



**HAL**  
open science

## Globally distributed root endophyte *Phialocephala subalpina* links pathogenic and saprophytic lifestyles

Markus Schlegel, Martin Munsterktter, Ulrich Guldener, Remy Bruggmann, Angelo Duo, Matthieu Hainaut, Bernard Henrissat, Christian M. K. Sieber, Dirk Hoffmeister, Christoph R. Grunig

► **To cite this version:**

Markus Schlegel, Martin Munsterktter, Ulrich Guldener, Remy Bruggmann, Angelo Duo, et al.. Globally distributed root endophyte *Phialocephala subalpina* links pathogenic and saprophytic lifestyles. BMC Genomics, 2016, 17 (1), pp.1015. 10.1186/s12864-016-3369-8 . hal-01439119

**HAL Id: hal-01439119**

**<https://hal.science/hal-01439119>**

Submitted on 19 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Open Access



# Globally distributed root endophyte *Phialocephala subalpina* links pathogenic and saprophytic lifestyles

Markus Schlegel<sup>1†</sup>, Martin Münsterkötter<sup>2†</sup>, Ulrich Güdener<sup>2,3</sup>, Rémy Bruggmann<sup>4</sup>, Angelo Duò<sup>1</sup>, Matthieu Hainaut<sup>5</sup>, Bernard Henrissat<sup>5</sup>, Christian M. K. Sieber<sup>2,6</sup>, Dirk Hoffmeister<sup>7</sup> and Christoph R. Grünig<sup>1,8\*</sup>

## Abstract

**Background:** Whereas an increasing number of pathogenic and mutualistic ascomycetous species were sequenced in the past decade, species showing a seemingly neutral association such as root endophytes received less attention. In the present study, the genome of *Phialocephala subalpina*, the most frequent species of the *Phialocephala fortinii* s.l. – *Acephala applanata* species complex, was sequenced for insight in the genome structure and gene inventory of these wide-spread root endophytes.

**Results:** The genome of *P. subalpina* was sequenced using Roche/454 GS FLX technology and a whole genome shotgun strategy. The assembly resulted in 205 scaffolds and a genome size of 69.7 Mb. The expanded genome size in *P. subalpina* was not due to the proliferation of transposable elements or other repeats, as is the case with other ascomycetous genomes. Instead, *P. subalpina* revealed an expanded gene inventory that includes 20,173 gene models. Comparative genome analysis of *P. subalpina* with 13 ascomycetes shows that *P. subalpina* uses a versatile gene inventory including genes specific for pathogens and saprophytes. Moreover, the gene inventory for carbohydrate active enzymes (CAZymes) was expanded including genes involved in degradation of biopolymers, such as pectin, hemicellulose, cellulose and lignin.

**Conclusions:** The analysis of a globally distributed root endophyte allowed detailed insights in the gene inventory and genome organization of a yet largely neglected group of organisms. We showed that the ubiquitous root endophyte *P. subalpina* has a broad gene inventory that links pathogenic and saprophytic lifestyles.

**Keywords:** Comparative genomics, Lifestyle, Root endophyte, Species complex, Parasitism-mutualism continuum

## Background

Plant roots are confronted with the colonization of symbiotic fungal species ranging from pathogens to mutualists [1]. While research has largely been focused on the symbiotic and pathogenic interactions, seemingly neutral associations with plant roots by endophytes received less attention [2, 3]. Endophytic fungi colonize functional roots tissues but disease symptoms do not develop at all or at least not for prolonged periods of time [4]. Despite their prevalence in many ecosystems, little is known about the

nature of interaction with their hosts [5, 6]. It is assumed that they behave along the mutualism - antagonism continuum depending on host conditions and environment, as shown for some mycorrhizal fungi [7–9].

Species belonging to the helotialean *Phialocephala fortinii* s.l. – *Acephala applanata* species complex (PAC) are the dominant root endophytes in woody plant species [5]. PAC shows a global distribution as PAC species colonize roots from arctic to subtropical plant species throughout the northern hemisphere [10–12]. Recently, a study proved the presence of PAC on the southern hemisphere [13]. In contrast to ectomycorrhizal species (EcM), which are usually confined to primary, non-lignified roots, PAC can be found in all parts of the root system, predominately on fine roots < 3 mm where up to

\* Correspondence: christoph.gruenig@microsynth.ch

†Equal contributors

<sup>1</sup>Institute of Integrative Biology (IBZ), Forest Pathology and Dendrology, ETH Zürich, 8092 Zürich, Switzerland

<sup>8</sup>Microsynth AG, Schützenstrasse 15, 9436 Balgach, Switzerland

Full list of author information is available at the end of the article



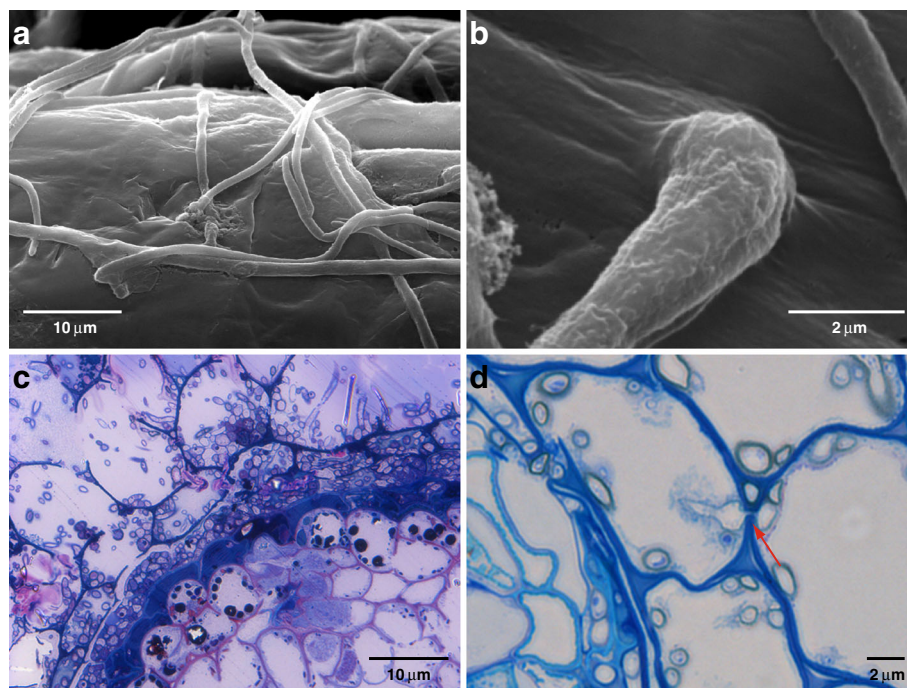
80% of randomly sampled roots were colonized [5]. In addition, PAC species are belonging to the first colonizers of tree seedlings in natural forest ecosystems, infecting them within a few weeks after germination [14].

PAC is composed of more than 15 cryptic species [10], eight of which were formally described [15, 16]. Among the strains sampled from a single study site, PAC species form communities of up to 10 sympatrically occurring species. In contrast to agricultural ecosystems, PAC communities remain stable for several years [17] although minor long-term changes in the frequency of species can be observed [18]. Most PAC communities are dominated by few species and additional species occur at low frequencies [5] as observed in many other community structures of biological systems [19]. Species diversity and community structure do not correlate with the tree community, geographical location, soil properties, management practices nor does climate, precipitation and temperature [10]. Host specificity of PAC species was not observed [5] except for *A. applanata* that was almost exclusively isolated from species belonging to the Pinaceae but rarely from ericaceous roots of the ground vegetation [14].

Despite the fact that PAC species are highly successful colonizers of plant roots and widely distributed in natural ecosystems, their ecological role is still poorly understood.

They were described as beneficial, neutral or pathogenic for different hosts, growing conditions and fungal strains [5, 20, 21]. New results comparing the effect of PAC species and strains on *Picea abies* indicate that the outcome of the interactions is mainly driven by the fungal genotype and follow the antagonism-mutualism continuum. Whereas some of the PAC/*P. subalpina* strains were shown to be pathogenic and killed most of the seedlings, others were benign [22]. Nevertheless, infection by PAC is costly for the plant since an increase in plant growth was never observed [22]. The outcome of PAC-host interactions depends on the ability of PAC strains to invade and colonize host root tissues. This is evident by the health status of Norway spruce seedlings, which negatively correlates with the biomass of the fungus in roots [22, 23]. However, negative effects of harmful PAC strains are reduced by the co-colonization of ectomycorrhizal fungi and other PAC strains [24].

The dynamics of PAC-host colonization was rarely studied [25–27], and data on intraspecific variation of colonization dynamics for different PAC species is missing completely. In general, PAC strains form hyphopodia- or appressoria-like structures to enter root hairs or epidermal cells (Fig. 1a, b). After entering the root, PAC strains grow inter- and intracellularly and colonize the cortex but rarely invade the vascular cylinder (Fig. 1c, d). Intercellular



**Fig. 1** Key features of the colonization of roots by PAC species. Key steps in the colonization of roots by PAC species (V. Queloz, unpublished). Figure 1a, b. Surface colonization and appressoria/hyphopodia formation of *P. subalpina* on *Cistus incanus* roots. Figure 1c. Cross-section of *Pinus strobus* root colonized by PAC stained using borax, methylene blue and toluidine blue and counterstained with basic fuchsin. The fungus completely colonizes the cortical tissue up to the endodermis. Accumulation of phenolic compounds in the vascular cylinder is evident by the presence of dark granular structures. Figure 1d. Example of intracellular colonization of *P. subalpina* in *C. incanus* cortex (arrow)

labyrinthine fungal tissue resembling the Hartig net in ectomycorrhizal fungi as well as mantel-like structures were occasionally observed for PAC [27–29].

Host defense reactions such as cell-wall appositions or lignituber formation were rarely observed [27]. Intracellular hyphae traverse host cells by narrow penetration hyphae without apparent lysis of the plant cell wall while the host cell cytoplasm disintegrates after colonization by PAC hyphae (Fig. 1d). Pertaining to the ultrastructural level, hyphae are not surrounded by host-derived perifungal membranes, which are regarded as hallmark for biotrophic fungal associations [27]. Finally, cortical cells of the plant are often associated with thick-walled, heavily melanized fungal cells, i.e. microsclerotia, which were shown to accumulate reserve substances [5, 25].

The availability of genomic sequences provides information on the gene inventory of species and identifies characteristic genomic structures and gene sets associated with different lifestyles [30–33]. Although the number of sequenced fungal genomes increases rapidly, genome sequencing of ascomycetes was mostly restricted to pathogenic, saprophytic or well-known mutualistic species. In contrast, only few endophytes were sequenced and restricted to endophytes in the Clavicipitaceae [34, 35]. Clavicipitalean endophytes are obligate biotrophs, colonize their hosts systemically and follow a very close symbiotic lifestyle with their hosts but roots are not colonized [36]. This sets them apart from non-clavicipitalean endophytes isolated from all parts of the plants at high frequencies. An exception is the genome of *Phialocephala scopiformis*, a foliar endophyte, for which the draft genome was recently published albeit with no analysis [37]. In the present study, the genome of *P. subalpina*, the most frequent species of the PAC, was sequenced and compared to the genomes of 13 ascomycetous species with different lifestyles to get first insights in the genome structure and gene inventory of non-clavicipitalean endophyte.

## Results

### *Phialocephala subalpina* holds a large and feature-rich genome

The genome of *P. subalpina* strain UAMH 11012 was sequenced at 25 x coverage using the Roche/454 GS FLX technology. Sequence data was assembled into 204 scaffolds (excluding rDNA repeat and mtDNA) with a genome size of 69.7 Mb and an average GC content of 45.9% (Table 1). The complete rRNA repeat (part of this assembly) and the mtDNA (<http://www.ebi.ac.uk/ena/data/view/JN031566>) [38] were assembled manually and validated by Sanger sequencing. Data sets are accessible at <http://pedant.helmholtz-muenchen.de/genomes.jsp?category=fungal>. The genome and annotation data was submitted to the European Nucleotide Archive (ENA, <http://www.ebi.ac.uk/ena/data/view/>

**Table 1** Genome statistics for *Phialocephala subalpina*

Genome size [Mb]	69.69
Scaffolds	205
Scaffolds N50 [kb]	1,449
N50 number of scaffolds	16
GC (%)	45.9
Predicted protein-coding genes	20,173
Average exon length [bp]	443.9
Average intron length [bp]	80.2
Average number of introns per gene	2.7
tRNAs	91
TEs content	5.70%
Other repeats <sup>a</sup>	8.10%

<sup>a</sup>tandem repeats, SSR, and low-complexity DNA

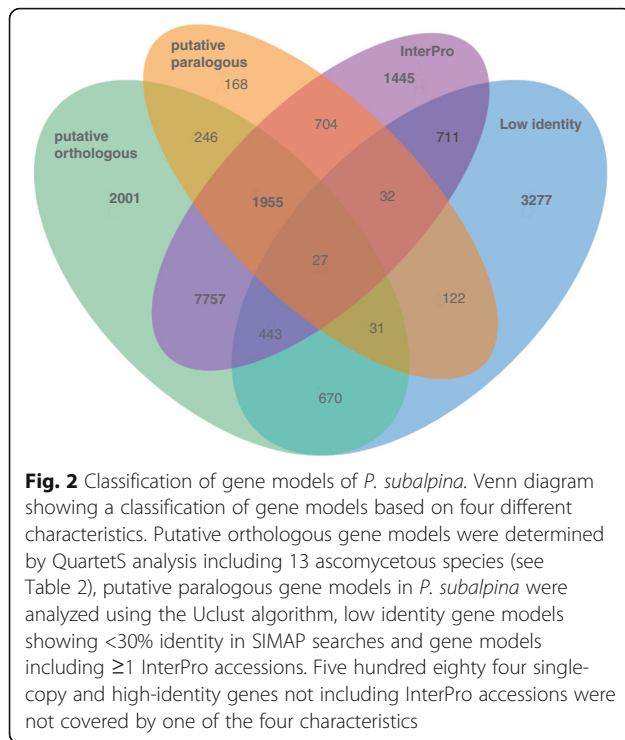
FJOG01000001-FJOG01000205). Mapping of reads data against the assembled genome did not reveal any significant deviations from the average 25x coverage except for the mtDNA (2,629x) and rDNA repeat (1,422x) indicating that no repetitive sequences were collapsed in short scaffolds leading to an underestimation of the genome size.

The annotation pipeline and manual validation resulted in 20,173 gene models. In addition, 91 tRNA genes coding for all amino acids and 20 5S rRNA loci were identified. None of the expected single-copy core orthologous genes found in eukaryotes (248 and 246 genes, see material and methods) was missing in the protein predications for *P. subalpina* indicating that the core gene inventory had completely been covered. This was supported by EST data as 28,045 out of 28,092 assembled transcripts (99.83%) mapped to the assembly with high coverage (Additional file 1). 73.6% of the 20,173 proteins showed an identity >30% with known proteins in the Similarity Matrix of Proteins database (SIMAP) [39]. The remaining 5,313 proteins with less than 30% identity included 4,233 species-specific *P. subalpina* proteins. No significant length difference was observed among the high identity and the low identity genes (Additional file 2A). Moreover, mapping of the 4,233 proteins against the 454 EST dataset and RNA-Seq data showed that 2,833 of these gene models were covered by ≥50% of the ORF length with EST/RNA-Seq data (Additional file 2B). A classification scheme of the gene models based on different analysis is given in Fig. 2.

### Classification of repetitive elements

Repeats that could be classified as transposable elements (TEs) accounted for 5.7% of the genome sequence (Table 2). In contrast, 1.8% of repeats identified in RepatScout analysis could not be classified to any TE family. TE were dominated by Class I elements of the Copia and Gypsy families attributing for 55% of all TEs.





In contrast, class II TEs were less abundant and were dominated by Tc1 Mariner and Helitron elements. A large fraction of the repeat consensus sequences of the RepeatScout analysis could not be assigned to any TE family. TEs of all classes/families were evenly scattered throughout the genome of *P. subalpina* and no evidence for TE-rich islands was observed (Fig. 3). Besides the TEs, the genome of *P. subalpina* also included low-complexity sequences, tandem as well as simple sequence repeats (total 8.1% of the genome).

**Table 2** Classification and frequency of the most important transposable elements in the genome of *P. subalpina*

TE class	TE family	In % of TEs	Cumulative length in the genome [in kb]	Proportion of the genome
Class I	Gypsy	28.9%	1,452	2.1%
	Copia	26.1%	1,308	1.9%
	non LTR	4.6%	229	0.3%
Class II	Tc1	8.2%	411	0.6%
	Mariner			
	Helitron	3.9%	194	0.3%
	MITE	1.2%	59	0.1%
	hAT	0.6%	32	0.0%
	Mutator	0.4%	18	0.0%
	Not classified	not classified	26.2%	1,284

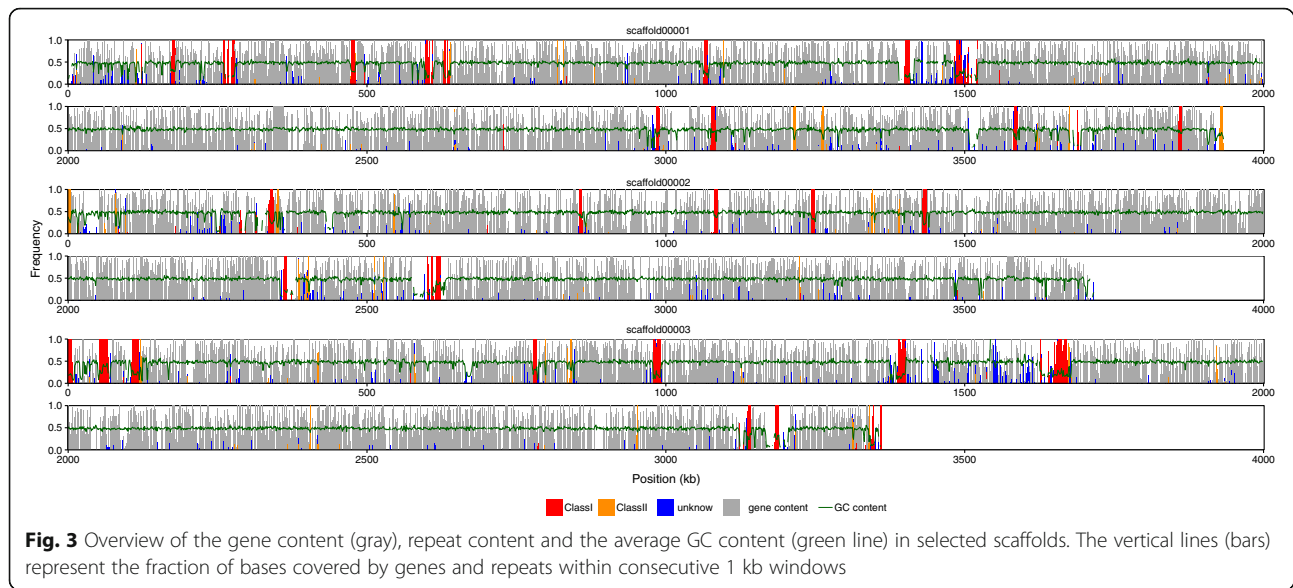
### Presence of RNAi pathway and evidence for RIP mechanisms

Homologs of DICER, ARGONAUT and RNA-dependent RNA polymerase genes were found in multiple copies in the genome of *P. subalpina* and comparison with other ascomycetes showed that the copy number of each of the three genes was higher than with other ascomycetes (Table 3). Similarly, several copies of cytosine methyltransferase gene of the Dnmt1 family were present in the genome of *P. subalpina* (Table 3). The cytosine methyltransferases encoded in *P. subalpina* were classified both based on the presence of InterPro domains and arrangements of the domains in comparison with *Neurospora crassa*, *Ascobolus immersus*, *Sclerotinia sclerotiorum* and *Botrytis cinerea* homologues. Whereas gene models PAC\_15881 and PAC\_01402 included only the C-5 cytosine methyltransferase domain (IPR001525) and showed homologies with the *RiD* gene of *N. crassa*, the other two genes (PAC\_07461, PAC\_02147) encoded the C-5 cytosine methyltransferase domain as well as BAH domains (IPR001025). Gene PAC\_07461 has a high similarity with *N. crassa Dim2* whereas gene PAC\_02147 has a high similarity with *Masc2* of *A. immersus*.

A clear difference in the dinucleotide frequencies was observed in repeat versus genomic control regions and the difference in dinucleotide frequencies was more pronounced in *P. subalpina* than in *S. sclerotiorum* (Fig. 4). TpA dinucleotides were significantly overrepresented in repeats whereas CpA/TpG were underrepresented suggesting a dominant mode of RIP for CpA to TpA mutations. In addition, repeat regions were generally rich in AT content as also ApA, ApT and TdT were more frequent in repeats than in non-repeat genomic regions (Fig. 3).

### The genome indicates horizontal gene transfer (HGT) events from bacteria co-occurring in the same habitat

Twenty one out of 163 genes, originally with a non-fungal best SIMAP hit, showed a skewed taxonomic distribution or higher bit scores with non-fungal taxa in BLAST searches against the NCBI non-redundant protein (nr) database (Table 4). Phylogenetic analysis for these 21 genes showed that they likely result of HGT as the gene trees significantly deviate from the expected species tree (for three examples see Fig. 5b-d; see also <http://purl.org/phylo/treebase/phylovs/study/TB2:S20196>). In contrast, the *RPB2* gene which was used as a control in the analysis, showed the expected ascomycetous phylogeny (Fig. 5a). In the majority of HGTs, the protein sequences taken from the phylogenetically closest non-fungal species were derived from soil-borne bacteria (i.e. *Collimonas arenae* or *Brevibacillus laterosporus*) or to live in the rhizosphere (i.e. *Frankia* sp.) and share therefore the same habitat as *P. subalpina*. In 12 of the phylogenies, one or few of the closest relatives were of fungal origin while most of the



remaining species were of prokaryotic origin. In several cases, *P. subalpina* clusters together with *O. maius* and/or *P. scopiformis* (Table 4 & Fig. 5d).

### Key secondary metabolite genes

The genome of *P. subalpina* encodes a large number of genes coding for putative secondary metabolite (SM) key enzymes (Table 5). 75.8% of the SM key genes in *P. subalpina* clustered with putative orthologous genes in the other 13 species without any obvious dominance of the closely related Leotiomycece species, i.e., *P. subalpina* shared most orthologous clusters with *Aspergillus flavus*. Numerous SM key genes are clustered in putative secondary metabolite loci including genes for acyl-, and methyltransferases, oxidoreductases, cytochrome P450 monooxygenases, or transcription factors (Table 5 & Additional file 3).

Among the eight genes for non-reducing Type I PKSs, two gene clusters were identified in *P. subalpina* probably involved in the melanin synthesis pathway. Whereas PAC\_11435 was placed in the same QuartetS cluster as *G. lozoyensis* *PKS1*, PAC\_07895 showed the highest

similarity with the gene of *A. alternata* encoding *alm* and the gene of *A. fumigatus* encoding *Alb1/PksP*. In addition, PAC\_07895 was placed in a second QuartetS cluster. Adjacent to both PKS genes, a putative hydroxynaphthalene reductase gene was found (PAC\_11432, PAC\_07896). However, the two putative scytalone dehydratases (PAC\_18365, PAC\_19872) in the *P. subalpina* genome were not located within either cluster. Presence of putative orthologous genes for the two PKS for the 13 ascomycetous species included in the comparative analysis (Table 6) showed that only melanized species were included in these two clusters. Further, genes of other Leotiomycece species, such as *S. sclerotiorum*, *B. cinerea* and *M. brunnea*, were represented in both clusters as *P. subalpina*. Two NRPS genes were identified that are likely encoding siderophore synthetases (PAC\_05248 and PAC\_13158). Besides key enzymes in secondary metabolism, the genome of *P. subalpina* also encodes a fucose-specific lectin (PAC\_07514) with similarities to the AAL protein of *Aleuria aurantia* that was shown to protect the fungus against predators and parasites [40].

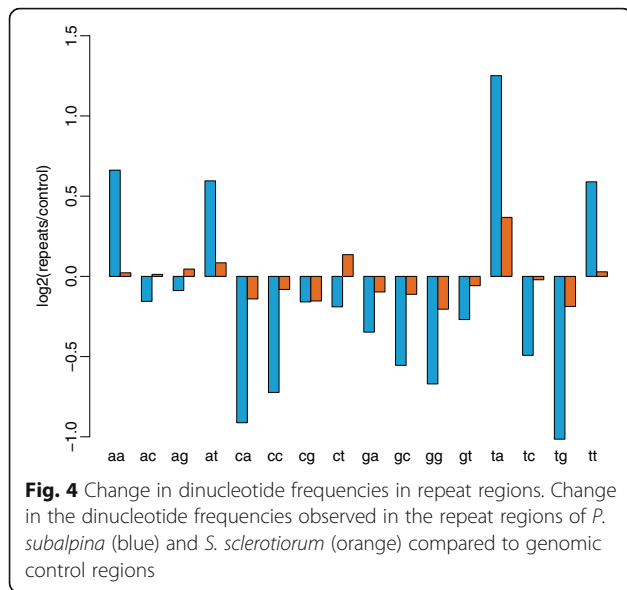
### Comparative genome analysis proves different lifestyles and enlarged gene families in *P. subalpina*

To explore the unexpectedly large gene set, the proteome was compared to 13 ascomycetous proteomes (Table 6) using orthologous cluster analysis, InterPro motif occurrence and a review of Carbohydrate-active enzymes (CAZymes). A total of 174,555 predicted proteins were grouped into 20,555 clusters of corresponding putative orthologous genes. Proteins of *P. subalpina* were present in 12,932 clusters (62.9%), significantly more than for any other investigated species. 1,408 clusters included proteins of all 13 species and

**Table 3** Presence of RNAi and RIP core proteins in different fungal genomes

Mechanism	Gene	Ps	Om	Ssc	Bc	Bg	Nc	Ca	Sc	Sp	Median
RNAi	Argo	4	2	2	3	2(3)	2	1	0	1	2
	Dicer	4	2	2	2	2	2	0	0	1	2
	RdRP	5	3	3	3	1	3	0	0	1	3
RIP	Dnmt1 family	4	3	3	2(3)	0	2	0	0	0	1

*Ps* *Phialocephala subalpina*, *Om* *Oidiodendron maius*, *Ssc* *Sclerotinia sclerotiorum*, *Bc* *Botrytis cinerea*, *Bg* *Blumeria graminis*, *Nc* *Neurospora crassa*, *Ca* *Candida albicans*, *Sc* *Saccharomyces cerevisiae*, *Sp* *Schizosaccharomyces pombe*

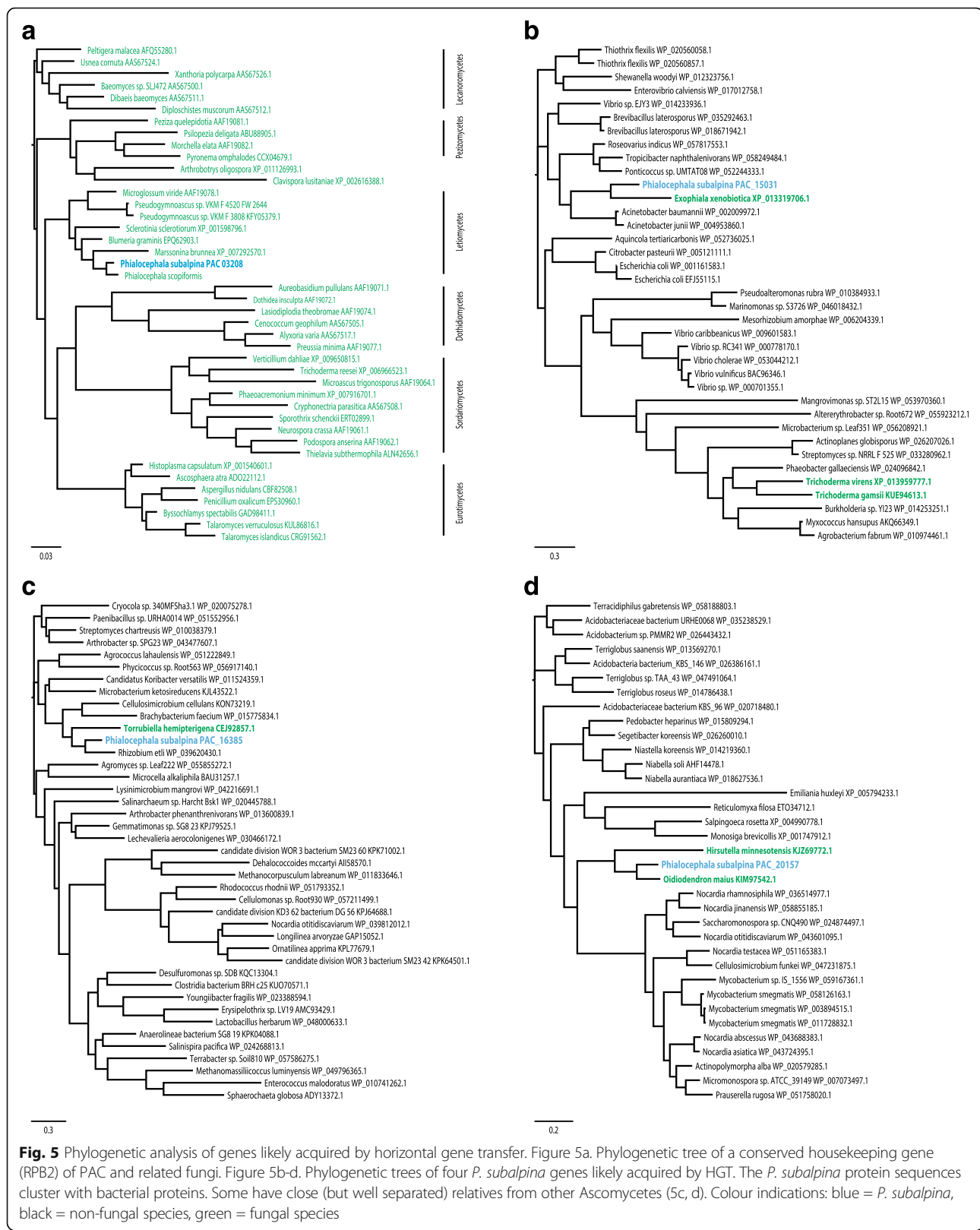


*P. subalpina* covering core functions in the primary metabolism, energy supply and cell cycle. 163 QuartetS clusters were mostly restricted to pathogens, whereas 61 clusters were predominantly found in saprophytic species. Principal component analysis based on these two datasets showed that *P. subalpina* was either placed within the pathogens or close to the saprophytes (Fig. 6a, b). Moreover, *P. subalpina* also shared the highest number of QuartetS clusters with the two mycorrhizal species (Table 7, Fig. 6c). The most frequent FunCat annotations for the pathogen- and saprophyte-related clusters showed that the secondary and C-metabolism was most often included but also several putative orthologous proteins classified as virulence and disease factors were recognized (Table 8).

A total of 6,556 InterPro accessions were annotated among the 14 species included in the analysis. A plateau of approx. 5,000–5,500 distinct InterPro accessions per

**Table 4** Genes of *P. subalpina* likely acquired by horizontal gene transfer from non-fungal species

Protein	Putative function	SignalP	Remarks	Order	Closest species	Habitat
PAC_13705	fatty acid desaturase			n.a.	<i>Sphaerosoma arcticus</i>	n.a.
PAC_15031	metallo-β-lactamase			Bacillales	<i>Brevibacillus laterosporus</i>	soil/water/insects
PAC_18575				Acidobacteriales	<i>Acidobacteria bacterium</i> KBS 146	soil
PAC_13085	β-lactamase			Caulobacteriales	<i>Phenylobacterium</i> sp.	plant root associated
PAC_16385	terpenoid cyclase			Rhizobiales	<i>Chelativorans</i> sp.	soil/rhizosphere
PAC_19778	glyoxalase		gene fragment	Burkholderiales	<i>Collimonas arenae</i>	soil
PAC_12936	glycoside hydrolase		with <i>P. scopiformis</i>	Burkholderiales	<i>Paraburkholderia</i> sp.	plant-associated bacteria
PAC_20157	hydrolase	secreted protein	with <i>O. maius</i>	Sphingobacteriales	<i>Pedobacter heparinus</i>	soil
PAC_01946	adenine deaminase		with <i>P. scopiformis</i>	Actinomycetales	<i>Frankia</i> sp. Eul1c	soil/plant symbiont
PAC_03408	kynurenine formamidase			Actinomycetales	<i>Frankia</i> sp. EUN1f	soil/plant symbiont
PAC_03301		secreted protein	with <i>O. maius</i> and <i>P. scopiformis</i>	Actinomycetales	<i>Streptomyces</i> sp.	mostly soil
PAC_17397	methyltransferase		with <i>O. maius</i>	Actinomycetales	<i>Streptomyces</i> sp.	mostly soil
PAC_19296				Actinomycetales	<i>Streptomyces</i> sp.	mostly soil
PAC_14321			with <i>O. maius</i>	Burkholderiales	<i>Curvibacter lanceolatus</i>	n.a.
PAC_07909	quinoprotein amine dehydrogenase	secreted protein	with <i>O. maius</i>	Burkholderiales	<i>Paraburkholderia udeis</i>	soil/rhizosphere
PAC_06755	carotenoid oxygenase		with <i>O. maius</i>	Actinomycetales	<i>Streptomyces</i> sp.	mostly soil
PAC_02359				Actinomycetales	<i>Mycobacterium avium</i>	water/soil
PAC_18364	peptidase		with <i>P. scopiformis</i>	Rhizobiales	<i>Bradyrhizobium</i> sp.	soil/plant symbiont
PAC_05940	oleate hydratase		with <i>P. scopiformis</i>	Spirochaetales	<i>Leptospira</i> sp.	n.a.
PAC_15362	lipase		with <i>P. scopiformis</i>	Actinomycetales	various species	n.a.
PAC_11525			with <i>P. scopiformis</i>	Ktedonobacteriales	<i>Ktedonobacter racemifer</i>	soil





**Table 5** Key secondary metabolite genes found in the genome of *P. subalpina*

Class	Type	Gene code	SM cluster <sup>a</sup>	Length (aa)	Domain structure <sup>b</sup>	Remarks
NRPS-like	L-alpha-aminoacidipate reductase	PAC_02750		1167	A-T-R	Lys2, L-lysine biosynthesis
NRPS-like	adenylating reductase	PAC_02944	8	1049	A-T-R	
NRPS-like	adenylating reductase	PAC_01324	3	1047	A-T-R	
NRPS-like	adenylating reductase	PAC_02391	6	1023	A-T-R	
NRPS-like	adenylating reductase	PAC_03877	10	1117	A-T-R	
NRPS-like	adenylating reductase	PAC_04959	12	1038	A-T-R	
NRPS-like	adenylating reductase	PAC_05733		1008	A-T-R	
NRPS-like	adenylating reductase	PAC_07003		1054	A-T-R	
NRPS-like	adenylating reductase	PAC_12811		1054	A-T-R	
NRPS-like	adenylating reductase	PAC_17490		1065	A-T-R	
NRPS-like	adenylating reductase	PAC_19780		1050	A-T-R	
NRPS-like	adenylating reductase	PAC_02202	5	1227	A-T-R-Kinase	
NRPS-like	adenylating reductase	PAC_08842		1345	A-T-R-KR	
NRPS-like	furanone synthetase	PAC_08910	17	1013	A-T-TE	
NRPS-like	quinone synthetase	PAC_14584		979	A-T-TE	
NRPS-like	unknown	PAC_19640		1394	A-T-DUF	
NRPS-like	unknown	PAC_12359		1655	A-T-DUF	
NRPS		PAC_05248	13	4676	A-T-C-A-T-C-A-T-C-T-C-T-C	Siderophore biosynthesis (Ferrichrome-like)
NRPS		PAC_13158		4838	A-T-C-A-T-C-T-C-A-T-C-T-C-T-C	Siderophore biosynthesis (Ferrichrome-like)
NRPS		PAC_15746		1639	A-T-C-T-C	
NRPS		PAC_20134		4692	C-A-T-C-A-T-C-A-T-C-A-T-C-T	
NRPS		PAC_16560		565	A-T-C	
NRPS		PAC_16746		1207	A-T-C	
NRPS		PAC_19282		1220	A-T-C	
PKS-NRPS		PAC_02326		3974	KS-AT-DH-MT-KR-T-C-A-T-R	
PKS-NRPS		PAC_08246	16	3965	KS-AT-DH-MT-KR-T-C-A-T-R	
Type I PKS	non-reducing PKS	PAC_01338	4	2542	SAT-KS-AT-PT-T-MT-TE	Citrinin biosynthesis-like
Type I PKS	non-reducing PKS	PAC_02435	7	2498	SAT-KS-AT-PT-T-MT-TE	Citrinin biosynthesis-like
Type I PKS	non-reducing PKS	PAC_03302		2223	KS-AT-DH-T-TE	
Type I PKS	non-reducing PKS	PAC_03589	9	1670	KS-AT-PT-T-TE	
Type I PKS	non-reducing PKS	PAC_07895	15	2171	SAT-KS-AT-PT-T-T-TE	Likely melanin biosynthesis
Type I PKS	non-reducing PKS	PAC_08751		2099	SAT-KS-AT-PT-T-T-TE	
Type I PKS	non-reducing PKS	PAC_10081		2125	SAT-KS-AT-PT-T-T-TE	
Type I PKS	non-reducing PKS	PAC_11435		2169	SAT-KS-AT-PT-T-T-TE	Likely melanin biosynthesis
Type I PKS	reducing PKS	PAC_00199	1	2694	KS-AT-DH-MT-ER-KR-T	Alternapyrone biosynthesis-like
Type I PKS	reducing PKS	PAC_00310	2	2283	KS-AT-DH-ER-KR-T	
Type I PKS	reducing PKS	PAC_01646		1687	AT-DH-MT-ER-KR-T	
Type I PKS	reducing PKS	PAC_04883	11	2239	KS-AT-DH-ER-KR-T	
Type I PKS	reducing PKS	PAC_06141		2258	KS-AT-DH-ER-KR-T	
Type I PKS	reducing PKS	PAC_10762		3173	KS-AT-DH-MT-ER-KR-T-Acyltransferase	

**Table 5** Key secondary metabolite genes found in the genome of *P. subalpina* (Continued)

Type I PKS	reducing PKS	PAC_11350	3203	KS-AT-DH-MT-ER-KR-T-Acyltransferase	
Type I PKS	reducing PKS	PAC_14253	2590	KS-AT-DH-MT-ER-KR-T	
Type I PKS	reducing PKS	PAC_14645	2274	KS-AT-DH-ER-KR-T	
Type I PKS	reducing PKS	PAC_16276	3140	KS-AT-DH-MT-ER-KR-T-Acyltransferase	
Type I PKS	reducing PKS	PAC_17799	2411	KS-AT-DH-ER-KR-T	Lovastatin-diketide synthase-like
Type I PKS	reducing PKS	PAC_19082	2581	KS-AT-DH-MT-ER-KR-T	
Type I PKS	reducing PKS	PAC_19990	2970	KS-AT-DH-MT-KR-T-C	Lovastatin-nonaketide synthase-like
Type I PKS	reducing PKS	PAC_10712	1430	KS-AT-TH-KR-T	6-Methylsalicylate synthase-like PKS
Type III PKS		PAC_02116	479	IPR012328/IPR001099	
DMATS		PAC_15749	538	IPR017795	
Terpene		PAC_05884	483		
Terpene		PAC_13844	593	IPR017825/IPR017825/IPR002060	Squalene synthase-like
Terpene		PAC_15298	557		
Terpene		PAC_01018	338	IPR008949	
Terpene		PAC_04028	331	IPR008949	
Terpene		PAC_11164	466	IPR008949	
Terpene		PAC_12198	392	IPR008949	
Terpene		PAC_16221	336	IPR008949	

<sup>a</sup> for details on the putative SM gene clusters see Additional file 3

<sup>b</sup> Abbreviations for domains: A adenylation, AT acyltransferase, C condensation, DH dehydratase, DUF terminal NRPS domain of unknown function, ER enoyl reductase, KR ketoreductase, KS ketosynthase, MT methyltransferase, PT product template, R reduction, SAT starter unit-ACP-transacylase, T thiolation (=acyl- or aryl- or peptidylcarrier protein), TE thioesterase, TH thiohydrolase

species was observed when plotting the number of distinct InterPro accessions per species against the total number of annotated InterPro accessions per species (Additional file 4). *P. subalpina* is represented by 5,232 distinct InterPro accessions, and the highest number was observed in the two saprophyte species *T. reesei* (5,529) and *A. flavus* (5,303). In contrast, mycorrhizal species and the obligate biotrophic pathogen *B. graminis* included a significantly lower number of InterPro accessions. 963 (14.6%) InterPro accessions were significantly overrepresented in the *P. subalpina* genome compared to the average count for the other 13 ascomycetous species and 386 (5.9%) InterPro accessions were encoded >3x in the *P. subalpina* genome. Mapping of the overrepresented InterPro accessions against the gene ontology (GO) showed that catabolic/metabolic processes, transporters, and InterPro accessions involved in binding events were most frequent (Table 9). 411 (43%) of the InterPro accessions were not linked with GO annotations and *P. subalpina* included among others 534 gene models with the fungal heterokaryon incompatibility domain (IPR010730), 549 gene models with the major facilitator superfamily domain (IPR020846) and 328 gene models with ankyrin repeats (IPR020683). Moreover, several classes of CAZymes contained InterPro accessions without GO annotation such as IPR017853 (glycoside

hydrolases) as well as IPR011050 (Pectin lyase fold/virulence factor) and IPR012334 (Pectin lyase fold). Only, 53 InterPro accessions were significantly underrepresented in *P. subalpina*. The underrepresented InterPro accessions frequently showed uneven distributions in the 13 ascomycetous genomes, and some of the accessions (IPR013762, IPR000477, IPR001584) are most probably related to transposable elements. Similarly to PCA based on QuartetS analysis, PCA based on InterPro accessions overrepresented in pathogenic and saprophytic species placed *P. subalpina* either in the pathogenic cluster or the saprophytic cluster (Fig. 7a, b).

Enrichment of GO terms for the overrepresented InterPro accessions in *P. subalpina* compared to average number of InterPro accessions found in 13 ascomycetous genomes. Of the 963 overrepresented InterPro accessions 57% mapped to one or several GO terms and the table summarizes the GO terms with the highest numbers of distinct InterPro accessions.

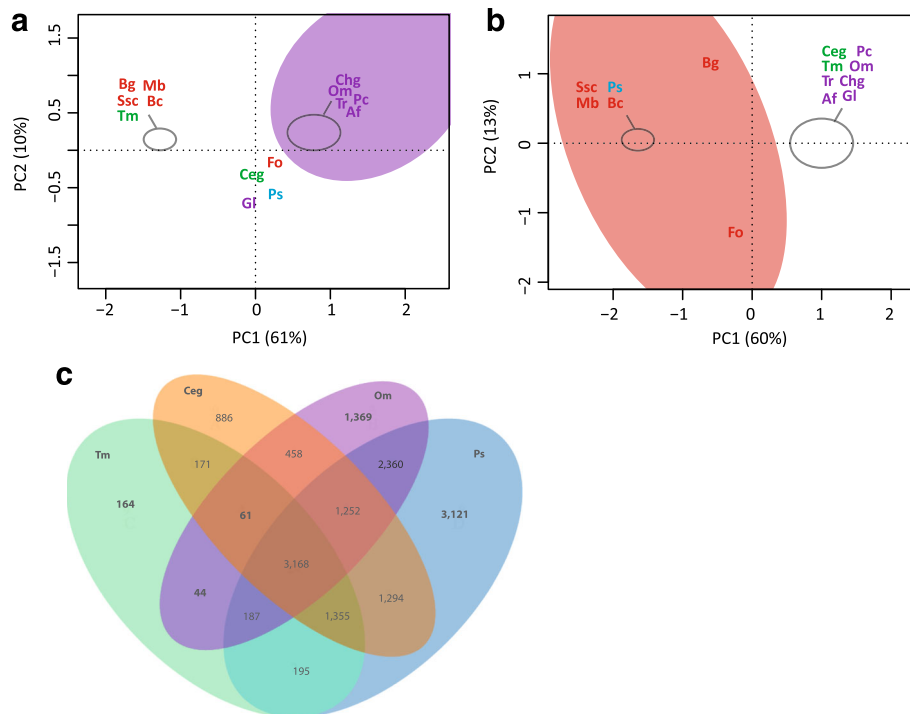
Eight hundred eighty one gene models in *P. subalpina* were classified in 138 different CAZyme families resulting in 998 CAZyme modules, which are significantly more than observed in the other ascomycetous species. The second most frequently found CAZyme modules (883 modules) were encoded by *F.*

**Table 6** Species included in the comparative genomic analysis

Species	Code	Lifestyle	Class	Order	Family	Reference
<i>Botrytis cinerea</i>	Bc	Pathogen (necrotroph)	Leotiomycetes	Helotiales	Sclerotiniaceae	[55]
<i>Sclerotinia sclerotiorum</i>	Ssc	Pathogen (necrotroph)	Leotiomycetes	Helotiales	Sclerotiniaceae	[55]
<i>Blumeria graminis</i>	Bg	Pathogen (obligate biotroph)	Leotiomycetes	Erysiphales	Erysiphaceae	[124]
<i>Marssonina brunnea</i>	Mb	Pathogen (hemi-biotroph)	Leotiomycetes	Helotiales	Dermateaceae	[125]
<i>Fusarium oxysporum</i> f. sp. <i>lycopersici</i>	Fo	Pathogen (hemi-biotroph)	Sordariomycetes	Hypocreales	Nectriaceae	[56]
<i>Aspergillus flavus</i>	Af	Saprophyte (soil & rhizosphere)	Eurotiomycetes	Eurotiales	Trichocomaceae	[126]
<i>Trichoderma reesei</i>	Tr	Saprophyte (soil) <sup>a</sup>	Sordariomycetes	Hypocreales	Hypocreaceae	[127]
<i>Chaetomium globosum</i>	Chg	Saprophyte (soil/plant debris)	Sordariomycetes	Sordariales	Chaetomiaceae	[128]
<i>Penicillium chrysogenum</i>	Pc	Saprophyte (soil)	Eurotiomycetes	Eurotiales	Aspergillaceae	[129]
<i>Glarea lozoyensis</i>	Gl	Saprophyte(soil) <sup>b</sup>	Leotiomycetes	Helotiales	Helotiaceae	[130]
<i>Oidiodendron maius</i>	Om	Saprophyte (peat bog)/ericoid mycorrhizal	Leotiomycetes	Leotiomycetes incertae sedis	Myxotrichaceae	[32]
<i>Tuber melanosporum</i>	Tm	Ectomycorrhizae	Pezizomycetes	Pezizales	Tuberaceae	[66]
<i>Cenococcum geophilum</i>	Ceg	Ectomycorrhizae	Dothideomycetes	Pleosporomycetidae incertae sedis	Gloniaceae	[131]
<i>Phialocephala subalpina</i>	Ps	Root endophyte	Leotiomycetes	Helotiales	Dermateaceae	this paper

<sup>a</sup>in contrast to other *Trichoderma* species, *T. reesei* does not show mycoparasitism

<sup>b</sup>ITS sequences from environmental samples often derived from soil/rhizosphere



**Fig. 6** Characterization of *P.* clusters enriched for pathogens or saprophytes. Principle component analysis (PCA) based on the presence/absence of species in putative orthologous gene clusters derived from QuartetS analysis. Figure 6a. PCA analysis based on 61 gene clusters enriched for saprophytic species. Figure 6b. PCA analysis based on 163 clusters enriched in pathogens. Figure 6c. Venn diagram showing the distribution of orthologous gene clusters for the two ectomycorrhizal species *Tuber melanosporum* and *Cenococcum geophilum*, the saprophyte/ericoid mycorrhizal species *Oidiodendron maius* as well as *P. subalpina*. Abbreviations of species are given in Table 2. Color code: green: mycorrhizal species, red: plant pathogens, purple: soil saprotrophs and blue: *P. subalpina*

**Table 7** Number of putative orthologous gene clusters shared by the two mycorrhizal species with the other species included in the analysis

Species	Total proteins	Total clusters	Bc	Ssc	Bg	Mb	Fo	Af	Tr	Chg	Pc	Gl	Om	Tm	Ceg	Ps
<i>Tm</i>	7,496	5,345	4,540	4,646	4,039	4,593	4,636	4,501	4,456	4,352	4,505	3,073	3,460	-	4,755	4,905
<i>Ceg</i>	14,561	8,645	5,797	5,795	4,492	5,556	6,038	5,795	5,576	5,439	5,704	3,733	4,939	4755	-	7,069

For abbreviation of species see Table 2

*oxysporum*. The genome of *P. subalpina* was especially rich in genes coding for glycoside hydrolases (471), glycosyltransferases (150) and redox enzymes (auxiliary activities enzymes, 157). Principal component analysis based on the frequency of CAZyme modules involved in plant cell wall degradation (PCWDEs) such as cellulose, hemicellulose, pectin, cutin, and enzymes likely acting on different substrates separated *P. subalpina* from all genomes (Fig. 8). Especially modules involved in pectin breakdown were encoded in high copy numbers in the genomes of both *P. subalpina* and *F. oxysporum* but *P. subalpina* included also a higher number of modules for cellulose and hemicellulose degradation (Additional file 5). The ectomycorrhizal species *T. melanosporum* and *C. geophilum* as well as the obligate biotrophic species *B. graminis* were separated due to the small number of PCWDEs (Fig. 8).

## Discussion

In the present study we sequenced the genome of the root endophyte *Phialocephala subalpina* belonging to the *Phialocephala fortinii* s.l. – *Acephala applanata* species complex one of the most prevalent species in forest ecosystems. By comparative genomic analysis with the gene inventory of 13 other ascomycetous species we show that *P. subalpina* links pathogenic and saprophytic lifestyles.

## Genome expansion due to a large set of distinct genes in gene families

With a genome size of approximately 69.7 Mb, the *P. subalpina* genome is significantly larger than the average genome size of previously sequenced ascomycetous species [41]. Genome expansions can be caused by various events including (i) genome duplications, (ii) invasion of autonomous elements such as TEs and expansion of repetitive sequences such as microsatellites and tandem repeats into the genome, (iii) the number and length of introns and/or (iv) the expansion of the gene inventory [42, 43]. No evidence for large segmental duplications were observed in the *P. subalpina* genome by genome-wide alignments using CoCoNUT [44] and the frequency of repeats in general and TEs in special was small, considering the genome size (Additional file 6). Proliferation of TE is counteracted by three processes which act at different stages of TE proliferation. Repeat induced point mutations (RIP) act on the DNA level by introducing C to T transitions and reversing CpG to TpA dinucleotides in repeated regions [45, 46]. MIP (Methylation induced premeiotically) *de-novo* methylates repeated DNA sequences [47] and RNA interference (RNAi) suppresses TE proliferation either by heterochromatin assembly or small interfering RNAs, which silence TE transcripts [48–50]. Indirect evidence that a RIP mechanism is or was active in *P. subalpina* stems from the skewed dinucleotide distribution in repeat

**Table 8** Results of FunCat enrichment analysis for the orthologous gene clusters including *P. subalpina* that were mainly restricted to pathogenic and saprophytic species

FunCat ID <sup>a</sup>	Description	Pathogen-related proteins	Saprophyte-related proteins	Total proteins in QuartetS clusters	Total proteins in Ps genome
01.05	C-compound and carbohydrate metabolism	12	6	18	2473
01.20	secondary metabolism	9	7	16	2652
01.06	lipid, fatty acid and isoprenoid metabolism	7	4	11	1387
20.03	transport facilities	7	3	10	1216
11.02.03.04	transcriptional control	7	2	9	923
20.01.03	C-compound and carbohydrate transport	5	4	9	504
16.01	protein binding	6	2	8	1904
16.03.01	DNA binding	6	2	8	577
20.09.18.07	non-vesicular cellular import	5	3	8	428
32.05.05	virulence, disease factors	5	-	5	350
01.06.06.11	tetracyclic and pentacyclic triterpenes cholesterol, steroids and hopanoids metabolism	-	2	2	173

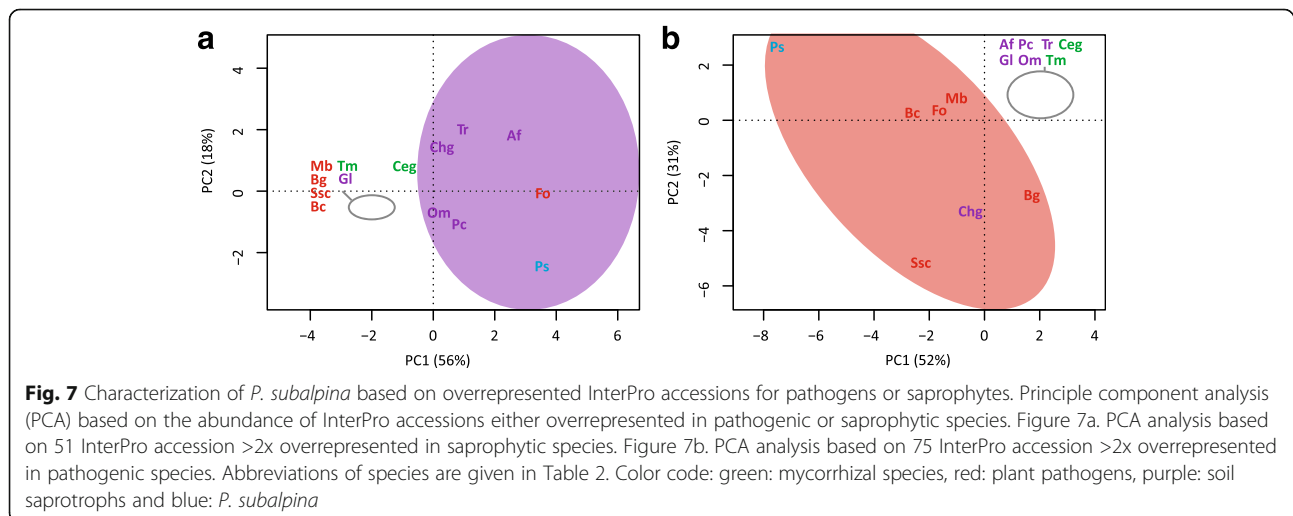
<sup>a</sup>see <http://mips.helmholtz-muenchen.de/funecatDB/>

**Table 9** Enrichment of GO terms for the overrepresented InterPro accessions encoded in *P. subalpina* (each InterPro accession was only considered once per gene model)

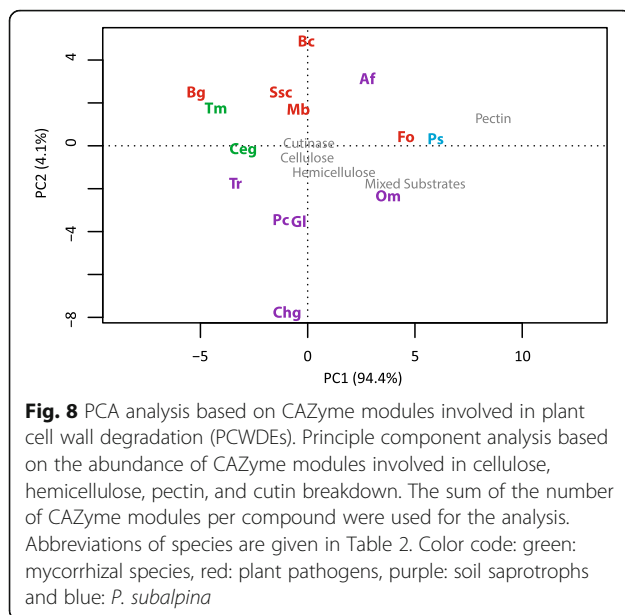
GOID	GO description	Number of InterPro accessions	Total proteins in Ps genome
NA	no GO annotation	411	6426
GO:0055114	oxidation-reduction process	103	1138
GO:0016491	oxidoreductase activity	54	1002
GO:0005975	carbohydrate metabolic process	43	368
GO:0008152	metabolic process	38	808
GO:0003824	catalytic activity	35	733
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	30	235
GO:0016020	membrane	28	459
GO:0016021	integral component of membrane	26	912
GO:0005515	protein binding	24	1065
GO:0005524	ATP binding	22	501
GO:0006508	proteolysis	22	209
GO:0055085	transmembrane transport	19	913
GO:0008270	zinc ion binding	17	882
GO:0016787	hydrolase activity	14	208
GO:0050660	flavin adenine dinucleotide binding	14	184
GO:0020037	heme binding	13	280

regions. RIP is active during the sexual cycle [51] but no sexual stage is known in *P. subalpina*. However, several lines of evidence suggest that sexual reproduction regularly occurs as (i) most PAC populations showed no gametic disequilibrium, (ii) mating types do not deviate from a 1:1 ratio in PAC populations and (iii) strong purifying selection was recorded in the mating type loci [52]. Moreover, teleomorphs are known for phylogenetically closely related species such as *Phaeomollisia piceae* or *Mollisia* spp. [53]. Therefore, it seems likely that field studies have overlooked the teleomorph of PAC species so far [52].

Common to the RIP/MIP process is that cytosine DNA methyltransferase of the Dnmt1 family play a key role [45, 54]. Homologs of cytosine DNA methyltransferases were present in *P. subalpina* including two gene models closely related to *N. crassa RiD* and one protein related to *N. crassa Dim2*. A fourth gene model found in *P. subalpina* (PAC\_02147) was closest related to MASC2 of *Ascobolus immersus* which was placed in a cluster exclusively with basidiomycete species in the study of Amselem et al. [54]. However, additional analysis showed that also other Leotiomycete species included MASC2 homologs (*B. cinerea*







(CCD54489), *Sclerotinia borealis* (ESZ90943) and *M. brunnea* (XP\_007291083)) adding some additional notable exceptions of ascomycetous proteins in this cluster. Beside the indirect evidence of RIP/MIP, *P. subalpina* included also RNAi key enzymes such as Dicer, Argonaute and RdRP in multiple copies. Based on these findings we hypothesize that *P. subalpina* has a well-developed arsenal of defense mechanisms in place to counteract the proliferation of TEs which may explain the comparative low frequencies of TEs.

No significant differences in intron length and the average intergenic distance were observed. However, the gene inventory was one of the largest among ascomycetes with 20,173 annotated gene models resulting in 2,500 to >10,000 more gene models than in other ascomycetes. In the light of the broad host range and the broad geographical distribution of species belonging to the *P. fortinii* s.l. – *A. applanata* species complex [1, 3, 5, 10] an enlarged gene repertoire could be expected. However, even compared to other fungal species with broad host ranges such as *S. sclerotiorum*, *B. cinerea* and *F. oxysporum* the number of gene models is large [55, 56]. The high numbers of gene models may be the result of the annotation of a large fraction of very short gene models [55] or gene models including TE fragments. In *P. subalpina*, no significant deviation in protein length distributions for low identity genes was observed and a large fraction of the low identity genes were covered by ESTs. Moreover, particular attention was paid to mask TEs. Therefore, the expansions in the gene inventory of *P. subalpina* do not result from annotation artefacts. Several factors contributed to the high number of gene models found in *P. subalpina*. First of all, *P. subalpina*

showed a slightly higher fraction of putative paralogous genes than observed on average for the other 13 ascomycetous species. *F. oxysporum* showed the highest fraction of putative paralogous genes and this observation is due to large segmental genome duplication [56]. Secondly, a significant number of gene models in *P. subalpina* were species-specific as they showed no significant hits in SIMAP analysis. The high fraction of species-specific genes may be the result of the lifestyle of *P. subalpina* as well as the missing genome data of closely related species in the *Loramyces* – *Vibrissea* clade [53, 57]. Indeed, if the supposedly species-specific genes of *P. subalpina* were blasted against the recently announced genome of the closely related *P. scopiformis* [37], 28% of the gene models showed significant blast hits ( $>1.0 E^{-14}$ ). Thirdly, significantly more InterPro accessions were annotated in *P. subalpina* including 13,074 gene models than in any other species used for comparison. The high number of annotated InterPro accessions in the gene inventory did, however, not result in an expanded functional catalogue as the number of distinct InterPro accessions per species reached a plateau at approx. 5,200–5,500 distinct accessions indicating that “more of the same” is present in *P. subalpina*.

Our analyses show that the gene inventory of *P. subalpina* was also expanded to some extent by HGT from non-fungal organisms. A systematic analysis of the acquired prokaryotic genes by 60 fungal genomes by Marcet-Houben & Gabaldón [58] showed that species in the Pezizomycotina exhibited a high incidence of inter-domain HGT including between 4 and 63 proteins per species. The 21 HGT events found in *P. subalpina* likely represents a lower limit of events and closer inspection of uncertain candidates will likely reveal additional HGT events. Moreover, HGT events from other fungal species were shown to be another significant source for HGT [59, 60]. The phylogenetically closest protein sequences often originated from species in the Burkholderiales or Actinomycetales that colonize soil and/or roots, i.e., they share their habitat with *P. subalpina* and rendering a HGT event likely.

In almost half of the cases, the HGT event was unique or occurred recently. In contrast, also older HGT events were detected that were characterized by the presence of multiple fungal species within a clade of non-fungal species [60]. Interestingly, *O. maius* and *P. scopiformis* shared several of the 21 studied HGT events. It is possible that the HGT event occurred in a common ancestor as both fungi are related to *P. subalpina* (Fig. 6a). Alternatively, it might be possible that these genes were introduced twice independently for *O. maius* as *O. maius* can also be found in similar habitats as *P. subalpina*. It is associated with roots of ericaceous shrubs and involved in the decomposition of sphagnum peat [61].

Genes transferred by HGT can drive important evolutionary innovations as shown for plant pathogens [60, 62–64]. This was also observed in *P. subalpina* as several proteins were hydratases or peptidases. Notably, two  $\beta$ -lactamases were included that are involved in the detoxification of  $\beta$ -lactam antibiotics [65] and may result in a competitive advantage against other microbes.

### The chameleonic genome of *P. subalpina*

The ecological role of members of the *P. fortinii* s.l.-*A. applanata* species complex is still poorly understood. They were described as beneficial, neutral or pathogenic for different hosts, growing conditions and fungal strains [5, 20] and even a role in mycorrhizal associations was hypothesized [21]. In order to shed light on the lifestyle of *P. subalpina* we compared the genome against 13 genomes of other ascomycetous species with different lifestyles. All our analyses showed that the gene inventory of *P. subalpina* supports multiple lifestyles. First, *P. subalpina* shared more putative orthologous genes with the two ectomycorrhizal species included in the analysis than any other species indicating some affinities with the two species. Nevertheless, an important difference is evident. EcM fungi as well as obligate biotrophic pathogens are generally characterized by a reduced gene inventory especially for plant cell wall degrading enzymes (PCWDEs) [32]. However, *P. subalpina* encodes a high number of PCWDEs. Moreover, several EcM fungi and obligate biotrophic pathogens show genome expansions due to TE proliferation [30, 66, 67] which was absent in *P. subalpina*.

In contrast to EcM fungi, saprophytes and necrotrophic/hemibiotrophic pathogens are well endowed with enzymes involved in the degradation of plant material [32]. The main difference between these two groups is that the pathogenic species must have a specific gene inventory allowing them to invade hosts and overcome plant immune response, i.e., pathogen-associated molecular patterns (PAMP) or damage-associated molecular patterns (DAMP) [68, 69]. Whereas necrotrophic pathogens kill the host tissue by secreting effectors like toxins and/or proteins, hemibiotrophic pathogens grow intracellularly, and form specialized structures such as haustoria for nutrient uptake [70]. Recent comparative genome analysis showed that there are only few genes associated with plant pathogens that are absent in non-pathogens [33, 71]. However, domains overrepresented in necrotrophic/hemibiotrophic pathogenic species compared to saprophytes were identified [71]. When our dataset was enriched for overrepresented InterPro accessions in pathogens only few of the accessions were in accordance with Soanes et al [71]. However, a closer look showed that also of the accessions listed by Soanes et al [71] were also overrepresented in our dataset yet with a factor <2x. Irrespective of which InterPro accession list was used for analysis, *P. subalpina* was placed close to the

pathogenic species (Fig. 7b and Additional file 7), indicating the robustness of the analysis. The list included genes with protease/peptidase domains, cutinases, pectinases, genes including a necrosis inducing domain or genes with chitin-binding modules. Many of these gene classes were shown to play a pivotal role in plant-pathogen interactions [68, 70]. For example, PAMP-induced host chitinases can either be overcome by the action of secreted proteases [72] or the secretion of LysM effectors that may be coupled with a chitin-binding module [73, 74]. The finding that *P. subalpina* includes the repertoire of pathogenic species fits well with the recent host-fungus interactions studies showing that PAC strains behave along the antagonism-mutualism continuum and are localized rather towards antagonistic interactions as the colonization results in reduced biomass accumulation of the host [22]. However, strong strain-specific differences were observed in the outcome of the interaction with some strains killing the majority of the seedlings and others only marginally affecting the host [22] and future studies are needed to understand the underlying mechanisms. Despite the pathogenic gene repertoire observed in *P. subalpina*, host defense mechanisms such as lignin formation and cell wall appositions are rarely observed during the invasion of PAC strains indicating that PAC can manipulate/suppress the plant immune response [25]. Although the precise mechanisms how *P. subalpina* suppresses host induced defense mechanisms are unknown, effectors such as small secreted proteins (SSPs) predicted in the genome of *P. subalpina* may be candidates as they were shown to function as effectors in plant-fungal interactions [70, 75] and genome-wide differential gene expression studies during host colonization will help to identify possible effectors.

Beside the pathogen-related gene repertoire, our analysis shows that *P. subalpina* has also affinities with saprophyte-specific genes indicating that the species includes the signature of both lifestyles in the genome. A very similar positioning in the analysis was observed for *F. oxysporum*. Indeed, *F. oxysporum* not only includes the >70 pathogenic variants but non-pathogenic strains of *F. oxysporum* were also isolated as endophytes from asymptomatic roots [76–78].

Both species share a high number of  $\beta$ -lactamase/ $\beta$ -lactamase-related genes with saprophytes involved in the detoxification of  $\beta$ -lactam antibiotics.  $\beta$ -lactamases are well known fungal defense effector proteins [79], and detoxifying  $\beta$ -lactam antibiotics help to persist against antagonists. Besides the  $\beta$ -lactamase genes, those coding for sulfatases and different hydrolases were also enriched in saprophytic species including *P. subalpina* and *F. oxysporum*. In contrast to most pathogenic species, *P. subalpina* showed also an enlarged repertoire of PCWDEs involved in cellulose and hemicellulose breakdown as well as proteins with auxiliary activities supposed to be involved in lignin

breakdown. Indeed, a strain of *P. fortinii* s.l. was shown to cause soft rot in autoclaved wood of beech and conifer species indicating that members of the PAC can degrade lignin, cellulose and hemicellulose [1].

#### Why does *P. subalpina* stand out of the crowd?

The answer of two fundamental questions is still pending: (i) why are PAC species so amazingly successful in colonizing their hosts and (ii) does the host also benefit from the colonization by PAC. At least the second question may be answered by the genome sequence of *P. subalpina*.

Although often hypothesized, endophytic fungi of forest trees were rarely shown to be mutualistic for their hosts [4]. However, the closely related needle endophyte *P. scopiformis* was shown to produce rugulosin, a potent secondary metabolite against herbivory [80, 81]. Similarly, interaction studies using pathogens (*Phytophthora pluri-vora* and *Elongisporangium undulatum*), *P. subalpina* strains and *P. abies* seedlings showed that some of the *P. subalpina* strains effectively reduced mortality and disease intensity caused by the pathogens [82]. In addition, secondary metabolites were identified in PAC species that inhibited *Phytophthora* spp. [82], and the genome of *P. subalpina* encodes a high number of secondary metabolite key enzymes. Some of these are compatible with known pathways and products, such as melanin or ferrichrome-like siderophores. For example, enzymes PAC\_05248 and PAC\_13158 resemble SidC-like (=type II) and NPS1-like (=type IV) synthetases, i.e., two distinct types of ferrichrome-like siderophore-producing enzymes [83]. Also, two non-reducing type I PKSs of *P. subalpina* flanked by hydroxynaphthalene reductases likely involved in the melanin synthesis were recognized. Whereas gene PAC\_1135 was placed in the same putative orthologous gene cluster as PKS1 of *G. lozoyensis*, gene PAC\_07895 was placed in a second orthologous gene cluster and showed high similarities with the *alm* gene of *A. alternata*/the *PksP/Alb1* gene of *A. fumigatus*. All three proteins were experimentally shown to be involved in the melanin production [84–86]. Several Leotiomycece species included genes in both clusters (i.e. *B. cinerea*, *S. sclerotiorum*, *M. brunnea*), whereas melanized non-helotialean species and *G. lozoyensis* were included in one of the clusters (i.e. *C. geophilum*, *O. maius* and *C.globosum*). Moreover, no gene models of non-melanized species were included in one of these two clusters. To the best of our knowledge this is the first report for the duplication of the melanin pathway in ascomycetous fungi, although the redundancy of important secondary metabolite genes was reported previously [87, 88]. Still, the products of many of the secondary metabolite key enzymes and clusters remain unresolved. However, they may represent the chemical language of *P. subalpina* to interact via small molecules

with host plants and other microorganisms in the rhizosphere.

Although the biomass and turnover rates of fine roots in forest ecosystems largely depend upon tree species, age of forest stands and climate, estimates indicate that as much as 30% of the net primary production is used for fine root production [89]. In an *Abies alba* stand for example >6.0 t/ha fine roots biomass with a cumulative length of > 20,000 km/ha was estimated [90] showing the importance of fine roots as carbon source. Given the fact that species of the *P. fortinii* s.l. – *A. applanata* species complex dominate many of the endophytic assemblages in fine roots of temperate and boreal forests [10], PAC species have access to a very substantial carbon source. *P. subalpina* can already colonize the future carbon source by entering healthy fine roots and may then switch to the saprophytic lifestyle as soon as the roots die off. Indeed, the signature of both lifestyles was observed in the genome although studies about the importance of PAC species in the fine root turnover are missing. Moreover, the fungus can escape the highly competitive soil community by colonizing the roots [91].

#### Conclusions

The analysis of a globally distributed root endophyte allowed for detailed insights in the gene inventory and genome organization of a yet largely neglected group of organisms. Our analysis showed that the genome of *P. subalpina* has a versatile genome including genes for both a pathogenic and saprophytic lifestyle but showed also some affinities with ectomycorrhizal species. The degree of pathogenicity among strains of *P. subalpina* is high, as observed in *F. oxysporum*. In *F. oxysporum* pathogenicity is driven by mobile pathogenicity chromosomes [56]. Re-sequencing of multiple strains of *P. subalpina* will help identify the molecular basis of its pathogenicity. In addition, one central question will be to understand the evolutionary trajectory of PAC, i.e. whether PAC will become more pathogenic in future which would have a severe effect on forest ecosystems health, or whether the PAC-host interaction gets less antagonistic.

#### Methods

##### Selection of *Phialocephala subalpina* strain and DNA isolation

Strain UAMH 11012 (UAMH Centre for Global Microfungal Biodiversity, Toronto, Canada) was used for genome sequencing. The strain was originally isolated as single hyphal tip culture from *P. abies* fine roots in an undisturbed forest in Switzerland [14] and was classified using multiple classes of molecular markers [16, 38, 92]. The strain was grown in malt extract broth (50 ml 2% (w/v)) for 10 days at 20 °C under constant shaking.

Mycelium was harvested by filtration, lyophilized, and total DNA was isolated using a CTAB-based protocol [93]. Strain identity was verified using microsatellite analysis [94] before the genome sequencing.

#### Sequencing, assembly and gap closing

A whole genome shotgun strategy (WGS) on the Roche/454 GS FLX (454) was used and sequencing was performed at the Functional Genomic Centre of University Zürich/ETH Zürich (FGCZ). In total, 1.3 Mio shotgun reads as well as 2.3 Mio reads from one 3 kb paired-end library, 6.2 Mio reads of three 8 kb paired-end libraries and 309,909 reads of a 20 kb library were included in the assembly. The assembly was performed using newbler 2.5 with a minimum overlap of 50 bases and 98% sequence similarity. We noticed that newbler tends to open gaps due to the high number of pair-end reads in the assembly. Therefore, reads mapping to both sides of the gaps were identified and gaps were closed after manual validation.

#### 454 sequencing of a normalized EST library

The same strain was used to generate a normalized EST library. The fungus was pre-cultivated in 50 ml of 2% malt broth (20 g l<sup>-1</sup> malt extract; Difco) for 14 days at 20 °C under constant shaking. Then, the mycelium was homogenized with a blender for 30 s and 5 ml of the homogenized mycelium was transferred to new 50 ml of 2% malt broth (20 g l<sup>-1</sup> malt extract; Hefe Schweiz). After 48 h the actively growing mycelium was harvested and immediately frozen in liquid nitrogen. Total RNA was isolated from approx. 75 mg fresh mycelium using the RNeasy plant mini kit (Qiagen, Hombrechtikon, Switzerland). Full-length cDNA was synthesized using the MINT kit (evrogen, Moscow, Russia) with a degenerated poly-T primer (5'-AAGCA GTGGTATCAACGCAGAGTAC (T)<sub>4</sub>G(T)<sub>9</sub>C(T)<sub>10</sub>VN-3') during the first strand cDNA synthesis [95] and polTM1 (5'-AAGCAGTGGTATCAACGCAGAGTACT TTTGCTTTTGTCTGTTTCTTTTVN-3') for the generation of dsDNA. The cDNA was normalized using the TRIMMER kit (evrogen) and the library was sequenced on the 454. Resulting reads were filtered for chimeras and then a whole transcriptome assembly was performed in newbler 2.3 with a minimum overlap of 50 bases and 98% sequence similarity.

#### Repeat library construction

A repeat library was constructed based on the final assembly of the genome (see Additional file 8). In brief, putative repeat sequences were derived from RepeatScout analysis [96]. Low-complexity sequences and microsatellites were removed using the default filtering options in RepeatScout. In addition, sequences <50 bases

and with <10 hit on the genome were excluded from further analysis. The remaining sequences were clustered using blastclust (-S 90 -L 0.9 -b F -p F) and only one sequence per cluster was used for further analysis. Blastx was used to exclude any sequences in the library not belonging to TEs (i.e. HET domain containing proteins or ubiquitin-like proteins). This draft library was then mapped against the genome using RepeatMasker and consensus sequences of complete TEs were derived. Finally, sequences were classified according to the systematic of Wicker et al. [97]. The manually curated TE library was used for genome masking before annotation.

#### Annotation of the *P. subalpina* genome

The annotation strategy is presented in Additional file 9. In brief, a reference dataset of 1,089 gene models/proteins covered by full-length EST sequences was established and used as training dataset for Augustus [98]. *Ab-initio* gene prediction was performed on the masked genome using GeneMark-ES [99], Augustus [100] and FGENESH (*Neurospora* and *Ustilago* matrices). GenomeThreader [101] was used to calculate spliced alignments for *P. subalpina* ESTs and protein data from related fungal species. The program was run based on a *P. subalpina* specific splicing model trained using 454 EST dataset with the software BSSM4GSQ [102]. For the training of the splicing model, ESTs were chosen that showed a high coverage >98%, a sequence similarity of 100% and had only one hit on the genome resulting in >4,000 intron/exon junctions.

A total of 28,092 assembled ESTs as well as 27,819 non-assembled but well matching 454 singleton reads were mapped. Protein sequences of *Rhynchospirium secalis*, *S. sclerotiorum*, *B. cinerea*, *F. graminearum*, *N. crassa*, *Saccharomyces cerevisiae* were mapped using GenomeThreader. Finally, Jigsaw [103] was trained based on the 1,089 high confident gene models and used to calculate the best gene model using all the available evidence (predictors, ESTs, and trans-alignments). Subsequently, all gene models were manually curated in Apollo [104] and functional annotation was performed in PEDANT [105].

#### Estimating the completeness of the *P. subalpina* genome and classification of gene models

The completeness of the *P. subalpina* genome and annotation was assessed by mapping two separate highly conserved core gene sets including 248 and 246 proteins respectively [106, 107]. In addition, the fraction of successfully and non-mapped ESTs was analyzed.

Gene models were classified based on the identity against the best hit in the Similarity Matrix of Proteins database (SIMAP) [39]. Proteins with identities ≥30% were considered as confidential gene prediction. The



quality of the low identity genes (proteins with <30% identity with a protein sequence in SIMAP) was assessed by mapping the protein sequences against the assembled 454 EST dataset using GenomeThreader and analyzing the coverage of the gene model by ESTs. In addition, RNA-Seq data was available following the genome sequencing/annotation [108] (<http://www.ebi.ac.uk/ena/data/view/PRJEB12610>) was mapped against the genome using tophat (<http://ccb.jhu.edu/software/tophat/index.shtml>), discarding reads with bad mapping quality. Read coverage was calculated using coverageBed (<http://bedtools.readthedocs.org/en/latest/content/tools/coverage.html>) for each exon within the coding sequence. The coverage graph coverage graph was calculated in R using the ggplot2 package [109].

#### Presence of RNAi pathway and analysis for the presence of RIP mechanism

A reference dataset of key proteins in the RNAi pathway of *Neurospora crassa* (ARGONAUT, DICER, and RNA-dependent RNA polymerases (RdRP)) was used to search for similar genes in *P. subalpina* as described in Laurie et al. [110]. In addition, the presence of repeat-induced point mutations (RIP) in the genome of *P. subalpina* was analyzed. As RIP results in C-to-T transitions at repetitive loci [45, 111] we analyzed di-nucleotide abundances of all predicted interspersed repeats and of non-repetitive control sequences using RIPCAL [112]. Over-represented dinucleotides were identified by determining the fold change of the dinucleotide abundance between repeats and controls. *S. sclerotiorum* was included as a reference in the RIP analysis. Gene, repeat, and GC content were calculated using a sliding window analysis (window size: 1,000 bp, step size: 1,000 bp) and plotted for each scaffold. In addition, the genome of *P. subalpina* was mined for cytosine DNA methyltransferase genes of the Dnmt1 family involved in RIP and classified as described in Amselem et al. [54].

#### Analysis of horizontal gene transfer from non-fungal species

A total of 163 proteins showed the best hit with non-fungal taxa (124 with bacteria, 20 with plants and 17 with metazoa) when mapped against the SIMAP database in PEDANT. The possibility of HGT for these genes was evaluated. In a first step, a BLAST search against the nr database (query coverage  $\geq 40\%$ ; identities  $\geq 20\%$ ; best 1,000 hits) was done and the taxonomic distribution and similarities of the hits were analyzed in R. Genes for HGT were selected as candidates if (i) they showed a biased taxonomic distribution of the hits (<15% fungal hits) and/or (ii) the fungal hits showed smaller bit scores in blast searches than non-fungal hits. Candidate genes were further analyzed using a phylogenetic approach. The full protein datasets of the  $\leq 1000$  best blastp hits against nr database for each

candidate protein were downloaded. In addition, each candidate gene was also blasted against the genome data of phylogenetically closely related fungal genomes (*G. lozoyensis*, *B. cinerea*, *M. brunnea*, *S. sclerotiorum*). Protein sequences for each candidate gene were clustered with USEARCH v8.0.1517 (<http://www.drive5.com/usearch/>) [113] using the -cluster\_fast option at an identity threshold of 0.95 (pre-sorted by length). If the resulting number of clusters exceeded 40, the threshold was sequentially reduced by 0.05 until cluster numbers were  $\leq 40$ , or the threshold reached 0.5. This procedure was only applied to non-fungal sequences. Fungal BLAST hits were clustered at 0.95 if the number of clusters was  $\leq 40$ . If not, the threshold was adjusted as described. The cluster representative sequence of each cluster was used for phylogenetic analysis. Protein sequences including the *P. subalpina* sequence were aligned using MAFFT [114] (E-INS-i method) and the alignment trimmed with TrimAl (<http://trimal.cgenomics.org/trimal>) [115] using the -strict setting. Maximum likelihood phylogenetic trees were inferred using FastTree 2.1 (<http://meta.microbesonline.org/fasttree/>) [116] with default settings. All trees were deposited in TreeBase (<http://purl.org/phylo/treebase/phylovs/study/TB2:S20196>). The DNA-dependent RNA polymerase II (*RPB2*) sequence that is routinely used in phylogenetic studies was used as control. Due to its high conservation, only proteins with > 75% identity were included and clustering done at a 90% threshold.

#### Secondary metabolism

Key proteins involved in the secondary metabolism of *P. subalpina* were searched by using conserved InterPro motifs of polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) and related enzymes such as terpene synthases (TPC), and dimethyl allyl tryptophan synthases (DMATs) followed by manual inspection of the genes to verify the domain arrangement in the case of PKSs and NRPSs. In addition, putative secondary metabolite clusters were identified by searching for genes encoding tailoring enzymes (e.g., acyl-, and methyltransferases, oxidoreductases, cytochrome P450 enzymes) up- and downstream of genes for key enzymes and searching for and promoters [117].

Gene cluster involved in melanin synthesis were predicted using protein (i) sequences of PKS genes experimentally shown to be involved in melanin synthesis such as *G. lozoyensis* *PKS1* (AAN59953.1; [85]), *Alternaria alternata* *alm* (BAK64048.1; [84, 118]) and *Alb1/PksP* of *Aspergillus fumigatus* (XP\_756095.1; [86]), (ii) additional genes involved in the melanin synthesis pathway such as scytalone dehydratases (BC1G\_144888) and hydroxynaphthalene reductases (BC1G\_04230) [55] and (iii) by comparing orthologous gene clusters including candidate PKS derived from QuartetS analysis for the 14



ascomycetous species with the presence of melanization in the respective species. In addition, the NRPS genes putatively involved in siderophore synthesis were annotated.

### Comparative genome analysis

Comparative genome analyses were performed against a selection of published genomes of species showing different lifestyles (i.e. saprophytes, bio- and necrotrophic parasites and mycorrhizal species, Table 1). Special emphasis was put on selecting ascomycetous genomes and, wherever possible, genomes that are closely related to *P. subalpina*. All genomes used for comparative analysis were functionally annotated in PEDANT.

Putative orthologous gene clusters were analyzed using QuartetS [119] and the total number of clusters in which a specific species was present as well as pairwise shared clusters was recorded. Orthologous gene clusters enriched in pathogenic species were searched (four out of the five pathogenic species show entries for the respective cluster and  $\leq 1$  species of the six saprotroph species and  $\leq 1$  species of the two mycorrhizal species are present in the clusters respectively). Gene clusters enriched in saprophytes were searched using the same strategy. Both matrices were then subjected to principle component analysis using the vegan package in R [120] to analyze the position of *P. subalpina* compared to the other species. Moreover, FunCat terms [121] were enriched for these lifestyle enriched clusters by mapping *P. subalpina* geneIDs against the FunCat annotations and selecting the ten most frequent FunCat terms (any geneID/FunCat category was only considered once). In addition, putative orthologous gene clusters shared among the two mycorrhizal species *T. melanosporum* and *C. geophilum* with *O. maius* and *P. subalpina* was analyzed separately as the limited number of mycorrhizal species did not allow to properly defining lifestyle enriched clusters.

In a second step, the non-redundant portion of InterPro accessions per gene model, i.e. counting each InterPro accession per gene model only once, was analyzed. Besides some general statistics such as the number of distinct InterPro accessions within a species or different lifestyles or the total number of InterPro accessions, the number of significantly over- or underrepresented InterPro accessions in *P. subalpina* was determined by comparing their abundance in *P. subalpina* against the average observed in the 13 genomes using a Z-test [122]. Significantly over- and underrepresented InterPro accessions were then mapped against the gene ontology annotations (<http://www.ebi.ac.uk/interpro/download.html>) and enriched for GO terms. Moreover, InterPro accessions  $>2x$  overrepresented in pathogenic species and saprophytic species were mined and the resulting

matrices were subjected to principle component analysis as described above.

In a third step carbohydrate-active enzymes were annotated using the CAZyme expert annotation pipeline [123] and the number of enzymes in the diverse CAZyme families were compared with a special emphasis on the CAZyme families likely involved in cellulose (GH6, GH7, GH45), hemicellulose (GH10, GH11, GH26, GH31, GH67, GH115, GH134), pectin (GH28, GH53, GH78, GH79, GH88, GH105, GH106, GH127, PL1, PL3, PL4, PL9, PL11, CE8, CE12), and cutin layer breakdown (CE5). Moreover, a fourth category of CAZyme families were included that likely act on different of the above mentioned substrates (GH12, GH30, GH43, GH5, GH51, GH54, GH62, GH74, GH93). The total number of CAZyme modules per substrate class and species were calculated and subjected to principal component analysis as described above. In addition, proteins with auxiliary activities (AAs) that are hypothesized to be involved in the degradation of lignin were analyzed.

### Additional files

**Additional file 1:** Mapping statics for the assembled 454 ESTs. (PDF 197 kb)

**Additional file 2:** Validation of low identity gene models. (PDF 190 kb)

**Additional file 3:** Putative secondary metabolite clusters identified in the *P. subalpina* genome (XLSX 12 kb)

**Additional file 4:** General statistics on the presence of InterPro accessions in the analyzed genomes. (XLSX 10 kb)

**Additional file 5:** CAZyme modules related to PCWD used for principle component analysis. (XLSX 12 kb)

**Additional file 6:** General genome statistics for the species included in the present study. (XLSX 15 kb)

**Additional file 7:** PCA analysis based on InterPro accessions. Placement of the 13 ascomycete species and *P. subalpina* in PCA based on InterPro accessions described in Soanes et al. 2008 to be enriched in pathogens. (PDF 359 kb)

**Additional file 8:** Strategy to construct the manually-curated repeat library. (PDF 114 kb)

**Additional file 9:** Overview of the annotation strategy applied. (PDF 120 kb)

### Abbreviations

CAZyme: Carbohydrate-active enzymes; HGT: Horizontal gene transfer; PAC: *Phialocephala fortinii* s.l. – *Acephala applanata* species complex; PCWDE: Plant cell wall degrading enzymes; TE: Transposable elements

### Acknowledgments

We thank Marzanna Kuenzler und Weihong Qi (FGCZ, Zürich) for running the 454/FLX sequencing and Thomas Wicker for help in the classification and construction of the TE library. We also thank Andrin Gross and Ottmar Holdenrieder (Institute of Integrative Biology (IBZ), ETH Zurich) for helpful comments on a previous version of the manuscript.

### Funding

This study was partially funded by Vontobel Stiftung, Zürich to CRG. The funding agency had no influence on the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

**Availability of data and materials**

The datasets supporting the conclusions of this article are available from the following repositories.

European Nucleotide Archive:

<http://www.ebi.ac.uk/ena/data/view/JN031566>

<http://www.ebi.ac.uk/ena/data/view/FJOG0100001-FJOG01000205>

Pedant Database:

<http://pedant.helmholtz-muenchen.de/genomes.jsp?category=fungal>.

CAZyme:

<http://www.cazy.org/>

TreeBase:

<http://purl.org/phylo/treebase/phylovs/study/TB2:S20196>

In addition, datasets further supporting the conclusions of this article are included within the article and its additional files.

**Authors' contributions**

AD and CRG designed the research project. AD and CRG performed all experiments and collected sequence data. MS, MM, UG, RB, MH, BH, CMKS, DH, CRG implemented analytical tools and performed analysis. UG, BH, CMKS, DH, CRG wrote the manuscript. All authors have read and approved the final manuscript.

**Authors' informations**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Consent for publication**

Not applicable.

**Ethics approval and consent to participate**

Not applicable.

**Author details**

<sup>1</sup>Institute of Integrative Biology (IBZ), Forest Pathology and Dendrology, ETH Zürich, 8092 Zürich, Switzerland. <sup>2</sup>Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany. <sup>3</sup>Department of Genome-oriented Bioinformatics, Technische Universität München, Wissenschaftszentrum Weihenstephan, 85354 Freising, Germany. <sup>4</sup>Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Berne, Baltzerstrasse 6, 3012 Bern, Switzerland. <sup>5</sup>Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257 CNRS, Université Aix-Marseille, 163 Avenue de Luminy, 13288 Marseille, France. <sup>6</sup>DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. <sup>7</sup>Friedrich-Schiller-Universität, Pharmazeutische Mikrobiologie, Winzerlaer Strasse 2, 07745 Jena, Germany. <sup>8</sup>Microsynth AG, Schützenstrasse 15, 9436 Balgach, Switzerland.

Received: 26 July 2016 Accepted: 2 December 2016

Published online: 09 December 2016

**References**

- Sieber TN. Fungal root endophytes. In: Waisel Y, Eshel A, Kafkafi U, editors. *Plant Roots: The hidden half*. 3rd ed. New York and Basel: Marcel Dekker; 2002. p. 887–917.
- Schulz BJE, Boyle CJC, Sieber TN. *Microbial Root Endophytes*. Berlin: Springer; 2006.
- Addy HD, Piercey MM, Currah RS. Microfungal endophytes in roots. *Can J Bot*. 2005;83(1):1–13.
- Sieber TN. Endophytic fungi in forest trees: are they mutualists? *Fungal Biol Rev*. 2007;21(2–3):75–89.
- Grünig CR, Queloz V, Sieber TN, Holdenrieder O. Dark septate endophytes (DSE) of the *Phialocephala fortinii* s.l. – *Acephala applanata* species complex in tree roots – classification, population biology and ecology. *Botany*. 2008; 86(12):1355–69.
- Rodriguez RJ, White JF, Arnold AE, Redman RS. Fungal endophytes: diversity and functional roles. *New Phytol*. 2009;182(2):314–30.
- Francis R, Read DJ. Mutualism and antagonism in the mycorrhizal symbiosis, with special reference to impacts on plant community structure. *Can J Bot*. 1995;73(51):1301–9.
- Johnson NC, Graham JH, Smith FA. Functioning of the mycorrhizal associations along the mutualism-parasitism continuum. *New Phytol*. 1997; 135(4):575–85.
- Schulz B, Boyle C. The endophytic continuum. *Mycol Res*. 2005;109(6):661–86.
- Queloz V, Sieber TN, Holdenrieder O, McDonald BA, Grünig CR. No biogeographical pattern for a root-associated fungal species complex. *Glob Ecol Biogeogr*. 2011;20(1):160–9.
- Piercey MM, Graham SW, Currah RS. Patterns of genetic variation in *Phialocephala fortinii* across a broad latitudinal transect in Canada. *Mycol Res*. 2004;108(8):955–64.
- Zhang C, Yin L, Dai S. Diversity of root-associated fungal endophytes in *Rhododendron fortunei* in subtropical forests of China. *Mycorrhiza*. 2009; 19(6):417–23.
- Bruzzone MC, Fontenla SB, Vohnik M. Is the prominent ericoid mycorrhizal fungus *Rhizoscypha ericae* absent in the Southern Hemisphere's Ericaceae? A case study on the diversity of root mycobionts in *Gaultheria* spp. from northwest Patagonia, Argentina. *Mycorrhiza*. 2015;25(1):25–40.
- Grünig CR, Duò A, Sieber TN. Population genetic analysis of *Phialocephala fortinii* s.l. and *Acephala applanata* in two undisturbed forests in Switzerland and evidence for new cryptic species. *Fungal Genet Biol*. 2006;43(6):410–21.
- Grünig CR, Sieber TN. Molecular and phenotypic description of the widespread root symbiont *Acephala applanata* gen. et sp. nov., formerly known as dark septate endophyte Type 1. *Mycologia*. 2005;97(3):628–40.
- Grünig CR, Duò A, Sieber TN, Holdenrieder O. Assignment of species rank to six reproductively isolated cryptic species of the *Phialocephala fortinii* s.l. – *Acephala applanata* species complex. *Mycologia*. 2008;100(1):47–67.
- Queloz V, Grünig CR, Sieber TN, Holdenrieder O. Monitoring the spatial and temporal dynamics of a community of the tree-root endophyte *Phialocephala fortinii* s.l. *New Phytol*. 2005;168(3):651–60.
- Stroheker S, Queloz V, Sieber TN. Spatial and temporal dynamics in the *Phialocephala fortinii* s.l. – *Acephala applanata* species complex (PAC). *Plant Soil*. 2016;407(1):231–41.
- McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, Dornelas M, Enquist BJ, Green JL, He FL, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol Lett*. 2007;10(10):995–1015.
- Zimmerman E, Peterson RL. Effect of a dark septate fungal endophyte on seed germination and protocorm development in a terrestrial orchid. *Symbiosis*. 2007;43(1):45–52.
- Jumpponen A, Mattson KG, Trappe JM. Mycorrhizal functioning of *Phialocephala fortinii* with *Pinus contorta* on glacier forefront soil: Interactions with soil nitrogen and organic matter. *Mycorrhiza*. 1998;7(5):261–5.
- Tellenbach C, Grünig CR, Sieber TN. Negative effects on survival and performance of Norway spruce seedlings colonized by fungal root endophytes are primarily isolate-dependent. *Environ Microbiol*. 2011;13(9):2508–17.
- Tellenbach C, Grünig CR, Sieber TN. Suitability of quantitative real-time PCR to estimate the biomass of fungal root endophytes. *Appl Environ Microbiol*. 2010;76(17):5764–72.
- Reininger V, Sieber TN. Mycorrhiza reduces adverse effects of dark septate endophytes (DSE) on growth of conifers. *PLoS One*. 2012;7(8), e42865.
- Yu T, Nassuth A, Peterson RL. Characterization of the interaction between the dark septate fungus *Phialocephala fortinii* and *Asparagus officinalis* roots. *Can J Microbiol*. 2001;47(8):741–53.
- Currah RS, Tsuneda A, Murakami S. Morphology and ecology of *Phialocephala fortinii* in roots of *Rhododendron brachycarpum*. *Can J Bot*. 1993;71(12):1639–44.
- Peterson RL, Wagg C, Pautler M. Associations between microfungal endophytes and roots: do structural features indicate function? *Botany*. 2008;86(5):445–56.
- O'Dell TE, Massicotte HB, Trappe JM. Root colonization of *Lupinus latifolius* Agardh. and *Pinus contorta* Dougl. by *Phialocephala fortinii* Wang & Wilcox. *New Phytol*. 1993;124(1):93–100.
- Fernando AA, Currah RS. A comparative study of the effects of the root endophytes *Leptodontidium orchidicola* and *Phialocephala fortinii* (Fungi Imperfecti) on the growth of some subalpine plants in culture. *Can J Bot*. 1996;74(7):1071–8.
- Wicker T, Oberhaensli S, Parlange F, Buchmann JP, Shatalina M, Roffler S, Ben-David R, Dolezel J, Simkova H, Schulze-Lefert P, et al. The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nat Genet*. 2013;45(9):1092–6.

31. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otiillar R, Spatafora JW, Yadav JS, et al. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science*. 2012;336(6089):1715–9.
32. Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canback B, Choi C, Ciochok N, Clum A, et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet*. 2015;47(4):410–5.
33. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillion 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. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, et al. Plant-Symbiotic Fungi as Chemical Engineers: Multi-Genome Analysis of the Clavicipitaceae Reveals Dynamics of Alkaloid Loci. *PLoS Genet*. 2013;9(2):e1003323.
35. Schardl CL, Young CA, Moore N, Krom N, Dupont P-Y, Pan J, Florea S, Webb JS, Jaromczyk J, Jaromczyk JW, et al. Chapter Ten - Genomes of Plant-Associated Clavicipitaceae. In: Francis MM, editor. *Advances in Botanical Research*. Volume 70. London: Academic Press; 2014. p. 291–327.
36. Schardl CL, Leuchtmann A, Spiering MJ. Symbiosis of grasses with seedborne fungal endophytes. *Annu Rev Plant Biol*. 2004;55(1):315–40.
37. Walker AK, Frasz SL, Seifert KA, Miller JD, Mondo SJ, LaButti K, Lipzen A, Dockter RB, Kennedy MC, Grigoriev IV, et al. Full genome of *Phialocephala scopiformis* DAOMC 229536, a fungal endophyte of spruce producing the potent anti-insectan compound rugulosin. *Genome Announc*. 2016;4(1):e01768–01715.
38. Duo A, Bruggmann R, Zoller S, Bernt M, Grunig CR. Mitochondrial genome evolution in species belonging to the *Phialocephala fortinii* s.l. - *Acephala applanata* species complex. *BMC Genomics*. 2012;13:17.
39. Arnold R, Goldenberg F, Mewes H-W, Rattei T. SIMAP—the database of all-against-all protein sequence similarities and annotations with new interfaces and increased coverage. *Nucleic Acids Res*. 2014;42(D1):D279–84.
40. Bleuler-Martinez S, Buttschi A, Garbani M, Wälti MA, Wohlschläger T, Potthoff E, Sabotic J, Pohleven J, Lüthy P, Hengartner MO, et al. A lectin-mediated resistance of higher fungi against predators and parasites. *Mol Ecol*. 2011; 20(14):3056–70.
41. Mohanta TK, Bae H. The diversity of fungal genome. *Biol Proced Online*. 2015;17:8.
42. Petrov DA. Evolution of genome size: new approaches to an old problem. *Trends Genet*. 2001;17(1):23–8.
43. Kelkar YD, Ochman H. Causes and consequences of genome expansion in fungi. *Genome Biol Evol*. 2012;4(1):13–23.
44. Abouelhoda M, Kurtz S, Ohlebusch E. CoCoNUT: an efficient system for the comparison and analysis of genomes. *BMC Bioinformatics*. 2008;9(1):476–93.
45. Galagan JE, Selker EU. RIP: the evolutionary cost of genome defense. *Trends Genet*. 2004;20(9):417–23.
46. Cambareri EB, Jensen BC, Schabtach E, Selker EU. Repeat-induced G-C to A-T mutations in *Neurospora*. *Science*. 1989;244:1571–5.
47. Goyon C, Faugeron G. Targeted transformation of *Ascoibolus immersus* and de novo methylation of the resulting duplicated DNA sequences. *Mol Cell Biol*. 1989;9(7):2818–27.
48. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc B*. 2009;364(1513):99–115.
49. Yamanaka S, Mehta S, Reyes-Turcu FE, Zhuang F, Fuchs RT, Rong Y, Robb GB, Grewal SIS. RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature*. 2013;493(7433):557–60.
50. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*. 2007;8(4):272–85.
51. Kasbekar DP. What have we learned by doing transformations in *Neurospora tetrasperma*? In: van den Berg MA, Maruthachalam K, editors. *Genetic Transformation Systems in Fungi*, vol. 2. Cham and Heidelberg: Springer; 2014. p. 47–52.
52. Zaffarano PL, Queloz V, Duo A, Grunig CR. Sex in the PAC: A hidden affair in dark septate endophytes? *BMC Evol Biol*. 2011;11:282.
53. Grünig CR, Queloz V, Duò A, Sieber TN. Phylogeny of *Phaeomollisia piceae* gen. sp. nov.: a dark-septate conifer-needle endophyte and its relationships to *Phialocephala* and *Acephala*. *Mycol Res*. 2009;113(2):207–21.
54. Amselem J, Lebrun M-H, Quesneville H. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC Genomics*. 2015;16(1):141.
55. Amselem J, Cuomo CA, van Kan JA, Viaud M, Benito EP. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet*. 2011;7(8), e1002230.
56. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. 2010;464(7287):367–73.
57. Wang Z, Johnston PR, Takamatsu S, Spatafora JW, Hibbett DS. Toward a phylogenetic classification of the Leotiomyces based on rDNA data. *Mycologia*. 2006;98(6):1065–75.
58. Marcet-Houben M, Gabaldón T. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet*. 2010;26(1):5–8.
59. Mallet LV, Becq J, Deschavanne P. Whole genome evaluation of horizontal transfers in the pathogenic fungus *Aspergillus fumigatus*. *BMC Genomics*. 2010;11(1):1–13.
60. Richards TA, Leonard G, Soanes DM, Talbot NJ. Gene transfer into the fungi. *Fungal Biol Rev*. 2011;25(2):98–110.
61. Rice A, Currah R. *Oidiodendron maius*: saprobe in Sphagnum peat, mutualist in ericaceous roots? In: Schulz BE, Boyle CC, Sieber T, editors. *Microbial Root Endophytes*, vol. 9. Berlin and Heidelberg: Springer; 2006. p. 227–46.
62. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr Biol*. 2006;16(18):1857–64.
63. Soanes D, Richards TA. Horizontal gene transfer in eukaryotic plant pathogens. *Annu Rev Phytopathol*. 2014;52(1):583–614.
64. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, Kazan K, Chakraborty S, McDonald BA, Manners JM. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathog*. 2012;8(9), e1002952.
65. Palzkill T. Metallo- $\beta$ -lactamase structure and function. *Ann N Y Acad Sci*. 2013;1277:91–104.
66. Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, et al. Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*. 2010; 464(7291):1033–8.
67. Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, et al. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*. 2008;452(7183):88–92.
68. Mengiste T. Plant immunity to necrotrophs. *Annu Rev Phytopathol*. 2012;50(1):267–94.
69. Bellincampi D, Cervone F, Lionetti V. Plant cell wall dynamics and wall-related susceptibility in plant–pathogen interactions. *Front Plant Sci*. 2014;5:228.
70. Presti LL, 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(1):513–45.
71. Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, Rattray M, Hubbard SJ, Oliver SG, Talbot NJ. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS One*. 2008;3(6), e2300.
72. Jashni MK, Dols IHM, Iida Y, Boeren S, Beenen HG, Mehrabi R, Collemare J, de Wit PJGM. 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.
73. Kombrink A, Thomma BPHJ. LysM effectors: secreted proteins supporting fungal life. *PLoS Pathog*. 2013;9(12), e1003769.
74. de Jonge R, Thomma BPHJ. Fungal LysM effectors: extinguishers of host immunity? *Trends Microbiol*. 2009;17(4):151–7.
75. Haquard S, Joly DL, Lin Y-C, Tisserant E, Feau N, Delaruelle C, Legué V, Kohler A, Tanguay P, Petre B, et al. A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (Poplar Leaf Rust). *Mol Plant-Microbe Interact*. 2011;25(3):279–93.
76. Sieber TN, Grunig CR. Fungal root endophytes. In: Eshel A, Beekman T, editors. *Plant Roots: The hidden half*. 4th ed. Boca Raton, London, New York: CRC Press; 2013. p. 1–49.
77. Gordon TR, Martyn RD. The evolutionary biology of *Fusarium oxysporum*. *Annu Rev Phytopathol*. 1997;35(1):111–28.
78. Gordon TR, Okamoto D, Jacobson DJ. Colonization of muskmelon and nonsusceptible crops by *Fusarium oxysporum* f. sp. *melonis* and other species of *Fusarium*. *Phytopathology*. 1989;79:1095–100.

79. Kunzler M. Hitting the sweet spot: glycans as targets of fungal defense effector proteins. *Molecules*. 2015;20(5):8144–67.
80. Sumarah MW, Puniani E, Blackwell BA, Miller JD. Characterization of polyketide metabolites from foliar endophytes of *Picea glauca*. *J Nat Prod*. 2008;71(8):1393–8.
81. Miller JD, Sumarah MW, Adams GW. Effect of a rugulosin-producing endophyte in *Picea glauca* on *Choristoneura fumiferana*. *J Chem Ecol*. 2008;34(3):362–8.
82. Tellenbach C, Sumarah MW, Grunig CR, Miller JD. Inhibition of *Phytophthora* species by secondary metabolites produced by the dark septate endophyte *Phialocephala europaea*. *Fungal Ecol*. 2013;6(1):12–8.
83. Bushley KE, Ripoll DR, Turgeon BG. Module evolution and substrate specificity of fungal nonribosomal peptide synthetases involved in siderophore biosynthesis. *BMC Evol Biol*. 2008;8(1):1–24.
84. Kimura N, Tsuge T. Gene cluster involved in melanin biosynthesis of the filamentous fungus *Alternaria alternata*. *J Bacteriol*. 1993;175(14):4427–35.
85. Zhang A, Lu P, Dahl-Roshak AM, Paresse PS, Kennedy S, Tkacz JS, An Z. Efficient disruption of a polyketide synthase gene (pks1) required for melanin synthesis through *Agrobacterium*-mediated transformation of *Glarea lozoyensis*. *Mol Gen Genomics*. 2003;268(5):645–55.
86. Langfelder K, Jahn B, Gehringer H, Schmidt A, Wanner G, Brakhage AA. Identification of a polyketide synthase gene (pksP) of *Aspergillus fumigatus* involved in conidial pigment biosynthesis and virulence. *Med Microbiol Immunol*. 1998;187(2):79–89.
87. Wick J, Heine D, Lackner G, Misiek M, Tauber J, Jagusch H, Hertweck C, Stallforth P, Hoffmeister D. A fivefold parallelized biosynthetic process secures chlorination of *Armillaria mellea* (honey mushroom) toxins. *Appl Environ Microbiol*. 2015;82(4):1196–204.
88. Braesel J, Götze S, Shah F, Heine D, Tauber J, Hertweck C, Tunlid A, Stallforth P, Hoffmeister D. Three redundant synthetases secure redox-active pigment production in the Basidiomycete *Paxillus involutus*. *Chem Biol*. 2015;22(10):1325–34.
89. Brunner I, Godbold LD. Tree roots in a changing world. *J For Res*. 2007;12(2):78–82.
90. Brunner I, Ruf M, Lüscher P, Sperisen C. Molecular markers reveal extensive intraspecific below-ground overlap of silver fir fine roots. *Mol Ecol*. 2004;13(11):3595–600.
91. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Micro*. 2010;8(1):15–25.
92. Queloz V, Duo A, Sieber TN, Grünig CR. Microsatellite size homoplasies and null alleles do not affect species diagnosis and population genetic analysis in a fungal species complex. *Mol Ecol Resour*. 2010;10:348–67.
93. Grünig CR, Linde CC, Sieber TN, Rogers SO. Development of single-copy RFLP markers for population genetic studies of *Phialocephala fortinii* and closely related taxa. *Mycol Res*. 2003;107(11):1332–41.
94. Queloz V, Duo A, Grünig CR. Isolation and characterization of microsatellite markers for the tree-root endophytes *Phialocephala subalpina* and *Phialocephala fortinii* s.s. *Mol Ecol Resour*. 2008;8(6):1322–5.
95. Beldade P, Rudd S, Gruber J, Long A. A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. *BMC Genomics*. 2006;7(1):130.
96. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics*. 2005;21 suppl 1:i351–8.
97. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;8(12):973–82.
98. Stanke M, Tzvetkova A, Morgenstern B. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol*. 2006;7 Suppl 1:S11.
99. Ter-Hovhannisyantsyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 2008;18(12):1979–90.
100. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*. 2006;7(1):1–11.
101. Gremme G, Brendel V, Sparks ME, Kurtz S. Engineering a software tool for gene structure prediction in higher organisms. *Inf Softw Technol*. 2005;47(15):965–78.
102. Sparks ME, Brendel V. Incorporation of splice site probability models for non-canonical introns improves gene structure prediction in plants. *Bioinformatics*. 2005;21(Suppl\_3):iii20–30.
103. Allen JE, Salzberg SL. JIGSAW: integration of multiple sources of evidence for gene prediction. *Bioinformatics*. 2005;21(18):3596–603.
104. Lee E, Harris N, Gibson M, Chetty R, Lewis S. Apollo: a community resource for genome annotation editing. *Bioinformatics*. 2009;25(14):1836–7.
105. Walter MC, Rattei T, Arnold R, Güldener U, Münsterkötter M, Nenova K, Kastenmüller G, Tischler P, Wölling A, Volz A, et al. PEDANT covers all complete RefSeq genomes. *Nucleic Acids Res*. 2009;37 suppl 1:D408–11.
106. Parra G, Bradnam K, Ning Z, Keane T, Korf I. Assessing the gene space in draft genomes. *Nucleic Acids Res*. 2009;37(1):289–97.
107. Aguilera G, Marthey S, Chiappello H, Lebrun M-H, Rodolphe F, Fournier E, Gendraud-Jacquemard A, Giraud T. Assessing the performance of single-copy genes for recovering robust phylogenies. *Syst Biol*. 2008;57(4):613–27.
108. Reininger V, Schlegel M. Analysis of the *Phialocephala subalpina* transcriptome during colonization of its host plant *Picea abies*. *PLoS One*. 2016;11(3), e0150591.
109. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2006.
110. Laurie JD, Ali S, Linning R, Mannhaupt G, Wong P, Güldener U, Münsterkötter M, Moore R, Kahmann R, Bakkeren G, et al. Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *Plant Cell*. 2012;24(5):1733–45.
111. Horns F, Petit E, Yockteng R, Hood ME. Patterns of repeat-induced point mutation in transposable elements of basidiomycete fungi. *Genome Biol Evol*. 2012;4(3):240–7.
112. Hane JK, Oliver RP. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics*. 2008;9:478.
113. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460–1.
114. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059–66.
115. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3.
116. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One*. 2010;5(3), e9490.
117. Sieber CMK, Lee W, Wong P, Münsterkötter M, Mewes H-W, Schmeitzl C, Varga E, Berthiller F, Adam G, Güldener U. The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One*. 2014;9(10), e110311.
118. Kheder AA, Akagi Y, Akamatsu H, Yanaga K, Maekawa N, Otani H, Tsuge T, Kodama M. Functional analysis of the melanin biosynthesis genes ALM1 and BRM2-1 in the tomato pathotype of *Alternaria alternata*. *J Gen Plant Pathol*. 2011;78(1):30–8.
119. Yu C, Zavaljevski N, Desai V, Reifman J. QuartetS: a fast and accurate algorithm for large-scale orthology detection. *Nucleic Acids Res*. 2011;39(13):e88.
120. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. vegan: Community Ecology Package. R package version 2.3-1. The Comprehensive R Archive Network. 2014. <https://CRAN.R-project.org/package=vegan>.
121. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Güldener U, Mannhaupt G, Münsterkötter M, et al. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res*. 2004;32(18):5339–45.
122. Stahel WA. Statistische Datenanalyse. 4th ed. Braunschweig: Vieweg; 2002.
123. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res*. 2014;42(Database issue):D490–5.
124. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EV L, Brown JKM, 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.
125. Zhu S, Cao Y-Z, Jiang C, Tan B-Y, Wang Z, Feng S, Zhang L, Su X-H, Breyova B, Vinar T, et al. Sequencing the genome of *Marssonina brunnea* reveals fungus-popular co-evolution. *BMC Genomics*. 2012;13(1):1–10.
126. Nierman WC, Yu J, Fedorova-Abrams ND, Losada L, Cleveland TE, Bhatnagar D, Bennett JW, Dean R, Payne GA. Genome sequence of *Aspergillus flavus* NRRL 3357, a strain that causes aflatoxin contamination of food and feed. *Genome Announc*. 2015;3(2):e00168–00115.
127. Martínez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, et al. Genome sequencing



- and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol.* 2008;26(5):553–60.
128. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science.* 2007;317(5843):1400–2.
  129. van den Berg MA, Albang R, Albermann K, Badger JH, Daran J-M, Driessen AJ M, Garcia-Estrada C, Fedorova ND, Harris DM, Heijne WHM, et al. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat Biotechnol.* 2008;26(10):1161–8.
  130. Youssar L, Grüning BA, Exleben A, Günther S, Hüttel W. Genome sequence of the fungus *Glarea lozoyensis*: the first genome sequence of a species from the Helotiaceae family. *Eukaryotic Cell.* 2012;11(2):250.
  131. Peter M, Kohler A, Ohm RA, Kuo A, Krützmann J, Morin E, Arend M, Barry KW, Binder M, Choi C, et al. Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nat Commun.* 2016;7:12662.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

