BMC Genomics

RESEARCH Open Access

Hybrid genome assembly of the cannabis powdery mildew agent *Golovinomyces ambrosiae* uncovers important resources for deciphering virulence factors

Félix-Antoine Roy[1], Yanick Asselin[1], Caroline Labbé[1], Guillaume Quang Henri Nguyen[2], Sima Mohammadi[2], Jeff Gauthier[2], Benjamin Cinget[1], Davoud Torkamaneh[2], Roger C. Levesque[2] and Richard R. Bélanger[1*]

Abstract

Background Cannabis powdery mildew, caused by the fungal pathogen *Golovinomyces ambrosiae*, poses a significant threat to licensed producers as its presence on marketable products can compromise product innocuity. While there is a focus in research for resistance genes within the *Cannabis sativa* germplasm, there is a lack of genetic information regarding the infecting agent, preventing validation of their effectiveness against different populations as the plant-pathogen interaction most likely follows a gene-for-gene relationship. In this paper, we assembled the first *G. ambrosiae* genome, providing insights into its genomic content and potential virulence determinants. The assembly was made using a hybrid approach, combining Oxford Nanopore Technologies long reads and Illumina short reads.

Results The resulting 155.2 Mb genome is composed of 73 contigs, 13 scaffolds and has a completeness score of 97.5%. Subsequent analysis highlighted the substantial transposable elements content of the pathogen, occupying 82.64% of its genomic composition. Prediction of protein-coding genes revealed 6995 highly confident gene models, including 169 candidate effector proteins. Among the latter, we highlighted 14 candidates sharing key characteristics of confirmed effectors in other powdery mildew pathosystems such as the presence of signal peptides, RALPH-like domains, and Y/F/WxC motifs.

Conclusions The result of this study provides valuable resources for future identification of avirulence genes within the *G. ambrosiae* species, responsible of conferring resistance when encoded effectors are recognized by a cognate host's resistance genes. Those findings will lead to guided breeding strategies and, ultimately, the selection of cultivars adapted to the pathogen's virulence profile.

Keywords Powdery mildew, Genome, *Cannabis sativa*, Genetic resistance, Effector, Avirulence

*Correspondence:
Richard R. Bélanger
richard.belanger@fsaa.ulaval.ca
[1]Département de Phytologie, Université Laval, Québec City, QC G1V 0A6, Canada
[2]Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec City, QC G1V 0A6, Canada

Introduction

Powdery mildews are part of the Ascomycota and are composed of an impressive diversity of species within the Erysiphales order [1]. To date, more than 900 species have been identified, and these obligate biotrophic fungi affect more than 16 000 angiosperms [2]. For cannabis (*Cannabis sativa* L.), powdery mildew (PM) is one of the most common diseases in production facilities. In Canada only, production sites sampled from 2013 to 2017 revealed the presence of the infecting agent in 100% of locations surveyed [3]. The disease can lead to major yield consequences when infected leaves, stems and flowers reduce photosynthesis and nutrient uptake [4]. More importantly, due to microbial testing guidelines for the commercialisation of legal products, the presence of the pathogen on cannabis flowers can severely impact market access/regulatory approval [5].

There is a notable discrepancy across the literature regarding the nomenclature of the fungal pathogens causing powdery mildew on cannabis. Many reports have been referring to the fungus as *Golovinomyces cichoracearum sensu lato* (*Erysiphe cichoracearum*) [3, 6, 7] while others have mentioned *Golovinomyces spadiceus* (*Erysiphe spadiceus*) [8, 9], *Golovinomyces ambrosiae* (*Erysiphe ambrosiae*) [10, 11] or *Podosphaera macularis* [12]. In the early 1900s, *E. cichoracearum* was considered to be a very cosmopolitan species, capable of infecting 280 different plant species divided into 27 families [13]. It was also noted that different biological forms were observed depending on the host. With this large host repertoire and the observation of different biological forms, it is likely that *E. cichoracearum* included many species presently classified under the genus *Golovinomyces* [14]. As of today, it is recognised that *G. cichoracearum* is a complex, now subdivided into 11 species including *G. ambrosiae* [2]. Furthermore, a recent multi-locus phylogenetic and taxonomic study revealed that despite *G. ambrosiae* and *G. spadiceus* being distinguished based on morphological traits [15], phylogenetic analysis using ITS, 28 S rDNA, IGS, TUB2 and CHS1 sequences confirm that both represent a single taxon [14]. There is now a consensus that the only powdery mildew species affecting cannabis are *G. ambrosiae* and *P. macularis*, with the latter being documented only under field conditions. It has also recently been documented that both species can co-infect *Cannabis sativa* hosts [12, 16].

Many strategies have been adopted to prevent or reduce the disease pressure. These include climate control in indoor production facilities by creating unfavorable conditions for the growth of the fungus, such as proper ventilation, humidity adjustments, reduced plant density or the removal of infected plants/leaves [4, 17]. The use of organic fungicides can also be a solution for managing the disease, as several products are registered for this purpose in cannabis production [18]. Out of the 19 registered products in Canada, only a few have published data on their efficacy. While some organic fungicides show promising control of the disease, efficacy results appear to vary among trials [4, 18]. Moreover, successful management of powdery mildew with these products requires rigorous weekly applications, which are time-consuming and directly increase production costs.

In many crops, powdery mildew is best managed by genetic approaches [4]. To our knowledge, two distinct mechanisms have been exploited to provide resistance to the disease [19]. One mechanism implicates resistance genes (*R genes*) with conserved nucleotide-binding site and/or leucine-rich repeat domains (NLRs), which encode proteins that function as intracellular immune receptors. These proteins are implicated in the direct or indirect recognition of specific secreted effector proteins encoded by the pathogen's avirulence gene (*Avr*) and are responsible for triggering a hypersensitive reaction (HR), inducing localized cell death to prevent the spread of the infection. This model implicates a gene-for-gene interaction, as typically one *R* gene is effective against one specific *Avr*. On this basis, *R* genes provide race-specific powdery mildew immunity to the host and have been exploited in various crops such as wheat, pea, and grape [20–22]. The second mechanism is related to powdery mildew susceptibility (*S*) genes, referred to as *Mildew resistance locus o* (MLO). Unlike *NLR* genes, these are only found in plants, and they provide broad-spectrum resistance to powdery mildew pathogens when subjected to loss-of-function due to natural or induced mutations. This mechanism has been widely exploited for barley, since the discovery of a naturally mutated *Mlo* gene in an Ethiopian landrace in the 1930s, which provides complete powdery mildew resistance [19, 23]. Ultimately, both mechanisms provide an economically and environmentally friendly solution to prevent powdery mildew.

Research for genetic resistance to powdery mildew in cannabis is currently advancing rapidly. In 2021, Mihalyov and Garfinkel were the first to identify a specific locus presumably containing an *R* gene, named PM1, in the Pacific Northwest experimental variety line "PNW39", providing complete resistance to a powdery mildew isolate [24]. In the same year, Pépin et al. identified and characterized 15 MLO candidate genes across five different *C. sativa* genomes. Among those candidates, it was observed that two of them, CsMLO1 and CsMLO4, were significantly upregulated during the infection, making them strong candidates for susceptibility [25]. In 2024, Stack et al. discovered a major quantitative trait locus (QTL) neighboring the previously identified CsMLO1 on *C. sativa* chromosome 1 (Chr01) from the cultivar "FL 58". By inspecting the CsMLO1 gene sequence on this cannabis variety, they discovered a 6.8 kb insertion that introduces

a premature stop codon, reinforcing evidence that this candidate could contribute to powdery mildew resistance because of loss-of-function [26]. More recently, another research team discovered a second locus thought to harbour an *R* gene, designated PM2, through a bulk-segregant analysis and high-throughput RNA sequencing [27].

While these discoveries are promising for the development of elite powdery mildew resistant cannabis cultivars, the lack of genetic knowledge of the pathogen itself poses a significant hindrance for the validation of their efficacy in large scale deployment. Even though *Mlo*-based resistance is said to provide broad spectrum protection, it has been reported in other crops that some strains of the fungal pathogen can bypass this mechanism possibly due to intrinsic genetic divergences in a RBR-family E3 ubiquitin ligase and a medA-like transcriptional regulator [28]. As for *R* genes, genetic knowledge of the pathogen is even more important given their intimate relationship with the pathogen's *Avr* genes. Since *Avrs* are rapidly evolving genes [29], significant diversity among them is expected in different geographical locations or time periods, directly impacting the effectiveness of *R* genes. Ultimately, a deeper understanding of the genetic features of the pathogen will make it possible to proceed to functionality testing of the newly identified *R* genes to validate their performance in various pathological scenarios.

In this study, we provide the first assembled genome of the cannabis powdery mildew pathogen *Golovinomyces ambrosiae*, produced by the combination of Oxford Nanopore Technologies (ONT) long reads and Illumina NovaSeq short reads sequencing platforms, offering an exhaustive description of its genetic composition such as transposable elements content, protein coding genes and potential virulence factors.

Materials and methods

Collection of fungal material

Heavily powdery mildew infected cannabis leaves were collected from the susceptible cultivar "Green dragon", grown in a greenhouse compartment at Laval University, Québec, Canada. Whole leaves were selected, cut into leaflets and quickly dipped in a 5% acetate acetone solution (5 g of cellulose acetate dissolved in 100 ml anhydrous acetone) [30]. Leaflets were dried for 8 min on two stainless steel cell spreaders placed side by side or until all acetone had completely evaporated leaving a hardened cellulose pellicle on the surface of the leaflets, trapping the mycelium and spores. The cellulose pellicles were peeled off the leaflet surfaces, stored in a 50-ml Falcon tube and placed immediately at -80 °C until further use.

DNA extraction

A total of 500 mg of cellulose pellicles containing the fungal material were placed in a mortar with liquid nitrogen and crushed with a pestle until a fine powder was obtained. The improved High molecular weight (HMW) DNA extraction method developed by Russo et al. [31, 32] was used for the current study. The final high-molecular-weight DNA concentration was quantified using the Qubit dsDNA BR Assay Kit on a Qubit Fluorometer (Thermo Fisher Scientific), and purity ratios were assessed with a NanoDrop spectrophotometer (Thermo Fisher Scientific). Fragment size distribution and DNA integrity were evaluated using the FemtoPulse system (Agilent Technologies) to confirm suitability for long-read sequencing.

Library preparation and DNA sequencing

For long-read sequencing, a library was prepared from 3 µg of HMW DNA using the ONT ligation sequencing kit (SQK-LSK114), following the manufacturer's protocol, and enriched for fragments ≥ 3 kb using ONT's Large Fragment Buffer. Sequencing was performed on a PromethION 2 solo (Oxford Nanopore Technologies, Oxford, UK) for 72 h using a FLO-PRO114M flow cell with R10.4.1 chemistry. From the same initial sample, 2 µg of DNA were sent to the Génome Québec Innovation Centre (Montréal, QC, Canada) where short-read library preparation was performed followed by a short-read Illumina NovaSeq sequencing to produce 2 × 150 paired-end reads.

Hybrid genome assembly

All steps used for genome assembly and subsequent analysis are summarized in Table 1. Long reads were basecalled using ONT's basecaller Dorado v1.0.2 within MinKNOW v25.04, with the basecalling model "dna_r10.4.1_e8.2_400bps_hac@v4.3.0". Adapter sequences were removed with Porechop v0.2.4 and reads were then filtered with NanoFilt v2.3.0 [33] to retain sequences with a Phred quality score equal or superior to 15 and a minimum length of 20 kb. To identify contaminants in the dataset, a sample of reads was queried against the BLASTn *nt* online database (E-value 1e-5). Species that could have been realistically present in the starting fungal sample, either by direct contact or present in the laboratory environment were selected for subsequent processing. Long and short reads were mapped using minimap2 v2.28 [34] to the selected contaminant genomes, (*C. sativa*, *Frankliniella occidentalis*, *Monilinia fructicola*, *Allantophomopsis cytisporea*) as well as genomes of closely related species of *G. ambrosiae* (*G. cichoracearum*, *Blumeria graminis* f.sp. *tritici* and *Erysiphe necator*) in order to capture reads closest to the target organism and eliminate unwanted sequences. Reads mapping only to powdery mildew genomes were used for the assembly using Flye v2.9.5 [35] (parameters: --nano-raw --iterations 3 --scaffold --min-overlap 8500). Illumina short reads were trimmed with Fastp 0.23.4 [36] and used for two rounds of polishing the long-read assembly using Pilon v1.23 [37]. To evaluate the completeness of

the genome assembly, a Benchmarking Universal Single-copy Orthologs (BUSCO) [38] analysis was run using the Ascomycota lineage database (ascomycota_odb10).

Phylogeny

Golovinomyces ambrosiae was included in a genome-scale phylogenetic analysis using conserved single-copy orthologs identified with the BUSCO v5.8.2 pipeline and the ascomycota_odb10 lineage dataset. The BUSCO lineage comprises 3235 conserved orthologous groups. Genomic assemblies from representative taxa within Ascomycota were retrieved from the NCBI database. Accession number for each genome is provided in Additional file 1. BUSCO screening identified single-copy orthologs across all genomes. From the total ortholog set, 1805 genes detected in at least 90% of the analyzed genomes were retained for downstream phylogenomic reconstruction. For each retained ortholog, predicted protein sequences were aligned independently using MAFFT v7.525 with the parameter --auto. Individual alignments were subsequently concatenated into a supermatrix using AMAS v1.0 [39] with the parameters concat -f fasta -d aa, resulting in a final concatenated amino acid (aa) matrix of 19.32 M sites for 18 taxa, including *G. ambrosiae* and the selected Leotiomycetes representatives, as well as an appropriate outgroup.

Maximum-likelihood phylogenetic inference was performed with IQ-TREE2 v2.1.3 [40]. Branch support was assessed using 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT tests. The best-fitting amino acid substitution mIQodel was selected automatically using ModelFinder-Plus (-m MFP), based on the Bayesian Information Criterion (BIC). According to BIC, Q.plant + F+R7 was chosen as best-fit model. The phylogenetic tree was visualized using Interactive Tree of Life (ITOL) online platform [41].

Repetitive elements and protein-coding gene predictions

The resulting genome was processed through RepeatModeler v2.0.6 [42] to identify repetitive regions which were then masked from the assembly using RepeatMasker v4.1.8 [43] to exclude those sequences from the gene prediction processes. Protein-coding genes were predicted using MAKER2 v2.31.10 [44], with provided evidence of gene annotations from the powdery mildew agent *Blumeria graminis* f.sp. *triticale* (GCA_905067625.1) and protein sequences of two closely related species of *G. ambrosiae* with genomic data available *E. necator* (GCA_024703715.1) and *G. cichoracearum* (GCA_003611235.1, GCA_003611195.1 and GCA_003611215.1). Ab initio gene predictions with Augustus v3.3.2 [45] were also integrated to the pipeline for additional gene models using the *Botrytis cinerea* species pretrained model parameter. The completeness of the gene prediction dataset was evaluated with BUSCO v5.8.2 against the Leotiomycetes database (leotiomycetes_odb10). Ultimately, a BLASTp search against the whole nonredundant protein database was ran to identify similar protein sequences.

Identification of candidate effector proteins (CEPs), homology analysis and orthogroup classification

Predicted protein sequences generated from the gene prediction pipeline were first filtered to keep only those ranging from 12 to 700 amino acids. The filtered dataset was submitted to the online SignalP 5.0 [46] servers to identify protein sequences with a predicted signal peptide. Sequences with the presence of a signal peptide were then processed by DeepTMHMM v2.0 [47] algorithm to predict transmembrane domains [48]. All protein sequences with a transmembrane domain were discarded and the remaining ones were submitted to the EffectorP 3.0 [49] online platform to ultimately predict effector-related proteins. To determine if any predicted sequences had homologues to known haustorium secreted proteins, a homology analysis was run using MMseqs2 [50] with publicly available data of effectors from *Golovinomyces cichoracearum* isolates GcM1 and GcM3 [51]. To validate if any of the predicted effectors in the entire dataset could be associated with RNase-Like Proteins associated with Haustoria (RALPH), a search was conducted in the protein annotations to identify CEPs with ribonuclease domains. An orthogroup classification of the RALPH-like CEPs was performed using OrthoFinder v2.5.5 [52] with predicted RALPH effector proteins of *Blumeria graminis* f. sp. *tritici* (GCA_900519115.1), *Blumeria graminis* f. sp. *hordei* (GCA_900237765.1), *Erysiphe necator* (GCA_024703715.1) and *G. cichoracearum* (GCA_003611235.1). Protein sequence identity between orthologues was evaluated by MAFFT and Clustal 2.1 [53].

Final gene prediction filtering and annotation

To obtain a final predicted gene dataset, all sequences validated by BUSCO analysis and all identified CEPs were initially kept. For the rest of the predicted genes, every sequence with an Annotation Edit Distance (AED) score superior to 0.1 was discarded to ensure high accuracy with evidence-based predictions. Functional annotation of this final gene set was run using eggNOG-mapper v2.1.12 [54] and InterProScan [55]. Additional search for specific domains was made using HHpred [56] online servers and HMMER from the EMBL-EBI website.

Identification and classification of duplicated genes

A BLASTp vs. all (E-value < 1E-5) search was performed on the final protein sequence dataset of predicted genes with a maximum of five hits per sequence to identify duplicated genes. The BLASTp output was then filtered to keep only hits with a minimum identity of 80% and a minimal subject coverage of 75% to retain sequences with relatively low sequence divergence. The classification of these duplicated genes as dispersed, proximal or tandem was done using MCScanX [57] with the *duplicate_gene_classifier* script.

Table 1 Summary of tools used for the assembly and characterization of the *Golovinomyces ambrosiae* genome

Workflow steps	Tools
Long read basecalling	Dorado v1.0.2 – MinKNOW v25.04
Adapter sequences removal (long reads)	Porechop v0.2.4
Adapter sequences removal (short reads)	Fastp v0.23.4
Long reads quality and size filtering	Nanofilt v2.3.0
Contaminant screening	BLASTn nt online database
Contaminant filtering	Minimap2 v2.28
Long reads genome assembly	Flye v2.9.5
Genome polishing	Pilon v1.23
Genome completeness validation	BUSCO v5.8.1
Phylogeny	BUSCO v5.8.1, MAFFT v7.525, AMAS v.1.0, IQ-TREE v2.1.3
Repetitive regions identification	RepeatModeler v2.0.6
Repetitive regions masking	RepeatMasker v4.1.8
Protein coding gene prediction	MAKER v2.31.10, AUGUSTUS v3.3.2
Protein coding gene annotation	eggNOG-mapper v2.2.12, InterProScan, HHpred, HMMHER
CEPs identification and analysis	SignalP v5.0, DeepTMHMM v2.0, EffectorP 3.0, OrthoFinder v.2.5.5, MAFFT & Clustal 2.1
Duplicated genes identification	BLASTp, MCScanX

Results

DNA sequencing and genome assembly

The PromethION 2 sequencing process yielded a total of 46.85 M reads, representing 99.42 Gb with an approximate N50 of 3.32 kb. After filtering for the desired Phred score, read length and the removal of sequences from other species, 163 035 reads (read length N50 = 33.12 kb) were used for the long-read assembly, representing 10.76 Gb of data. For the Illumina paired-end sequencing, a total of 70.3 M read pairs were obtained, representing a total of 21.5 Gb of data. After all assembly processes, the final *G. ambrosiae* genome sequence was obtained with 33x coverage and a total size of 155.2 Mb, consisting of 73 contigs and 13 scaffolds. The assembly N50 is 3.5 Mb, and the largest sequence assembled is 12.0 Mb long. The BUSCO completeness analysis revealed that the assembly contains 97.5% ($n = 1665$) (96.4% single, $n = 1645$; 1,1% duplicated, $n = 19$) of the expected orthologs in the Ascomycota lineage (ascomycota_odb10), with a low proportion of fragmented (0.1%, $n = 1$) and missing genes (2.4%, $n = 41$). A summary of all assembly statistics is shown in Table 2.

Phylogeny

The phylogenetic relationship of *Golovinomyces ambrosiae* with other Ascomycetes is displayed in Fig. 1. Every node of the tree scored 100% bootstrap support, confirming excellent robustness for the grouping of each species. The species *G. ambrosiae* forms a monophyletic clade with other *Golovinomyces* species, and according to branch lengths, the tree shows that the newly assembled species is slightly less divergent from *G. magnicellulatus* compared to *G. cichoracearum*. This result supports the claim that *G. ambrosiae* is genetically distinct from *G. cichoracearum*. All species within the *Golovinomyces* genus share a most common ancestor with *Erysiphe* spp., which also form a monophyletic clade. Compared to all other powdery mildew species sampled, *Golovinomyces* spp. had the highest level of divergence with grass powdery mildews of the *Blumeria genus*. As for the dicotyledonous powdery mildew pathogens only, species within the *Podospharea* genus are the most distant from *Golovinomyces* species.

Transposable elements

The *G. ambrosiae* genome was largely composed of transposable elements (TE), with 82.64% repetitive regions. Retroelements dominate the TE content (57.62%), with Long Terminal Repeat (LTR) elements being the major class (29.34%), followed by Long Interspersed Nuclear Elements (LINEs) (28.28%). Among LTRs, Ty1/Copia and Ty3/mgd4/DIRS1 are the only two families present in this class. DNA transposons, including Tc1-IS630-Pogo, MULE-MuDR and hobo-Activator families make up 11.17% of the repetitive regions. A very small fraction accounted for Rolling-circles (0.02%), low complexity (0.02%) and simple repeats (0.48%). The remaining repeat content remained unclassified (13.32%). The complete relative T.E content and non-repetitive DNA of *G. ambrosiae* is illustrated in Fig. 2.

Identification of candidate effector proteins (CEPs), homology analysis and orthogroup classification

A total of 169 CEPs were identified in the *G. ambrosiae* genome (Additional File 2). Of those, 122 were predicted to be primarily cytoplasmic, while 29 are expected to be apoplastic. The remaining 18 candidates were classified as dual-localized, meaning they received a prediction for both localizations. The smallest sequence predicted was 29 amino acid long, while the longest had

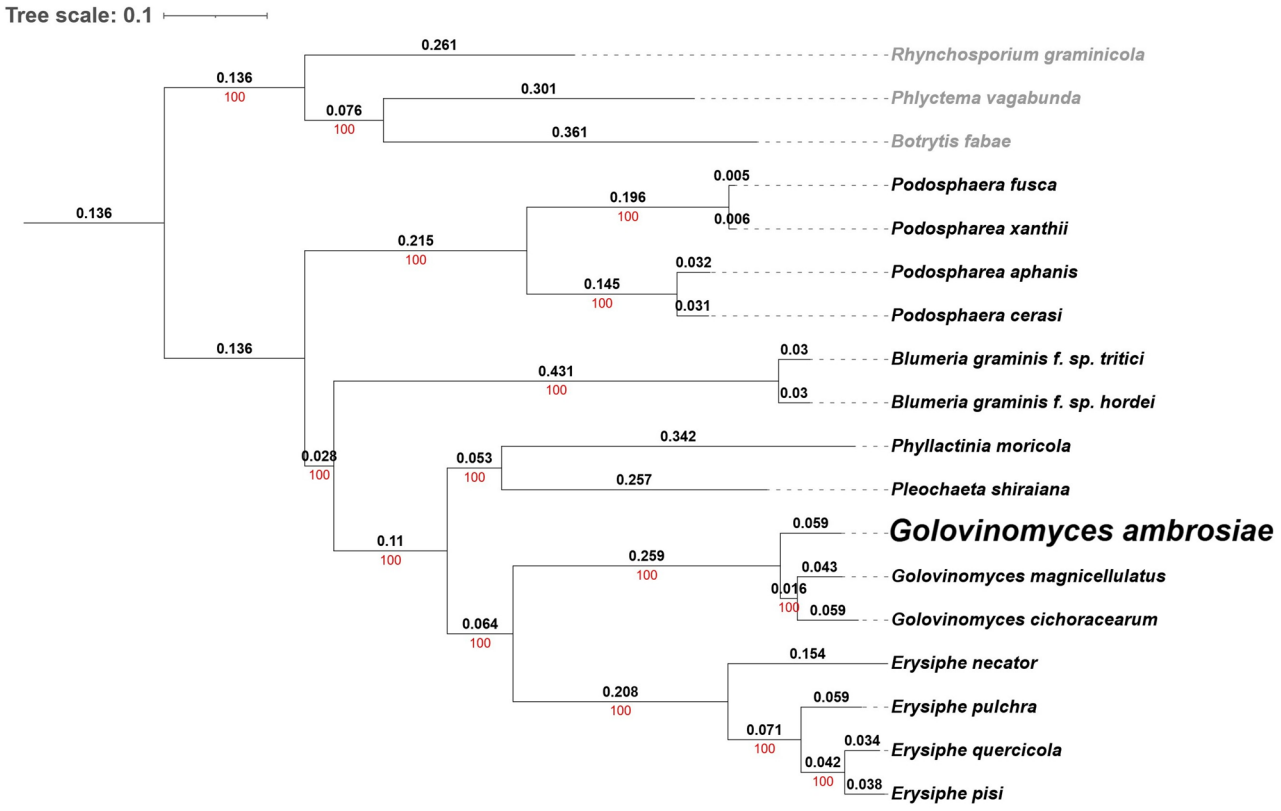

Fig. 1 Maximum-likelihood phylogenetic tree of 18 Ascomycetes including *Golovinomyces ambrosiae*. The phylogenetic tree illustrates the evolutionary relationship of *G. ambrosiae* with 17 other Ascomycetes. Species' accession numbers are displayed in Additional file 1. The genome of every species was processed with BUSCO v5.8.1 using the ascomycota_10odb database. A total of 1805 genes, present in at least 90% of all genomes were retrieved and aligned using MAFFT v7.525. Alignments were concatenated by AMAS 1.0 to generate a supermatrix. The resulting amino acid supermatrix, consisting of 19.32M sites for all taxa was used for the generation of the maximum-likelihood phylogenetic tree using IQ-TREE v2.1.3. Grey labeled species are part of the outgroup used to root the phylogenetic tree, composed of three Ascomycetes (*Rhynchosporium graminicola, Phlyctema vagabunda and Botrytis fabae*) that are not part of the Leotiomycetes. All remaining species are related to the powdery mildew disease. Bootstrap support is shown in red below branches. Branches lengths, representing the number of amino acid substitutions per site are displayed over branches

a length of 687. The average size of all predicted CEPs was 232 aa long, with a median of 194. Among all predicted CEPs, 60 harbored a Y/F/WxC motif in their protein sequence. The frequencies of the three possible combinations of the motif were 32% for the YxC, 65% for the FxC and 3% for the WxC. Analysis performed with sequences of predicted effectors upregulated in haustoria by *G. cichoracearum* revealed 23 homologues among the *G. ambrosiae* CEPs, with sequence identity ranging from 23.6% to 82.0% (Additional files 3). Almost half of all those homologous predicted effectors were found on contig 11. The search for RALPH-like predicted effector revealed 24 CEPs sharing similarities with ribonucleases. More than half of these predicted sequences were found to harbor the Y/F/WxC motif, and additionally, 14 of them (CEP11_33.146, CEP11_58.17, CEP11_59.30, CEP11_60.3, CEP11_61.149, CEP11_62.1, CEP11_63.133, CEP11_63.94, CEP11_64.126, CEP11_65.8, CEP119_0.92, CEP119_18.91, CEP145_1.67, CEP255_7.25) were also part of the homologues to haustoria secreted CEPs of *G. cichoracearum*. In all these CEPs, at least one Y/F/WxC motif is present in the predicted RNase domain location. Based on the HHpred outputs, the predicted RNase domain for these CEPs shares homology with ribonuclease T1 of *Aspergillus oryzae.* The protein architecture of those candidates is illustrated in Fig. 3. Upon analysis, the 14 RALPH candidates were grouped into six distinct orthogroups with other similar predicted effectors from different powdery mildew species. All orthogroups and the protein sequence similarity percentages between orthologues are displayed in Additional File 4. The analysis revealed that 10 RALPH candidates (CEP11_63.94, CEP11_64.126, CEP11_65.8, CEP11_62.1, CEP11_58.17, CEP11_60.3, CEP11_63.133,

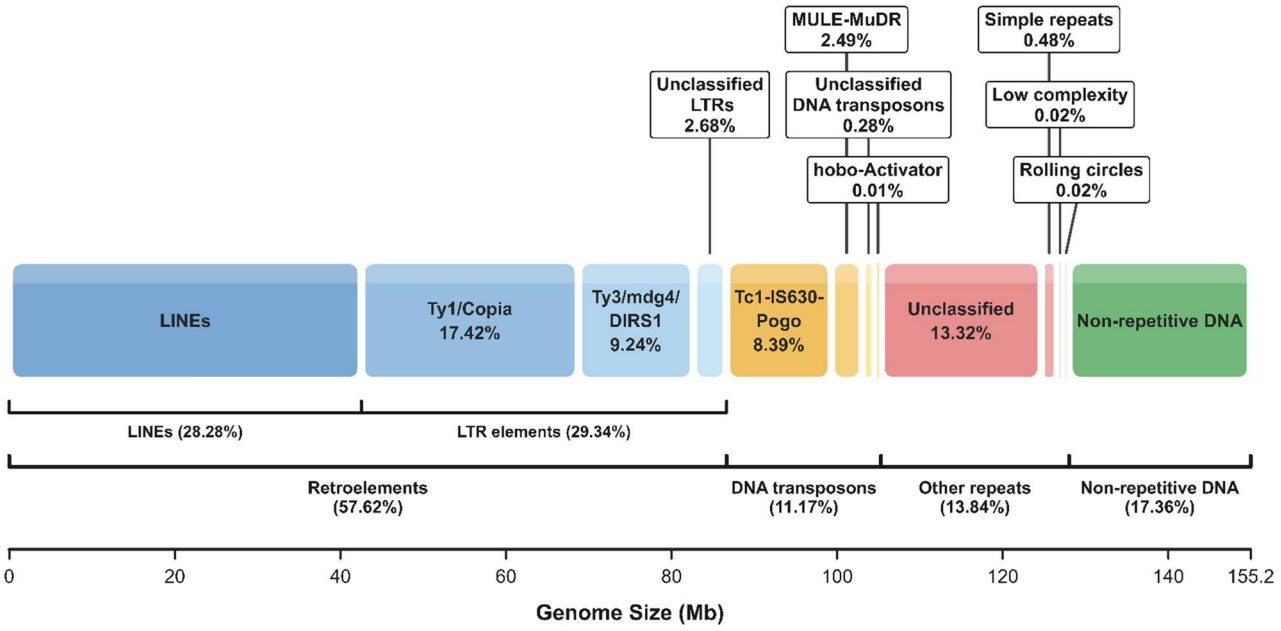

Fig. 2 Genome composition of *Golovinomyces ambrosiae*. Relative proportion of transposable elements (TE) and non-repetitive DNA within the *Golovinomyces ambrosiae* genome. Repetitive regions were predicted with RepeatModeler v2.0.6. Retroelements occupy more than half of the genome size (57.62%), divided almost equally between Long Interspersed Nuclear Elements (LINEs) and Long Terminal Repeat (LTR) elements. DNA transposons are composed of Tc1-IS630-Pogo, MULE-MuDR, hobo-Activator and a small fraction of unclassified elements. A significant proportion (13.32%) of repetitive elements remain unclassified. In total, TEs make up 82.64% of the genome size (155.2 Mb), representing 128.27 Mb

CEP11_33.146, CEP119_0.92, and CEP255_7.25) have orthologues across all selected powdery mildew species. One candidate (CEP119_18.91) has orthologues only within dicot PM species, while the remaining three (CEP145_1.67, CEP11_61.149, CEP11_59.30) only have orthologues with *G. cichoracearum*. Overall, predicted RALPH protein sequences of *G. ambrosiae* show very low similarity to orthologues from *B. graminis* f. sp. *tritici* and *B. graminis* f. sp. *hordei*, and higher similarity is observed with orthologues of dicot PM species. Protein sequence similarity with orthologues from *G. cichoracearum* ranges from 26.90% to 64.49%, indicating poor conservation even among species within the same genus.

Predicted protein coding genes and annotation

Based on the prediction analysis, *G. ambrosiae* harbors a total of 6995 protein coding genes. Among those, 2767 are part of the 3234 core Leotiomycetes genes (leotiomycetes_odb10). The functional annotation pipeline successfully characterized a total of 6021 predicted protein sequences by having an assigned eggNOG description (5477), GO terms (4692), KEGG pathways (1755) and/or Pfam domains (4710). Furthermore, InterProScan matched 5702 of the predicted protein sequences to known protein families, domains or conserved sites. The eggNOG-mapper tool additionally classified 3302 predicted protein sequences in Clusters of Orthologous Genes (COGs) functional categories. The category with the most abundant number of genes was Replication, recombination and repair, of which 76.72% were predicted to be implicated in transposable elements activity. This high number of genes in this category is consistent with the abundance of transposable elements composing the *G. ambrosiae* genome. Other categories with substantial number of genes are related to information storage and processing, cellular processes and signaling, and global metabolism. The 10 COG categories exhibiting the highest gene numbers are shown in Fig. 4. Overall, a total of 974 predicted genes (13.92%) remained unclassified. The BLAST search against the nonredundant protein database matched 6748 (95.38%) of the total predicted protein sequences to proteins from other powdery mildew species, with the main ones being *G. cichoracearum*, *G. magnicellulatus* and *Podosphaera aphanis*. All protein annotation results are displayed in Additional file 5.

Duplicated genes

Identification of duplicated genes offers relevant information about the pathogen's adaptation and evolutionary transformations. A total of 626 genes were found

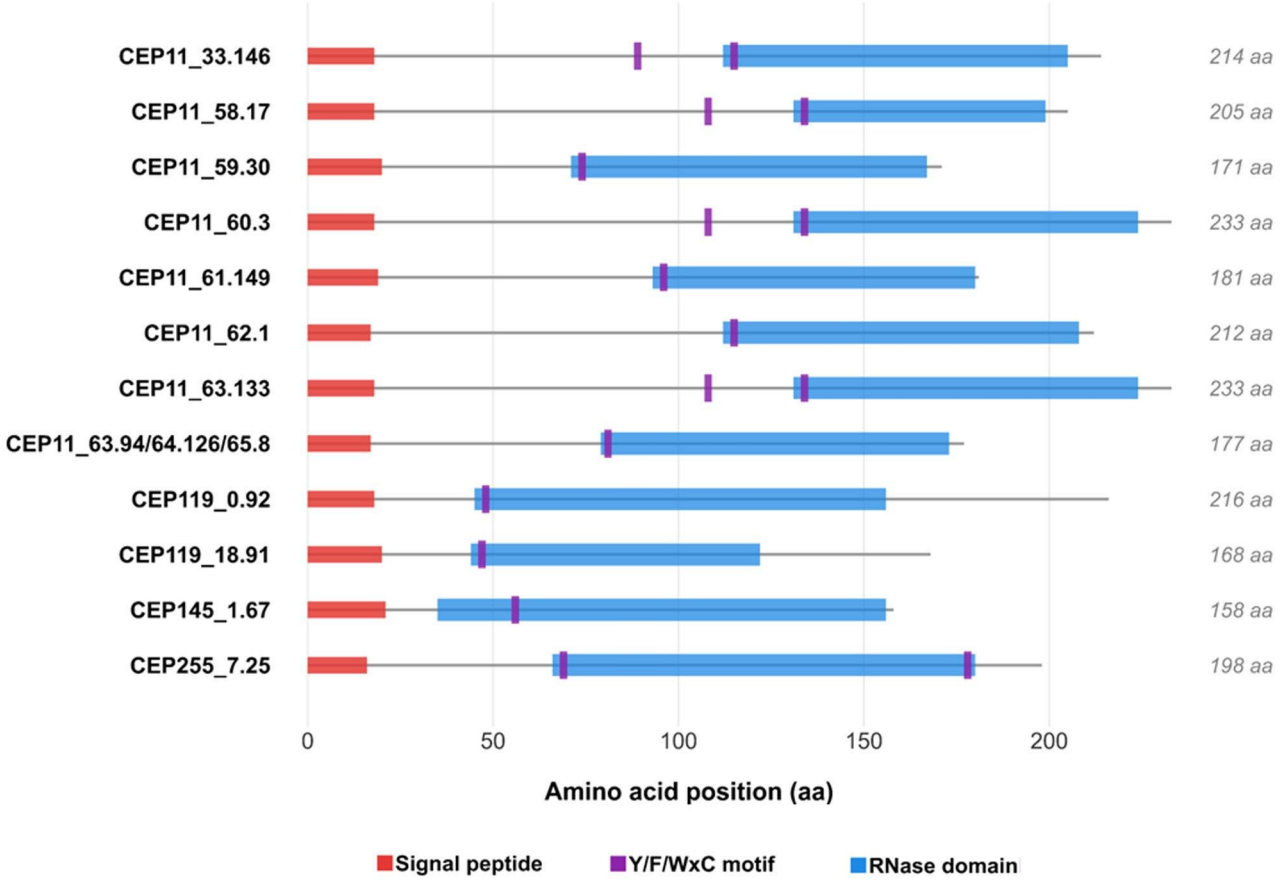

Fig. 3 Protein architecture of RALPH-like candidate for effector proteins (CEPs) of *Golovinomyces ambrosiae*. Among predicted (CEPs), 14 shared the presence of a signal peptide, an RNase domain homologous to GUANYL-SPECIFIC RIBONUCLEASE T1 (E-value > 1e-5) and at least one Y/F/WxC motif. Three CEPs (CEP11_63.94, CEP11_64.126, CEP11_65.8) show identical protein sequences. Every CEP shown is also homologous to candidate effectors secreted by haustoria of *Golovinomyces cichoracearum* isolates GcM1 and GcM3. Signal peptide position was predicted by SignalP 5.0. RNase domains were predicted using HHpred online servers (E-value > 1e-5)

to be duplicated within the *G. ambrosiae* genome, representing 8.95% of the whole predicted gene set. Most of them are dispersed (86.10%, n = 539) while a fraction is categorised as proximal (9.74%, n = 61) or tandem duplicated (4.15%, n = 26). Among the duplicated genes, 8.14% have a role in information storage and processing, cellular processes and signaling, and global metabolism. We also found many occurrences of duplicates harboring a ribonuclease domain, and genes required for fungal penetration, such as genes encoding proteins with the DUF3129 domain [58]. In total, 53.04% (n = 332) of the identified duplicates remain unclassified, having no match with a corresponding eggNOG description, GO term, KEGG pathway, Pfam domain or InterProScan ID.

Among CEPs, the proportions of duplicate classes were different from other protein coding genes. A total of 12.43% (n = 21) of CEPs showed duplication, with 52.38% (n = 11) being classified as proximal and the remaining 47.62% (n = 10) as dispersed. For proximal duplicates, there is a cluster on contig 11 of three CEPs (CEP11_63.94, CEP11_64.126, CEP11_65.8) with 100% sequence similarity. Still on the same contig, another CEP cluster showed four copies (CEP11_58.17, CEP11_33.146, CEP11_60.3, CEP11_63.133), of which two were proximal, and two dispersed. The predicted protein sequences of these four copies are 205 to 233 aa long and vary in similarity between 87.0% and 97.4%, possibly indicating a more ancient duplication. All these candidates are part of the RAPLH-like proteins, harbor the Y/F/WxC motif and are homologous to haustoria secreted candidate effector proteins of *G. cichoracearum* isolates GcM1 and GcM3. Other proximal duplications appear on contig 31, with four CEPs of near equal length and sequence similarity ranging from 80.25 to 95.12%.

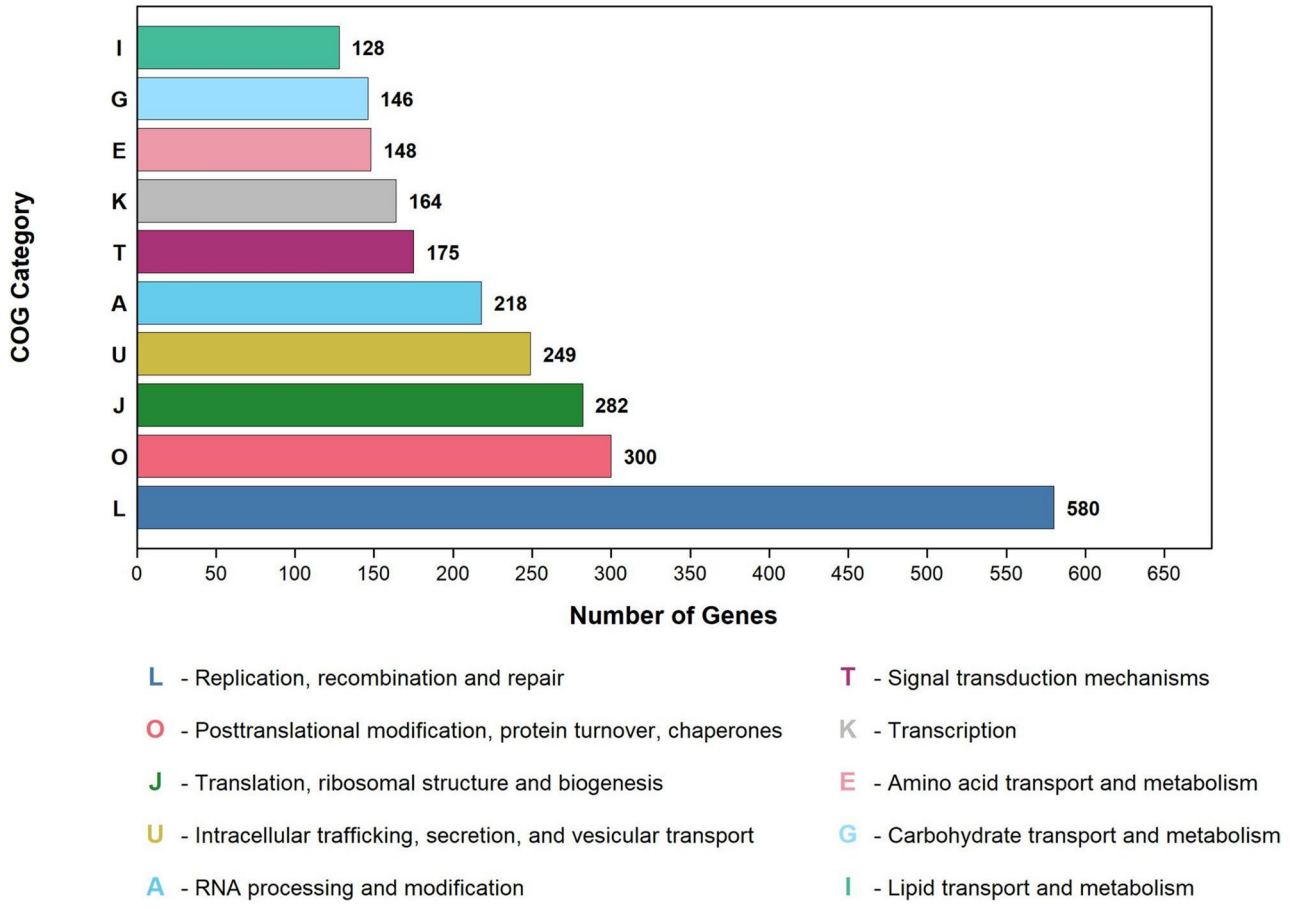

Fig. 4 Clusters of Orthologous Groups (COGs) functional categories with the highest gene numbers in *Golovinomyces ambrosiae*. The protein annotation performed by eggNOG-mapper v2.2.12 classified 3302 genes in various COG categories. Predicted protein-coding genes grouped in the Replication, recombination and repair category are the most represented among all groups. Other groups with the highest gene counts are related to information storage and processing [J, A, K], cellular processes and signaling [T, U, O] and global metabolism [G, E, I]

Table 2 Assembly statistics of the *Golovinomyces ambrosiae* genome

Assembly metric	Value
Genome size (bp)	155 192 055
Number of total fragments	86
Number of contigs	73
Number of scaffolds	13
Total of gap bases (bp)	842
Assembly N50 (bp)	3 548 631
Longest fragment (bp)	12 043 756
Repeat content (%)	82.64
GC content (%)	41.57
Protein coding genes	6995
CEPs	169

Discussion

The assembled genome presented in this study is among the most complete PM genomes available in terms of contiguity and provides the first genomic insights into the cannabis PM pathogen. *Golovinomyces ambrosiae* displays a very similar genome composition to other characterized powdery mildew species. Its size of 155.2 Mb is comparable to the 152.7 Mb of *Podosphaera xanthii* YZU573 [59], 129.9 Mb of *G. magnicellulatus* [60] and the predicted 173.8-221.8 Mb of *G. cichoracearum* [51]. The number of genes in the assembly is also consistent with other published PM genomes, ranging mostly from 6046 to 9372 [61–63]. Transcriptomic analyses would be a valuable complement to provide additional accuracy and information into the gene model predictions, gene expression, and functional genes that could be missing from in silico predictions. While it seems common for these genomes to be dominated by transposable elements, and specifically by LINEs and LTRs, *G. ambrosiae* was found to possess the highest repetitive elements content among all published dicot PM genomes to date. With no evidence from the annotations of crucial genes required for repeat-induced point mutation (RIP), a mechanism widely employed among fungi for the disabling of TE [61, 64], their absence clearly supports the impressive invasion of repetitive regions in this cannabis pathogen.

The abundance of TEs show remarkable genome plasticity potential in *G. ambrosiae*. In general, TEs can have negative impacts on genetic activities due to their capability of replicating in functionally essential DNA regions; on the other hand, they can also be drivers of beneficial adaptations and novel genes [65–67]. Their expansion can be accentuated by a lack of sexual recombination, which is typical of powdery mildew fungi that rely on asexual reproduction, a dominant condition in the context of indoor cannabis production. Pathogenicity may also be affected by the activity of TEs. Indeed, effector coding genes present in repeat-rich regions are more susceptible to accumulating mutations, which often improve fitness of the pathogen [65]. Congruent with the high TE content in the *G. ambrosiae* genome, numerous TE-related genes encoding proteins such as reverse transcriptase and transposase were identified and retained in the total gene set. We believe these genes could inform about the pathogen's genome dynamics. Further analyses of those genes would be interesting to validate if they are truly active and how they affect genomic composition. In consideration of the abundance of TEs composing the assembled genome, it is reasonable to suggest that there is likely substantial genetic diversity within the *G. ambrosiae* species, potentially leading to isolates with differential virulence profiles.

The genetic data regarding potential effector proteins provided by the assembly is also of great interest. Their abundance is in range with other identified CEPs of dicot PMs, particularly with the 159–175 found in *G. cichoracearum* [51]. Additionally, many candidate effectors of *G. ambrosiae* harbor the Y/F/WxC motif, which is a characteristic shared among powdery mildew effectors [68]. This motif has also been found to be an important structural element for RALPH proteins [69]. For CEPs involved in pathogenicity, the focus is on proteins that are secreted by the haustoria during pathogen-host interactions. The presence of a signal peptide, the Y/F/WxC motif and protein homology with haustoria secreted candidate effectors are important features that facilitate the filtration of the most promising candidates based on this characteristic. Additionally, since all known PM *Avrs* recognized by a corresponding host's *R* gene are linked to RALPH proteins [61], CEPs sharing sequence similarity with ribonucleases is another significant characteristic worthy of consideration. With identified sequences sharing either one or more of these key attributes, this study may provide a strong foundation for future identification of bona fide *Avr* genes. Among those, the 14 CEPs sharing every of the *Avr*-like features mentioned as depicted in Fig. 3 appear to be the most promising. These candidates should be carefully studied in future transcriptomic analyses to first confirm their expression. Functional experiments should follow to confirm their role in pathogenicity. Of these CEPs, three were identified as proximal duplicates sharing perfect sequence identity. If these predicted candidates were found to be bona fide effectors implicated in virulence, this particular strain of *G. ambrosiae* may have great fitness advantages by benefiting from an increased effector production [70]. In contrast, it could be a significant drawback if those sequences are related to avirulent phenotypes, requiring gain-of-virulence mutations in every copy to counteract resistance [71]. The other CEPs showing four non-identical copies are also interesting. If implicated in virulence, this could be a sign that these genes were subject to diversifying selection, potentially as a result of an arms race against the host [72].

Direct *Avr*-R protein interactions have been validated in the *Blumeria graminis* f. sp *hordei* (*Bgh*) – barley pathosystem [73–75]. In total, seven effector proteins of *Bgh* are recognized by their matching Mildew Locus a (MLA) gene, all part of the *Mla* locus, encoding multiple allelic NLRs. In all studies, the selection of top-ranking candidate effectors has been made through transcriptome-wide association studies (TWAS) and/or comparative genome analysis, with the goal of finding significant sequence polymorphisms matching the distinct phenotypes among numerous *Bgh* isolates. For all effectors, the avirulent and virulent genotypes were either differentiated by nonsynonymous SNPs, presence/absence of transcript, a transposon insertion in the gene sequence or splice site mutations leading to intron retention. The process of matching *Avr* genes to their corresponding resistance protein in the host has also been performed in the wheat-powdery mildew *Blumeria graminis* f. sp *tritici* pathosystem. In the host's genome, the *Pm3* gene is the most diverse resistance gene studied, with 17 functional allelic variants identified to date. To those, three matching *Avr* genes have been identified, and gain-of-virulence mutations have been attributed to the presence of a suppressor of avirulence gene (*Svr*) and/or amino-acid polymorphisms [20, 76]. Other *Avrs* associated with *Pm1a*, *Pm2*, *Pm17*, *Pm8 and Pm60* were also identified through QTL-mapping, genetic fine mapping, map-based cloning, genome-wide association studies, and avirulence depletion assay [71, 77–81]. In one of those studies, another gain-of-virulence mechanism such as complete deletion of the *Avr* gene has been highlighted [78]. These types of studies and their findings are critical for the comprehension of plant-pathogen interactions regarding virulence.

Genetic characterisation of how resistance can hold or break down within certain pathogen populations is necessary for enlightened breeding strategies. In the context of cannabis powdery mildew, it is essential to extend the genomic profiling of multiple *G. ambrosiae* strains, which will allow researchers to better understand the diversity of genetic determinants implicated in

virulence. Future work on the identification and cloning of the *G. ambrosiae Avrs* will allow functional validation against the currently identified *R* genes in the *C. sativa* germplasm. Subsequently, the unravelling of the distinct pathotypes will make it possible to confidently make use of the genetic resistance sources against powdery mildew. For wheat, it has been proposed to collectively create an interactive online *R*-gene atlas, which would provide information on exploited *R* genes and the corresponding *Avr* genotypes against which they provide protection [82]. Additionally, population survey data would be included to track changes in the effectorome of pathogens from different locations, informing on potential resistance breakdown threats. Since research on the cannabis powdery mildew resistance is still in its infancy, it would be of great interest to generate such data for each new resistance source discovered before findings get scattered. At least, for each newly discovered *R* gene in the host, equivalent efforts should be invested into deciphering the cognate *Avr* gene of the pathogen. In this context, this study sets a strong foundation for such future characterisations of host-pathogen interactions, enabling the most efficient approach of counteracting the disease.

Abbreviations

aa	Amino acid
Avr	Avirulence gene
Bgh	*Blumeria graminis* f. sp. *hordei*
BIC	Bayesian Information Criterion
BUSCO	Benchmarking Universal Single-copy Orthologs
CEPs	Candidate effector proteins
Chr01	Chromosome 1
COGs	Clusters of Orthologous genes
HMW	High molecular weight
HR	Hypersensitive reaction
ITOL	Interactive Tree of Life
MLA	Mildew Locus a
MLO	Mildew resistance locus o
ONT	Oxford Nanopore Technologies
PM	Powdery mildew
R genes	Resistance genes
NLRs	Nucleotide-binding site and/or leucine-rich repeat domains
QTL	Quantitative trait locus
RALPH	RNase-Like Proteins associated with Haustoria
RIP	Repeat-induced point mutation
S	Susceptibility
Svr	Suppressor of avirulence
TWAS	Transcription-wide association studies

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12995-2.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions

F.A.R. performed all bioinformatics workflows, analyses and wrote the initial draft and manuscript. Y.A. and B.C. provided guidance for the bioinformatics workflows and reviewed the manuscript. C.L. and F.A.R. performed the DNA extraction. R.R.B. wrote, reviewed and edited the manuscript. D.T. edited the manuscript. Funding acquisition and supervision was provided by R.R.B. and D.T. R.C.L. devised the Nanopore strategies in genome analysis and revised the manuscript. S.M., J.G. and G.Q.H.N. revised the manuscript, used the PromethION Solo 2 for WGS and bioinformatics analyses.

Funding
This research was supported by the Greenhouse Research Chair in Plant Protection-MAPAQ-Premier Tech and the program NSERC-Alliance awarded to RRB. RCL is funded by Genome Canada, Genome Quebec, CIHR and NFRF. FAR was supported by an NSERC Scholarship.

Data availability

The genome assembly, Illumina paired-end reads, and Oxford Nanopore PromethION 2 reads have been deposited in the NCBI databases under BioProject accession number PRJNA1438116. The assembly is accessible at DDBJ/ENA/GenBank under the accession number JBWDLV000000000. The Illumina paired-end reads and Oxford Nanopore PromethION 2 have been deposited in NCBI Sequence Read Archive (SRA) under accession numbers SRR37663630 and SRR37669284.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 18 March 2026 / Accepted: 25 May 2026
Published online: 28 May 2026

References

1. Belanger RR, Bushnell WR, Dik AJ, Carver TL. The Powdery Mildews: a Comprehensive Treatise. APS Press - The American Phytopathological Society, St. Paul, Minnesota, USA, 292 pp. ISBN 0-89054-291-0.
2. Park JH, Choi Y-J, Shin H-D. Revisiting Golovinomyces Species (Erysiphaceae) in Korea: Re-identification, New Records, and Description of Golovinomyces physalidis sp. nov. Mycobiology. 2025;53(4):450–65. https://doi.org/10.1080/12298093.2025.2517424.
3. Pépin N, Punja ZK, Joly DL. Occurrence of Powdery Mildew Caused by Golovinomyces cichoracearum sensu lato on Cannabis sativa in Canada. Plant Dis. 2018;102(12):2644. https://doi.org/10.1094/PDIS-04-18-0586-PDN.
4. Scott, Punja. Evaluation of disease management approaches for powdery mildew on Cannabis sativa L. (marijuana) plants. Can J Plant Pathol. 2021;43(3):394–412. https://doi.org/10.1080/07060661.2020.1836026.
5. Health Canada. (2025). *Good production practices guide for cannabis*. Government of Canada. http://epe.lac-bac.gc.ca/100/201/301/weekly_acquisitions_list-ef/2019/19–30/publications.gc.ca/collections/collection_2019/sc-hc/H14-271-2018-fra.pdf
6. Maymon M, Jerushalmi S, Freeman S. First report of Golovinomyces cichoracearum sensu lato on Cannabis sativa in Israel. New Disease Rep. 2020;42(1):11. https://doi.org/10.5197/j.2044-0588.2020.042.011.
7. Punja ZK, Collyer D, Scott C, Lung S, Holmes J, Sutton D. Pathogens and Molds Affecting Production and Quality of Cannabis sativa L. Front Plant Sci. 2019;10:1120. https://doi.org/10.3389/fpls.2019.01120.
8. Szarka D, Tymon L, Amsden B, Dixon E, Judy J, Gauthier N. First Report of Powdery Mildew Caused by Golovinomyces spadiceus on Industrial Hemp (Cannabis sativa) in Kentucky. Plant Dis. 2019;103(7):1773. https://doi.org/10.1094/PDIS-01-19-0049-PDN.
9. Farinas C, Hand FP. First Report of Golovinomyces spadiceus Causing Powdery Mildew on Industrial Hemp (Cannabis sativa) in Ohio. Plant Dis. 2020;104(10):2727.
10. Brochu A-S, Labbe C, Bélanger RR, Pérez-López E. First Report of Powdery Mildew Caused by Golovinomyces ambrosiae on Cannabis sativa L. (marijuana) in Quebec, Canada. Plant Dis. 2022. https://doi.org/10.1094/PDIS-02-22-0350-PDN.

11. Wiseman MS, Bates TA, Garfinkel AR, Ocamb CM, Gent DH. First Report of Powdery Mildew Caused by Golovinomyces ambrosiae on Cannabis sativa in Oregon. Plant Dis. 2021;105(9):2733.
12. Punja ZK. First report of the powdery mildew pathogen of hops, Podosphaeria macularis, naturally infecting cannabis (Cannabis sativa L., marijuana) plants under field conditions. Can J Plant Pathol. 2022;44(2):235–49. https://doi.org/10.1080/07060661.2021.1960424.
13. Salmon ES. A MONOGRAPH OF THE ERYSIPHACEAE. Mem Torrey Bot Club. 1900;9:1–292.
14. Qiu P-L, Liu S-Y, Bradshaw M, Rooney-Latham S, Takamatsu S, Bulgakov TS, Tang S-R, Feng J, Jin D-N, Aroge T, Li Y, Wang L-L, Braun U. Multi-locus phylogeny and taxonomy of an unresolved, heterogeneous species complex within the genus Golovinomyces (Ascomycota, Erysiphales), including G. ambrosiae, G. circumfusus and G. spadiceus. BMC Microbiol. 2020;20(1). https://doi.org/10.1186/s12866-020-01731-9.
15. Braun U, Cook RTA. Taxonomic Manual of the Erysiphales (Powdery Mildews), CBS-KNAW Fungal Biodiversity Centre.; 2012. ISBN: 978-90-70351-89-2.
16. Mieslerová B, Kitner M, Šenkyřík JB, Křivánková T, Sedlářová M, Braun U, Lebeda A. The co-occurrence of two powdery mildew species on Cannabis sativa in the Czech Republic. J Phytopathol. 2024;172(3). https://doi.org/10.1111/jph.13348.
17. Buirs L, Punja ZK. Integrated Management of Pathogens and Microbes in *Cannabis sativa* L. (Cannabis) under Greenhouse Conditions. Plants. 2024;13(6):786. https://doi.org/10.3390/plants13060786.
18. Akinrinlola RJ, Hansen ZR. Efficacy of Organic Fungicides Against Hemp Powdery Mildew Caused by Golovinomyces ambrosiae in a Greenhouse in Tennessee. Plant Disease. 2023;107(6):1867–1873. https://doi.org/10.1094/PDIS-09-22-2094-RE.
19. Sirangelo TM. NLR- and mlo-Based Resistance Mechanisms against Powdery Mildew in Cannabis sativa. Plants (Basel Switzerland). 2023;13(1). https://doi.org/10.3390/plants13010105.
20. Bourras S, Kunz L, Xue M, Praz CR, Müller MC, Kälin C, Schläfli M, Ackermann P, Flückiger S, Parlange F, Menardo F, Schaefer LK, Ben-David R, Roffler S, Oberhaensli S, Widrig V, Lindner S, Isaksson J, Wicker T, et al. The AvrPm3-Pm3 effector-NLR interactions control both race-specific resistance and host-specificity of cereal mildews on wheat. Nat Commun. 2019;10(1):2292. https://doi.org/10.1038/s41467-019-10274-1.
21. Sun S, Fu H, Wang Z, Duan C, Zong X, Zhu Z, Singh J. Discovery of a Novel er1 Allele Conferring Powdery Mildew Resistance in Chinese Pea (Pisum sativum L.) Landraces. PLoS ONE. 2016;11(1). https://doi.org/10.1371/journal.pone.0147624.
22. Massonnet M, Riaz S, Pap D, Figueroa-Balderas R, Andrew Walker M, Cantu D. The grape powdery mildew resistance loci Ren2, Ren3, Ren4D, Ren4U, Run1, Run1.2b, Run2.1, and Run2.2 activate different transcriptional responses to Erysiphe necator. Front Plant Sci. 2022;13. https://doi.org/10.3389/fpls.2022.1096862.
23. Jørgensen IH. Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. Euphytica: Int J Plant Breed. 1992;63(1–2):141–52. https://doi.org/10.1007/BF00023919.
24. Mihalyov PD, Garfinkel AR. Discovery and Genetic Mapping of PM1, a Powdery Mildew Resistance Gene in Cannabis sativa L. Front Agron. 2021. https://doi.org/10.3389/fagro.2021.720215.
25. Pépin N, Hebert FO, Joly DL. Genome-Wide Characterization of the MLO Gene Family in Cannabis sativa Reveals Two Genes as Strong Candidates for Powdery Mildew Susceptibility. Front Plant Sci. 2021;12:729261. https://doi.org/10.3389/fpls.2021.729261.
26. Stack GM, Cala AR, Quade MA, Toth JA, Monserrate LA, Wilkerson DG, Carlson CH, Mamerto A, Michael TP, Crawford S, Smart CD, Smart LB. Genetic Mapping, Identification, and Characterization of a Candidate Susceptibility Gene for Powdery Mildew in Cannabis sativa. Mol Plant-Microbe Interactions: MPMI. 2024;37(1):51–61. https://doi.org/10.1094/MPMI-04-23-0043-R.
27. Seifi S, Leckie KM, Giles I, O'Brien T, MacKenzie JO, Todesco M, Rieseberg LH, Baute GJ, Celedon JM. Mapping and characterization of a novel powdery mildew resistance locus (PM2) in Cannabis sativa L. Front Plant Sci. 2025;16:1543229. https://doi.org/10.3389/fpls.2025.1543229.
28. Kusch S, Frantzeskakis L, Lassen BD, Kümmel F, Pesch L, Barsoum M, Walden KD, Panstruga R. A fungal plant pathogen overcomes mlo-mediated broad-spectrum disease resistance by rapid gene loss. New Phytol. 2024;244(3):962–79. https://doi.org/10.1111/nph.20063.
29. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol. 2015;66:513–45. https://doi.org/10.1146/annurev-arplant-043014-114623.
30. Li L, Collier B, Pietro Spanu. Isolation of Powdery Mildew Haustoria from Infected Barley. Bio-Protocol. 2019;9(14). https://doi.org/10.21769/BioProtoc.3299.
31. Mayjonade B, Gouzy J, Donnadieu C, Pouilly N, Marande W, Callot C, Langlade N, Muños S. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. Biotechniques. 2016;61(4):203–5.
32. Russo A, Mayjonade B, Frei D, Potente G, Kellenberger RT, Frachon L, Copetti D, Studer B, Frey JE, Grossniklaus U, Schlüter PM. Low-Input High-Molecular-Weight DNA Extraction for Long-Read Sequencing From Plants of Diverse Families. Front Plant Sci. 2022;13:883897. https://doi.org/10.3389/fpls.2022.883897.
33. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: Visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666–9. https://doi.org/10.1093/bioinformatics/bty149.
34. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100. https://doi.org/10.1093/bioinformatics/bty191.
35. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6. https://doi.org/10.1038/s41587-019-0072-8.
36. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2023;2(2):e107. https://doi.org/10.1002/imt2.107.
37. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963. https://doi.org/10.1371/journal.pone.0112963.
38. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2. https://doi.org/10.1093/bioinformatics/btv351.
39. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4:e1660. https://doi.org/10.7717/peerj.1660.
40. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74. https://doi.org/10.1093/molbev/msu300.
41. Letunic and Bork. Nucleic Acids Res. 2024. https://doi.org/10.1093/nar/gkae268.
42. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020;117(17):9451–7. https://doi.org/10.1073/pnas.1921046117.
43. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protocols Bioinf. 2004;4:Unit410.
44. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491. https://doi.org/10.1186/1471-2105-12-491.
45. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinf (Oxford England). 2011;27(6):757–63. https://doi.org/10.1093/bioinformatics/btr010.
46. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–3. https://doi.org/10.1038/s41587-019-0036-z.
47. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros A, Marcatili JJ, Nielsen P, Krogh H, A., Winther O. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022. https://doi.org/10.1101/2022.04.08.487609.
48. Sonah H, Deshmukh RK, Bélanger RR. Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges. Front Plant Sci. 2016;7:126. https://doi.org/10.3389/fpls.2016.00126.
49. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. Molecular Plant-Microbe Interactions : MPMI. 2022;35(2):146–156. https://doi.org/10.1094/MPMI-08-21-0201-R.
50. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026–8.
51. Wu Y, Ma X, Pan Z, Kale SD, Song Y, King H, Zhang Q, Presley C, Deng X, Wei CI, Xiao S. Comparative genome analyses reveal sequence features reflecting distinct modes of host-adaptation between dicot and monocot powdery mildew. BMC Genomics. 2018;19(1):705. https://doi.org/10.1186/s12864-018-5069-z.
52. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. https://doi.org/10.1186/s13059-019-1832-y.
53. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S. The EMBL-EBI Job Dispatcher sequence analysis tools framework

in 2024. Nucleic Acids Res. 2024;52(W1):W521–W525. https://doi.org/10.1093/nar/gkae241. PMID: 38597606; PMCID: PMC11223882.
54. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825–9. https://doi.org/10.1093/molbev/msab293.
55. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40. https://doi.org/10.1093/bioinformatics/btu03.
56. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol. 2018;430(15):2237–43. https://doi.org/10.1016/j.jmb.2017.12.007.
57. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. https://doi.org/10.1093/nar/gkr1293.
58. Wei Huang S, Hong G, Tang Y, Lu, Chengshu Wang. Unveiling the function and regulation control of the DUF3129 family proteins in fungal infection of hosts. Philosophical Trans Royal Soc B: Biol Sci. 2019;374(1767):1–9. https://doi.org/10.1098/rstb.2018.0321.
59. Wang Z, Du Y, Li S, Xu X, Chen X. A Complete Genome Sequence of *Podosphaera xanthii* Isolate YZU573, the Causal Agent of Powdery Mildew Isolated from Cucumber in China. Pathogens (Basel Switzerland). 2023;12(4):561. https://doi.org/10.3390/pathogens12040561.
60. Farinas C, Gluck-Thaler E, Slot JC, Peduto Hand F. Whole-Genome Sequence of the Phlox Powdery Mildew Pathogen Golovinomyces magnicellulatus Strain FPH2017-1. Microbiol resource announcements. 2019;8(36):e00852–19. https://doi.org/10.1128/MRA.00852-19.
61. Kusch S, Qian J, Loos A, Kümmel F, Spanu PD, Panstruga R. Long-term and rapid evolution in powdery mildew fungi. Mol Ecol. 2024;33(10):e16909. https://doi.org/10.1111/mec.16909.
62. Frantzeskakis L, Németh MZ, Barsoum M, Kusch S, Kiss L, Takamatsu S, Panstruga R. The *Parauncinula polyspora* Draft Genome Provides Insights into Patterns of Gene Erosion and Genome Expansion in Powdery Mildew Fungi. mBio. 2019;10(5):e01692–19. https://doi.org/10.1128/mBio.01692-19.
63. Gañán L, White RA 3rd, Friesen ML, Peever TL, Amiri A. A Genome Resource for the Apple Powdery Mildew Pathogen *Podosphaera leucotricha*. Phytopathology. 2020;110(11):1756–8. https://doi.org/10.1094/PHYTO-05-20-0158-A.
64. Spanu PD et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science (New York, N.Y.)*, *330*(6010), 1543–1546. https://doi.org/10.1126/science.1194573.
65. Möller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens. Nat Rev Microbiol. 2017;15(12):756–71. https://doi.org/10.1038/nrmicro.2017.76.
66. Nottensteiner M, Zechmann B, McCollum C, Hückelhoven R. A barley powdery mildew fungus non-autonomous retrotransposon encodes a peptide that supports penetration success on barley. J Exp Bot. 2018;69(15):3745–58. https://doi.org/10.1093/jxb/ery174.
67. Sabelleck B, Panstruga R. Novel jack-in-the-box effector of the barley powdery mildew pathogen? J Exp Bot. 2018;69(15):3511–4. https://doi.org/10.1093/jxb/ery192.
68. Godfrey D, Böhlenius H, Pedersen C, Zhang Z, Emmersen J, Thordal-Christensen H. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. BMC Genomics. 2010;11:317. https://doi.org/10.1186/1471-2164-11-317.
69. Cao Y, Kümmel F, Logemann E, Gebauer JM, Lawson AW, Yu D, Uthoff M, Keller B, Jirschitzka J, Baumann U, Tsuda K, Chai J, Schulze-Lefert P. Structural polymorphisms within a common powdery mildew effector scaffold as a driver of coevolution with cereal immune receptors. Proc Natl Acad Sci USA. 2023;120(32):e2307604120. https://doi.org/10.1073/pnas.2307604120.
70. Pedersen C, Loren V, van Themaat E, McGuffin LJ, Abbott JC, Burgis TA, Barton G, Bindschedler LV, Lu X, Maekawa T, Wessling R, Cramer R, Thordal-Christensen H, Panstruga R, Spanu PD. Structure and evolution of barley powdery mildew effector candidates. BMC Genomics. 2012;13:694. https://doi.org/10.1186/1471-2164-13-694.
71. Müller MC, Kunz L, Schudel S, Lawson AW, Kammerecker S, Isaksson J, Wyler M, Graf J, Sotiropoulos AG, Praz CR, Manser B, Wicker T, Bourras S, Keller B. Ancient variation of the *AvrPm17* gene in powdery mildew limits the effectiveness of the introgressed rye *Pm17* resistance gene in wheat. Proc Natl Acad Sci USA. 2022;119(30). https://doi.org/10.1073/pnas.2108808119. Article e2108808119.
72. Liu L, Xu L, Jia Q, Pan R, Oelmüller R, Zhang W, Wu C. Arms race: diverse effector proteins with conserved motifs. Plant Signal Behav. 2019;14(2):1557008. https://doi.org/10.1080/15592324.2018.1557008.
73. Lu X, Kracher B, Saur IML, Bauer S, Ellwood SR, Wise R, Yaeno T, Maekawa T, Schulze-Lefert P. Allelic barley MLA immune receptors recognize sequence-unrelated avirulence effectors of the powdery mildew pathogen. Proc Natl Acad Sci USA. 2016;113(36):E6486–95. https://doi.org/10.1073/pnas.1612947113.
74. Saur IML, Bauer S, Kracher B, Lu X, Franzeskakis L, Müller MC, Sabelleck B, Kümmel F, Panstruga R. Multiple pairs of allelic MLA immune receptor-powdery mildew AVRA effectors argue for a direct recognition mechanism. eLife. 2019;8. https://doi.org/10.7554/eLife.44471. Article e44471.
75. Bauer S, Yu D, Lawson AW, Saur IML, Frantzeskakis L, Kracher B, Logemann E, Chai J, Maekawa T, Schulze-Lefert P. The leucine-rich repeats in allelic barley MLA immune receptors define specificity towards sequence-unrelated powdery mildew avirulence effectors with a predicted common RNase-like fold. PLoS Pathog. 2021;17(1). https://doi.org/10.1371/journal.ppat.1009223. Article e1009223.
76. Bourras S, McNally KE, Ben-David R, Parlange F, Roffler S, Praz CR, Oberhaensli S, Menardo F, Stirnweis D, Frenkel Z, Schaefer LK, Flückiger S, Treier G, Herren G, Korol AB, Wicker T, Keller B. Multiple avirulence loci and allele-specific effector recognition control the *Pm3* race-specific resistance of wheat to powdery mildew. Plant Cell. 2015;27(10):2991–3012. https://doi.org/10.1105/tpc.15.00171.
77. Hewitt T, Müller MC, Molnár I, Mascher M, Holušová K, Šimková H, Kunz L, Zhang J, Li J, Bhatt D, Sharma R, Schudel S, Yu G, Steuernagel B, Periyannan S, Wulff B, Ayliffe M, McIntosh R, Keller B, Lagudah E, Zhang P. A highly differentiated region of wheat chromosome 7AL encodes a Pm1a immune receptor that recognizes its corresponding AvrPm1a effector from Blumeria graminis. New Phytol. 2021;229:2812–26. https://doi.org/10.1111/nph.17075.
78. Praz CR, Bourras S, Zeng F, Sánchez-Martín J, Menardo F, Xue M, Yang L, Roffler S, Böni R, Herren G, McNally KE, Ben-David R, Parlange F, Oberhaensli S, Flückiger S, Schäfer LK, Wicker T, Yu D, Keller B. AvrPm2 encodes an RNase-like avirulence effector which is conserved in the two different specialized forms of wheat and rye powdery mildew fungus. New Phytol. 2017;213:1301–14. https://doi.org/10.1111/nph.14372.
79. Manser B, Koller T, Praz CR, Roulin AC, Zbinden H, Arora S, Steuernagel B, Wulff BBH, Keller B, Sánchez-Martín J. Identification of specificity-defining amino acids of the wheat immune receptor Pm2 and powdery mildew effector AvrPm2. Plant J. 2021;106(4):993–1007. https://doi.org/10.1111/tpj.15214.
80. Kunz L, Sotiropoulos AG, Graf J, Razavi M, Keller B, Müller MC. The broad use of the Pm8 resistance gene in wheat resulted in hypermutation of the AvrPm8 gene in the powdery mildew pathogen. BMC Biol. 2023;21(1):29. https://doi.org/10.1186/s12915-023-01513-5.
81. Kunz L, Jigisha J, Menardo F, Sotiropoulos AG, Zbinden H, Zou S, Tang D, Hückelhoven R, Keller B, Müller MC. Avirulence depletion assay: Combining R gene-mediated selection with bulk sequencing for rapid avirulence gene identification in wheat powdery mildew. PLoS Pathog. 2025;21(1):e1012799. https://doi.org/10.1371/journal.ppat.1012799.
82. Hafeez AN, Arora S, Ghosh S, Gilbert D, Bowden RL, Wulff BBH. Creation and judicious application of a wheat resistance gene atlas. Mol Plant. 2021;14(7):1053–70. https://doi.org/10.1016/j.molp.2021.05.014.

Publisher's Note