

RESEARCH ARTICLE

Open Access



# Network analysis exposes core functions in major lifestyles of fungal and oomycete plant pathogens

Eswari PJ Pandaranayaka<sup>1</sup>, Omer Frenkel<sup>2</sup>, Yigal Elad<sup>2</sup>, Dov Prusky<sup>3</sup> and Arye Harel<sup>1\*</sup>

## Abstract

**Background:** Genomic studies demonstrate that components of virulence mechanisms in filamentous eukaryotic pathogens (FEPs, fungi and oomycetes) of plants are often highly conserved, or found in gene families that include secreted hydrolytic enzymes (e.g., cellulases and proteases) and secondary metabolites (e.g., toxins), central to the pathogenicity process. However, very few large-scale genomic comparisons have utilized complete proteomes from dozens of FEPs to reveal lifestyle-associated virulence mechanisms. Providing a powerful means for exploration, and the discovery of trends in large-scale datasets, network analysis has been used to identify core functions of the primordial cyanobacteria, and ancient evolutionary signatures in oxidoreductases.

**Results:** We used a sequence-similarity network to study components of virulence mechanisms of major pathogenic lifestyles (necrotroph (ic), N; biotroph (ic), B; hemibiotroph (ic), H) in complete pan-proteomes of 65 FEPs and 17 saprobes. Our comparative analysis highlights approximately 190 core functions found in 70% of the genomes of these pathogenic lifestyles. Core functions were found mainly in: transport (in H, N, B cores); carbohydrate metabolism, secondary metabolite synthesis, and protease (H and N cores); nucleic acid metabolism and signal transduction (B core); and amino acid metabolism (H core). Taken together, the necrotrophic core contains functions such as cell wall-associated degrading enzymes, toxin metabolism, and transport, which are likely to support their lifestyle of killing prior to feeding. The biotrophic stealth growth on living tissues is potentially controlled by a core of regulatory functions, such as: small G-protein family of GTPases, RNA modification, and cryptochrome-based light sensing. Regulatory mechanisms found in the hemibiotrophic core contain light- and CO<sub>2</sub>-sensing functions that could mediate important roles of this group, such as transition between lifestyles.

**Conclusions:** The selected set of enriched core functions identified in our work can facilitate future studies aimed at controlling FEPs. One interesting example would be to facilitate the identification of the pathogenic potential of samples analyzed by metagenomics. Finally, our analysis offers potential evolutionary scenarios, suggesting that an early-branching saprobe (identified in previous studies) has probably evolved a necrotrophic lifestyle as illustrated by the highest number of shared gene families between saprobes and necrotrophs.

**Keywords:** Fungus–plant interaction, Network, Core function, Plant pathogen, Necrotroph, Biotroph, Hemibiotroph, Virulence, Comparative genomics

\* Correspondence: [aryeharel@volcani.agri.gov.il](mailto:aryeharel@volcani.agri.gov.il)

<sup>1</sup>Department of Vegetable and Field Crops, Institute of Plant Sciences, Volcani Center, Agricultural Research Organization, Rishon LeZion, Israel  
Full list of author information is available at the end of the article



## Background

Filamentous eukaryotic pathogens (FEPs; i.e., fungi and oomycetes) of plants cause extensive losses in annual yields of staple crops worldwide [1, 2]. The danger posed by these pathogens is enhanced by accelerated pathogen evolution, mainly due to the continuous use of fungicides in monoculture practice, and human- or climate-dependent dispersal [2, 3]. Understanding the genetic basis of fungal and oomycete pathogenicity mechanisms may provide new avenues for the development of revamped disease-control strategies. Despite the increasing number of sequenced FEP genomes (e.g., the 1000 fungal genomes from the Joint Genome Institute (JGI) [4]), there are very few large-scale genomic comparisons that make use of complete proteomes from at least a few dozen FEP genomes, which could reveal novel and niche-specific virulence mechanisms (e.g., study of proteases in [5, 6], and cell wall-degrading enzymes in [7]).

Genomic studies have shown that components of virulence mechanisms in FEPs are often highly conserved, or found in gene families that are potentially generated due to their association with the high genomic plasticity found in many of these pathogens [8–15]. One example is the conserved signaling module in fungi and oomycetes (i.e., the phosphorylative regulation machinery), which is pivotal for sensing environmental cues, and for regulating infection-associated morphogenetic transitions in pathogens [13, 16–19]. Comparative genomic studies have also pinpointed the dispersal of conserved effector families and domains across FEP species: (i) LysM domain-containing effectors that sequester chitin oligosaccharides from host defense [20]; (ii) toxins (TOXB, TOX2, HC, and Nep1-like proteins) [21–23]; (iii) the RXLR sequence motif mediating host translocation in oomycete effectors [24]; (iv) CRN effectors, cell death-inducing oomycete effectors [8]; and (v) Hce2s effectors potentially involved in adaptation to stress [25]. In addition, the capacity to generate and coordinately secrete proteins and secondary metabolites is prevalent in these pathogens, and central to their pathogenicity process [26]. These secreted components include a large arsenal of hydrolytic enzymes (e.g., cellulases, pectinases, proteases, lipases), oxidoreductases [27–29], and metabolites (e.g., polyketides, terpenes, and nonribosomal peptide (NPS)) effectors, some of them diverse, and tailored to a specific host [21, 24, 30]. Despite their high diversity and host specificity, over half of the predicted effectors are part of gene families- in 3 studied species of *Pucciniomycetes* (51 to 68% of the effectors), 2 *Phytophthora* species (77% of the effectors), and 18 *Dothideomycetes* (79% of the total count of effectors from all 18 species) [9, 31–33]. The correlation of certain gene families to specific lifestyles has facilitated defining metabolic activity, and the pathogenicity mechanisms required for different ecological niches [9, 33].

Three major lifestyles are known in fungal and oomycete phytopathogens. The necrotrophic lifestyle (hereafter, N is used to refer to necrotrophs), which is characterized by killing of the host cell before feeding on its dead tissue, is involved in utilizing host-selective toxin effectors (e.g., ToxA, Tox1/2/4, and Nep1-like proteins) (in) directly interacting with a host-susceptibility gene product, and ultimately leading to cell death [21, 24, 34, 35]. One example in this category is the broad host range fungus *Botrytis cinerea*, capable of infecting over 1400 plant species (including 200 cultivars) [36]. The biotrophic lifestyle (B will refer to biotrophs), which is characterized by nutrition and growth on living tissue, requires avoidance of plant defense mechanisms while feeding on the host compounds. One example in this category is *Erysiphe necator*, known to cause powdery mildew in grapes [37]. The hemibiotrophic lifestyle (H will refer to hemibiotrophs) is characterized by an initial biotrophic infection mode, followed by a transition to the necrotrophic stage. One example in this category is the fungus *Colletotrichum gloeosporioides* which causes significant damage to subtropical and tropical fruit before and after harvest [38]. In contrast to the pathogenic lifestyles, the saprotrophic lifestyle (Sap) is characterized by nutrition and growth on organic matter or decaying tissue [39]. One example in this category, is the model filamentous fungus *Neurospora crassa* [40]. Processing of organic/decaying tissue is typically associated with extracellular enzymatic degradation and subsequent absorption of nutrients. A fundamental aspect of the plant-pathogen interaction is induction of plant defense as a result of recognition of often conserved pathogen-associated molecular patterns (PAMPs, e.g., glucans, and chitin) by pathogen recognition receptors (PRRs) [24, 41–45], which is often termed PAMP-triggered immunity (PTI). Pathogens secrete effectors which suppress this primary defense mechanism (i.e., the PTI) and allow them to infect plants [24, 41–45]. In turn, plants evolved to produce R proteins (mainly nucleotide binding–leucine-rich repeat (NB-LRR) receptors) which invoke the plant defense upon (in) direct recognition of pathogen effectors, termed effector-triggered immunity (ETI) [43–45]. The effectors participate in both the (hemi) biotrophic and necrotrophic virulence processes, and their activity is important for avoidance of plant defense mechanisms in biotrophs.

To the best of our knowledge, there are very few large-scale genomic comparisons that make use of complete proteomes from dozens of FEP genomes, to discover novel, and niche-specific virulence mechanisms (e.g., study of proteases in [5, 6], and cell wall-degrading enzymes in [7]). The power of such analyses was demonstrated in the study of 18 *Dothideomycetes* genomes with diverse lifestyles (3 Sap, 6 N, 2 B, and 7 H) compared to outgroup genomes. That study identified 3 K core gene families (comprised of 66 K genes) of *Dothideomycetes* having at

least one representative in each *Dothideomycetes* genome, containing 233 Pfam domains that were expanded in *Dothideomycetes* compared to the control. These core gene families also contained 69 Pfam domains that were expanded in *Dothideomycetes* pathogens vs. outgroup pathogens [33]. Empowered by diverse multiple genomes of *Dothideomycetes*, that analysis highlighted gene families potentially playing a role in necrotrophic, hemibiotrophic, and saprotrophic lifestyles, primarily within the *Dothideomycetes* class of Ascomycota. Following observation of the conserved pathogenicity mechanisms mentioned above, and common characteristics of the major pathogenic lifestyles, we hypothesize that it is feasible to deploy the power of comparative genomics analysis in a large set of FEPs to identify core functions of pathogenicity for those lifestyles, as partially demonstrated in the *Dothideomycetes* class.

Networks offer a fashionable methodology for studying large-scale multifaceted genomic and functional genomic data [46]. Enabling integration of metadata [47, 48], it can facilitate the correlation of genomic elements and pathways with diverse pathogenic lifestyles (e.g., (hemi) biotrophic, and necrotrophic). Supported by a mathematical background for analysis and validation of the results, it provides a powerful means for exploration, and the discovery of trends in large-scale datasets, such as multiple genomes of FEPs. In our current study, we used sequence-similarity network analysis [46, 48] encompassing available complete pan-proteomes of 82 fungi and oomycetes (18 B, 20 H, 22 N, and 17 Sap; Additional file 2: Table S1) to identify components of virulence mechanisms. Our comparative analysis highlights approximately 190 significantly enriched core functions found in 70% of the genomes of a pathogenic lifestyle (e.g., core necrotrophic functions are shared by 70% of the genomes in this lifestyle). This includes functions that are specifically enriched in one lifestyle, and functions that are shared between pathogenic lifestyles. We show that these core functions can assist in discriminating the different pathogenic lifestyles. Finally, empowered by network analysis, our study of shared families in the entire set of 82 genomes illustrates potential evolutionary routes between these lifestyles.

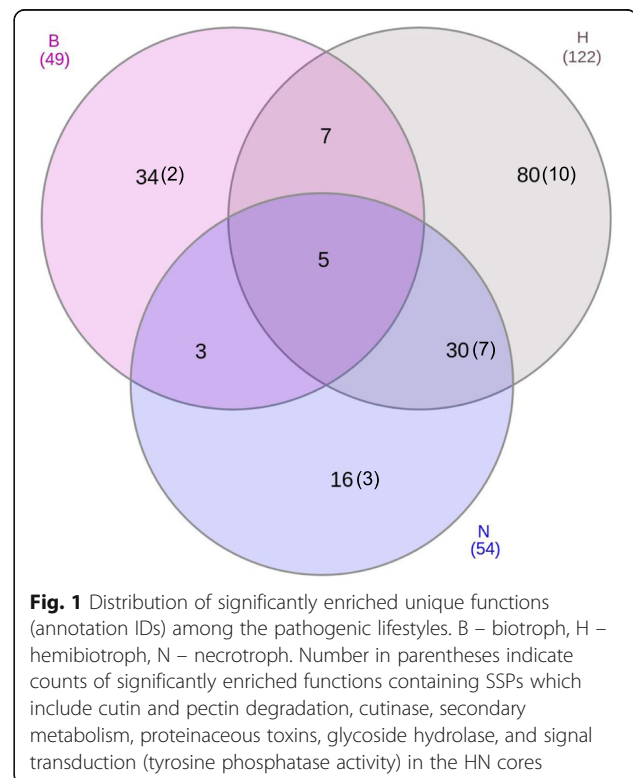
## Results

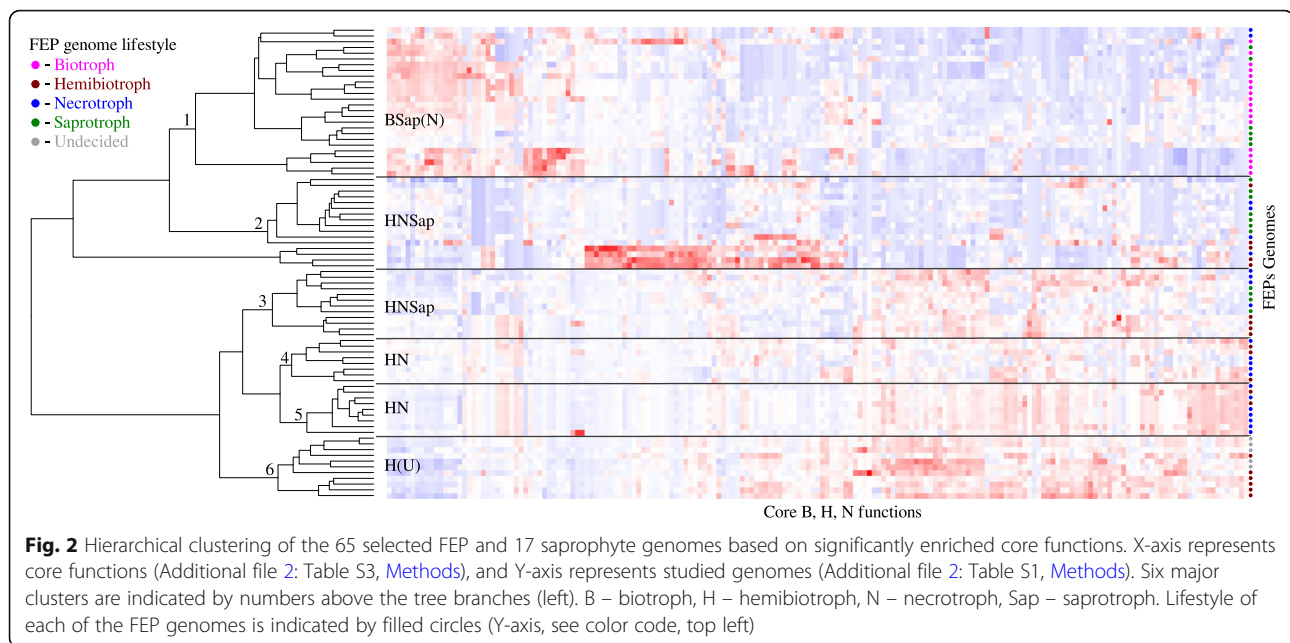
Our pan-proteome network consisted of approximately 3.9 K core gene families shared by at least 70% of a lifestyle (see section “Core components”, Methods). Approximately 40% of these core families were shared among all four lifestyles, i.e., core of all lifestyles (center of the Venn diagram, Additional file 1: Figure S1), and 25% were unique core families of only one lifestyle. The highest number of cores was found in H, followed by N, and B (Additional file 1: Figure S1). Hereafter, Ncore refers to the core of N (Bcore to core of B, Hcore to core of H, and Sapcore to core of Sap). Most of the proteins

(≥89%) in the core gene families were annotated for having either KEGG orthologs, InterPro domains or MEROPS proteases (Additional file 2: Table S2). Based on these annotations, we identified approximately 190 significantly enriched functions (see section “Calculation of enrichment and significance of core pathogenic functions”, Methods) in the core gene families of lifestyles H, B, and N (Fig. 1). All downstream analyses, unless otherwise specified, were based on these functions (often referred to as core functions). These core functions consisted of annotations which were enriched in only one lifestyle (i.e., B, H, or N), and annotations shared between several pathogenic lifestyles (Fig. 1). Around 4% of the core families did not contain proteins with the above-specified annotations, and only 6% of these unannotated families contained small secreted proteins (SSPs).

### Core gene families may assist in discriminating between the pathogenic lifestyles

To test whether the identified core functions can be used to differentiate between pathogenic lifestyles, we utilized hierarchical clustering (Fig. 2). The clustering analysis showed separation of B genomes from other pathogenic lifestyles (with the exception of 2 N genomes; see cluster 1 in Fig. 2). N and H genomes appeared in 5 clusters (clusters 2–6 in Fig. 2): clusters 2 and 3 also contained Sap, while clusters 4–6, which contained most (55%) of the N and H genomes, did not contain Sap.





Cluster 6 contained only H, along with all of the ambiguously characterized genomes (indicated by U, undecided, Fig. 2). The shared clustering of H and N corresponded with the highest number of shared functions within this lifestyle pair (HN column, Table 1).

### Core gene families of pathogenic lifestyles

In the analysis of core functions, most were found to belong to 7 abundant functional categories (bold in Table 1) which contained at least 10% of the annotations of a pathogenic lifestyle, and at least 5 significantly enriched annotations. These abundant functional categories included: transport in H, N, and Bcores; carbohydrate metabolism, secondary metabolite synthesis, and protease in H and Ncores; nucleic acid metabolism, and signal transduction in Bcores; and amino acid metabolism in Hcores. Other less abundant functional categories that contained significantly enriched annotations in B, H, and Ncores included trafficking, light-mediated functions, signal transduction, uncharacterized oxidoreductases, CO<sub>2</sub> sensing, and chaperones. In addition, we identified several significantly enriched uncharacterized domains or KEGG orthologs in the cores of each of the pathogenic lifestyles (designated as unknown in Table 1 and Additional file 2: Table S3). Some abundant functional categories (bold in Table 1) characterized only one pathogenic lifestyle (e.g., amino acid metabolism in Hcores), whereas others were abundant in more than one lifestyle (e.g., transport). Hereafter, functions shared by more than one lifestyle are referred to as lifestyle1lifestyle2cores; e.g., HNcores which contain functions enriched in both H and N cores.

### Core gene families shared between pathogenic lifestyles

HNcores contained significantly enriched functions abundant in (Table 1, and corresponding detailed annotations

in Additional file 2: Table S3): (i) carbohydrate metabolism related to cell wall-associated (i.e., including the cuticle) degradation and remodeling, such as pectinase, cutinase, and glycoside hydrolase family 28; (ii) secondary metabolite synthesis related mainly to toxins, and xenobiotic compound degradation and toxin synthesis; (iii) transport related to toxins and phospholipids; (iv) proteases related to serine peptidases of families 8–10, and metallopeptidase family M28. BHcores were significantly enriched in cryptochrome/photolyase-based DNA-repair functions, and in less abundant functions, such as transporters of glycerol, urea, and CO<sub>2</sub>; and glucanases (carbohydrate metabolism). Our analysis also identified a few functions that were significantly enriched in the cores of all three pathogenic lifestyles (BHNcores), such as members of serine peptidase family 8, and acyl-CoA oxidase participating in protein kinase A (PKA)-mediated beta lipid metabolism.

### Core gene families enriched in a specific pathogenic lifestyle

The network analysis also enabled the identification of abundant functional categories that contained functions enriched in the core of only one pathogenic lifestyle (Table 1, and corresponding detailed annotations in Additional file 2: Table S3).

Bcores – highly enriched functions were found mainly in: (i) nucleic acid metabolism; and (ii) signal transduction (GTPase, lysophospholipase, and tyrosine kinase activity). Less abundant specific Bcore-enriched functions included translation (t-RNA synthesis and ribosomal domain), and a KEGG ortholog with unknown function.

Hcores – highly enriched functions were found mainly in: (i) carbohydrate metabolism (certain glycoside hydrolase



**Table 1** Frequency of annotation IDs that are significantly enriched in core components of pathogenic lifestyles within selected functional categories (see section “Calculation of enrichment and significance of core pathogenic functions”, [Methods](#)). Number in parentheses indicates percentage of annotations in a lifestyle, e.g., there are 10 annotations related to nucleic acids in B which represent 29.4% of the annotations of B. Numbers in bold represent abundant functional categories. Detailed annotations are illustrated in Additional file 2: Table S3

Functional category/Core <sup>a</sup>	B (%) <sup>b</sup>	H (%)	N (%)	HN (%)	BH (%)	NB (%)	BHN (%)
<b>Amino acid metabolism</b>	2 (5.9)	<b>12 (15.0)</b>	1 (6.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
<b>Carbohydrate metabolism</b>	2 (5.9)	<b>22 (27.5)</b>	1 (6.3)	<b>11 (36.7)</b>	1 (14.3)	0 (0.0)	0 (0.0)
<b>Nucleic acid related</b>	<b>10 (29.4)</b>	4 (5.0)	0 (0.0)	2 (6.7)	4 (57.1)	1 (33.3)	0 (0.0)
<b>Protease</b>	3 (8.8)	5 (6.3)	3 (18.8)	<b>5 (16.7)</b>	0 (0.0)	1 (33.3)	4 (80.0)
<b>Secondary metabolites</b>	3 (8.8)	<b>19 (23.8)</b>	2 (12.5)	<b>7 (23.3)</b>	0 (0.0)	0 (0.0)	1 (20.0)
<b>Signal transduction</b>	<b>6 (17.7)</b>	3 (3.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
<b>Transporters</b>	<b>6 (17.7)</b>	<b>16 (20.0)</b>	<b>7 (43.8)</b>	<b>6 (20.0)</b>	2 (28.6)	0 (0.0)	0 (0.0)
Chaperone	0 (0.0)	1 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
CO <sub>2</sub> sensing	0 (0.0)	1 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Energy	1 (2.9)	1 (1.3)	0 (0.0)	1 (3.3)	0 (0.0)	0 (0.0)	0 (0.0)
Light sensing and light-responsive nucleic acid functions	1 (2.9)	1 (1.3)	0 (0.0)	0 (0.0)	4 (57.1)	0 (0.0)	0 (0.0)
Oxidoreductases	0 (0.0)	2 (2.5)	1 (6.3)	1 (3.3)	0 (0.0)	0 (0.0)	0 (0.0)
Symbiosis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	0 (0.0)
Trafficking	3 (8.8)	7 (8.8)	0 (0.0)	1 (3.3)	0 (0.0)	1 (33.3)	0 (0.0)
Translation	3 (8.8)	1 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Unknown	1 (2.9)	4 (5.0)	2 (12.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Vitamin	0 (0.0)	2 (2.5)	0 (0.0)	1 (3.3)	0 (0.0)	0 (0.0)	0 (0.0)
Total annotation ID count <sup>c</sup>	34	80	16	30	7	3	5

<sup>a</sup> B biotroph, H hemibiotroph, N necrotroph, HN hemibiotroph and necrotroph, BH biotroph and hemibiotroph, NB necrotroph and biotroph, BHN biotroph, hemibiotroph, and necrotroph

<sup>b</sup> Counts of annotations (e.g., KEGG ortholog ID) associated with a function

<sup>c</sup> One annotation ID is counted only once even if it occurs in multiple functions

families, glucanoyltransferase, lactate dehydrogenase, expansin, and fucosidase); and (ii) amino acid metabolism (Gly, Ser, and His metabolism). Less abundant enriched Hcore functions included chaperones, CO<sub>2</sub> sensing, rhodopsin-based light sensing, and 4 unknown function.

Ncores – highly enriched functions were found mainly in the transporters, and contained 2 domains with unknown functions. Less abundant annotations consisted of different proteases subfamilies in different pathogenic lifestyles (e.g., ubiquitin related-degradation in the Ncores).

#### Identification of SSPs in core functions of pathogenic lifestyles

Predicted SSPs were found in 14% of the significantly enriched core functions (indicated in parentheses in Fig. 1, and in SSP column of Additional file 2: Table S3). In line with their role in pathogen–host interactions, most of the SSP functions were related to cutin and pectin degradation, cutinase, secondary metabolism, proteinaceous toxins, glycoside hydrolase, and signal transduction (tyrosine phosphatase activity) in the HN cores. In addition, complete genomic analysis (regardless of the network) showed that H contain significantly (40%) more predicted

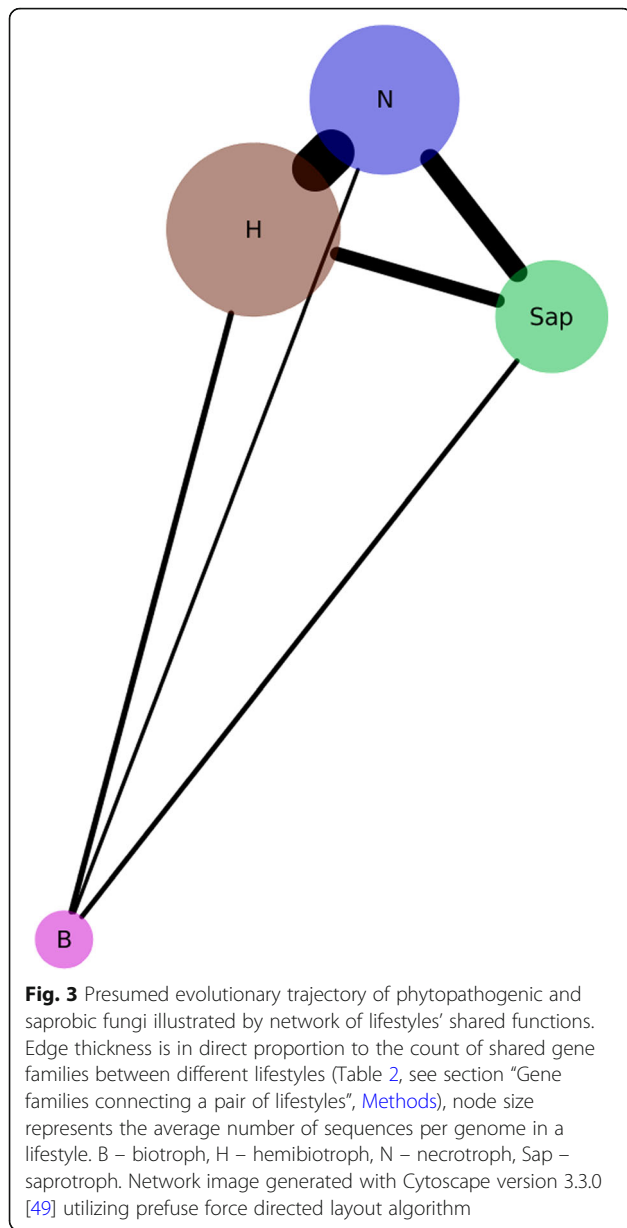
SSPs per genome than N, and pathogens have 66% more SSPs than saprophytes ( $P < 0.05$ , t-test).

#### Evolutionary trajectory of fungal pathogens

To study potential evolutionary trajectories of plant pathogens, we used a genomic approach to assess the number of gene families connecting a pair of lifestyles ([Methods](#)). This section encompassed all gene families (including core). Our results demonstrated (Fig. 3 and Table 2) a central place for N and H. Each of them shared the highest number of gene families with other groups. Accordingly, the highest number of gene families was shared between the N–H lifestyle pair, followed by N–Sap, H–Sap, H–B, and Sap–B.

#### Discussion

In this work, we focused on the core gene families that are predominant in the major lifestyles of filamentous fungal (and oomycete) plant pathogens. The network analysis used in our work illustrated that H has more significantly enriched core functions than N and B (in that order, Fig. 1). This is in line with the requirement of H to have both necrotrophic and biotrophic capabilities, in addition



**Table 2** Counts of gene families (components) connecting a pair of lifestyles (see Methods). Related to Fig. 3. Numbers in parentheses are mean values obtained from 10,000 random simulations for each lifestyle. All simulations were found significant ( $P < 0.0001$ , non-parametric rank test, see Methods)

Lifestyle	Biotrophs	Hemibiotrophs	Necrotrophs	Saprotrophs
Biotrophs		360 (256)	261 (376)	292 (229)
Hemibiotrophs	233 (164)		2027 (708)	570 (418)
Necrotrophs	144 (207)	2020 (598)		846 (542)
Saprotrophs	228 (119)	701 (327)	957 (491)	

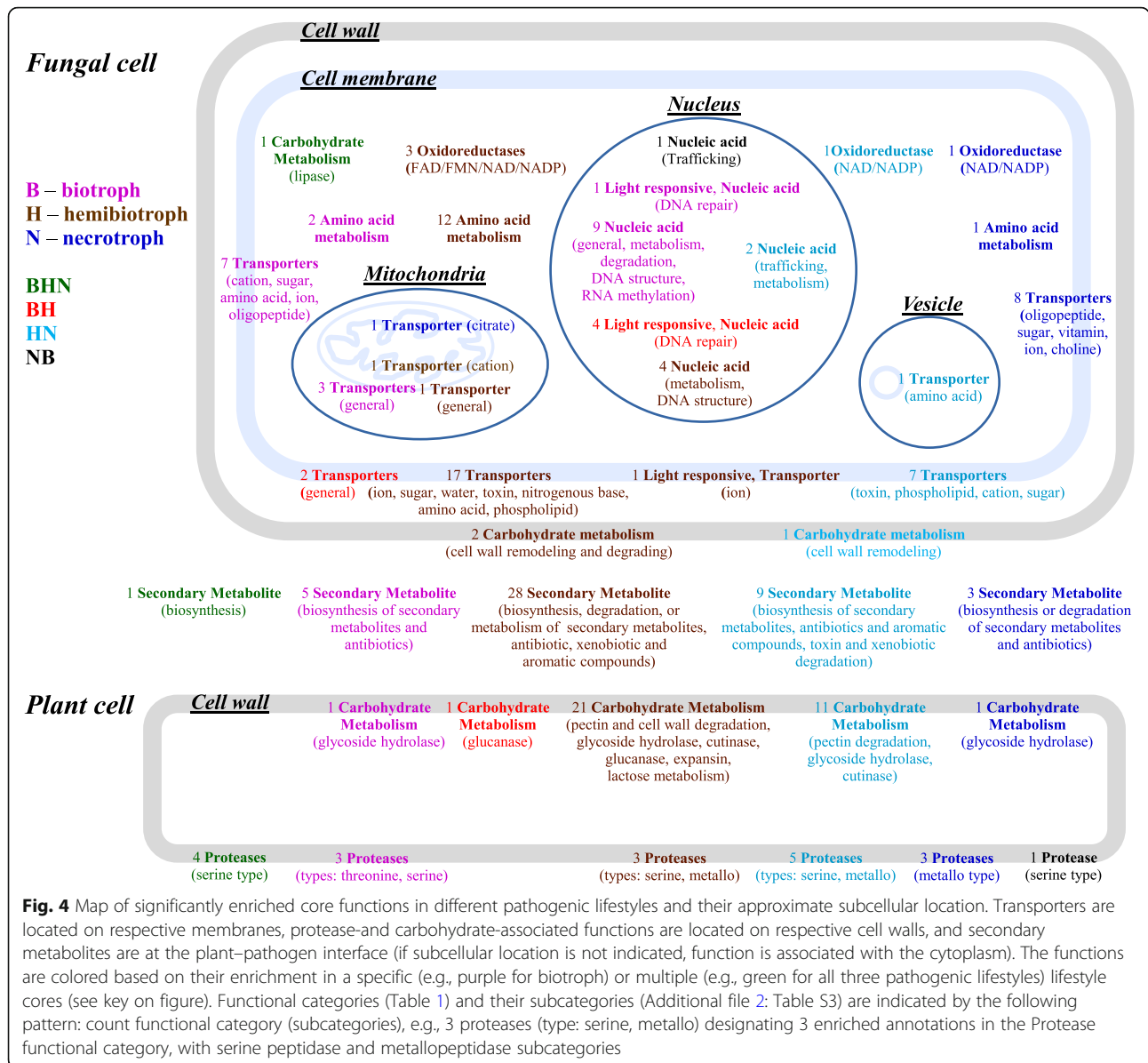
to functions associated with the transition between these lifestyles. This result is also in agreement with the higher number of SSPs per genome in H (regardless of core functions). The lowest number of biotrophic core functions can be explained by previous studies demonstrating that many functions which are required for virulence in this group have diversified, i.e., they are restricted to a specific taxonomic level or niche, and they are therefore not found in the core (see examples in [9, 22, 31, 33, 50]).

**Differentiating between lifestyles is empowered by core gene functions**

One potential use of the core functions identified in this work was demonstrated by hierarchical clustering (Fig. 2). This analysis enabled differentiating B (together with some Sap genomes) from other pathogenic groups, obtaining most of the N and H genomes in HN clusters (2 with and 2 without Sap genomes), and obtaining a separate cluster of H. A comparative genomics study of 18 *Dothideomycetes* species (4 Sap, 5 N, 7 H, and 2 B), illustrated that clustering of these genomes using annotations of all genomic: (i) carbohydrate activity enzymes (CAZymes), showed 2 major clusters of HNSap and BHSap lifestyles; (ii) proteases, yielded mainly a separate H cluster, and a mixed HN cluster; (iii) lipases, showed mainly 2 HNSap clusters (the latter 2 contained also 21 outgroup genomes within Ascomycota and Basidiomycota). Thus, all genomic annotations of these 3 enzyme classes (CAZyme, proteases, and lipases) enabled obtaining similar (or less differentiating) separation between lifestyles compared to the use of selected core functions in the current work. All of the genomes with ambiguously characterized lifestyle (i.e., referred to as both H and N in the literature) were clustered with H (cluster 6, Fig. 2). Unfortunately, most of the work in the related literature does not include a detailed characterization or description of these lifestyles in each pathogen. However, as both necrotrophic and hemibiotrophic lifestyles are illustrated for a fungal pathogen in those studies, it is more likely to be hemibiotrophic, as its biotrophic stage could be more elusive (short or only appearing under specific conditions) and not identified in all studies. The distribution of saprophytes in biotrophic and necrotrophic lifestyles is in line with some studies suggesting that early diverging fungi were saprotrophic (see discussion below).

**Mapping core functions in pathogenic lifestyles**

Our analysis provided a map of the core functions in the H, B, and N pathogenic lifestyles (Fig. 4, and Additional file 2: Table S3) derived from significantly enriched annotations in core gene families of these lifestyles.



### Functions enriched in all pathogenic lifestyles

Our analysis identified enrichment of members of serine protease family S8, and acyl-CoA oxidase in all three pathogenic lifestyles (indicated in green, BHN, Fig. 4). A previous computational study showed that the S8 serine proteases (subtilisin, identified in the BHNcore) are abundant across fungal lineages, and are highly correlated with pathogenic lifestyle in both animals and plants [5, 6]. A few studies illustrated the role of subtilisins (or subtilisin-like) in virulence, mediated mainly by cuticle degradation in fungal pathogens of insects (e.g., [51, 52]). Acyl-CoA oxidase (identified in the BHNcore) mediates the first step of beta oxidation which may be invoked by PKA, contributing to the pathogenicity process of phytopathogenic fungi [16]. The acetyl-CoA product of beta oxidation

could enter the citric acid cycle to produce energy; alternatively, it is known to participate in the formation of metabolites such as glycerol, melanin, and glucose (via gluconeogenesis), known to contribute to virulence processes such as appressorium-mediated plant infection, in phytopathogenic fungi [53–56].

### The necrotrophic lifestyle

This section refers to fungal pathogenic functions that were enriched in the Ncore or in both N and Hcores, the latter attributed to the necrotrophic stage of H (indicated in blue, N; and in light blue, HN; Fig. 4). Our analysis revealed that the Ncore is enriched in functions associated with cell wall-associated degrading enzymes (e.g., pectinase, cutinase, and glycoside hydrolase family

28), toxin metabolism, proteases, and transport. These functions are probably needed to support necrotrophic growth, involving maceration of the host cell barriers (e.g., cell wall), and induction of host cell death followed by sequestering of nutritional compounds (e.g., amino acids and carbohydrates). For example, comparative analysis of mostly necrotrophic *Botrytis* species highlighted multiple cell wall- (carbohydrate-) degrading enzymes such as pectinases [57]. Toxin synthesis and degradation were abundant in the HNcore functions; these are known to mediate plant cell death and protection against plant defense mechanisms in the necrotrophic process [21, 24, 34, 35]. Accordingly, toxin transport was found to be enriched in these cores, in agreement with the previously identified arsenal of toxins in necrotrophs that mediate killing of the host cell prior to feeding on it [21, 34, 35]. Import of other compounds enriched in N (e.g., phospholipid and choline) could further support nutrition of the pathogen during the course of infection. Serine and metalloproteases: Roles in nutrient acquisition, host cell degradation and host-fungus interactions (e.g., neutralization of defense mechanism) have been previously illustrated for proteases in general [58, 59], and for serine proteases in particular [5, 60]. A computational study of serine proteases (found in the MEROPS database) illustrated that families S9 and S10 are abundant in fungal genomes, partially supporting their identification in the HNcore in our study; however, no correlation with pathogenic lifestyles was found for these families [5]. Comparative analysis of mostly necrotrophic *Botrytis* species facilitated the identification of a clade of 8 species with shared proteases (1 serine-type peptidase, 1 hydrolase acting on glycosyl bonds, 1 asparaginase, and 1 G1 endopeptidase) [57]. Pathogen proteases can participate in inhibiting plant defense components such as pathogenesis-related proteins (e.g., antimicrobial chitinases), and  $\beta$ -1,3-glucanases which mediate fungal cell wall hydrolysis (e.g., [61–63]). Metalloprotease activity (identified as enriched in Ncores) has been previously correlated with fungal phytopathogenic activity, directed mainly against plant chitinase used for defense [59]. One example is the FoMep1 protease secreted by *Fusarium oxysporum* f. sp. *lycopersici* which (together with a serine protease) was responsible for the degradation of chitinases of tomato [64]. To the best of our knowledge, the role of metalloprotease family M28, identified as enriched in HNcores in the current study, in fungal virulence against plants has not been previously demonstrated.

### The biotrophic lifestyle

This section refers to fungal pathogenic functions enriched in Bcores or in both B and Hcores, as the latter are attributed to the biotrophic stage of H (indicated in pink, B; and in red, BH; Fig. 4). The abundance of enriched functions related to signal-transduction processes (e.g., GTPase,

lysophospholipase, and tyrosine phosphatase), and nucleotide metabolism (in specific DNA/RNA structures and recognition) could facilitate the tight regulation required for biotrophs to control their avoidance of plant defense mechanisms while feeding on the host compounds. The effect of Bcore functions of suppression of the plant defense system was demonstrated by a secreted tyrosine phosphatase of the bacteria *Pseudomonas syringae* that suppressed the immune responses of *Arabidopsis* by dephosphorylating a plant pattern recognition receptor [65]. Some of the small G-protein family of GTPases, such as Rac, Rho, and Rab, participate in regulating the mitogen-activated protein (MAP) kinase cascade in eukaryotes [66], which plays an important role in environmental sensing and consequent morphogenesis in phytopathogenic fungi [67, 68]. An example is the CDC42 Rho GTPase, which is involved in vegetative differentiation and is required for pathogenicity in the biotrophic wheat pathogen *Claviceps purpurea* [69]. A few functions associated with carbohydrate metabolism and secondary metabolism functions were enriched in the Bcore (Fig. 4, and Additional file 2: Table S3). This is in line with the comparative analysis of 4 downy mildew species and 3 *Phytophthora* species that also identified a few functions related to carbohydrates (such as pectin lyase and cutinase) and secondary metabolism (e.g., necrosis-inducing proteins) [70]. Comparative genomic analysis of various powdery mildew-causing pathogens also illustrated a reduced set of carbohydrate active enzymes devoted to plant cell wall depolymerization and secondary metabolites [12, 71]. Nucleotide metabolism: The potential role of RNA metabolism in the Bcore is supported by a recent study of the biotrophic obligate fungal pathogen *Plasmopara viticola*, which identified positive selective pressure (indicated by pairwise dN/dS values) in genes coding for RNA modification and processing [72]. One of these genes was the DEAD box helicase [72] (involved in transcription, splicing, and RNA transport), observed in the Bcore, which is known to regulate multiple virulence genes in the fungal pathogen of mammals, *Cryptococcus neoformans* [73]. Analysis of genes under positive selection in the biotroph *Plasmopara viticola* also highlighted genes associated with RNA metabolism, mRNA maturation and processing, or rRNA and tRNA modification, and DEAD/DEAH RNA helicase [72]. Specific histone residues are known to undergo posttranslational modification (mainly methylation and acetylation) [74], and therefore Bcore histone variants could affect histone modification, which might ultimately affect transcription and epigenetic-based regulation. The role of histone modification (i.e., methylation and acetylation) has been demonstrated in the pathogenicity process of several phytopathogenic fungi [75, 76]. For example, deletion of *gcn5* histone acetyltransferase in the biotrophic fungal pathogen *Ustilago maydis* significantly reduced the infection process on maize [77].



### Cryptochromes and photolyases in the Bcore

Cryptochromes are photoreceptors that are closely related to photolyases, but they do not necessarily exhibit DNA-repair functionality and may possess regulatory functions [78]. In the biotrophic fungal pathogen *Blumeria graminis* f. sp. *hordei*, UV-C irradiation inhibited conidial germination and appressorium formation (participating in host penetration), while upregulation of 3 putative photolyases was observed, suggesting their potential role in protection from UV-C [79]. Disruption of PHL1 (a cryptochrome/photolyase homolog) in the hemibiotrophic phytopathogenic fungus *Cercospora zea-maydis* inhibited light-dependent DNA repair (photoreactivation) activity, and exhibited reduced expression of another cryptochrome, and of genes involved in nucleotide excision and chromatin remodeling during DNA-damage repair [80, 81]. Although cryptochromes were not enriched in the Ncore, they are found in several N genomes. An interesting example is the necrotrophic fungal pathogen *B. cinerea*, where the cryptochrome BcCRY1 acts as the major photolyase in photoprotection, and the cryptochrome BcCRY2 participates in regulating photomorphogenesis (repression of conidiation) [82]. Although this function, may appear in a different path in biotrophs, it could play an important role in fungal plant pathogens. These findings, together with their position in the Bcores, suggest that cryptochromes mediated photoprotection, and photomorphogenesis could play a central role in the biotrophic lifestyle.

### The hemibiotrophic lifestyle

One intriguing role for functions enriched only in the Hcores (indicated in brown, Fig. 4) might be participation in the shift between lifestyles. Degradation of lignocellulose compounds: While some GH families identified in the Hcore are active on a narrow range of substrates (e.g., xylanase for GH12 and galactanase for GH53), others (e.g., GH 1, 3, and 11) have diverse activities [83, 84]. Comparative analysis of plant cell wall-degrading enzymes in fungal genomes also showed that the GH3 family is significantly more abundant in hemibiotrophs (and in necrotrophs) than in biotrophs [7]. Lactate dehydrogenase (identified in the Hcore) may support pyruvate production, during infection, from plant-based lactate, generated as a byproduct of plant primary metabolism [85], and the resulting pyruvate could support energy needs of the infection. The observed differences between the lifestyles in a profile of carbohydrate metabolism-related functions could be the result of adaptation of fungal pathogens to different plant biomass (e.g., composition of plant cell walls affecting penetration). Alternatively, different profiles of these functions could generate changes in environmental conditions (e.g., changes in the composition of soluble compounds or pH) that would serve as a cue for related functions,

such as transition between lifestyles. It is known, for example, from several phytopathogenic fungal systems that favorable pH conditions promote the infection process in the necrotrophic stage (e.g., in *Sclerotinia sclerotiorum* [86] or *C. gloeosporioides* [38, 87]). Expansins are cell wall-loosening proteins that are abundant in plant-associated microbes, including plant pathogens (according to a genomic search in NR, NCBI [88]). A few studies have explored the role of expansins in phytopathogen virulence [89, 90]. In the hemibiotrophic cacao pathogen *Moniliophthora perniciosa*, aggregated MpCP2 with cellulose-loosening activity was shown to promote spore (basidiospore) germination and subsequent tube growth, whereas the MpCP2-encoding gene was expressed in necrotic seeds; thus, MpCP2 had a potential role in both biotrophic (spore germination) and necrotrophic (seed) stages [89]. Despite the observed abundance of expansin in plant pathogens, there are very few genetic studies suggesting a potential role for fungal expansins in the virulence of phytopathogenic fungi. Thus, the current work, highlighting its position in the Hcore, suggests that functional studies of expansins' involvement in virulence are likely to be fruitful.

### Light and CO<sub>2</sub> perception in the Hcores

Carbonic anhydrase facilitates CO<sub>2</sub> sensing and subsequent differentiation, and virulence in the two human pathogens *Candida albicans* and *C. neoformans* [91–93]. These studies, together with its enrichment in the Hcores, suggest a similar role in sensing alterations in CO<sub>2</sub> level during plant infection, followed by induction of processes such as the transition between lifestyles. Rhodopsins: Fungi contain bacteriorhodopsins/microbial opsins that are light-driven ion pumps generating proton gradients across membranes [94]. Infection of rice plants with the rhodopsin-deficient mutant homolog (CarO) of *Fusarium fujikuroi* (ambiguously referred to as H or N, see Additional file 2: Table S1) showed more severe symptoms than the control strain, indicating a potential role of rhodopsin in the regulation of plant infection [95]. Silencing of the opsin ortholog Sop1 in the necrotroph *Sclerotinia sclerotiorum* resulted in reduced necrotic growth on oilseed rape leaves, and higher sensitivity to osmotic stress [96]. This illustrates the role of rhodopsins in light sensing and photomorphogenesis of phytopathogenic fungi, and along with its identification in the Hcores, suggests that alterations in light regime could play a role in virulence functions of hemibiotrophs, such as transitioning between lifestyles.

### Evolutionary trajectory of fungal pathogens

Early diverging fungal lineages (e.g., Blastocladiomycota and Chytridiomycota) identified in phylogenetic studies [4, 97–100] contain mainly saprobes and obligate biotrophs (and some endosymbionts) [98, 101]. Our analysis complements these observations by suggesting scenarios that presumably followed the emergence of these two lifestyles. Primordial fungal saprobes, able to both

decompose organic compounds and degrade debris of ancestral plants, presumably evolved a necrotrophic lifestyle as suggested by the highest number of shared gene families between Sap and N (Fig. 3). Although an alternative route could be suggested from the high number of Sap families shared with H, the evolution into N supplies a simpler explanation, which could have been followed by the subsequent emergence of H. The necrotrophic lifestyle could have initially evolved by acquisition of a relatively small number of toxins and lytic enzymes able to cause cell death. In this regard, the study of shared pectinase families in Dikarya and early diverging *Gonapodya prolifera*, a saprobe (member of the Chytridiomycota) able to grow on pectin as a carbon source, provides evidence for a common fungal ancestor able to feed on ancestral plant/algal pectin-containing debris [97]. Alternatively, or simultaneously, a primordial B could have acquired necrotrophic mechanisms, shifting to a hemibiotrophic lifestyle as illustrated by the highest number of gene families shared by B and H; subsequent loss of functions could have generated the necrotrophic-only lifestyle. The initial step of this scenario, starting with an ancestral biotrophic lifestyle, is more complicated than the aforementioned saprobic origin, as it requires acquisition of functions regulating the hemibiotrophic shift in addition to necrotrophic functions. However, it is supported by phylogenetic studies which have identified an early diverging sister clade of fungi (the Cryptomycota and Microsporidia taxa) that is made up of obligate biotrophic endoparasites [98, 100]. In both scenarios, acquiring a new lifestyle would have been advantageous in competition for niche/food resources.

## Conclusions

Our network analysis provides a map of the core functions in three major lifestyles of phytopathogenic fungi and oomycetes. The core functions highlighted in this work, which have not been previously associated with studied pathogenic lifestyles, including several enriched orthologs or domains with unknown function and some core families that cannot be annotated (Additional file 2: Table S2), open new avenues for future research that will enable a better understanding of these pathogens, and the discovery of novel functions associated with pathogenicity. It would make sense to start with core families with unknown function that contain SSPs, as the latter are often associated with pathogenicity. Regulatory mechanisms found in the Hcore functions include light- and CO<sub>2</sub>-sensing functions that could mediate important roles in this group, such as transition between lifestyles. These roles could also be regulated by changes in environmental composition resulting from the different core of lignocellulose-degrading enzymes found in this lifestyle. The presence of photoreceptors

(cryptochrome and rhodopsin) in the cores of plant pathogens raises the novel possibility of their central role in virulence, which is in agreement with the understanding that FEPs coevolved with photoautotrophic plant hosts. Our finding of light-sensing functions in the pathogen cores is partially supported by a survey of 22 Ascomycota which showed that they contain light-sensing mechanisms. These should confer better adaptation (protection, phototropism, morphogenesis, and circadian clock activity) under different light regimes [94].

The selected set of enriched core functions identified in our work can be used in other studies and applications. For example, these core functions can assist in identifying the pathogenic potential of samples analyzed by metagenomics or single-cell genomics. An interesting application in this direction would be to facilitate advanced agrotechnical practice, which is based on soil and leaf metagenomics (in addition to chemical monitoring) in future “next generation agriculture” [102, 103]. Last, empowered by the whole genomic network methodology, our analysis offers potential evolutionary scenarios following the emergence of an early branching saprobe and/or the obligate biotroph described in previous works.

## Methods

### Selected organisms

The data sets analysed in this study (downloaded at February 2016) can be found mainly in the National Center for Biotechnology Information, and in the Ensembl genomes databases using the accession numbers (and links) listed in Additional file 2: Table S1. The 82 selected genomes fungi and oomycetes represent the following lifestyles: 18 B, 20 H, and 22 N, 17 Sap (control or non-pathogens), and 5 pathogens ambiguously annotated as N or H (Additional file 2: Table S1). All biotrophs were treated uniformly in downstream analyses. The lifestyle of an organism was determined from either the respective database from which the sequences were collected, or the literature.

### Construction of the pan-proteome network

The pan-proteome sequence-similarity network was computed using EGN [104] for the 82 genomes with their 1,041,984 predicted protein sequences (hereafter, protein sequences) aligned using all-vs.-all BLASTP. Each node in the network represents a protein sequence from the 82 proteomes, and edges represent sequence similarity between pairs of protein sequences above a selected threshold that is accepted in the field [48] with minor modifications: minimal sequence length of 40 residues,  $E$  value  $< 10^{-4}$ , sequence identity  $\geq 35\%$  and minimal match coverage  $\geq 70\%$ . Only subgraphs with  $\geq 5$  nodes were included in further analyses (covering 30% of the subgraphs, and 85% of the sequences, Additional file 1: Figure S2).

The resulting network contained a large number of separate (unconnected) subgraphs (referred to as components), representing an operational gene family (referred as, gene family) whose sequences do not share significant similarity with other components [48]. When stringent criteria (larger thresholds) are enforced, most of these families putatively have the same or a closely related major function. Our pan-proteome network consisted of approximately 53 K gene families comprising 704 K protein-encoding genes (Additional file 1: Figure S2). Most (94%) of these were relatively small families (5 to 82 proteins per family, containing 1 protein per organism on average), that together comprised 61% of the sequences.

Python scripts were used to identify components of interest (e.g., having selected lifestyles), and to calculate relevant features (e.g., number of nodes per lifestyle, edges per each pair of lifestyles, or edges connecting selected lifestyles).

#### Core components (i.e., core gene families)

In this work, we defined a core component of a lifestyle as a component containing proteins derived from  $\geq 70\%$  of the organisms of that lifestyle. For example, a core component of N will have proteins from at least 70% of the 22 N used to construct the network (Additional file 2: Table S1). A Venn diagram was used to demonstrate the uniqueness of the core components (Additional file 1: Figure S1) using the web tool InteractiVenn [105].

#### Annotation of protein-encoding genes

All the 1,041,984 protein sequences of all the 82 organisms (genomes) were annotated based on several established platforms (Additional file 2: Table S2). KEGG orthologs and related pathways were identified using web-based KEGG Automatic Annotation Server (KASS [106]) based on BLAST single-directional best hit (SBH) against a selected list of genomes (Additional file 2: Table S4) using default thresholds and parameters. Protein domains were searched with standalone InterProScan [107] against 6 databases (Pfam, ProDom, Gene3D, TIGRFAM, ProSitePatterns, and PRINTS; recommended by InterPro, personal communication) using default thresholds and parameters. Protease families were identified using BLASTP ( $E$  value  $< 0.001$ ) against the MEROPS database [108] downloaded in November 2017. Carbohydrate-degrading enzymes were also identified using a Hidden Markov Model (HMM) search against the CAZymes HMM database (dbCAN-fam-HMM) [109], downloaded in October 2017, using protein sequences (with default steps and parameters as suggested for fungi by the developers). However, functions found by this database (which were significantly enriched, see below) were also found by other approaches (e.g., KEGG orthologs, and InterPro domain search), and therefore

they were not included in this final report. Possible secreted effectors were predicted by choosing proteins with sequence length  $\leq 300$  residues [9, 110], at least 2% Cysteines in the protein sequence (for short proteins, at least 2 Cys) [111], signal peptide based on standalone SignalP 4.1 [112], and no transmembrane region based on standalone TMHMM [113].

#### Calculation of enrichment and significance of core pathogenic functions

Enrichment of an annotation (i.e., KEGG, IPR, MEROPS, CAZyme) in the core components was calculated using the following equation  $(CPSA\_SPC/CPAA\_SPC)/(CPSA\_B/CPAA\_B)$  [48, 114], where CPSA\_SPC stands for Count of Proteins with Specific Annotation ID in Selected Pathogenic Core (SPC), CPAA\_SPC stands for Count of Proteins with Any Annotation ID in SPC, CPSA\_B stands for Count of Proteins with Specific Annotation ID in Background, CPAA\_B stands for Count of Proteins with Any Annotation ID in Background. We calculated enrichment for several combinations of Background and SPC: (i) general background comparing selected core to the entire network. Background - entire network excluding SPC; SPC - core components of selected lifestyle. For example  $(\text{Proteins with selected KEGG ID in Ncore}/\text{Proteins with Any KEGG ID in the rest of the network})/(\text{Proteins with selected KEGG ID in the rest of the network})$ ; (ii) to validate enrichment relative to non-pathogens, we compared the core of a selected pathogenic lifestyle to the non-pathogenic lifestyle of Sap. Background - Sap in core components of selected pathogenic lifestyle; SPC - selected pathogenic lifestyle in core components of selected lifestyle. For example -  $(\text{Proteins with selected KEGG ID in N organisms of Ncore}/\text{Proteins with Any KEGG ID in N organisms of Ncore})/(\text{Proteins with selected KEGG ID in Sap organisms of Ncore}/\text{Proteins with Any KEGG ID in Sap organisms of Ncore})$ . In addition we have validated that none of the significantly enriched functions identified in the pathogenic cores were significantly enriched in the non-pathogenic Sap cores (e.g., none of the KEGG IDs enriched in Ncore were enriched in the Sapcore); (iii) to further confirm that the functions of a selected pathogenic lifestyle are not highly abundant in the non-pathogenic Sap lifestyle (i.e., missed in calculations above because they were omitted from the network due to the thresholds selected to generate edges), we also tested the organisms' background regardless of the network. Background - proteins from Sap; SPC - all annotated proteins from the selected lifestyle. For example  $(\text{Proteins with selected KEGG ID in N}/\text{Proteins with Any KEGG ID in N})/(\text{Proteins with selected KEGG ID in Sap}/\text{Proteins with Any KEGG ID in Sap})$ .

Significance of each annotation (i.e., KEGG, IPR, MEROPS, CAZyme, Fig. 1, Table 1 and Additional file 2: Table S3) was calculated using Fisher's exact test in the *scipy* module of python [114]; only annotations with  $P < 0.05$  and enrichment  $> 1.1$  in all combinations of Background and SPC (above) were considered significantly enriched, and used in further analyses. The core functions were hierarchically clustered (Fig. 2) based on their abundance in each of the 82 organisms with the *hclust* package in R based on Euclidian distance using Ward's hierarchical agglomerative clustering [115].

### Gene families connecting a pair of lifestyles

Gene families connecting a pair of lifestyles were defined as components (i.e., operational gene families, see above) containing at least 1 direct connection (network edge) between proteins of 2 lifestyles (other lifestyles may exist in that component), and that the lifestyle selected as the focus of the analysis (see random simulated lifestyles below) does not have a direct connection with any other lifestyle. For example, for the B–N connection, such a component would contain at least 1 network edge connecting proteins from these two lifestyles, and if the B lifestyle is at the focus of the analysis, then it will not be connected to any other lifestyle (i.e., only B–N, and B–B edges are allowed for B). In contrast to the analysis described in previous section which is aimed to identify core functions of fungal and oomycete pathogens by utilizing genome sequences of representatives from both of these groups, the analysis described here is focused on potential evolutionary connections between lifestyles (i.e., B, N, H, and Sap). However, taxonomic classification and phylogenetic analyses have suggested that oomycetes form a clade that is distinct from fungi [97, 98]. To prevent possible mix between clade (i.e., oomycetes and fungi) and lifestyle connections in the evolutionary analysis we have excluded the oomycetes from this analysis.

### Significance of counts of gene families connecting a pair of lifestyles

To assess the significance of counts of gene families connecting a pair of lifestyles, we used random simulated lifestyles, in which labels (i.e., the lifestyle) for the organisms were randomly shuffled while preserving the numbers of genomes belonging to each lifestyle. In Table 2, the lifestyle in columns (the lifestyle being focused on) was kept intact, whereas lifestyles in rows were randomly shuffled. Once a new label was attributed to a genome, all of its proteins (nodes) were labeled as this new lifestyle. For each lifestyle, numbers of gene families connecting a pair of lifestyles were enumerated. The process was repeated  $4 \times 10,000$  times for each focused-on lifestyle (Table 2). All simulated counts were averaged and are reported in parentheses in Table 2. Significance of

the values was computed using non-parametric, empirical  $P$ -values, based on ranked real values of the simulations. When the real value was higher than 1 out of 10,000 simulated values, we attributed a  $P$ -value of  $< 0.0001$ . Normalized values were also significant for most cases (Additional file 2: Table S5), although they contained few non-significant differences between real data and simulations (indicated by asterisks in Additional file 2: Table S5), it did not affect the results. In the other cases, there were significantly more or less exclusive gene families between two groups in the real data than in the simulations. Normalization was generated by dividing the count of the components of interest by the maximum number of components that can exist in a studied network; e.g., if there are 15K components of N, and 10K components of B, the maximum possible number of components with an N–B connection (network edge) is 10K. Based on the counts of gene families connecting a pair of lifestyles, a network was generated (Fig. 3), using Cytoscape version 3.3.0 [49].

Interpretation of the simulations. For a pair of column X and row Y that shows more gene families than in the random simulation (e.g., column B and row H), genomes or gene families from lifestyle X are more similar to genomes or gene families from lifestyle Y than to genomes or gene families from any other lifestyle (in the dataset). Accordingly, when there are significantly less gene families in the real data than in the simulations (e.g., column B and row N), genomes or gene families in lifestyle X are more dissimilar to genomes or gene families in lifestyle Y than to genomes or gene families from any other lifestyle (in the dataset).

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12864-019-6409-3>.

**Additional file 1 : Figure S1.** Distribution of core components among all four lifestyles. B, H, N and Sap stands for Biotroph, Hemibiotroph, Necrotroph, and Saprotrorphs respectively. **Figure S2.** Cumulative frequency of components and sequences in the network.

**Additional file 2 : Table S1.** Organisms used in this study. **Table S2.** Frequency of annotated core families and their proteins (see sections "Core components" and "Construction of the pan-proteome network", **Methods**). **Table S3.** Significantly enriched annotations in core components of Biotroph (B), Hemibiotroph (H) and Necrotroph (N). \* Annotations starting with: K, KEGG ortholog; IPR, InterPro domain; MER, MEROPS proteases. \*\* Functional categories: N, nucleic acid related; L, light responsive; A, amino acid metabolism; T, transporters; C, carbohydrate metabolism; P, protease and peptidases; SM, secondary metabolites (including toxins); O, oxidoreductases (with no other function); Tf, trafficking; E, energy; ST, signal transduction. \*\*\* Following procedure described in methods, and subsequent manual curation. **Table S4.** Selected organisms used in KAAS analysis for identification of KEGG orthologs. **Table S5.** Normalized counts of gene families (components) connecting a pair of lifestyles. Numbers in parenthesis are the normalized mean values obtained from random 10,000 simulations for each lifestyle. A single asterisk indicates nonsignificant values ( $P > 0.01$ , non-parametric rank test, see **Methods**) (XLS 3237 kb)



## Abbreviations

B: biotroph; CAZyme: carbohydrate-active enzymes; FEP: filamentous eukaryotic pathogen; H: hemibiotroph; HMM: hidden Markov Model; KEGG: Kyoto Encyclopedia of Genes and Genomes; MAPK: mitogen-activated protein kinase; N: necrotroph; PKA: protein kinase A; SPC: selected pathogenic core, SSP – small secreted protein

## Acknowledgements

We are grateful to Dr. Slim Karkar (Neurospina, CEA, Paris) for interesting discussions. Contribution of Agricultural Research Organization (ARO), Volcani Center No: 18/2019.

## Author's contributions

EPJP performed the analyses and led the writing of the manuscript; AH guided the analyses and writing of the manuscript; OF, YE, and DP collaborated in interpreting the results and writing the manuscript. All authors read and approved the final manuscript.

## Funding

Supported by a grant from the Israeli Chief Scientist, Ministry of Agriculture (grant number 20-10-0069); ARO postdoctoral fellowship. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Availability of data and materials

The data sets analysed in this study can be found in The National Center for Biotechnology Information Taxonomy database using the accession numbers listed in Additional file 2: Table S1.

## Ethics approval and consent to participation

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Vegetable and Field Crops, Institute of Plant Sciences, Volcani Center, Agricultural Research Organization, Rishon LeZion, Israel. <sup>2</sup>Department of Plant Pathology and Weed Research, Institute of Plant Protection, Volcani Center, Agricultural Research Organization, Rishon LeZion, Israel. <sup>3</sup>Department of Postharvest Science, Institute of Postharvest and Food Sciences, Volcani Center, Agricultural Research Organization, Rishon LeZion, Israel.

Received: 14 August 2019 Accepted: 17 December 2019

Published online: 26 December 2019

## References

- Oerke EC. Crop losses to pests. *J Agr Sci*. 2005;144(1):31–43.
- Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. Emerging fungal threats to animal, plant and ecosystem health. *Nature*. 2012;484(7393):186–94.
- Brown JK, Hovmoller MS. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*. 2002;297(5581):537–41.
- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otililar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res*. 2014;42(Database issue):D699–704.
- Muszewska A, Stepniewska-Dziubinska MM, Steczkiewicz K, Pawlowska J, Dziedzic A, Ginalski K. Fungal lifestyle reflected in serine protease repertoire. *Sci Rep*. 2017;7(1):9147.
- Li J, Gu F, Wu R, Yang J, Zhang KQ. Phylogenomic evolutionary surveys of subtilase superfamily genes in fungi. *Sci Rep*. 2017;7:45456.
- Zhao Z, Liu H, Wang C, Xu JR. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*. 2013;14:274.
- Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*. 2009;461(7262):393–8.
- Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A*. 2011;108(22):9166–71.
- Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RH, Zody MC, Kunjeti SG, Donofrio NM, et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*. 2010;330(6010):1540–3.
- Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol*. 2012;10(6):417–30.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stuber K, Ver Loren van Themaat E, Brown JK, Butcher SA, Gurr SJ, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*. 2010;330(6010):1543–6.
- Perez-Nadales E, Nogueira MF, Baldin C, Castanheira S, El Ghalid M, Grund E, Lengeler K, Marchegiani E, Mehrotra PV, Moretti M, et al. Fungal model systems and the elucidation of pathogenicity determinants. *Fungal Genet Biol*. 2014;70:42–67.
- Nino-Sanchez J, Tello V, Casado-Del Castillo V, Thon MR, Benito EP, Diaz-Minguez JM. Gene expression patterns and dynamics of the colonization of common bean (*Phaseolus vulgaris* L.) by highly virulent and weakly virulent strains of *Fusarium oxysporum*. *Front Microbiol*. 2015;6:234.
- Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev*. 2015;35:57–65.
- Harel A, Gorovits R, Yarden O. Changes in protein kinase a activity accompany sclerotial development in *Sclerotinia sclerotiorum*. *Phytopathology*. 2005;95(4):397–404.
- Harel A, Bercovich S, Yarden O. Calcineurin is required for sclerotial development and pathogenicity of *Sclerotinia sclerotiorum* in an oxalic acid-independent manner. *Mol Plant-Microbe Interact*. 2006;19(6):682–93.
- Bahn YS, Xue C, Idnurm A, Rutherford JC, Heitman J, Cardenas ME. Sensing the environment: lessons from fungi. *Nat Rev Microbiol*. 2007;5(1):57–69.
- Dickman MB, Yarden O. Serine/threonine protein kinases and phosphatases in filamentous fungi. *Fungal Genet Biol*. 1999;26(2):99–117.
- de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MH, Thomma BP. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science*. 2010;329(5994):953–5.
- Oliva R, Win J, Raffaele S, Boutemy L, Bozkurt TO, Chaparro-Garcia A, Segretin ME, Stam R, Schornack S, Cano LM, et al. Recent developments in effector biology of filamentous plant pathogens. *Cell Microbiol*. 2010;12(6):705–15.
- Van der Does HC, Rep M. Virulence genes and the evolution of host specificity in plant-pathogenic fungi. *Mol Plant-Microbe Interact*. 2007;20(10):1175–82.
- Wight WD, Labuda R, Walton JD. Conservation of the genes for HC-toxin biosynthesis in *Alternaria jesenskae*. *BMC Microbiol*. 2013;13:165.
- Vleeshouwers VG, Oliver RP. Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Mol Plant-Microbe Interact*. 2014;27(3):196–206.
- Stergiopoulos I, Kourmpetis YA, Slot JC, Bakker FT, De Wit PJ, Rokas A. In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Mol Biol Evol*. 2012;29(11):3371–84.
- Keller NP, Turner G, Bennett JW. Fungal secondary metabolism - from biochemistry to genomics. *Nat Rev Microbiol*. 2005;3(12):937–47.
- Budak SO, Zhou M, Brouwer C, Wiebenga A, Benoit I, Di Falco M, Tsang A, de Vries RP. A genomic survey of proteases in *Aspergilli*. *BMC Genomics*. 2014;15:523.
- Zhao Z, Liu H, Wang C, Xu JR. Correction: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*. 2014;15:6.
- Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, Figueroa M, Freitag M, Hane JK, Henrissat B, et al. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3 (Bethesda)*. 2013;3(1):41–63.
- de Jonge R, Bolton MD, Thomma BP. How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr Opin Plant Biol*. 2011;14(4):400–6.
- Pendleton AL, Smith KE, Feau N, Martin FM, Grigoriev IV, Hamelin R, Nelson CD, Burleigh JG, Davis JM. Duplications and losses in gene families of rust pathogens highlight putative effectors. *Front Plant Sci*. 2014;5:299.
- Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, et al. *Phytophthora* genome sequences

- uncover evolutionary origins and mechanisms of pathogenesis. *Science*. 2006;313(5791):1261–6.
33. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLoS Pathog*. 2012;8(12):e1003037.
  34. Manning VA, Chu AL, Steeves JE, Wolpert TJ, Ciuffetti LM. A host-selective toxin of *Pyrenophora tritici-repentis*, Ptr ToxA, induces photosystem changes and reactive oxygen species accumulation in sensitive wheat. *Mol Plant-Microbe Interact*. 2009;22(6):665–76.
  35. Ottmann C, Luberacki B, Kufner I, Koch W, Brunner F, Weyand M, Mattinen L, Pirhonen M, Aderlüh G, Seitz HU, et al. A common toxin fold mediates microbial attack and plant defense. *Proc Natl Acad Sci U S A*. 2009;106(25):10359–64.
  36. Elad Y, Pertot I, Cotes A, Stewart A. Plant hosts of *Botrytis* spp. In: Fillingier S, Elad Y, editors. *Botrytis – the Fungus, the Pathogen and its Management in Agricultural Systems*. London: Springer; 2016. p. 413–86.
  37. Gadoury DM, Cadle-Davidson L, Wilcox WF, Dry IB, Seem RC, Milgroom MG. Grapevine powdery mildew (*Erysiphe necator*): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. *Mol Plant Pathol*. 2012;13(1):1–16.
  38. Alkan N, Friedlander G, Ment D, Prusky D, Fluhr R. Simultaneous transcriptome analysis of *Colletotrichum gloeosporioides* and tomato fruit pathosystem reveals novel fungal pathogenicity and fruit defense strategies. *New Phytol*. 2015;205(2):801–15.
  39. Moore D, Robson GD, Trinci APJ. 21st Century Guidebook to Fungi. Cambridge: Cambridge University Press; 2000.
  40. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. 2003;422(6934):859–68.
  41. Dou D, Kale SD, Wang X, Chen Y, Wang Q, Jiang RH, Arredondo FD, Anderson RG, Thakur PB, McDowell JM, et al. Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. *Plant Cell*. 2008;20(4):1118–33.
  42. Bos JL, Armstrong MR, Gilroy EM, Boevink PC, Hein I, Taylor RM, Zhendong T, Engelhardt S, Vetukuri RR, Harrower B, et al. *Phytophthora infestans* effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proc Natl Acad Sci U S A*. 2010;107(21):9909–14.
  43. de Wit PJ. How plants recognize pathogens and defend themselves. *Cell Mol Life Sci*. 2007;64(21):2726–32.
  44. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet*. 2010;11(8):539–48.
  45. Thomma BP, Numberger T, Joosten MH. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell*. 2011;23(1):4–15.
  46. Baptiste E, Lopez P, Bouchard F, Baquero F, McInerney JO, Burian RM. Evolutionary analyses of non-genealogical bonds produced by introgressive descent. *Proc Natl Acad Sci U S A*. 2012;109(45):18266–72.
  47. Harel A, Bromberg Y, Falkowski PG, Bhattacharya D. Evolutionary history of redox metal-binding domains across the tree of life. *Proc Natl Acad Sci U S A*. 2014;111(19):7042–7.
  48. Harel A, Karkar S, Cheng S, Falkowski PG, Bhattacharya D. Deciphering primordial cyanobacterial genome functions from protein network analysis. *Curr Biol*. 2015;25(5):628–34.
  49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
  50. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*. 2006;444(7115):97–101.
  51. Yu G, Liu JL, Xie LQ, Wang XL, Zhang SH, Pan HY. Characterization, cloning, and heterologous expression of a subtilisin-like serine protease gene VIPr1 from *Verticillium lecanii*. *J Microbiol*. 2012;50(6):939–46.
  52. Yang J, Liang L, Li J, Zhang KQ. Nematicidal enzymes from microorganisms and their applications. *Appl Microbiol Biotechnol*. 2013;97(16):7081–95.
  53. Gu Q, Yuan Q, Zhao D, Huang J, Hsiang T, Wei Y, Zheng L. Acetyl-coenzyme a synthetase gene ChAcs1 is essential for lipid metabolism, carbon utilization and virulence of the hemibiotrophic fungus *Colletotrichum higginsianum*. *Mol Plant Pathol*. 2019;20(1):107–23.
  54. Foster AJ, Ryder LS, Kershaw MJ, Talbot NJ. The role of glycerol in the pathogenic lifestyle of the rice blast fungus *Magnaporthe oryzae*. *Environ Microbiol*. 2017;19(3):1008–16.
  55. Chen XL, Wang Z, Liu C. Roles of peroxisomes in the rice blast fungus. *Biomed Res Int*. 2016;2016:9343417.
  56. Wang ZY, Soanes DM, Kershaw MJ, Talbot NJ. Functional analysis of lipid metabolism in *Magnaporthe grisea* reveals a requirement for peroxisomal fatty acid beta-oxidation during appressorium-mediated plant infection. *Mol Plant-Microbe Interact*. 2007;20(5):475–91.
  57. Valero-Jimenez CA, Veloso J, Staats M, Van Kan JAL. Comparative genomics of plant pathogenic *Botrytis* species with distinct host specificity. *BMC Genomics*. 2019;20(1):203.
  58. Sexton AC, Howlett BJ. Parallels in fungal pathogenesis on plant and animal hosts. *Eukaryot Cell*. 2006;5(12):1941–9.
  59. Jashni MK, Mehrabi R, Collemare J, Mesarich CH, de Wit PJ. The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant-pathogen interactions. *Front Plant Sci*. 2015;6:584.
  60. Reddy PV, Lam CK, Belanger FC. Mutualistic fungal endophytes express a proteinase that is homologous to proteases suspected to be important in fungal pathogenicity. *Plant Physiol*. 1996;111(4):1209–18.
  61. Naumann TA, Wicklow DT, Price NP. Identification of a chitinase-modifying protein from *Fusarium verticillioides*: truncation of a host resistance protein by a fungalysin metalloprotease. *J Biol Chem*. 2011;286(41):35358–66.
  62. Olivier F, Eugenia Zanetti M, Oliva CR, Covarrubias AA, Casalongué CA. Characterization of an extracellular serine protease of *Fusarium eumartii* and its action on pathogenesis related proteins. *Eur J Plant Pathol*. 2002;108(1):63–72.
  63. Langner T, Gohre V. Fungal chitinases: function, regulation, and potential roles in plant/pathogen interactions. *Curr Genet*. 2016;62(2):243–54.
  64. Jashni MK, Dols IH, Iida Y, Boeren S, Beenen HG, Mehrabi R, Collemare J, de Wit PJ. Synergistic action of a metalloprotease and a serine protease from *Fusarium oxysporum* f. sp. *lycopersici* cleaves chitin-binding tomato chitinases, reduces their antifungal activity, and enhances fungal virulence. *Mol Plant-Microbe Interact*. 2015;28(9):996–1008.
  65. Macho AP, Schwessinger B, Ntoukakis V, Brutus A, Segonzac C, Roy S, Kadota Y, Oh MH, Sklenar J, Derbyshire P, et al. A bacterial tyrosine phosphatase inhibits plant pattern recognition receptor activation. *Science*. 2014;343(6178):1509–12.
  66. Hall A. Rho GTPases and the actin cytoskeleton. *Science*. 1998;279(5350):509–14.
  67. Jiang C, Zhang X, Liu H, Xu JR. Mitogen-activated protein kinase signaling in plant pathogenic fungi. *PLoS Pathog*. 2018;14(3):e1006875.
  68. Chen C, Harel A, Gorovoits R, Yarden O, Dickman MB. MAPK regulation of sclerotial development in *Sclerotinia sclerotiorum* is linked with pH and cAMP sensing. *Mol Plant-Microbe Interact*. 2004;17(4):404–13.
  69. Scheffer J, Chen C, Heidrich P, Dickman MB, Tudzynski P. A CDC42 homologue in *Claviceps purpurea* is involved in vegetative differentiation and is essential for pathogenicity. *Eukaryot Cell*. 2005;4(7):1228–38.
  70. Fletcher K, Klosterman SJ, Derevnina L, Martin F, Bertier LD, Koike S, Reyes-Chin-Wo S, Mou B, Michelmore R. Comparative genomics of downy mildews reveals potential adaptations to biotrophy. *BMC Genomics*. 2018;19(1):851.
  71. Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics*. 2018;19(1):381.
  72. Dussert Y, Mazet ID, Couture C, Gouzy J, Piron MC, Kuchly C, Bouchez O, Rispé C, Mestre P, Delmotte F. A high-quality grapevine downy mildew genome assembly reveals rapidly evolving and lineage-specific putative host adaptation genes. *Genome Biol Evol*. 2019;11(3):954–69.
  73. Panepinto J, Liu L, Ramos J, Zhu X, Valyi-Nagy T, Eksi S, Fu J, Jaffe HA, Wickes B, Williamson PR. The DEAD-box RNA helicase Vad1 regulates multiple virulence-associated genes in *Cryptococcus neoformans*. *J Clin Invest*. 2005;115(3):632–41.
  74. Brosch G, Loidl P, Graessle S. Histone modifications and chromatin dynamics: a focus on filamentous fungi. *FEMS Microbiol Rev*. 2008;32(3):409–39.
  75. Elias-Villalobos A, Barrales RR, Ibeas JI. Chromatin modification factors in plant pathogenic fungi: insights from *Ustilago maydis*. *Fungal Genet Biol*. 2019;129:52–64.
  76. Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent MH, Connolly LR, Freitag M, Rouxel T, et al. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet*. 2014;10(3):e1004227.
  77. Gonzalez-Prieto JM, Rosas-Quijano R, Dominguez A, Ruiz-Herrera J. The UmGcn5 gene encoding histone acetyltransferase from *Ustilago maydis* is involved in dimorphism and virulence. *Fungal Genet Biol*. 2014;71:86–95.

78. Chaves I, Pokorny R, Byrdin M, Hoang N, Ritz T, Brettel K, Essen LO, Van der Horst GT, Batschauer A, Ahmad M. The cryptochromes: blue light photoreceptors in plants and animals. *Annu Rev Plant Biol.* 2011;62:335–64.
79. Zhu M, Riederer M, Hildebrandt U. UV-C irradiation compromises conidial germination, formation of appressoria, and induces transcription of three putative photolyase genes in the barley powdery mildew fungus, *Blumeria graminis* f. sp. *hordei*. *Fungal Biol.* 2019;123(3):218–30.
80. Bluhm BH, Dunkle LD. PHL1 of *Cercospora zeae-maydis* encodes a member of the photolyase/cryptochrome family involved in UV protection and fungal development. *Fungal Genet Biol.* 2008;45(10):1364–72.
81. Kim H, Ridenour JB, Dunkle LD, Bluhm BH. Regulation of stomatal tropism and infection by light in *Cercospora zeae-maydis*: evidence for coordinated host/pathogen responses to photoperiod? *PLoS Pathog.* 2011;7(7):e1002113.
82. Cohrs KC, Schumacher J. The two cryptochrome/photolyase family proteins fulfill distinct roles in DNA photorepair and regulation of conidiation in the gray mold fungus *Botrytis cinerea*. *Appl Environ Microbiol.* 2017;83(17):e00812–17.
83. Murphy C, Powlowski J, Wu M, Butler G, Tsang A. Curation of characterized glycoside hydrolases of fungal origin. Database (Oxford). 2011;2011:bar020.
84. Harvey AJ, Hrmova M, De Gori R, Varghese JN, Fincher GB. Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins.* 2000;41(2):257–69.
85. Maurino VG, Engqvist MK. 2-Hydroxy acids in plant metabolism. *Arabidopsis Book.* 2015;13:e0182.
86. Criscitiello MF, Dickman MB, Samuel JE, de Figueiredo P. Tripping on acid: trans-kingdom perspectives on biological acids in immunity and pathogenesis. *PLoS Pathog.* 2013;9(7):e1003402.
87. Bi F, Barad S, Ment D, Luria N, Dubey A, Casado V, Glam N, Minguez JD, Espeso EA, Fluhr R, et al. Carbon regulation of environmental pH by secreted small molecules that modulate pathogenicity in phytopathogenic fungi. *Mol Plant Pathol.* 2016;17(8):1178–95.
88. Nikolaidis N, Doran N, Cosgrove DJ. Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. *Mol Biol Evol.* 2014;31(2):376–86.
89. Barsottini MR d O, de Oliveira JF, Adamoski D, Teixeira PJ, do Prado PF, Tiezzi HO, Sforca ML, Cassago A, Portugal RV, de Oliveira PS, et al. Functional diversification of cerato-platanins in *Monilophthora perniciosa* as seen by differential expression and protein function specialization. *Mol Plant-Microbe Interact.* 2013;26(11):1281–93.
90. Quarantin A, Castiglioni C, Schafer W, Favaron F, Sella L. The *Fusarium graminearum* cerato-platanins loosen cellulose substrates enhancing fungal cellulase activity as expansin-like proteins. *Plant Physiol Biochem.* 2019;139:229–38.
91. Elleuche S, Poggeler S. Carbonic anhydrases in fungi. *Microbiology.* 2010; 156(Pt 1):23–9.
92. Mogensen EG, Janbon G, Chaloupka J, Steegborn C, Fu MS, Moyrand F, Klengel T, Pearson DS, Geeves MA, Buck J, et al. *Cryptococcus neoformans* senses CO<sub>2</sub> through the carbonic anhydrase Can2 and the adenylyl cyclase Cac1. *Eukaryot Cell.* 2006;5(1):103–11.
93. Klengel T, Liang WJ, Chaloupka J, Ruoff C, Schroppe K, Naglik JR, Eckert SE, Mogensen EG, Haynes K, Tuite MF, et al. Fungal adenylyl cyclase integrates CO<sub>2</sub> sensing with cAMP signaling and virulence. *Curr Biol.* 2005;15(22):2021–6.
94. Schumacher J. How light affects the life of *Botrytis*. *Fungal Genet Biol.* 2017; 106:26–41.
95. Adam A, Deimel S, Pardo-Medina J, Garcia-Martinez J, Konte T, Limon MC, Avalos J, Terpitz U. Protein activity of the *Fusarium fujikuroi* rhodopsins CarO and OpsA and their relation to fungus-plant interaction. *Int J Mol Sci.* 2018;19(1):215.
96. Lyu X, Shen C, Fu Y, Xie J, Jiang D, Li G, Cheng J. The microbial opsin homolog Sop1 is involved in *Sclerotinia sclerotiorum* development and environmental stress response. *Front Microbiol.* 2016;6:1504.
97. Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, LaButti KM, Lindquist EA, Yee Ngan C, Ohm RA, et al. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biol Evol.* 2015;7(6):1590–601.
98. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Guaidan C, Fraker E, Miadlikowska J, et al. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature.* 2006;443(7113):818–22.
99. Spatafora JW, Aime MC, Grigoriev IV, Martin F, Stajich JE, Blackwell M. The fungal tree of life: From molecular systematics to genome-scale phylogenies. *Microbiol Spectr.* 2017;5(5). <https://doi.org/10.1128/microbiolspec.FUNK-0053-2016>.
100. Berbee ML, James TY, Strullu-Derrien C. Early diverging fungi: diversity and impact at the dawn of terrestrial life. *Annu Rev Microbiol.* 2017;71:41–60.
101. Herivaux A, Duge de Bernonville T, Roux C, Clastre M, Courdavault V, Gastebois A, Bouchara JP, James TY, Latge JP, Martin F, et al. The identification of phytohormone receptor homologs in early diverging fungi suggests a role for plant sensing in land colonization by fungi. *MBio.* 2017; 8(1):e01739–16.
102. Schlaeppli K, Bulgarelli D. The plant microbiome at work. *Mol Plant-Microbe Interact.* 2015;28(3):212–7.
103. Harel A, Arya GC. Phyllosphere and its potential role in sustainable agriculture. In: Tripathi V, Kumar P, Tripathi P, Kishore A, editors. *Microbial Genomics in Sustainable Agroecosystems*. Singapore: Springer; 2019. p. 21–38.
104. Halary S, McInerney JO, Lopez P, Bapteste E. EGN: a wizard for construction of gene and genome similarity networks. *BMC Evol Biol.* 2013;13:146.
105. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics.* 2015;16:169.
106. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35(Web Server issue):W182–5.
107. Jones P, Binns D, Chang HY, 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.
108. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 2014; 42(Database issue):D503–9.
109. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 2009;37(Database issue):D233–8.
110. 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.
111. Lu S, Edwards MC. Genome-wide analysis of small secreted cysteine-rich proteins identifies candidate effector proteins potentially involved in *Fusarium graminearum*-wheat interactions. *Phytopathology.* 2016;106(2):166–76.
112. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011;8(10):785–6.
113. Chen Y, Yu P, Luo J, Jiang Y. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm Genome.* 2003;14(12):859–65.
114. Rivals I, Personnaz L, Taing L, Potier MC. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics.* 2007;23(4):401–7.
115. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J Classif.* 2014;31(3):274–95.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

