oxysporum	CF00159	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), root	Marion County, FL	JAOQBI000000000	SRR21444570	4,130,577	43.7 (24.7)[a]	136	50.6	11	1,462	47.38	98.9	17,802	13F, 7R
F. oxysporum (ONT)	CF00159	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), root	Marion County, FL		SRR21444560	391,895[b]	19.03				5.912[c]				
F. oxysporum	CF00160	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), leaf	Marion County, FL	JAOQBE000000000	SRR21444556	4,257,800	26.27	259	44.9	11	1,126	47.41	94.1	15,276	12F, 8R
F. oxysporum	CF00161	*Citrus sinensis* L. Osbeck (Parson Brown)/ *Citrus aurantium* (sour orange), leaf	Marion County, FL	JAOQBF000000000	SRR21444557	6,329,714	40.21	59	41.9	9	1,492	47.56	92.3	14,538	10F, 9R
F. oxysporum	CF00165	*Citrus sinensis* L. Osbeck (Hamlin)/× *Citroncirus* spp. (Swingle), leaf	Lake County, FL	JAOQBG000000000	SRR21444558	3,918,330	23.96	245	44.9	11	1,552	47.38	93.6	15,470	0F, 3R
F. torreyae	CF00136	*Citrus sinensis* L. Osbeck (Valencia)/× *Citroncirus* spp. (Swingle), leaf	Martintown County, FL	JAOQAZ000000000	SRR21444568	6,378,752	39.89	92	46.5	13	1,014	47.92	99.8	14,845	17F, 17R

[a] Summary statistics shown are for hybrid genome assembly with Illumina and ONT sequence reads.

[b] ONT reads are single ended.

[c] N_{50} for ONT read lengths.

[d] Host/tissue indicates host material from which strain was isolated. Location indicates United States location, either from California (CA) or from Florida (FL); full location description is available in NCBI BioSample. The GenBank accession number of the deposited genome assembly and the SRA accession number for individual sequencing runs are listed for each isolate. The number of reads was used to help determine the coverage values for almost all the genomes except the two Nanopore genomes mentioned. Oxford Nanopore Technologies (ONT) sequencing data coverage is calculated using average depth of sequencing for only Nanopore reads. Genomes CF159 and CF175 had combined Illumina and ONT coverage calculated and indicated in the table, where Illumina-only coverage is in parentheses next to the total coverage. Genome assembly calculations include number of contigs, genome size, N_{50} (longest in length in 50% of genome), L_{50} (number of contigs that are longest in length in 50% of genome), and G+C content, while genome annotation results include number of genes predicted and annotated. BUSCO completion statistics and comparisons were determined using the sordariomycetes_odb10 database with 3,817 genes. Telomeres were calculated on completed genomes using find_telomere.py script (41).

[e] ID, identifier.

[f] F, forward; R, reverse.

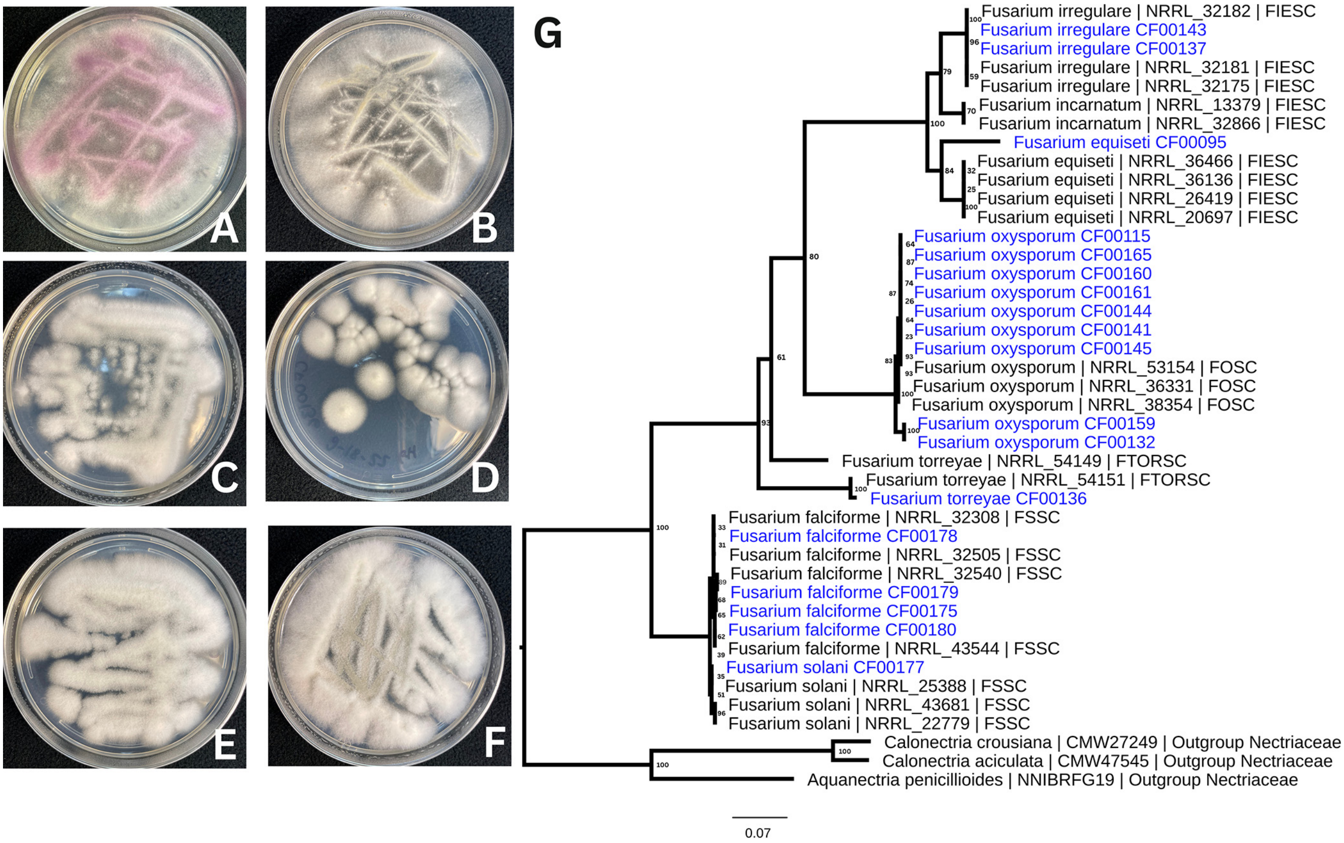

FIG 1 Agar culture of *Fusarium* species isolated from citrus and phylogenetic tree describing 18 strains with respective NCBI isolates. (A) Isolate CF115, *Fusarium oxysporum*. (B) Isolate CF00179, *Fusarium falciforme*. (C) Isolate CF00136, *Fusarium torreyae*. (D) Isolate CF00137, *Fusarium irregulare*. (E) Isolate CF00177, *Fusarium solani*. (F) Isolate CF00095, *Fusarium equiseti*. (G) Species assignments were inferred from the phylogenetic tree constructed from TEF1 sequences for the 18 isolates from this study (blue) and 21 reference sequences of identified *Fusarium* species and rooted with sequences of three outgroup taxa. Identification abbreviations for *Fusarium* species complexes are FIESC (*Fusarium incarnatum-equiseti* complex), FOSC (*Fusarium oxysporum* species complex), FTORSC (*Fusarium torreyae* species complex), and FSSC (*Fusarium solani* species complex).

of identified TEF1 genes and those available in *FUSARIUM*-ID v 3.0 (36) was created with MUSCLE (5.1) (37). The alignment was trimmed with ClipKIT (38), and the phylogenetic relationships of the strains were inferred with IQ-TREE 2 (39). The 18 isolates were placed among six known *Fusarium* species (Fig. 1), and their position was used to assign the taxonomic identification presented in Table 1.

Data availability. This whole-genome project has been deposited at DDBJ/ENA/GenBank under the BioProject accession no. PRJNA855134. The individual SRA read accession numbers and genome accession numbers for each isolate are listed in Table 1. Genome assembly, annotation, and TEF1 phylogenetic assessment pipeline and related code are archived in Zenodo (40).

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Agriculture, National Institute of Food and Agriculture Hatch projects CA-R-PPA-211-5062-H (J.E.S.), CA-R-PPA-5020-H (M.C.R.), and CA-R-BPS-5071-H (P.E.R.) and USDA-NIFA (CA-R-PPA-5220-CG) to M.C.R., P.E.R., and J.E.S. Genome assembly and annotation were performed on the IIGB High-Performance Computing Cluster supported by NSF DBI-1429826 and NIH S10-OD016290 grants.

Illumina library preparation and Oxford Nanopore library preparation were performed at the Microbial Genome Sequencing Center (MiGS), now Sequencing Center (SeqCenter), Pittsburgh, PA.

REFERENCES

1. White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. *In* Innis MA, Gelfand DH, Sninsky JJ, White TJ (ed), PCR protocols. Academic Press, San Diego, CA.

2. Carter-House D, Stajich JE, Unruh S, Kurbessoian T. 2020. Fungal CTAB DNA extraction. Protocolsio https://doi.org/10.17504/protocols.io.bhx8j7rw.
3. Jain M, Olsen HE, Paten B, Akeson M. 2016. Erratum to: The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:256. https://doi.org/10.1186/s13059-016-1122-x.
4. Palmer JM, Stajich JE. 2022. Automatic assembly for the fungi (AAFTF): genome assembly pipeline. https://zenodo.org/record/6326242.
5. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
6. Bushnell B. 2022. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. Joint Genome Institute, Berkeley, CA.
7. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418.
8. Brown CT, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. J Open Source Softw 1:27. https://doi.org/10.21105/joss.00027.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
10. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
11. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
12. Oxford Nanopore Technologies. 2018. Nanoporetech/medaka. Sequence correction provided by ONT Research. https://github.com/nanoporetech/medaka.
13. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
14. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38:4647–4654. https://doi.org/10.1093/molbev/msab199.
15. Palmer JM, Stajich J. 2020. Funannotate v1.8.1: eukaryotic genome annotation. https://zenodo.org/record/4054262.
16. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A 117:9451–9457. https://doi.org/10.1073/pnas.1921046117.
17. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11. https://doi.org/10.1186/s13100-015-0041-9.
18. Smit AFA. 2004. Repeat-Masker Open-3.0. http://www.repeatmasker.org.
19. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59.
20. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439. https://doi.org/10.1093/nar/gkl200.
21. Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform 2:lqaa026. https://doi.org/10.1093/nargab/lqaa026.
22. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20:2878–2879. https://doi.org/10.1093/bioinformatics/bth315.
23. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
24. Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31. https://doi.org/10.1186/1471-2105-6-31.
25. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol 9:R7. https://doi.org/10.1186/gb-2008-9-1-r7.
26. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
27. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
28. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.
29. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314. https://doi.org/10.1093/nar/gky1085.
30. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res 46:D624–D632. https://doi.org/10.1093/nar/gkx1134.
31. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res 46:W200–W204. https://doi.org/10.1093/nar/gky448.
32. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701.
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
34. Madden T. 2003. The BLAST sequence analysis tool. The NCBI handbook. National Center for Biotechnology Information, Bethesda, MD.
35. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
36. Torres-Cruz TJ, Whitaker BK, Proctor RH, Broders K, Laraba I, Kim H-S, Brown DW, O'Donnell K, Estrada-Rodríguez TL, Lee Y-H, Cheong K, Wallace EC, McGee CT, Kang S, Geiser DM. 2022. FUSARIUM-ID v.3.0: an updated, downloadable resource for Fusarium species identification. Plant Dis 106:1610–1616. https://doi.org/10.1094/PDIS-09-21-2105-SR.
37. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113.
38. Steenwyk JL, Buida TJ, III, Li Y, Shen X-X, Rokas A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol 18:e3001007. https://doi.org/10.1371/journal.pbio.3001007.
39. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015.
40. Kurbessoian T, Stajich J. 2023. Citrus-HLB-Micro/Genome_Fungal_Culture: MRA release. https://zenodo.org/record/7530760.
41. Hiltunen M, Ament-Velásquez SL, Johannesson H. 2021. The assembled and annotated genome of the fairy-ring fungus Marasmius oreades. Genome Biol Evol 13:evab126. https://doi.org/10.1093/gbe/evab126.