Journal of Plant Pathology (2025) 107:2345–2350
https://doi.org/10.1007/s42161-025-01981-4

GENOME RESOURCE PAPER

Genome sequencing and assembly of *Phytophthora oleae*, isolate VK10A, causative agent of rot of olive drupes

Sebastiano Conti Taguali[1,2] · **Mario Riolo**[1] · **Federico La Spada**[1] · **Santa Olga Cacciola**[1] · **Giuseppe Dionisio**[3]

Received: 28 February 2025 / Accepted: 21 July 2025 / Published online: 11 August 2025

Keywords *Phytophthora oleae* · Genome assembly · Whole-genome sequencing · Oomycete genomics · Pathogenicity effectors · Pathogenicity effectors · RxLR · Crinkler effectors · Emerging olive pathogen

Phytophthora, often referred to as 'plant destroyers,' is a genus of oomycetes comprising numerous species known for causing destructive plant diseases (Davison 1998). The genus is rapidly expanding (Brasier et al. 2022). Despite inconsistencies between molecular phylogeny and morphology-based taxonomy, *Phytophthora* is well-recognized and classified into multiple phylogenetic clades encompassing more than 223 taxa (Abad et al. 2023).

The genus includes many economically important species that affect a wide range of plant hosts, including vegetables, fruit and nut trees, forest trees, and nursery crops. *Phytophthora oleae* is a newly identified species within the *Phytophthora* genus (NCBI: txid2107226), known for causing significant diseases in plants, particularly in olives.

This pathogen has been isolated from the rhizosphere soil and roots of olive trees with root rot, as well as from rotting fruit (drupes). *Phytophthora oleae* infects both wild and cultivated olive trees across Mediterranean regions, particularly in Italy and Spain (Ruano-Rosa et al. 2018; González et al. 2019).

In Spain, *P. oleae* has been described as the causal agent of the decline of wild olive (*Olea europaea* var. *sylvestris*) forests, which are ecologically important and serve as a genetic reservoir for olive breeding programs (González et al. 2019). The pathogen was repeatedly isolated from rootlets of symptomatic wild olives and identified based on morphological traits and ITS and cox1 sequence analyses (González et al. 2019).

Notably, *P. oleae* exhibits optimal growth at a relatively low temperature of 19.9 °C on carrot agar medium, distinguishing it from other *Phytophthora* species affecting olives (González et al. 2019). The pathogenicity of *P. oleae* was confirmed on healthy wild olive seedlings, indicating its potential threat not only to wild olive populations, but also to cultivated olive orchards, particularly under changing climate conditions and shifting rainfall patterns in southern Spain (González et al. 2019).

In southern Italy, *P. oleae* has been associated with rot of mature olive drupes in two local cultivars of olive (Ruano-Rosa et al. 2018; Brasier et al. 2022), indicating its role in determining significant damages to olive production. The identification of *P. oleae* as a pathogen of both fruit and roots underscores its versatility and the potential of this species to jeopardize the olive cultivation in the Mediterranean.

In this study, *P. oleae* isolate VK10A (CBS 153549) has been isolated from a rotten olive drupe collected from a wild olive (*Olea europaea* var. *sylvestris*) tree belonging to the "Complesso Speleologico Villasmundo S. Alfio Nature Reserve (NR)" (Melilli, Siracusa, Italy; DATUM WGS 84, 37°13'17.2"N 15°06'22.3"E). The species identity of isolate VK10A was confirmed by amplification, sequencing, and BLAST analysis of the Internal Transcriber Spacer regions of the ribosomal DNA. The ITS sequence of the isolate VK10A showed 100% identity with *P. oleae* isolate Po2a (CBS7670, GenBank accession No. KY982934.1), paratype

✉ Santa Olga Cacciola
olga.cacciola@unict.it

✉ Giuseppe Dionisio
giuseppe.dionisio@agro.au.dk

[1] Department of Agriculture, Food and Environment (Di3A), University of Catania, Via Santa So-fia 100, Catania 95123, Italy

[2] Department of Agricultural Science, Mediterranea University of Reggio Calabria, Localitá Feo di Vito, Reggio Calabria 89122, Italy

[3] Department of Agroecology, section of Crop Biotechnology and Genetics, Aarhus University, Forsoegsvej 1, Slagelse 4200, Denmark

of the *P. oleae type* isolate Po1a (ex-type: CBS7669), belonging to clade 2d (Ruano-Rosa et al. 2018; Jung et al. 2024).

General information of the sequencing approach and methods include a genomic DNA (gDNA) preparation of high quality and length for long length sequencing by Oxford Nanopore Sequencing (ONT) with about 150 x as expected genome coverage and Illumina short read sequencing (150 PE) of gDNA (with approximately 90 x coverage) to polish the ONT reads. Furthermore, to facilitate the mapping of the coding sequences (CDSs) onto the newly predicted genes, a de novo transcriptome assembly was performed using cDNA and the Trinity assembler (Haas et al. 2013).

Detailed information for genomic sequencing is described starting from a four-day-old culture of *P. oleae* isolate VK10A grown in Potato Dextrose Agar (PDA) which was subjected to gDNA extraction by using DNeasy PowerSoil Pro kit (Qiagen, Denmark) according to the manufacturer's protocol. The obtained gDNA fragments were in the range of 14–22 kbp with no discernible small fragments as determined by Bioanalyzer analysis (Agilent Technologies, Denmark). DNA concentrations and purities were measured with the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, USA) and the NanoDrop One (Thermo Fisher Scientific, USA). DNA size distributions were evaluated using the Genomic DNA ScreenTapes on the Agilent Tapestation 4200 (Agilent, USA).

Library preparation was carried out using the ligation sequencing kits (Oxford Nanopore Technologies, UK) on a barcoded SQK-LSK114.24 DNA library prepared according to the manufacturer's protocol, with minor modifications (DNAsense ApS, Aalborg, Denmark) (Sereika et al. 2022). The barcoded DNA library (approximately 10–20 fmol) was loaded onto a primed FLO-MIN114 flow cell and sequenced on a MinION Mk1b device, running MinKNOW v. 23.04.5. FAST5 signal data was base-called and demultiplexed with Guppy v. 6.5.7 (Oxford Nanopore Technologies, Oxford, United Kingdom) using the super-accurate algorithm (dna_r10.4.1_e8.2_400bps_5khz_sup.cfg). Removal of low-quality reads and generation of basic sequencing data statistics were obtained using Nanoq v. 0.9.0 (Steinig and Coin 2022).

Filtered reads were de novo assembled with Flye (Kolmogorov et al. 2019, 2020) polished once with the help of the Medaka v. 1.8.0 software (https://github.com/nanoporetech/medaka). Subsequently, the ONT data errors were polished with Illumina 350 PE, 1-2Gb (approximately 4-7 M reads) data using Racon (via Minimap and Samtools). The draft assembly was subsequently polished twice with Medaka v. 1.8.0. Assembly graphs were inspected with Bandage v. 0.8.1 (Wick et al. 2015). Contigs below 1000 bp were removed with SeqKit v. 2.2.0 (Shen et al. 2016). BUSCO v. 5.2.2 (Manni et al. 2021) was used to evaluate genome completion. Quality-filtered Illumina and Nanopore reads were classified with Kaiju v. 1.9.2 (Menzel et al. 2016) against the subset of GenBank proteins *nr* eukaryotic database.

For Illumina RNAseq, *P. oleae* VK10A was grown on PDA and V8 medium and incubated at 20 °C respectively for 10 and 5 days. Total RNA was extracted from mycelia using 1.5 ml of TRI Reagent (Zymo Research, Nordic BioSite, Sweden) using ZR BashingBead Lysis (0.5 mm Y_2O_3 Stabilized Zirconia Beads; Zymo Research, Nordic BioSite, Sweden) and the FastPrep-24™ 5G bead beating grinder and lysis system (MP Biomedicals, Fischer Scientific, Denmark). Genomic DNA (gDNA) and Total RNA samples were sequenced at Novogene (Cambridge, UK) with Illumina Hiseq 4000 and 150 pair end read length (PE), at 30 million of read depths. Data quality control revealed a Q30 [(Base count of Phred value > 30) / (Total base count)] average of 93% for a total of 98% of clean reads.

The quantitative assessment of the assembly was carried out by calculating assembly statistics by using the BUSCO (Benchmarking Universal Single-Copy Orthologs) scores (Manni et al. 2021). BUSCO evaluates the completeness of assemblies by looking for the presence or absence of highly conserved genes (BUSCO orthologs). The *P. oleae* genome assembly was 99% complete relative to an identified ortholog (*P. plurivora*) that covers the entire length of the alignment sequence of the corresponding BUSCO benchmarking. *P. oleae* genome assembly was 1% fragmented which is related to identified orthologs that only partially covers the length of the sequence alignment. Figure 1 shows the Flye v. 2.9.1-b1780 generated *P. oleae* genome draft assembly graph. One contig clearly represented the complete mitochondrial genome, which was annotated to contain 65 genes, including 38 CDS, 2 rRNA, and 25 tRNA genes, and submitted to GenBank (PQ510212). However, due to many repetitive regions we could not split the assembly into chromosomes. The Flye Quality-filtered Illumina and Nanopore reads were also classified with Kaiju v. 1.9.2, against the GenBank nr protein database (2023-05-10 database, 321 M protein sequences from both prokaryotes and microbial eukaryotes). When blasted against known *Phytophthora* genomes using GenBank deposited genomes, the *P. oleae* genome gave as top blastn hits the *P. plurivora* (NCBI: txid639000) genome (with around 93% of genome similarity. This latter also possesses a small genome of 46.9 Mbp (ASM3002794v1).

P. oleae genome annotation was performed using Funannotate 1.8.17 (Stanke and Waack 2003; Jensen et al. 2008; Ter-Hovhannisyan et al. 2008; Eddy 2011; Palmer and Stajich 2020), which started with a filtering out of scaffolds containing mitochondrial DNA sequences and finally 156 scaffolds remained. Afterwards, a RepeatMasking step of

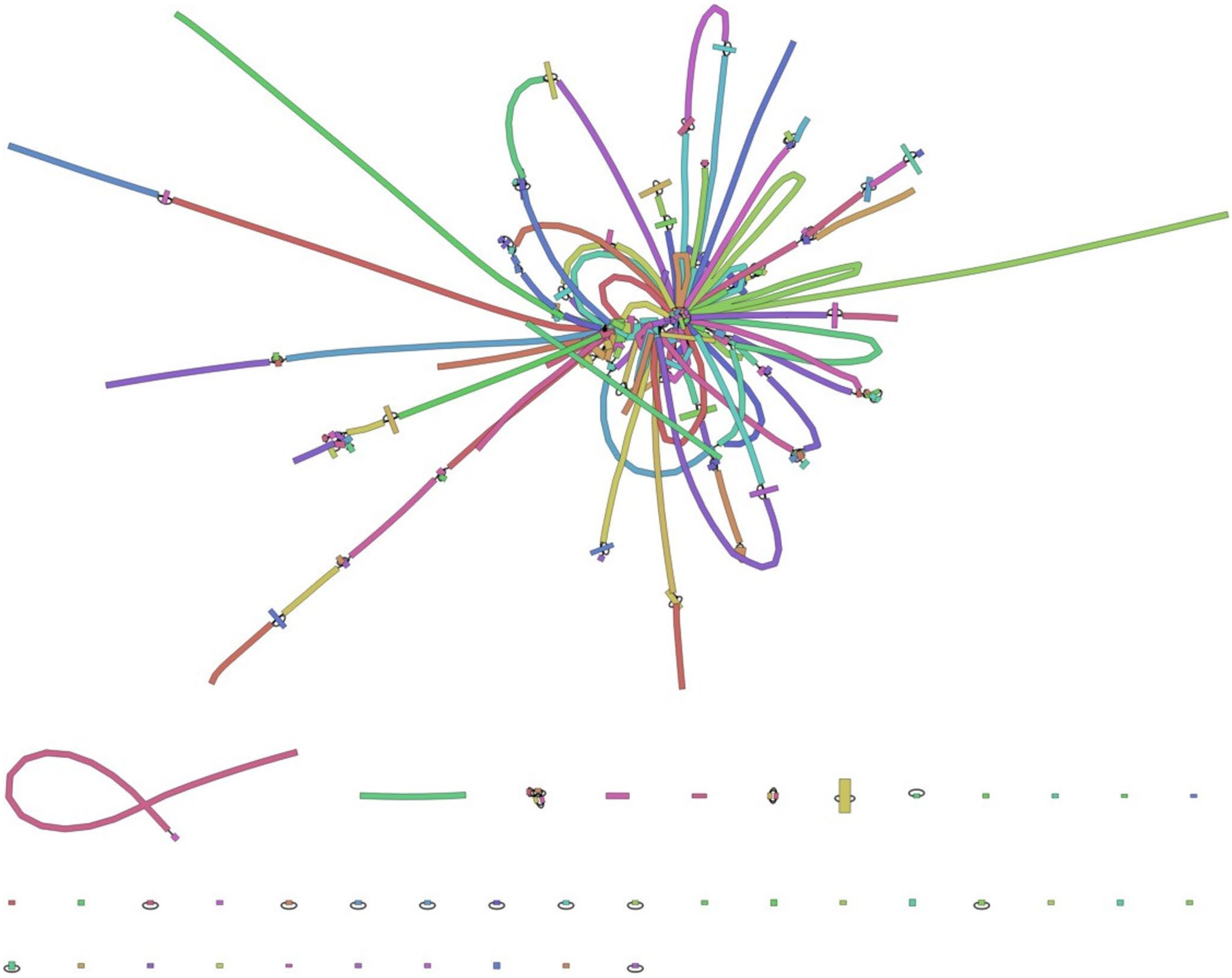

Fig. 1 *P. oleae* genome draft assembly graph. Draft assembly graph resulting from the draft *de novo* genome assembly using Flye v. 2.9.1-b1780, highlighting assembly content and contiguity. Only 167 contigs were found in this first draft. However, due to the known *Phytophthora* microsatellite highly repetitive Simple Sequence Repeat (SSR) regions, the contigs in future assembly might be shorter and hopefully better organized into chromosomes. Coloured lines denote size-scaled contiguous DNA elements. Line thickness is proportional to contig coverage. Thin black lines denote connectivity between segments

the genome was made and these soft masked scaffolds were used for the ab initio prediction using AUGUSTUS (Stanke and Waack 2003; Stanke and Morgenstern 2005) trained with *Fusarium graminearum* intron-exon was made. The funannotate "annotate" pipeline was combined with the Trinity *de novo* mRNA assembly of the cDNA. Prediction of open reading frames (ORFs) from *de novo* mRNA assembly was performed both on the sense and antisense strand by GeneMark (Borodovsky and McIninch 1993a, b) and GeneMark-ES, version 2 (Ter-Hovhannisyan et al. 2008). For getting the genomic coding sequences (gCDS) we found that besides the (A/C)AG|*GT*(A/G)AGT consensus sequence (the exon|intron boundary) whereas the first two nucleotides of the intron are underlined, another rare exon|intron consensus, CG|GC as boundary with exons|intron was observed while GT|TG, GT|AC, and AT|AG rare intron|exon consensus was also found at low percentage.

In total for the *P. oleae* first draft assembled genome, represented by 43.7 Mbp, we have annotated 19,089 genes, of which 18,841 gCDS/mRNA and 865 tRNA. A comparison with Trinity *de novo* assembly (Haas et al. 2013) and GeneMark prediction, less proteins were predicted using the cDNA (Table 1).

The *P. oleae* VK10A genome (43.7 Mbp) as genome size in comparison to other *Phytophthora* species genome size spanning from 228.5 Mbp (e.g. *P. infestants*, reference genome ASM14294v1), to 165.6 Mbp (e.g. *P. palmivora*, PpalZC01v1) or from 101.5 Mbp (e.g. *P. megakarya*, ASM221536v1) to 59 Mbp (e.g. *P. nicotianae*,

Table 1 *P. oleae* VK10A ONT genomic DNA plus *de Novo RNA* sequencing and assembly statistics together gene prediction and proteome annotation. Using a combination of ONT sequencing and illumina DNA and RNA sequencing we have produced the first draft of *P. oleae* VK10A genome and proteome annotation. The table shows both DNA sequencing and assembly data as well as gene prediction by AUGUSTUS (Funanotate) or Trinity (in both sense and antisense orientation). Legend: the data (Mb) denotes the data yield and the number of generated reads. Read N50 (bp) is the value where half of the data is contained within reads of length N50 or greater. Quality score denotes the median Phred-scaled read quality score. Contigs denote the number of contiguous DNA elements associated with the assembly. Assembly size (Mbp) and largest contig (Mbp) denote the size of the final assembly and the length of the longest contig, respectively. Contig n50 (kbp) is the value where half of the assembly is contained within contigs of length N50 or greater. GC content (%) denotes the mean content of the G and C nucleotides

	Raw			Trimmed		
ONT DNA sequencing data	**Data (Mbp)**	**Read N50 (bp)**	**Quality score**	**Data (Mbp) 150 X**	**Read N50 (bp)**	**Quality score**
	10,358	7364	18.6	6701	7562	20.5
Assembly data	**Scaffolds**	**Assembly length (Mbp)**	**Largest contig (Mbp)**	**GC content (%)**	**Contig N50 (kbp)**	
	156	43.16	2.05	51.82	684	
Gene prediction data	**Funannotate nuclear genes**	**TRINITY de novo mRNA**	**Funannotate gCDS/mRNA**	**tRNA**	**Secretome**	**Apoplastic effectors**
	19,089	13,547	18,841	865	1995	605
	Funannotate mitochondrion genes	**Mitochondrion CDS**	**Mitochondrion rRNA**	**Mitochondrion tRNA**		
	65	38	2	25		
Proteome annotation data	**CAZy enzymes**	**Glycoside Hydrolases (GHs)**	**Glycosyl Transferases (GTs)**	**Polysaccharide Lyases (PLs)**	**Carbohydrate Esterases (CEs)**	**Auxiliary Activities (AAs)**
	553	248	117	45	29	104
	Secreted (SignalP)	**Proteolytic enzymes (MEROPS)**	**Peptidases**	**Proteases**	**InterPro annotated**	**GO annotated**
	1995	472	270	202	11,526	8769
	NPP1 proteins	**RxLR effectors**	**Elicitins**	**Crinkler**		
	39	236	78	264		

ASM332846v1) places it among the smaller *Phytophthora* genomes.

Initially, the *P. oleae* VK10A related proteome was annotated by blasting it against most of the GenBank *Phytophthora* species proteins, Mercator4 v7.0 (Lohse et al. 2014). With Funannotate "annotate" function, most of the proteins were hierarchical, functionally and phylogenetically annotated using hmmscan/search, diamond and using eggnog database ver. 5.0 (Jensen et al. 2008; Huerta-Cepas et al. 2018). Furthermore, by the blasting into the automated Carbohydrate-active enzyme ANnotation (dbCAN) we identified 553 CAZy enzymes (Yin et al. 2012) which details are presented in different Tabs of the excel file related to the annotation of *P. oleae* NCBI genome (GCA_046127205.1) on Zenodo (https://doi.org/10.5281/zenodo.13254492). InterProScan ver 5.0 (Jones et al. 2014) and gene ontology GO annotation was also performed by the Funannotate annotate pipeline. About 472 proteases (270 peptidases and 202 proteases) were detected belonging to MEROPS databases of which some of those correspond to known pathogenic cysteine and subtilisin proteases. From the secretome point of view, 1995 proteins were predicted to be having a secretion type leader peptide using signalp version 4.0 (Petersen et al. 2011) (Table 1).

From the effectors and pathogenicity proteins point of view, we have blasted among several *Phytophthora* species with annotated protein our *P. oleae* proteins and discovered 236 RxLR effectors, of which 26 were top blast hits to known avirulence (Avh) proteins, 78 Elicitins, 264 Crinkler proteins and 39 necrosis inducing proteins (NPP). This genome represents the first publicly available assembly of *P. oleae*, an emerging olive pathogen recently reported in multiple geographic regions. The availability of this resource will enable in-depth studies on its population structure, host adaptation, and pathogenic mechanisms, and will support the development of diagnostic tools and sustainable disease management strategies.

Acknowledgements The authors acknowledge the projects Emerge-Now- Emergent plantpest digital services"; HORIZON-CL6–2024-FARM2FORK-02–3-two-stage; Innovation Action; 36 months-Codice 101182461. The authors acknowledge the projects "Nuove soluzioni tecnologiche per la filiera degli agrumi: trasferimento di innovazione per il miglioramento degli standard qualitativi delle produzioni agrumicole e per fronteggiare la minaccia di malattie infettive di origine fungina degli agrumi nel comprensorio del GAL Eloro – NewCitrusTech" (cod. 1.3.1" - Misura 19 PSR 2014/2022 Sostegno allo sviluppo locale leader, sottomisura 19.2 "Sostegno all'esecuzione degli interventi nell'ambito della strategia di Sviluppo locale di tipo partecipativo - CLLD); "Azioni Innovative per la Produttività del Distretto dell'Ortofrutta di Qualità – INNOVAPROD" (cod. 1.3.2" - Misura 19 PSR 2014/2022 "Sostegno allo sviluppo locale leader, sottomisura 19.2; Sostegno all'esecuzione degli interventi nell'ambito della strategia di Sviluppo locale di tipo partecipativo - CLLD); the INTERREG NEXT Italia–Tunisia project "PROMESERVAGRI 4.0 - Réseau de transfert des résultats de la recherche pour la transition vers des MPME dans la culture et la mouture des olives qui adoptent l'innovation et les technologies avancées (CUP: E23C25000110007, Reference No. A1–1.3–228).

Funding Open access funding provided by Università degli Studi di Catania within the CRUI-CARE Agreement. Novo Nordisk Foundation grant number: NNF22OC0076891.

Data availability The Whole Genome Shotgun (WGS) of *P. oleae* genome assembly project has been deposited at DDBJ/ENA/GenBank under the accession number JBIMZQ000000000, associated with BioProject PRJNA1080363 and BioSample SAMN43902879. Raw sequencing reads have been submitted to the NCBI Sequence Read Archive under accession numbers SRX29220164 (Illumina data) and SRX29220163 (ONT data).
The proteome annotation table has been deposited to Zenodo databases: https://doi.org/10.5281/zenodo.13254492.
The *Phytophthora oleae* isolate VK10A has been deposited in the CBS Collection of the Westerdijk Fungal Biodiversity Institute (Utrecht, Netherlands) under the accession number CBS 153549.

Declarations

Ethical approval Not applicable.

Conflict of interest Authors declare no conflict of interest.

References

Abad ZG, Burgess TI, Bourret T, Bensch K, Cacciola SO, Scanu B, Mathew R, Kasiborski B, Srivastava S, Kageyama K, Bienapfl JC, Verkleij G, Broders K, Schena L, Redford AJ (2023) *Phytophthora*: taxonomic and phylogenetic revision of the genus. Stud Mycol 106:259. https://doi.org/10.3114/SIM.2023.106.05
Borodovsky M, McIninch J (1993a) GENMARK: parallel gene recognition for both DNA strands. Comput Chem 17:123–133. https://doi.org/10.1016/0097-8485(93)85004-V
Borodovsky M, McIninch J (1993b) Recognition of genes in DNA sequence with ambiguities. BioSystems 30:161–171. https://doi.org/10.1016/0303-2647(93)90068-N
Brasier C, Scanu B, Cooke D, Jung T (2022) *Phytophthora*: an ancient, historic, biologically and structurally cohesive and evolutionary successful generic concept in need of preservation. IMA Fungus 13:1–25. https://doi.org/10.1186/S43008-022-00097-Z
Davison EM (1998) *Phytophthora* diseases worldwide. Plant Pathol 47(2):224–225. https://doi.org/10.1046/j.1365-3059.1998.0179a.x

Eddy SR (2011) Accelerated profile Hmm searches. PLoS Comput Biol 7(10):e1002195

González M, Pérez-Sierra A, Sánchez ME (2019) *Phytophthora oleae*, a new root pathogen of wild olives. Plant Pathol 68(5):901–907. https://doi.org/10.1111/ppa.13024

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) De Novo transcript sequence reconstruction from rna-seq using the trinity platform for reference generation and analysis. Nat Protoc 8(8):1494–1512

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 (2018) Eggnog 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47(D1):D309–D314

Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P (2008) Eggnog: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res 36(Database issue):D250–254

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) Interproscan 5: Genome-scale protein function classification. Bioinformatics 30(9):1236–1240

Jung T, Milenković I, Balci Y, Janoušek J, Kudláček T, Nagy ZÁ, Baharuddin B, Bakonyi J, Broders KD, Cacciola SO, Chang T-T, Chi NM, Corcobado T, Cravador A, Đorđević B, Durán A, Ferreira M, F C-H, Garcia L, Hieno A, H H-H, Hong C, Junaid M, Kageyama K, Kuswinanti T, Maia C, Májek T, Masuya H, Magnano G di San Lio, Mendieta-Araica B, Nasri B, Oliveira LSS, Pane A, Pérez-Sierra A, Rosmana A, Sanfuentes, von Stowasser E, Scanu B, Singh R, Stanivuković Z, Tarigan M, Thu PQ, Tomić Z, Tomšovský M, Uematsu S, Webber JF, Zheng H-C, Zheng F-C, Brasier CM, Horta Jung M (2024) Worldwide forest surveys reveal forty-three new species in phytophthora major clade 2 with fundamental implications for the evolution and biogeography of the genus and global plant biosecurity. Stud Mycol 107:251–388. https://doi.org/10.3114/SIM.2024.107.04

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

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA (2020) MetaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17:1103–1110. https://doi.org/10.1038/s41592-020-00971-x

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B (2014) Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ 37:1250–1258. https://doi.org/10.1111/PCE.12231

Manni M, Berkeley MR, Seppey M, Zdobnov EM (2021) BUSCO: assessing genomic data quality and beyond. Curr Protoc 1:e323. https://doi.org/10.1002/CPZ1.323

Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257. https://doi.org/10.1038/ncomms11257

Palmer JM, Stajich J (2020) Funannotate v1. 8.1: eukaryotic genome annotation. Zenodo 28. https://doi.org/10.5281/zenodo.4054262

Petersen TN, Brunak S, Von Heijne G, Nielsen H (2011) Signalp 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786

Ruano-Rosa D, Schena L, Agosteo GE, Magnano di San Lio G, Cacciola SO (2018) *Phytophthora oleae* sp. Nov. Causing fruit rot of Olive in Southern Italy. Plant Pathol 67:1362–1373. https://doi.org/10.1111/ppa.12836

Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, Albertsen M (2022) Oxford nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference Polishing. Nat Methods 19:823–826. https://doi.org/10.1038/s41592-022-01539-7

Shen W, Le S, Li Y, Hu F (2016) SeqKit: A Cross-Platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE 11:e0163962. https://doi.org/10.1371/JOURNAL.PONE.0163962

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 (Suppl. 2) ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33:W465–W467. https://doi.org/10.1093/NAR/GKI458

Steinig E, Coin L (2022) Nanoq: ultra-fast quality control for nanopore reads. J Open Source Softw 7:2991. https://doi.org/10.21105/joss.02991

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an Ab initio algorithm with unsupervised training. Genome Res 18:1979–1990. https://doi.org/10.1101/GR.081612.108

Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de Novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/BIOINFORMATICS/BTV383

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) Dbcan: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40(W1):W445–W451. https://doi.org/10.1093/nar/gks479