

Comparative Genomics Including the Early-Diverging Smut Fungus *Ceraceosorus bombacis* Reveals Signatures of Parallel Evolution within Plant and Animal Pathogens of Fungi and Oomycetes

Rahul Sharma^{1,2,3,4}, Xiaojuan Xia^{1,2,3}, Kai Riess⁵, Robert Bauer⁵, and Marco Thines^{1,2,3,4,*}

¹Biodiversity and Climate Research Centre (BIK-F), Frankfurt (Main), Germany

²Department for Biological Sciences, Institute of Ecology, Evolution and Diversity, Goethe University, Frankfurt (Main), Germany

³Senckenberg Gesellschaft für Naturforschung, Frankfurt (Main), Germany

⁴Cluster for Integrative Fungal Research (IPF), Frankfurt (Main), Germany

⁵Plant Evolutionary Ecology, Institute of Evolution and Ecology, University of Tübingen, Germany

*Corresponding author: E-mail: marco.thines@senckenberg.de.

Accepted: August 3, 2015

Data deposition: This project has been deposited at the European Nucleotide Archive (ENA) under the accession PRJEB6935.

Abstract

Ceraceosorus bombacis is an early-diverging lineage of smut fungi and a pathogen of cotton trees (*Bombax ceiba*). To study the evolutionary genomics of smut fungi in comparison with other fungal and oomycete pathogens, the genome of *C. bombacis* was sequenced and comparative genomic analyses were performed. The genome of 26.09 Mb encodes for 8,024 proteins, of which 576 are putative-secreted effector proteins (PSEPs). Orthology analysis revealed 30 ortholog PSEPs among six Ustilaginomycotina genomes, the largest groups of which are lytic enzymes, such as aspartic peptidase and glycoside hydrolase. Positive selection analyses revealed the highest percentage of positively selected PSEPs in *C. bombacis* compared with other Ustilaginomycotina genomes. Metabolic pathway analyses revealed the absence of genes encoding for nitrite and nitrate reductase in the genome of the human skin pathogen *Malassezia globosa*, but these enzymes are present in the sequenced plant pathogens in smut fungi. Interestingly, these genes are also absent in cultivable oomycete animal pathogens, while nitrate reductase has been lost in cultivable oomycete plant pathogens. Similar patterns were also observed for obligate biotrophic and hemi-biotrophic fungal and oomycete pathogens. Furthermore, it was found that both fungal and oomycete animal pathogen genomes are lacking cutinases and pectinesterases. Overall, these findings highlight the parallel evolution of certain genomic traits, revealing potential common evolutionary trajectories among fungal and oomycete pathogens, shaping the pathogen genomes according to their lifestyle.

Key words: comparative genomics, convergent evolution, evolutionary biology, metabolic pathways, positive selection, smut fungi, Ustilaginomycotina.

Introduction

In both fungal and oomycete pathogens, several biological and ecological characteristics, such as biotrophy, animal and plant parasitism, as well as colonizing terrestrial and marine habitats have evolved in parallel and often multiple times (Göker et al. 2007; McLaughlin et al. 2009; Thines and Kamoun 2010; Kemen and Jones 2012; Sharma et al. 2014). Several studies have been performed for investigating

lifestyle signatures in the genomes of plant and animal pathogens of both fungi and oomycetes (Kämper et al. 2006; Haas et al. 2009; Baxter et al. 2010; Schirawski et al. 2010; Kemen et al. 2011; Lamour et al. 2012; Laurie et al. 2012; Yoshida et al. 2013; Que et al. 2014). These studies provided insights into the adaptation of fungal and oomycete pathogens to necrotrophy (Levesque et al. 2010), hemibiotrophy (Tyler et al. 2006; O'Connell et al. 2012), and obligate

biotrophy (Baxter et al. 2010; Duplessis et al. 2011). Yet, there are only few studies addressing the consequences of common evolutionary trajectories in fungi and oomycetes with convergent lifestyles (Duplessis et al. 2011; McDowell 2011). In addition, sequencing efforts for basidiomycete pathogens, especially in Ustilaginomycotina have so far been focused on only few related species, which renders comparative analyses difficult.

Many pathogens of the Ustilaginomycotina cause severe diseases to economically important plants (Martinez-Espinoza et al. 2002). An example of this is *Ustilago maydis*, the causal agent of maize smut, which is responsible for losses of millions of dollars annually (Boa 2001). Thus it is critical to study the evolution and pathogenicity mechanisms of this group of pathogens, in order to discover ways on how to avoid the negative effects. Genome resources are available for *U. maydis*, which is also a model pathosystem for elucidating the molecular basis of pathogenicity (Kämper et al. 2006). Later genome sequencing and comparative genomic analyses were performed within smut pathogens (Xu et al. 2007; Schirawski et al. 2010; Laurie et al. 2012; Que et al. 2014; Sharma et al. 2014); however, all were focused on a single family of the Ustilaginomycotina, the Ustilaginaceae.

Ceraceosorus bombacis is a plant pathogen of the Ustilaginomycotina affecting the cotton tree, *Bombax ceiba* (Cunningham et al. 1976). The pathogen is largely unrelated to the other plant pathogenic genera of the Ustilaginomycotina, which were so far included in genomic studies (Begerow et al. 2006).

Most of the pathogenic fungi in the Ustilaginomycetidae switch between two forms; a saprotrophic unicellular yeast growth form and a pathogenic filamentous growth form (Bölker 2001). Dimorphism in smut fungi is thus a key feature associated with the switch from saprotrophic to pathogenic growth (Mitchell 1998; Bölker 2001; Sanchez-Martinez and Perez-Martin 2001; Klein and Tebbets 2007). In plant pathogenic Ustilaginomycotina, filamentous growth is needed to invade the host, enabled by the secretion of a plethora of effector proteins into the host. Many such putative secreted effector proteins (PSEPs) have been identified in the genomes of Ustilaginomycotina (Kämper et al. 2006; Xu et al. 2007; Schirawski et al. 2010; Laurie et al. 2012; Sharma et al. 2014). Genome-wide positive selection studies suggested that the genes encoding for PSEPs are under higher selection pressure than the nonsecreted proteins (Sharma et al. 2014).

Genes involved in metabolic pathways have been studied in several fungi (Keller and Hohn 1997; Duplessis et al. 2011). Because of the availability of increasing amounts of genomic data for pathogens from different phylogenetic groups, it has become possible to relate the absence or presence of metabolic pathways to their lifestyles. Previous studies have reported the absence of some enzymes that play a key role in the nitrite metabolism in the genomes of animal infecting oomycete pathogens (Jiang et al. 2013), and as well as in

genomes of obligate biotrophic plant pathogens of fungi (Duplessis et al. 2011) and oomycetes (Jiang et al. 2013). To the best of our knowledge, there are no studies that explicitly address these findings in terms of parallel evolution of different lifestyles in the fungal and oomycete plant and animal pathogens.

Apart from proteins involved in primary metabolism, fungal and bacterial genomes encode for proteins producing a variety of low molecular mass compounds, known as nonribosomal peptides and polyketides (Keller et al. 2005; Bölker et al. 2008; Bode 2009). These metabolites are not vital for the growth of these organisms, but are involved in other cellular activities (Keller et al. 2005) and sometimes play an important role in pathogenicity (Stergiopoulos et al. 2013). Several studies have reported corresponding genes in fungal and bacterial genomes (Dean et al. 2005; Bölker et al. 2008; Inglis et al. 2013). In Ustilaginomycotina, these genes have not been the subject of detailed analyses, so far, despite their potential role in pathogenicity.

Thus, it was the aim of this study to perform comparative genome analyses in fungal and oomycete pathogens using a set of representative species to unravel common evolutionary trajectories determining metabolic capacities in both primary and secondary metabolism. This includes the prediction of the potentially pathogenicity-related genes for *C. bombacis* and other Ustilaginomycotina pathogens, genome wide positive selection studies on these, the search for core effectors in Ustilaginomycotina genomes, and the screening of fungal and oomycete genomes for common patterns related to their metabolism.

Materials and Methods

DNA Isolation from Spores and Library Preparation for Genomic Sequencing

Liquid cultures of *C. bombacis* (strain ATCC 22867) were grown in SAM medium (Agar-Agar 7.5 g/l, Potato Dextrose Broth 7.5 g/l, yeast extract 2.5 g/l, malt extract 2.5 g/l, clarified V8 juice 8 ml/l, rifampicin 15 mg/l) at 25 °C for 4 days. The pellet from 100 ml of the culture was resuspended with 5 ml lysis buffer (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM ethylenediaminetetraacetic acid- Na_2 , 1% sodium dodecyl sulfate, 2% Triton X-100) (Hoffman and Winston 1987) and the suspension was transferred to five 2-ml Eppendorf tubes (700 μl each). In addition, 300 μl glass beads and 700 μl phenol/chloroform/isoamylalcohol (PCI; 25:24:1 pH 7.5–8.0) were added to the suspension and vortexed for 15 min. After 2-min centrifugation at 12,000 \times g, the supernatant was collected and extracted once again with PCI. The genomic DNA (gDNA) was precipitated by adding two times the volume of the collected supernatant of 100% ice-cold ethanol and 1/10 of the volume of the collected supernatant of a 3 M sodium acetate solution and incubating for 30 min at -20°C . Finally

the DNA was pelleted by centrifugation at $6,000 \times g$ for 10 min at 4°C , washed with 70% ice-cold ethanol, and dried at room temperature. Tris-EDTA buffer with DNase-free RNase was used to dissolve the DNA pellet. The quantity of the isolated DNA was evaluated by a Qubit Fluorometer (Invitrogen, Darmstadt, Germany) and its integrity was checked on a 1.5% agarose gel in ultraviolet radiation after staining with ethidium bromide.

RNA Isolation

RNA was extracted from 100 ml of liquid culture using a NucleoSpin Plant RNA kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany) according to the instructions of the manufacturer. The sample cultivation conditions for the RNA isolation are the same as above. The RNA quantity was measured using a Qubit Fluorometer (Invitrogen) and the quality was evaluated on a 1.5% agarose gel under ultraviolet radiation after staining with ethidium bromide.

Processing of Raw gDNA and RNA-Seq Data

Sequencing was carried out by a commercial sequencing provider (LGC Genomics). Illumina adapter-clipped and primer-clipped gDNA reads of 300, 600 bp, 3, and 8 kb insert libraries were further processed by quality and length filters using FastQFS (Sharma and Thines 2015). Reads having an average quality below 26 phred score, a length shorter than 72 bp, and having ambiguous bases were discarded from the analyses.

RNA-Seq data were processed using Trimmomatic v0.32 (Bolger et al. 2014) with a quality score threshold of 15 in a window of 5 bases.

Assembly of gDNA, CEGMA Analysis, and Repeat Element Masking

Assemblies with the Velvet v1.2.09 short read assembler (Zerbino and Birney 2008) were optimized considering several different k -mer lengths and k -mer coverage cut-offs. Genome assembly with a k -mer length of 69 and the k -mer coverage cut-off "auto" generated the best results considering the N50 scaffold size and the size of the largest scaffold. Integrity and completeness of the assembled scaffolds was assessed by using the CEGMA pipeline v2.4 (Parra et al. 2007). Repeat elements were predicted using RepeatModeler v1.0.4 (<http://www.repeatmasker.org/RepeatModeler.html>, last accessed September 16, 2015). RepeatModeler uses RECON (Bao and Eddy 2002) and RepeatScout v1 (Price et al. 2005) to perform *de novo* repeat predictions. The latest Repbase library version 20130422 (Jurka et al. 2005) was used for reference-based repeat element searches. Tandem repeat elements were predicted using trf (Benson 1999) from within the RepeatModeler pipeline. RepeatMasker (<http://www.repeat-masker.org/>, last accessed September 16, 2015) was used to mask the predicted repeat elements.

Gene Prediction, Annotation, and Expression Value Calculation

Gene model prediction was done using both *ab initio*-based methods and RNA-Seq-derived transcript mapping onto the assembled genome (supplementary fig. S1, Supplementary Material online), as described before (Sharma et al. 2015). For *ab initio*-based methods, gene models were predicted using both GeneMark-ES v2.3 (Borodovsky and Lomsadze 2011) and Augustus v2.7 (Stanke et al. 2006). GeneMark-ES was run on the repeat masked genome. Predicted coding sequences (CDS) from GeneMark-ES were then used for RNA-Seq mapping to generate a training set for Augustus. For this purpose, CDS and 100 bp both upstream and downstream of the predicted CDS coordinates were extracted and RNA-Seq reads were mapped using Bowtie2 v2.1.0 (Langmead and Salzberg 2012). Only those gene models were used for Augustus training that were having a more than 10-fold coverage. To generate intron/exon hint files for Augustus, Tophat2 v2.0.10 (Kim et al. 2013) was used to map RNA-Seq reads on the repeat-masked genome. This mapping file was further processed by Samtools v0.1.18 (Li et al. 2009) to generate final intron/exon hints. Using the training set and the intron/exon hint file Augustus v2.7 was run on the genome scaffolds.

In addition, gene models were predicted using PASA r2013-09-07 (Haas et al. 2003) and GMAP r2013-11-27 (Wu and Watanabe 2005). Transcripts generated using Cufflinks v2.1.1 (Trapnell et al. 2010) were aligned to the genome. All four gene models were subjected to EvidenceModeler r2012-06-25 (Haas et al. 2008), to generate consensus gene models. A high weightage of 5 was given to transcript-mapping gene model predictions, whereas a value of 1 was assigned to *ab initio*-based predictions. In a second round of gene prediction, Tophat2 mapping was done on the repeat and gene-masked genome and a transcript file was generated using Cufflinks. Intron-exon boundaries were defined using the Transdecoder package r2013-11-17 (<http://transdecoder.sourceforge.net/>, last accessed September 16, 2015) (supplementary fig. S1, Supplementary Material online). After generating the final set of genes, the Cufflinks tool was used for the calculation of gene expression values, using RNA-Seq, in terms of fragments per kilobase of exon per million of fragments mapped (FPKM). FPKM values for the PSEP-encoding genes were also calculated. Genes that were having an FPKM value greater than or equal to 1 have been reported in this study.

Functional Annotation of the Proteome

Functional analyses of the predicted protein sequences were done using the BLAST2GO (Conesa et al. 2005) annotation package. The whole set of protein sequences was also aligned against eukaryotic orthologous group (KOG) sequences (Koonin et al. 2004) by using an e -value cut-off of e^{-5} in BLASTP 2.2.28+ (Altschul et al. 1990). KEGG (Kanehisa and

Goto 2000) analysis was performed using the online KAAS server (Moriya et al. 2007) and functional assignments were done using the KO file distributed from the KEGG webpage. Panther protein family analyses were performed locally by using Panther v1.03 (Mi et al. 2005) and the data version of PANTHER 9.0. Protein sequences were clustered by using SCPS v0.9.8 (Nepusz et al. 2010) with TribeMCL (Enright et al. 2002) on the predicted proteome.

Secretome Prediction

Protein sequences with a secretion signal were predicted using SignalP v4.1 (Petersen et al. 2011). Those sequences which were having cleavage site beyond the 40th amino acid position were not considered. These predictions were further refined using TargetP v1.1 (Emanuelsson et al. 2000). Only those proteins with a secretion signal according to TargetP v1.1 were kept, those which were predicted to be targeted to mitochondria were discarded. For further refinement, transmembrane domains were predicted using TMHMM v2.0c (Krogh et al. 2001). Only those proteins were kept which were predicted to have at most one transmembrane domain, all others were discarded from the secretome set.

KEGG Pathway Analysis

KEGG pathway analyses were done by using the KAAS online server applying bidirectional testing. The list of resulting protein ids with KO ids was processed to assign pathways and EC numbers using perl scripts. Protein sequences of all other species included in this study were also analyzed using the KAAS server and the presence and absence of pathways were tested using shell and perl scripts.

Panther Protein Family Analyses for Under- and Overrepresentation of Protein Families

Panther protein family analyses were done by running Panther locally on all protein sequences of six genomes using an e-value cut-off of e^{-5} . Protein ids associated with a Panther id were uploaded to the panther server (<http://www.pantherdb.org/>, last accessed September 16, 2015) for analyzing under- and overrepresentation of protein families. Multiple hypothesis testing was done using Bonferroni corrections (BC) as in Levesque et al. (2010).

Prediction of Orthologous Genes

OrthoMCL v2.0.3 (Li et al. 2003) was run to identify orthologs and paralogs within the six genomes. As in a previous study (Sharma et al. 2014), the OrthoMCL parameters "percentMatchCutoff" and "evalueExponentCutoff" were assigned a value of 50% and e^{-5} , respectively. The output files generated by OrthoMCL were processed by using perl and shell scripts. The 1:1 orthologs were extracted from the

list of orthologous gene groups. Orthologs present in the six genomes in terms of PSEPs were extracted and functional annotations were done using Panther, Gene Ontology (GO), and Interproscan v5.3-46.0 (Quevillon et al. 2005) outputs. Orthologs present in the five plant pathogens but absent in *Malassezia globosa* were extracted in addition and functional annotations were done as discussed above.

Phylogenomic Analyses

Phylogenomic analyses were performed considering the six smut genomes and including *Saccharomyces cerevisiae* as an outgroup. Orthologous genes were predicted using protein sequences of all six species (as described above) and 1:1 orthologs were extracted for multiple sequence alignments. MAFFT (Katoh and Standley 2013) v6.903 was used with the G-INS-i algorithm, to perform multiple sequence alignments. Alignments were concatenated using perl scripts. Maximum-likelihood phylogenetic inference was done using RAXML (Stamatakis 2006) v7.3.0 with 1,000 bootstraps replicates with the GTRGAMMA model.

Genes Involved in the Dimorphic Switch from Yeast to Filamentous Growth

Genes involved in the dimorphic switch from yeast to filamentous growth were predicted using orthology information generated by OrthoMCL. Genes which could not be assigned using orthology information were further searched using BLASTP and tBLASTn alignments by using an e-value cut-off of e^{-10} .

Secondary Metabolism Genes Prediction

Genes involved in secondary metabolism were predicted using SMURF (Khaldi et al. 2010). Domains of polyketide synthase (PKS) genes were further investigated by using InterPro domain information generated from BLAST2GO. These methods were performed on all six genomes analyzed in this study. The resulting PKS and nonribosomal peptide synthetases (NRPS) genes were aligned using MAFFT (Katoh and Standley 2013) v6.903 with the G-INS-i algorithm and RAXML (Stamatakis 2006) v7.3.0 was used for maximum-likelihood phylogenetic inference with 1,000 bootstrap replicates and the GTRGAMMA as substitution model.

Positive Selection Analyses

Positive selection analyses were performed considering the 1:1 orthologs among the six smut genomes. RAXML was used for generating a phylogenetic tree based on concatenated multiple sequence from the six smut genomes alignments of 1:1 orthologous sequences, which were done using MAFFT. Prank v.121002 (Loytynoja and Goldman 2008) codon alignments were performed on all 1:1 orthologous genes, omega values

were calculated using the Codeml module of the PAML package, v4.8 (Yang 2007). The branch-site model (Zhang et al. 2005), using test2 as outlined in the PAML manual was used to compare null and alternate hypotheses. In the branch-site model, positively selected branches (selection test among lineages) and sites (selection estimation at codon level within a protein sequence) were estimated by marking one branch as foreground and other five as background. This analysis was performed for all the six branches of the phylogenetic tree. Statistical testing was performed by applying a likelihood ratio test (LRT) (Anisimova et al. 2001) at a 5% level of significance. Multiple hypothesis testing was done by conducting BC and false discovery rate (FDR) tests, both at a 5% level of significance. Only those genes were considered under positive selection, which had at least one site under selection with $\geq 95\%$ Bayes Empirical Bayes (BEB) confidence at less than 5% FDR.

Functional annotations and protein family analyses of the positively selected genes were performed using Pfam r27.0 (Finn et al. 2008), InterProScan, and Panther.

Data Access

Sequences of all libraries of gDNA have been submitted to the European Nucleotide Archive (ENA) database (Study accession number PRJEB6935). The four genomic read libraries can be accessed from ENA under the accession numbers ERR583964–ERR583967. The RNA-Seq library can be found under the ENA accession number ERR583968. The assembled genome and gene models of *C. bombacis* have been filed under the accession numbers CCYA01000001–CCYA01000485. Genomic sequences and other related files are also accessible from our local server: dx.doi.org/10.12761/SGN.2015.6 (last accessed September 16, 2015).

Results

Genome Assemblies

In this study comparative genome analyses in Ustilaginomycotina were performed using the genome of *C. bombacis*, reported in this study, together with five additional representative smut genomes. These were the genomes of the human skin pathogen *M. globosa*, the smut fungi parasitizing maize, *Sporisorium reilianum* and *U. maydis*, the barley smut pathogen *Ustilago hordei* and the dicot-infecting *Melanopsichium pennsylvanicum*. The genome of *C. bombacis* was assembled in 485 scaffolds, with a total assembled genome size of 26.09 Mb. The N_{50} th scaffold size was 819.82 kb and longest scaffold was 1.61 Mb in length. To assess the quality of the assembled genome of *C. bombacis*, *N*-classes were generated from N_{10} to N_{100} . The number of scaffolds and length of each *N*-class were calculated and plotted to illustrate the quality of the genome assemblies (fig. 1A). CEGMA revealed 93.95% of complete gene mapping, using 258 core eukaryotic genes.

Genomes of *M. globosa*, *Me. pennsylvanicum*, *Sp. reilianum*, *U. hordei* and *U. maydis* had values of 96.77%, 97.18%, 97.98%, 97.18% and 97.98%, respectively. Repeat elements made up for a total of 558.54 kb (2.14%) and were masked for further genome analyses. The repeat content of the genomes of *M. globosa*, *Me. pennsylvanicum*, *Sp. reilianum*, *U. hordei* and *U. maydis* is 1%, 3%, 1%, 8% and 1%, respectively.

A total of 8,024 protein-coding genes were predicted in the genome of *C. bombacis*, out of which 576 were predicted as PSEP-encoding genes. FPKM values of all protein-encoding genes and PSEP-encoding genes were calculated. A total of 20.73% and 14.23% of all protein-encoding genes (supplementary file S1, Supplementary Material online) and PSEP-encoding genes (supplementary file S2, Supplementary Material online), respectively, were having FPKM values equal to or greater than 1. The genomes of *M. globosa*, *Me. pennsylvanicum*, *Sp. reilianum*, *U. hordei* and *U. maydis* encode for 4,286, 6,280, 6,673, 7,111 and 6,787 protein-encoding genes, respectively. Applying the same PSEP-encoding genes finding methods as for *C. bombacis* on other sequenced genomes resulted in the prediction of 193, 394, 535, 461 and 536 PSEP-encoding genes in the genome of *M. globosa*, *Me. pennsylvanicum*, *Sp. reilianum*, *U. hordei* and *U. maydis*, respectively.

Genes Potentially Involved in Pathogenicity

Putative pathogenicity-related genes have been reported in all available Ustilaginomycotina genomes (Xu et al. 2007). Generally these play a key role during the host colonization and actively interact with the host proteins during host–microbe interactions. For the comparison of these genes from *C. bombacis* with those of other Ustilaginomycotina, a functional annotation of the PSEP-encoding genes was performed in this study (table 1). These findings suggest that the *C. bombacis* genome encodes the highest number of pathogenicity-related genes compared other smut fungi, even though the overall amount is still similar (table 1). *Malassezia globosa* has a much smaller secretome than other Ustilaginomycetidae and does not have a gene encoding for Cutinase, which is present in the plant pathogenic smut fungi. As expected, no pectin esterase-encoding gene was found in the genome of *M. globosa*, whereas all sequenced plant-infecting species encode this gene.

For linking the presence or absence of certain pathogenicity-related genes with the lifestyle of oomycete and fungal pathogens, several sequenced genomes of fungal and oomycete were investigated with respect to their pathogenicity-related genes. Oomycete and fungal pathogens tested in this study have been listed in the supplementary file S3, Supplementary Material online. Similar to the animal-infecting species *M. globosa*, cutinase and pectin esterase were absent

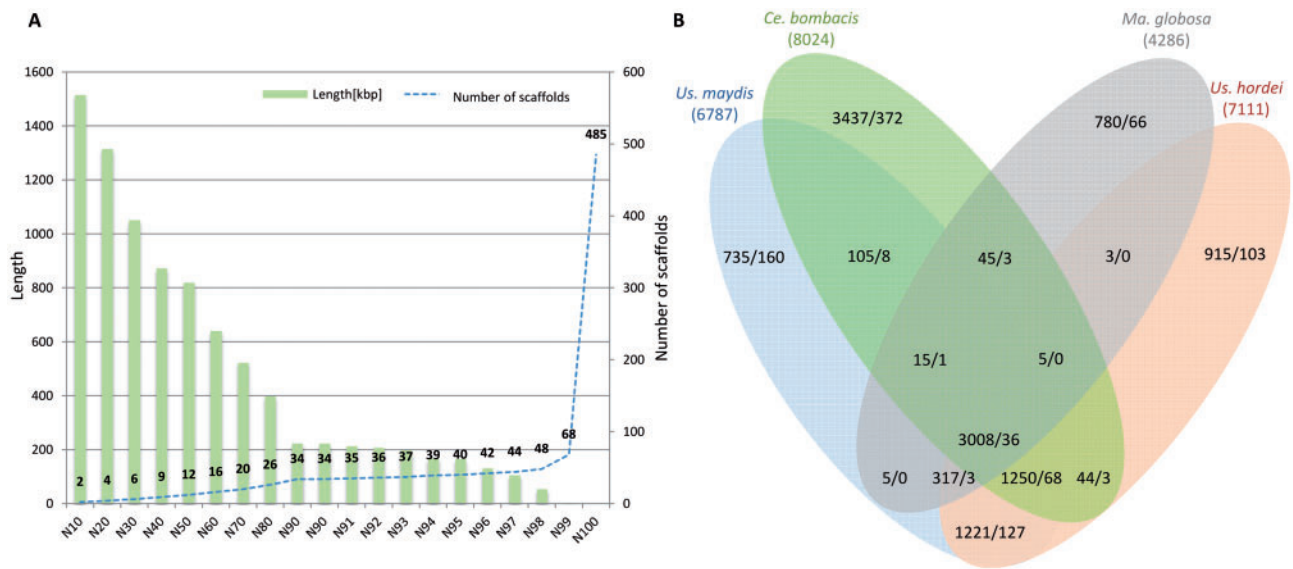


FIG. 1.—Genome assembly quality assessment and orthology analyses. (A) Genome assembly quality was assessed by calculating the number of scaffolds and length of the smallest scaffold of each N-class from N10 to N100. Each N-class was defined by considering the percentage of genome covered by sorting the assembled scaffolds from largest to smallest. Numbers on the line plot represent the number of scaffolds in each N-class. (B) Orthology analyses were performed considering all protein-coding genes and PSEP-encoding genes in the genomes of *C. bombacis*, *M. globosa*, *U. hordei*, and *U. maydis*. Numbers in parentheses represent the total number of protein-encoding genes predicted in these genomes. Orthologs have been represented considering all genes/PSEP-encoding genes. Orthology analyses of PSEP-encoding genes were performed by running OrthoMCL on the PSEP-encoding genes of all four genomes.

Table 1
Candidate Pathogenicity-Related Genes in the Six Smut Genomes

| | <i>Ceraceosorus bombacis</i> | <i>Malassezia globosa</i> | <i>Ustilago maydis</i> | <i>Sporisorium reilianum</i> | <i>Ustilago hordei</i> | <i>Melanopsichium pennsylvanicum</i> |
|-----------------------------------------------------|------------------------------|---------------------------|------------------------|------------------------------|------------------------|--------------------------------------|
| ATP-binding cassette (ABC) transporter ^a | 15 | 13 | 22 | 19 | 18 | 19 |
| Protease inhibitor ^a | 3 | 2 | 3 | 4 | 4 | 4 |
| Phospholipase ^a | 13 | 8 | 11 | 12 | 13 | 12 |
| Lipase ^a | 33 | 17 | 33 | 33 | 31 | 32 |
| Cysteine protease inhibitor ^a | 0 | 0 | 1 | 1 | 1 | 1 |
| Serine protease ^a | 54 | 26 | 51 | 54 | 47 | 46 |
| Aspartic protease ^a | 13 | 21 | 11 | 10 | 10 | 10 |
| Glycosyl hydrolases ^b | 28 | 10 | 27 | 31 | 28 | 23 |
| Cutinase ^c | 3 | 0 | 4 | 3 | 2 | 2 |
| Pectin esterase ^c | 2 | 0 | 1 | 1 | 1 | 1 |
| Cytochrome P450 ^c | 22 | 6 | 21 | 16 | 15 | 13 |
| Pectin lyase fold/virulence factor ^c | 4 | 2 | 4 | 5 | 4 | 4 |

^aPredictions were done by using PANTHER.

^bPredictions were done by using the KEGG pathway information.

^cPredictions were done by using Interpro domain information. Highest values are highlighted in bold.

in the genome of the oomycete fish pathogen *Saprolegnia parasitica*, as reported before (Jiang et al. 2013). Even though *M. globosa* encodes less than the half of PSEPs compared with other smut genomes, it encodes the highest number of Aspartic proteases. Likewise, the highest number of Aspartic proteases was predicted in the genome of *Sap.*

parasitica in comparison to plant-infecting oomycetes. Expansion of kinases has been demonstrated in the genome of *Sap. parasitica* over the other oomycete pathogens (Jiang et al. 2013). Such an expansion of kinases was not observed in the genome of *M. globosa* in comparison to the plant pathogen of this subphylum.

Table 2

Candidate Secreted Pathogenicity-Related Genes—Under- and Overrepresentation

| | <i>Malassezia globosa</i> | <i>Ceraceosorus bombacis</i> | Expected | +/- | P-value |
|---------------------------------------|---------------------------|------------------------------|----------|-----|----------|
| Carbohydrate metabolic process | 4 | 23 | 6.18 | + | 1.36E-05 |
| Polysaccharide metabolic process | 1 | 8 | 1.55 | + | 3.28E-02 |
| Protein metabolic process | 40 | 43 | 61.82 | - | 3.01E-01 |
| Cellular amino acid metabolic process | 1 | 6 | 1.55 | + | 8.59E-01 |
| Serine protease | 2 | 17 | 3.09 | + | 3.58E-06 |
| Esterase | 1 | 8 | 1.55 | + | 3.37E-02 |
| Hydrolase | 33 | 68 | 51 | + | 7.39E-01 |
| Aspartic protease | 15 | 9 | 23.18 | - | 6.97E-02 |
| Serine-type peptidase activity | 2 | 17 | 3.09 | + | 3.14E-06 |
| Aspartic-type endopeptidase activity | 15 | 9 | 23.18 | - | 6.12E-02 |

Protein Family Under- and Overrepresentation

The two members of the Exobasidiomycetes s.l., *C. bombacis*, which is a plant pathogen, and *M. globosa*, which is an animal pathogen, were investigated for protein family under- and overrepresentation. As the hosts are highly divergent, such analyses can provide insights into the enrichment or depletion of certain PSEP-encoding genes between these two genomes. These analyses revealed that the PSEP-encoding genes involved in carbohydrate, polysaccharide, and cellular amino acid metabolic processes are overrepresented in the genome of *C. bombacis*, but PSEP-encoding genes involved in protein-related processes are underrepresented in the genome of *C. bombacis* (table 2). In terms of PANTHER protein classes, glycosidase, serine proteases, esterases, and hydrolase are overrepresented in PSEP-encoding genes of *C. bombacis* in comparison to *M. globosa*. However, genes for aspartic proteases are underrepresented in the genome of *C. bombacis*. GO molecular function analyses confirmed that PSEP-encoding genes involving aspartic-type endopeptidase activity are underrepresented in the genome of *C. bombacis* (table 2).

Orthology Analyses

Orthology analyses were carried out for examining the presence and absence of orthologous genes, in particular PESP-encoding genes, among the genomes of Ustilaginomycotina pathogens. PSEP-encoding genes are often fast-evolving to adapt to changes in the targets they are operating. However, there might be some core effectors that are the hallmark of an entire group of pathogens and which are crucial for pathogenicity (Sharma et al. 2014). Doing such analyses on the genomes of two Exobasidiomycetes s.l., and two Ustilaginomycetes it was revealed that the genomes of *C. bombacis*, *M. globosa*, *U. hordei* and *U. maydis* share 3,008 orthologs, out of which 36 are genes encoding for secreted proteins (fig. 1B), which for consistency are here referred to as PSEPs, even though the authors are aware that several of these proteins might only be important for acquiring nutrients

in the yeast stage. Similar analyses were performed considering two additional members of the Ustilaginomycotina. These analyses revealed that the six genomes share 2,942 orthologs, out of which 30 were PSEP-encoding genes (supplementary file S4, Supplementary Material online). Functional annotation of these suggest that the majority of these genes are genes encoding for enzymes such aspartic peptidase, glycoside hydrolase, and thioredoxin/protein disulfide isomerase (supplementary file S4, Supplementary Material online), which might contribute to nutrient acquisition in the yeast stage.

Orthologs that are present in the plant pathogens but absent in *M. globosa* were also assessed. These investigations revealed 1,152 orthologs absent in *M. globosa* out of which 53 were PSEP-encoding genes (supplementary file S5, Supplementary Material online). Functional annotations using Pfam, Interpro, and Panther analyses of these 53 genes revealed that the majority of these genes are hypothetical proteins, but another larger fraction is predicted to be involved in the breakdown of carbohydrates (supplementary file S5, Supplementary Material online). Orthology analyses revealed the presence of all RNA interference (RNAi) genes in the genome of *C. bombacis* (supplementary file S6, Supplementary Material online). Consistent with the previous studies, RNAi-related genes were not found in the genomes of *U. maydis* (Laurie et al. 2012) and *M. globosa* (Xu et al. 2007).

Phylogenomic Analysis

Extensive phylogenetic work has been performed to infer the phylogenetic relationships of the Ustilaginomycotina, but the majority of these studies was based on a handful of genes only (Begerow et al. 2006). For elucidating the phylogenetic relationships using all 1:1 orthologous genes among these pathogens, genome-wide orthology and phylogenetic analyses were carried out in this study. A total of 1,629 1:1 orthologs were identified among seven genomes, including *S. cerevisiae* as an outgroup. These analyses inferred a sister-group

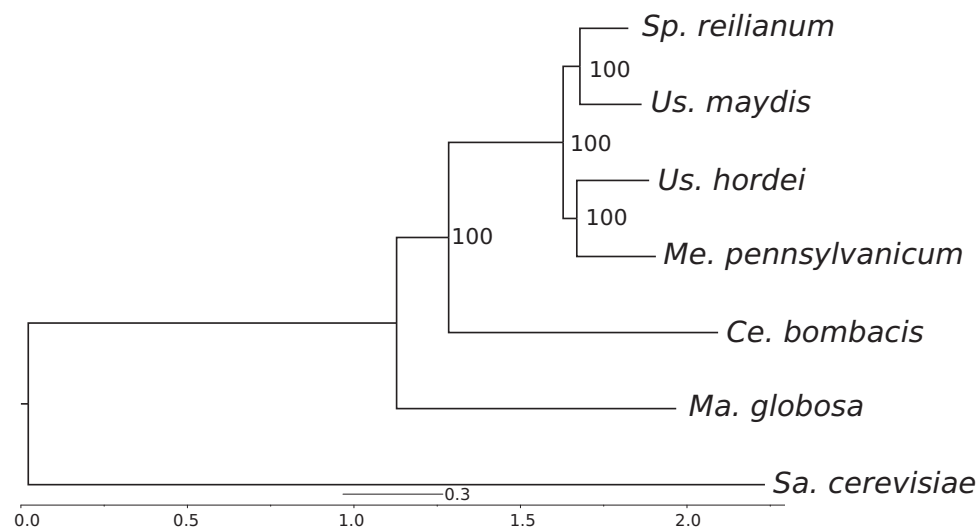


Fig. 2.—Phylogenetic relationships among the six Ustilaginomycotina pathogens. Phylogenetic tree based on alignments of all 1,629 1:1 orthologs as inferred by maximum-likelihood analysis using RAxML. Support values from bootstrap replicates are indicated on the branches.

relationship of *U. maydis* and *Sp. reilianum* with maximum support, as did the sister-group relationship of *U. hordei* and *Me. pennsylvanicum* (fig. 2). These four species were grouped together with maximum support. *Cereceosorus bombacis* was the next-diverging lineage and *M. globosa* was placed basal to all five plant pathogenic smut fungi investigated.

Genes Involved in the Dimorphic Switch from Yeast to Hyphal Growth

Most Ustilaginomycotina species have the capability to switch from yeast to hyphal growth, known as dimorphic switch. This transition is often associated with the transition from saprotrophic to pathogenic growth. The six Ustilaginomycotina genomes were scanned for the genes involved in the dimorphic switch, using the genes of *U. maydis* as a reference. Even though the type strain of *C. bombacis* grew hyphal, the genome contains all key genes involved in the dimorphic switch that have been identified in *U. maydis*. The same methods were applied to identify these genes in other pathogens (supplementary file S7, Supplementary Material online), interestingly, *Prf1* (Pheromone response factor) and *Rop1* (Rhoptry protein 1) could not be identified in *M. globosa* suggesting an impairment in filamentation in this animal pathogen.

Conservation of Validated Secreted Effectors

In this study, sequences of several validated *U. maydis* effectors were searched within the five other Ustilaginomycotina genomes, with the aim of understanding the evolutionary history of these effector proteins. Past studies have reported that *Pep1* (protein essential during penetration-1) of *U. maydis* plays a vital role during plant–microbe interaction, being crucial for the penetration into the maize epidermis

(Doehlemann et al. 2009). In this study, the six smut genomes were investigated for the presence of *Pep1* orthologs. *Pep1* could not be found in the genomes of *C. bombacis* and *M. globosa*, but were present in the other smut genomes investigated, in line with previous investigations (Sharma et al. 2014; Hemetsberger et al. 2015). The *Tin2* effector of *U. maydis* triggers the activation of the ZmTTK1 kinase in maize, which further affects anthocyanin biosynthesis (Tanaka et al. 2014). An orthologous gene of *U. maydis Tin2* is present in *Sp. reilianum* (*sr10057*), but only absent in the other smut genomes investigated.

Comparative analyses revealed the absence of a sucrose transporter (*Srt1*), which is localized in the host plasma membrane, in the genome of *M. globosa*. However, the *U. maydis Srt1* gene *um02374* (Wahl et al. 2010) has orthologs in all plant pathogenic smuts investigated. Similar analyses considering the *O*-mannosyltransferase *Pmt4* gene of *U. maydis* (Fernandez-Alvarez et al. 2012) revealed orthology in all plant pathogens (supplementary file S8, Supplementary Material online).

The conservation of the *U. maydis Pit* cluster (Doehlemann et al. 2011) was checked in other pathogen genomes. Interestingly, orthologous genes of all four genes of the *Pit* cluster were found only in *Me. pennsylvanicum* and *U. hordei*. In the genome of *Sp. reilianum*, an ortholog of *Pit2*, which encodes a secreted protein, could not be found using orthology searches. However, further analyses suggested the presence of *Pit2* in the *Pit* cluster of *Sp. reilianum* with high sequence divergence compared with to *Pit2* genes of other smut fungi, suggesting a high evolutionary pressure on this gene in this species; however, positive selection analyses did not reveal a signature of selection pressure on this gene. Only

Pit3 of the *Pit* cluster could be found in *C. bombacis* and *M. globosa* (fig. 3).

Hum3 and *Rsp1* of *U. maydis*, which are known to have an essential function during the pathogenic development (Müller et al. 2008), have no orthologous genes in the genome of *C. bombacis* and *M. globosa*; however, orthologs of these genes were present in the other three smut pathogen genomes (supplementary file S8, Supplementary Material online).

Positive Selection Analyses

The survival of the fittest, a process often referred to as natural “selection”, leads to the diversification of organisms and the fixation of mutations that are beneficial for the genome in which they are present (Wallace 1858; Darwin 1859). This process also plays a key role during the evolutionary adaptation of pathogens (Kemen et al. 2011; Dong et al. 2014; Sharma et al. 2014). For unravelling the patterns of positive selection in the evolution of Ustilaginomycotina pathogens, genome-wide positive selection analyses were carried out among the six genomes. The codeml module of the PAML package was used for the prediction of positively selected lineages and protein sites. These analyses revealed the highest percentage of positively selected genes in the genome of *Me. pennsylvanicum* (fig. 4A), but with respect to PSEP-encoding genes, *C. bombacis* showed the highest percentage of genes selected, outranking the dicot-infecting smut *Me. pennsylvanicum* (fig. 4B). Generally, PSEP-encoding genes had a higher proportion of positively selected genes than proteins predicted not to be secreted and thus in less intimate contact with the host environment. Positively selected PSEP-encoding genes of the six species are listed in the supplementary file S9, Supplementary Material online. Category enrichment analyses gave no statistically supported results, due to the high proportion of unannotated positively selected PSEP-encoding genes. However, Panther and GO annotations (supplementary file S9, Supplementary Material online) suggest that protease-encoding genes are under comparatively high selection pressure.

Genes Involved in Secondary Metabolism

Secondary metabolites can be key factors in pathogenicity and previous studies have revealed the presence of genes encoding for proteins involved in secondary metabolite production, such as polyketides and nonribosomal proteins in fungal genomes (Xu et al. 2007; Waskiewicz et al. 2010). Evolutionary studies of such genes in pathogens are so far largely limited to ascomycetes. To fill this knowledge gap, the six Ustilaginomycotina genomes were scanned for genes involved in secondary metabolism. The genome of *C. bombacis* encodes five PKS-like genes (table 3). Out of these five PKS-like genes, three have a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP) domain, essential for the functioning of classical PKSs (fig. 5A). Out of these three PKSs, two have a ketoreductase (KR)

domain and one is featuring both a dehydratase (DH) and a KR domain. Similar analyses on other Ustilaginomycetidae genomes show that *M. globosa* genome encodes only one classical PKS, proteins with an enoylreductase (ER) and proteins with a KR domain. The Ustilaginales plant pathogens including *Me. pennsylvanicum*, *Sp. reilianum*, *U. hordei* and *U. maydis* encode three, three, three and five PKS-like proteins, respectively.

Phylogenetic inference and orthology studies suggest that two of the three PKS genes (both with a KR domain) of *C. bombacis* do not have any orthologous gene in the other five genomes (fig. 5B). One PKS, which has both a DH and a KR domain, has an orthologous gene only in *M. globosa*. The PKSs of *Me. pennsylvanicum*, *Sp. reilianum*, *U. hordei* and *U. maydis* are orthologous to each other (fig. 5B and C).

Interestingly, 24 and 23 genes in the two dicot-infecting Ustilaginomycotina, *C. bombacis* and *Me. pennsylvanicum*, encode for proteins carrying only an ER domain (fig. 5A). Orthologs of these proteins predicted by OrthoMCL were detected in the genomes of the other pathogens, which are parasitic to monocots, but none of these features an ER domain according to Interpro domain searches (fig. 5D).

KEGG Pathway and Protein Family Comparisons

KEGG pathway analyses were carried out for evaluating the presence and absence of pathways within the genomes of fungal and oomycete pathogens. The aim of these analyses was to link the presence and absence of such pathways to the lifestyle of a particular pathogen. By this common evolutionary trajectories of phylogenetically distant pathogens with similar lifestyles can be investigated. Enzymes which play a key role in nitrogen and sulfur metabolism were searched for in the genomes of six Ustilaginomycotina genomes, and the genomes of several other pathogens, representing different phylogenetic groups in fungi and oomycetes (supplementary file S3, Supplementary Material online).

It was revealed that all fungal biotrophic non-obligate plant pathogens are having all five genes involved in nitrogen metabolism (fig. 6A). Nitrite reductase was not present in any of the hemibiotrophic *Phytophthora* spp. genomes, whereas the remaining four enzymes of the corresponding pathway were present. Similarly, the nitrite reductase-encoding gene was also absent in all fungal hemibiotrophic pathogens included. Also, obligate biotrophic pathogen genomes from both fungi and oomycetes lack the genes encoding for nitrite (all) and nitrate (with the exception of *Melampsora laricis-populina*) reductase. Interestingly, both fungal and oomycete animal nonobligate pathogen genomes lack both nitrate and nitrite reductase-encoding genes.

Similarly, key enzymes required for sulfur metabolism were also searched for. It was revealed that the genomes of oomycete animal pathogens and oomycete obligate biotrophic

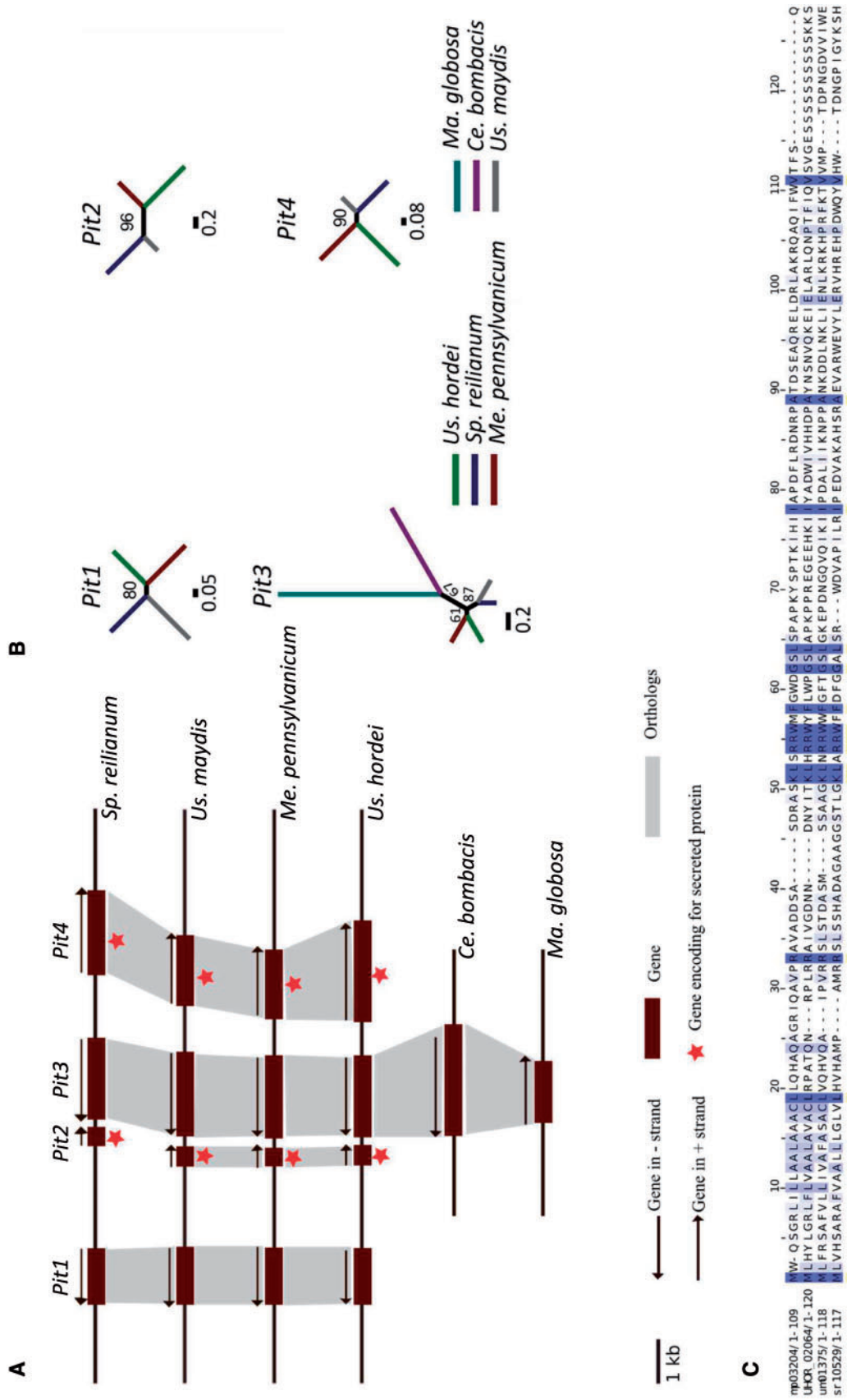


Fig. 3.—Conservation of the *U. maydis* Pit (proteins important for tumors) cluster. (A) Conservation of the *U. maydis* Pit cluster was tested within all six pathogen genomes. The Pit cluster is composed of four genes in total; out of these four genes two genes (*Pit2* and *Pit4*) are predicted to be PSEP-encoding genes. (B) Phylogenetic trees representing the divergence of Pit genes. The phylogenetic inference using the maximum-likelihood algorithm was calculated using RAxML. Numbers represent bootstrap support values from 1,000 bootstrap replicates. (C) Multiple sequence alignments of *Pit2* of the four Ustilaginaceae pathogen genomes. The *Pit2* gene was found conserved in four smut genomes; however, *Pit2* of *Sp. reilianum* showed a high sequence divergence and was therefore not predicted as an ortholog.

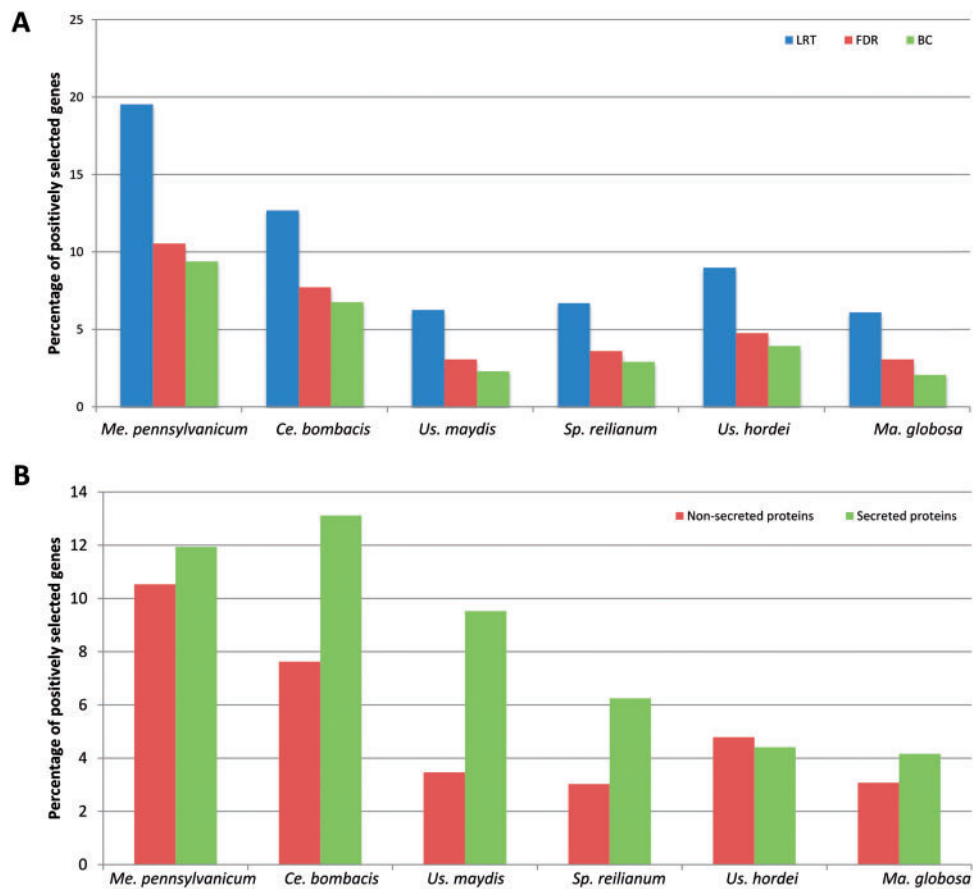


Fig. 4.—Positively selected genes within Ustilaginomycotina. (A) Percentage of positively selected genes among the six genomes. Statistical significance was inferred by LRT, FDR, and BC. All tests were done at a 5% level of significance and only those genes with at least one positively selected site with a BEB confidence $\geq 95\%$ were considered. (B) Comparisons of positively selected non-PSEP-encoding genes and PSEP-encoding genes. Only those genes were considered which were having at least one positively selected site with a BEB confidence $\geq 95\%$ and a FDR at a 5% level of significance.

Table 3

Number of Genes Encoding Secondary Metabolite Synthases

| | DMAT | NRPS | PKS |
|--------------------------------------|------|------|------|
| <i>Ceraceosorus bombacis</i> | 2 | 7(2) | 5(3) |
| <i>Malassezia globosa</i> | 1 | 3(1) | 1(1) |
| <i>Melanopsichium pennsylvanicum</i> | 1 | 5(1) | 3(2) |
| <i>Sporisorium reilianum</i> | 2 | 6(2) | 3(2) |
| <i>Ustilago hordei</i> | 0 | 7(3) | 3(2) |
| <i>Ustilago maydis</i> | 2 | 9(3) | 5(3) |

NOTE.—Numbers in brackets represent high confidence values (candidate genes harboring all canonical domains).

plant pathogens lack a sulfite reductase (fig. 6B). In contrast to this, all key genes required for sulfur metabolism were present in the genomes of fungal animal and fungal obligate biotrophic plant pathogens (fig. 6B). All investigated fungal and oomycete hemibiotrophs harbor all essential genes for the sulfur metabolism.

Discussion

General Genome Characteristics

Extensive comparative genomic studies have been conducted on plant pathogens of the Ustilaginomycetes (Kämper et al. 2006; Xu et al. 2007; Schirawski et al. 2010; Que et al. 2014; Sharma et al. 2014). But these comparative studies were limited to the crown group of smut fungi, the Ustilaginaceae. Moreover, several evolutionary studies of plant and animal pathogens were focused either on certain pathogenic lineages, that is, oomycetes (Haas et al. 2009; Baxter et al. 2010; Levesque et al. 2010) or on fungi (Kämper et al. 2006; Xu et al. 2007; Schirawski et al. 2010; Que et al. 2014; Sharma et al. 2014), or were limited to certain lifestyles (McDowell 2011). Thus, this study was performed to investigate patterns of parallel evolution resulting from common evolutionary trajectories across several lifestyles. As there was no genome available for an early diverging plant pathogenic member of the Ustilaginomycotina, the genome of *C.*

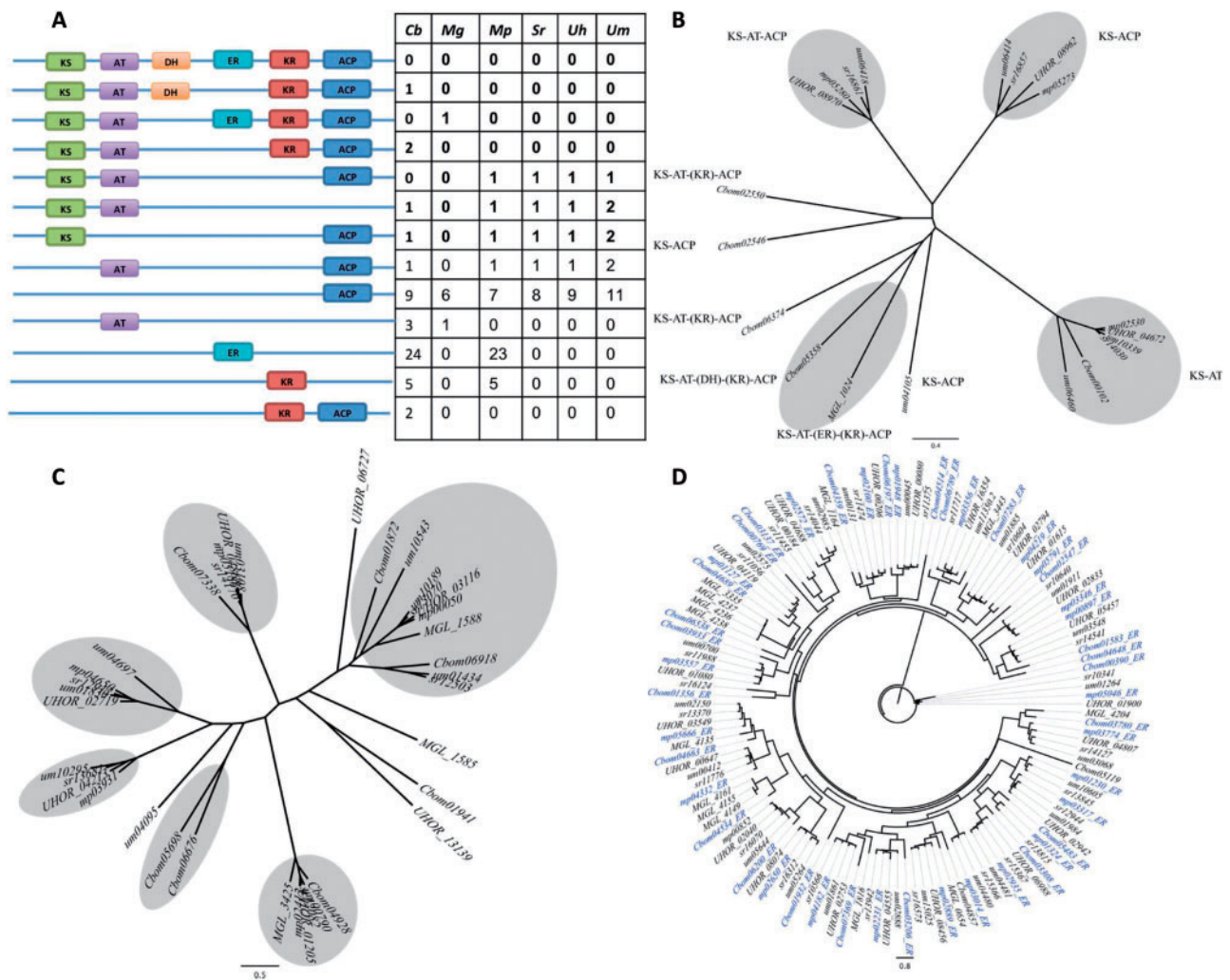


Fig. 5.—Secondary metabolism. (A) Numbers of proteins featuring the PKS domains KS, AT, ACP, KR, DH, and ER. *Cb*, *Mg*, *Mp*, *Sr*, *Uh* and *Um* represent the presence of a respective domain combination in the genomes of *C. bombacis*, *M. globosa*, *Me. pennsylvanicum*, *Sp. reilianum*, *U. hordei* and *U. maydis*, respectively. Number in bold represents the total number of PKS and PKS-like protein-encoding genes. (B) Orthology and phylogenetic tree of PKS and PKS-like genes of all six species. Gray shading indicates orthologs. Phylogenetic analysis was conducted using maximum-likelihood inference as implemented in RAxML with 1,000 bootstrap replicates; support values are indicated on the branches. (C) Phylogenetic analyses conducted by including NRPS and NRPS-like genes of all six pathogens. (D) Phylogenetic tree based on all genes with an ER domain as well as their orthologs in other genomes. Proteins with an ER domain have been represented in blue font.

bombacis, a member of the Exobasidiomycetes, has been sequenced.

With 26.09 Mb the assembled genome size of *C. bombacis* is bigger than that of other Ustilaginomycotina. However, the genome of *C. bombacis* contains less repeat elements than *U. hordei* (Laurie et al. 2012) and *Me. pennsylvanicum* (Sharma et al. 2014). The genome of *C. bombacis* encodes for slightly higher number of protein-encoding genes in comparison to the other sequenced members of Ustilaginomycotina. Notably, the genome of *M. globosa* contains the lowest number of protein-encoding genes in comparison to the sequenced Ustilaginomycotina genomes,

with only 53% of the amount of protein-encoding genes as compared with *C. bombacis*. The low amount of protein-coding genes is probably a result of the loss of genes required for plant pathogenicity after the jump to an animal host, but most likely also additionally affected by the loss of the hyphal stage. However, this hypothesis can only be tested with the availability of more genomic sequences of the Exobasidiomycetes s.l. and Malasseziales.

The genome of *C. bombacis* harbors all genes that play a key role in the RNAi. Such genes were absent in the genomes of *M. globosa* (Xu et al. 2007) and *U. maydis*

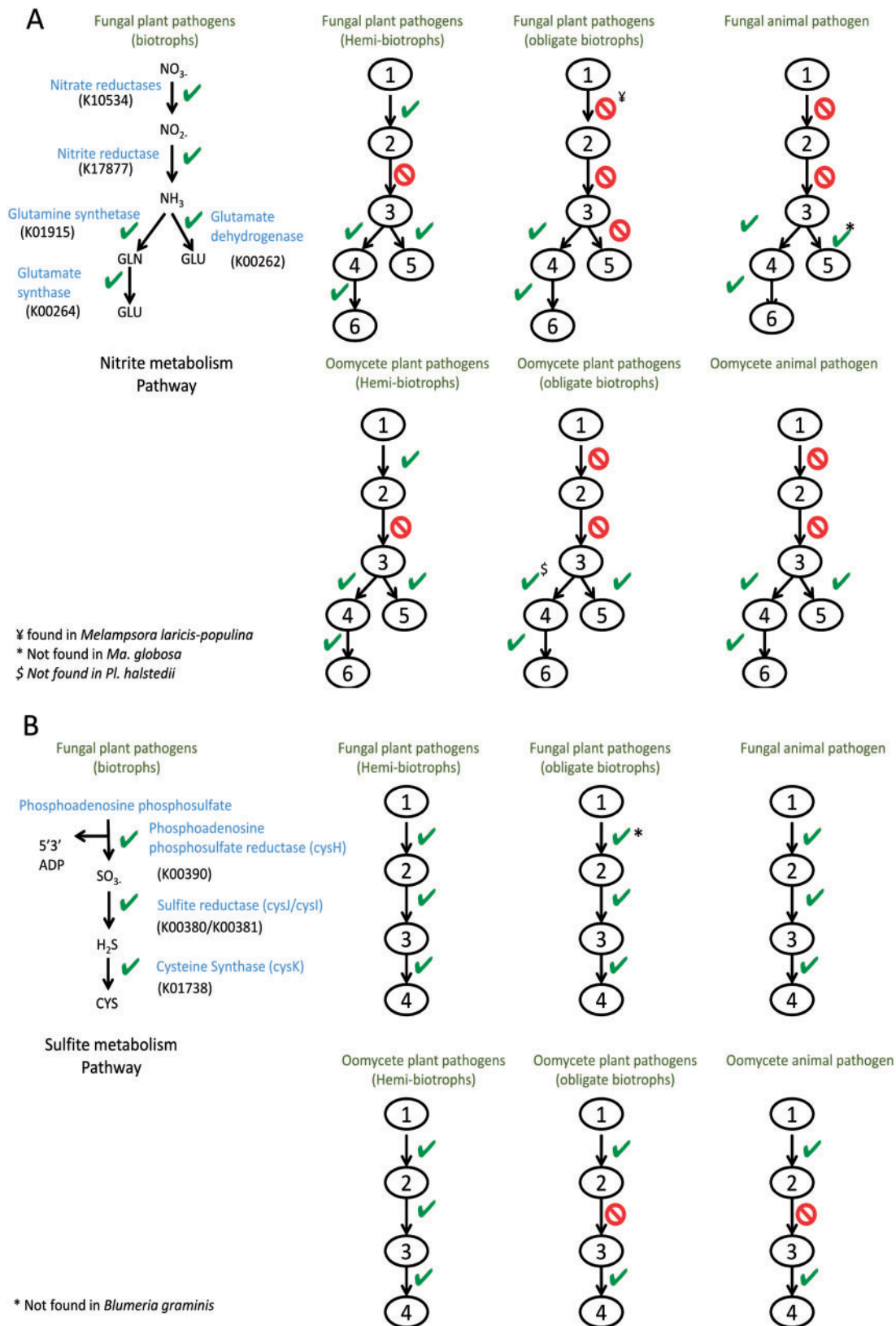


FIG. 6.—Nitrogen and sulfur metabolism pathways within fungal and oomycete pathogens, adapted from Jiang et al. (2013). Nitrogen (A) and sulfur (B) metabolism pathway genes were predicted with the fungal and oomycete plant and animal pathogens in relation to the lifestyle of these pathogens.

(Laurie et al. 2008, 2012), suggesting that this mechanism is not dispensable in *C. bombacis*.

Phylogenomic Analyses

Many phylogenetic studies using internal transcribed spacer (ITS), large ribosomal subunit (LSU), and a few protein sequences have been performed to study the phylogeny of Ustilaginomycotina (Suh and Sugiyama 1993; Begerow et al. 2006; Xu et al. 2007). In this study, maximum bootstrap support for a sister-group relationship of *U. maydis* and *Sp. reilianum* was observed for the first time. Also the sister-group relationship of *U. hordei* and *Me. pennsylvanicum* received maximum support, highlighting the difficulties of the current generic concept in the Ustilaginaceae. *Ustilago hordei*, the type species of the genus, is more closely related to *Me. pennsylvanicum* than to *U. maydis* (Begerow et al. 2006; McTaggart et al. 2012). The latter species is possibly more closely related to *Sporisorium sorghi*, the type species of *Sporisorium*, and to *Sp. reilianum* than to *U. hordei* (Begerow et al. 2006; McTaggart et al. 2012). This leaves the possibility to transfer *U. maydis* to the genus *Sporisorium*, which is however undesirable, because *U. maydis* is the most widely known smut fungus and the combination *Sporisorium maydis* would not even be available, as an arguably *Aspergillus*-like fungus has been described as *Sp. maydis* in the 19th century (Cesari and Dante 1845). Detailed taxonomic studies have to reveal, whether the name *U. maydis* should be retained in the future.

Genes Involved in Yeast–Filament Transition

Dimorphism is very common in fungal pathogens and a switch from yeast to filamentous growth is often needed for their pathogenicity. Many studies have reported and discussed the genes involved in this switching of unicellular to multicellular forms of fungal pathogens (Bölker 2001; Sanchez-Martinez and Perez-Martin 2001; Nadal et al. 2008; Elias-Villalobos et al. 2011). Here, the presence of these genes within all five plant pathogens including *C. bombacis* was revealed, whereas many genes which play key role in dimorphism were absent in the genome of *M. globosa*, in line with a previous study (Xu et al. 2007). As *M. globosa* apparently has lost pathogenicity to plants and grows only superficially, hyphal growth probably became dispensable in this organism.

Polyketide Synthases

Fungi encode several genes involved in producing secondary metabolites (Yu, Bhatnagar, et al. 2004; Yu, Chang, et al. 2004; Dean et al. 2005; Ehrlich et al. 2005; Xu et al. 2007). Type 1 PKS-encoding genes are a well-studied class of secondary-metabolite-related genes and phylogenomic studies have provided insights into the evolution of these genes in fungi and bacteria (Kroken et al. 2003). So far, little is known regarding the evolution of secondary metabolite genes in

Ustilaginomycotina (Xu et al. 2007). To shed light on the evolution of these genes, type 1 PKS and NRPS genes within six Ustilaginomycotina genomes were predicted and subjected to evolutionary analyses. Only PKSs from Exobasidiomycetes s.l. had additional domains in addition to the PKS-defining KS, AT, and ACP domains, reflecting either the loss of additional domains in the PKS genes of Ustilaginales or the gain of additional DH, KR, and ER domains by PKS genes of *C. bombacis* and *M. globosa* for producing reduced polyketides. The finding that only few simple PKSs are encoded by the genomes of Ustilaginales is in line with the only few polyketides so far reported from this group (Xu et al. 2007; Bölker et al. 2008).

It was found that the genomes of the dicot-infecting species *C. bombacis* and *Me. pennsylvanicum* were encoding for proteins which were featuring only an ER domain known from PKS genes. Further analyses suggest that the orthologs of these genes were present in other monocot pathogens (*U. maydis* and *U. hordei*) and the animal pathogen *M. globosa*, but without carrying ER domains. It could thus be speculated that the presence of these proteins with ER domains coincides with pathogenicity on dicot plant hosts, even though the function of these proteins remains enigmatic.

Orthology and Positive Selection Analyses

The predicted set of 30 genes encoding secreted proteins in all smut genomes reflects the conservation of a core set of genes that might play an essential role for survival in the yeast stage or for infecting another organism. As functional analyses showed that the majority of these genes belongs to either peptidase, glycoside hydrolase, or thioredoxin/protein disulfide isomerase families, the larger part of these might be necessary for nutrient acquisition in the yeast stage. But at present it cannot be excluded that some of the encoded proteins also play a pivotal role during host infection.

A total of 53 genes encoding for secreted proteins were absent in *M. globosa* but present in all plant pathogens. This suggests the conservation of 53 orthologs in plant infecting Ustilaginomycotina genomes, which are probably not required for basic functioning of the yeast stage but are important for infecting plant hosts. Thus, as expected, a gene encoding for cutinase was present among the 53 orthologs. The largest group of these 53 orthologs consisted of hypothetical proteins of unknown function, followed by glycosyl hydrolase family proteins, which are depleted in the genome of *M. globosa*. A low number of glycosyl hydrolase protein family encoding genes in the genome of *M. globosa* has also been described before (Xu et al. 2007), corroborating the present depletion analysis. A total of 35% of the 53 orthologs encoding secreted proteins exclusive to the plant pathogenic Ustilaginomycotina were lacking functional annotations and probably encode for effector genes needed for basic plant pathogenicity. This set of genes might be a valuable source

for future functional studies which might provide further insights into plant pathogenicity.

All 1:1 orthologous genes of the six Ustilaginomycotina genomes investigated in this study were investigated for signatures of positive selection. As reported before in fungal and oomycete pathogens (Schirawski et al. 2010; Kemen et al. 2011; Sharma et al. 2014), genes encoding secreted proteins are under higher selection pressure than genes encoding non-secreted proteins. *Melanopsichium pennsylvanicum* showed the highest percentage of positively selected genes among the Ustilaginales pathogens, possibly because of the huge host jump of this pathogen from grasses to a dicot host (Sharma et al. 2014). However, *C. bombacis* had the highest proportion of PSEP-encoding genes. Functional analyses of positively selected PSEP-encoding genes of all species show that genes encoding for hypothetical proteins, serine protease, phospholipase, phosphodiesterase, and chaperone proteins are under high selection pressure, probably owing to the adaptation to the specific host environments (supplementary file S9, Supplementary Material online).

PSEP-Encoding Genes

The genome of *C. bombacis* was predicted to encode for 8,048 proteins, out of which 576 are PSEP-encoding genes. The number of protein-coding genes in the genome of *C. bombacis* is higher than the other sequenced Ustilaginomycotina genomes, and also the amount of PSEP-encoding genes is slightly higher. The comparatively low number of PSEP-encoding genes in the genome of *M. globosa* might be the result of an expansion of few gene families as an adaptation to infecting animal hosts, but a loss of most PSEP required for plant pathogenicity, such as cutinase or intracellular effectors that are targeting plant-specific defense pathways. Also *Me. pennsylvanicum* has lost several PSEP-encoding genes as a result of a host jump to a dicot host (Sharma et al. 2014), and the impact of a host jump from plant to animal hosts, with the complete loss of haustoria might have resulted in an even more reduced secretome.

To unravel common patterns in oomycete and fungal pathogens with respect to pathogenicity-related genes, all six smut genomes were investigated and compared with the situation in oomycetes. The apparent absence of cutinase and pectin esterase-encoding genes in the genome of *M. globosa* is in line with a previous study (Xu et al. 2007). Similarly the absence of these genes were observed in the genome of the oomycete animal pathogens *Sap. parasitica* and *Aphanomyces astaci* whereas these genes were present in the genomes of related plant-infecting hemibiotrophic *Phytophthora* species (Jiang et al. 2013) and the obligate biotrophic species *Hyaloperonospora arabidopsidis* (Baxter et al. 2010). The lack of these genes might constitute a major hurdle for the switch from animal to plant parasitism, at least in case of (hemi-)biotrophic species.

No cysteine protease inhibitor gene could be identified in the genomes of *C. bombacis* and *M. globosa*; however, these were present in all other plant pathogenic Ustilaginomycotina and have also been reported as gene families in plant pathogenic oomycetes (Tian et al. 2007), suggesting a parallel evolution of this feature. In total 21 Aspartic proteases were predicted in the genome of *M. globosa* which is the highest amount among smut genomes, which is probably an adaptation to animal pathogenicity, as also in the animal-pathogenic oomycete *Sap. parasitica*, this family is larger than in the plant pathogenic hemibiotrophic and obligate biotrophic oomycetes.

In contrast, secreted serine proteases were underrepresented in the animal pathogen *M. globosa* compared with plant pathogenic Ustilaginomycetidae. However, this was not the case for the oomycete plant and animal pathogens, probably reflecting that the depletion of such enzyme-encoding genes within the genome of *M. globosa* is not specifically associated with its presence on an animal host.

Validated Effector Conservation

Secreted effectors are usually considered as fast-evolving and thus often show very low sequence conservation. However, some studies have reported the orthology of effectors across species in some pathogen groups (Quinn et al. 2013; Sharma et al. 2014). In this study, it was found that *Pep1* conservation (Doehlemann et al. 2009; Hemetsberger et al. 2012) probably does not extend to the genomes of *C. bombacis* and *M. globosa*. This suggests that *Pep1* might be a synapomorphic trait for the Ustilaginales.

The *U. maydis* *Pit* cluster is composed of four genes, out of which two are having a secretion signal and are required for tumor formation (Doehlemann et al. 2011). Conservation of this cluster was observed in the genomes of the four Ustilaginales genomes (fig. 2A), whereas this gene cluster was found to be absent in the genomes of both *C. bombacis* and *M. globosa*, with the exception of the nonsecreted *Pit3* gene, which was present in all Ustilaginomycotina genomes.

The conservation of some pathogenicity-related proteins, including some PSEPs, suggests that although the majority of PSEP-encoding genes is fast-evolving and thus difficult to track across species or genera, few pathogenicity-related genes remain conserved within larger phylogenetic entities, thus probably constituting the core arsenal of pathogenicity-related proteins that are the hallmark of the evolutionary success of the corresponding pathogen groups.

Absence and Presence of Enzymes of Key Metabolic Pathways

Detailed studies have been performed on the metabolic pathways in some fungal and oomycete plant and animal pathogens (Xu et al. 2007; Baxter et al. 2010; Jiang et al. 2013). These studies have reported the absence of genes

involved in nitrite metabolism in the genome of obligate biotrophic plant pathogenic oomycetes (Baxter et al. 2010; Kemen et al. 2011), the oomycete fish pathogen *Sap. parasitica* (Jiang et al. 2013), and the fungal obligate biotrophic plant pathogen *Melampsora laris-populina* (Duplessis et al. 2011). To relate this absence of genes to the lifestyle of oomycete and fungal pathogens, the analyses in this study included oomycete hemibiotrophic plant pathogens, obligate biotrophic plant pathogens, and animal pathogens, as well as plant pathogenic fungal biotrophs, hemi-biotrophs, obligate biotrophs, and animal pathogens (supplementary file S3, Supplementary Material online). These analyses revealed the absence of both nitrate reductases and nitrite reductase in both fungal and animal facultative pathogens. This is probably reflecting the abundant nitrogen source which is in the proteins of their hosts in conjunction with a specialization for the colonization of protein-rich debris in the absence of a compatible host.

Interestingly, nitrite reductase was found absent in the hemi-biotrophic pathogens of both fungi and oomycetes, which can probably dispense the energy-consuming conversion from nitrite to ammonium as they rely primarily on the acquisition of reduced nitrogen from their host plants.

Additional key enzymes for nitrite metabolism were absent in the genomes of obligate biotrophs of both fungi and oomycete pathogens. Similar to the situation in facultative animal pathogens this is probably reflecting the fact that they fully depend on the host to supply nutrients and thus do no longer require the energy-consuming pathways for reducing inorganic nitrogen.

However, all nitrite metabolism enzymes were found in the genomes of the five biotrophic plant pathogens of the Ustilaginomycotina included in this study, suggesting that these pathogens are still capable of reducing inorganic nitrogen. This is most likely due to the fact that the Ustilaginomycotina have retained a saprotrophic yeast stage, which might depend on nitrate as a nitrogen source. A recent study on the genome of *Sporisorium scitamineum* (Que et al. 2014) showed that the sugarcane smut can use nitrate as a source of nitrogen.

The parallel loss of key enzymes for nitrogen acquisition in both oomycetes and fungi hints at common evolutionary trajectories that have shaped the genome on correlation with different lifestyles. The finding that of the organisms included in this study only obligate biotrophic oomycetes have lost the ability to use sulfate as a source of sulfur suggests that this pathway is not as easily lost, probably due to some advantage in the ability to reduce sulfur even reduced sulfur can be acquired from the hosts. As an alternative explanation, additional vital reactions might be catalyzed by the enzyme, apart from those for sulfur acquisition. Future studies on a broader set of pathogens will reveal, whether more metabolic pathways are impacted across eukaryotic kingdoms in a similar manner as nitrogen metabolism.

Overall, it was found that there are several signatures of parallel evolutions among the plant and animal pathogens of fungi and oomycetes. Thus, it can be concluded that although fungal and oomycete pathogens diverged long back, there are several common evolutionary trajectories which shaped their genomes in relation to a certain lifestyle. In addition to this, it was found that the plant pathogenic Ustilaginomycotina share a core set of PSEPs absent from the animal pathogenic *M. globosa*. This suggests that these genes might play crucial roles in plant pathogenicity, a hypothesis which can be tested in future functional studies.

Supplementary Material

Supplementary files S1–S9 and figure S1 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Claus Weiland for support with respect to cluster access and Eva-Maria Gerstner for providing them a DOI for data files. They also thank Imran Khaliq for providing useful comments on the manuscript. This work was supported by the research funding program LOEWE “Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz” of Hesse’s Ministry of Higher Education, Research, and the Arts in the framework of the LOEWE research cluster for Integrative Fungal Research (IPF). This manuscript is dedicated to the memory of Robert Bauer, who passed away during this study. The authors declare that no competing interest exists.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol.* 18:1585–1592.
- Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12:1269–1276.
- Baxter L, et al. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549–1551.
- Begerow D, Stoll M, Bauer R. 2006. A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data. *Mycologia* 98:906–916.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.
- Boa E. 2001. Compendium of corn diseases. St Paul (MN): American Phytopathological Society Press.
- Bode HB. 2009. Entomopathogenic bacteria as a source of secondary metabolites. *Curr Opin Chem Biol.* 13:224–230.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bölker M. 2001. *Ustilago maydis*—a valuable model system for the study of fungal dimorphism and virulence. *Microbiology* 147: 1395–1401.
- Bölker M, Basse CW, Schirawski J. 2008. *Ustilago maydis* secondary metabolism—from genomics to biochemistry. *Fungal Genet Biol.* 45(Suppl 1):S88–S93.

- Borodovsky M, Lomsadze A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics*. Chapter 4:Unit 4 6 1-10.
- Cesari A, Dante A. 1845. *Bellezze della Commedia di Dante* Alighieri. Parma (Italy): Fiaccadori.
- Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
- Cunningham OL, Bakshi BK, Lentz PL, Gilliam MS. 1976. Two new genera of leaf-parasitic fungi (Basidiomycetidae: Brachybasidiaceae). *Mycologia* 68: 640–654.
- Darwin C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: W. Clowes and Sons.
- Dean RA, et al. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434:980–986.
- Doehlemann G, et al. 2009. Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. *PLoS Pathog.* 5:e1000290.
- Doehlemann G, Reissmann S, Assmann D, Fleckenstein M, Kahmann R. 2011. Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. *Mol Microbiol.* 81:751–766.
- Dong S, et al. 2014. Effector specialization in a lineage of the Irish potato famine pathogen. *Science* 343:552–555.
- Duplessis S, et al. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A.* 108:9166–9171.
- Ehrlich KC, Yu J, Cotty PJ. 2005. Aflatoxin biosynthesis gene clusters and flanking regions. *J Appl Microbiol.* 99:518–527.
- Elias-Villalobos A, Fernandez-Alvarez A, Ibeas JI. 2011. The general transcriptional repressor Tup1 is required for dimorphism and virulence in a fungal plant pathogen. *PLoS Pathog.* 7: e1002235.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting sub-cellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 300:1005–1016.
- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30:1575–1584.
- Fernandez-Alvarez A, et al. 2012. Identification of O-mannosylated virulence factors in *Ustilago maydis*. *PLoS Pathog.* 8:e1002563.
- Finn RD, et al. 2008. The Pfam protein families database. *Nucleic Acids Res.* 36:D281–D288.
- Göker M, Voglmayr H, Riethmüller A, Oberwinkler F. 2007. How do obligate parasites evolve? A multi-gene phylogenetic analysis of downy mildews. *Fungal Genet Biol.* 44:105–122.
- Haas BJ, et al. 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31:5654–5666.
- Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9:R7.
- Haas BJ, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461: 393–398.
- Hemetsberger C, et al. 2015. The fungal core effector Pep1 is conserved across smuts of dicots and monocots. *New Phytol.* 206:1116–1126.
- Hemetsberger C, Herrberger C, Zechmann B, Hillmer M, Doehlemann G. 2012. The *Ustilago maydis* effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity. *PLoS Pathog.* 8:e1002684.
- Hoffman CS, Winston F. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 57:267–272.
- Inglis DO, et al. 2013. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiol.* 13:91.
- Jiang RH, et al. 2013. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. *PLoS Genet.* 9:e1003272.
- Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110:462–467.
- Kämper J, et al. 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444:97–101.
- Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Keller NP, Hohn TM. 1997. Metabolic pathway gene clusters in filamentous fungi. *Fungal Genet Biol.* 21:17–29.
- Keller NP, Turner G, Bennett JW. 2005. Fungal secondary metabolism—from biochemistry to genomics. *Nat Rev Microbiol.* 3:937–947.
- Kemen E, et al. 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol.* 9:e1001094.
- Kemen E, Jones JD. 2012. Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends Plant Sci.* 17:448–457.
- Khalidi N, et al. 2010. SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol.* 47:736–741.
- Kim D, et al. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36.
- Klein BS, Tebbets B. 2007. Dimorphism and virulence in fungi. *Curr Opin Microbiol.* 10:314–319.
- Koonin EV, et al. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5:R7.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 305:567–580.
- Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG. 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc Natl Acad Sci U S A.* 100:15670–15675.
- Lamour KH, et al. 2012. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Mol Plant Microbe Interact.* 25:1350–1360.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359.
- Laurie JD, et al. 2012. Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *Plant Cell* 24:1733–1745.
- Laurie JD, Linning R, Bakkeren G. 2008. Hallmarks of RNA silencing are found in the smut fungus *Ustilago hordei* but not in its close relative *Ustilago maydis*. *Curr Genet.* 53:49–58.
- Levesque CA, et al. 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* 11:R73.
- Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li L, Stoekert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Loytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320:1632–1635.

- Martinez-Espinoza AD, Garcia-Pedrajas MD, Gold SE. 2002. The Ustilaginales as plant pests and model systems. *Fungal Genet Biol.* 35:1–20.
- McDowell JM. 2011. Genomes of obligate plant pathogens reveal adaptations for obligate parasitism. *Proc Natl Acad Sci U S A.* 108:8921–8922.
- McLaughlin DJ, Hibbett DS, Lutzoni F, Spatafora JW, Vilgalys R. 2009. The search for the fungal tree of life. *Trends Microbiol.* 17:488–497.
- McTaggart AR, Shivas RG, Geering AD, Vanky K, Scharaschkin T. 2012. A review of the *Ustilago-Sporisorium-Macalpinomyces* complex. *Persoonia* 29:55–62.
- Mi H, et al. 2005. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 33:D284–D288.
- Mitchell AP. 1998. Dimorphism and virulence in *Candida albicans*. *Curr Opin Microbiol.* 1:687–692.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAA5: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185.
- Müller O, Schreier PH, Uhrig JF. 2008. Identification and characterization of secreted and pathogenesis-related proteins in *Ustilago maydis*. *Mol Genet Genomics.* 279:27–39.
- Nadal M, Garcia-Pedrajas MD, Gold SE. 2008. Dimorphism in fungal plant pathogens. *FEMS Microbiol Lett.* 284:127–134.
- Nepusz T, Sasidharan R, Paccanaro A. 2010. SCPS: a fast implementation of a spectral method for detecting protein families on a genome-wide scale. *BMC Bioinformatics* 11:120.
- O'Connell RJ, et al. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat Genet.* 44:1060–1065.
- Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8:785–786.
- Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.
- Que Y, et al. 2014. Genome sequencing of *Sporisorium scitamineum* provides insights into the pathogenic mechanisms of sugarcane smut. *BMC Genomics* 15:996.
- Quevillon E, et al. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33:W116–W120.
- Quinn L, et al. 2013. Genome-wide sequencing of *Phytophthora lateralis* reveals genetic variation among isolates from Lawson cypress (*Chamaecyparis lawsoniana*) in Northern Ireland. *FEMS Microbiol Lett.* 344:179–185.
- Sanchez-Martinez C, Perez-Martin J. 2001. Dimorphism in fungal pathogens: *Candida albicans* and *Ustilago maydis*—similar inputs, different outputs. *Curr Opin Microbiol.* 4:214–221.
- Schirawski J, et al. 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 330:1546–1548.
- Sharma R, et al. 2015. The genome of the basal agaricomycete *Xanthophyllomyces dendrorhous* provides insights into the organization of its acetyl-CoA derived pathways and the evolution of Agaricomycotina. *BMC Genomics* 16:233.
- Sharma R, Mishra B, Runge F, Thines M. 2014. Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the smut fungus *Melanopsichium pennsylvanicum*. *Genome Biol Evol.* 6:2034–2049.
- Sharma R, Thines M. 2015. FastQFS-A tool for evaluating and filtering paired-end sequencing data generated from high throughput sequencing. *Mycol Progress* 14:60.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stanke M, Schoffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62.
- Stergiopoulos I, Collemare J, Mehrabi R, De Wit PJ. 2013. Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. *FEMS Microbiol Rev.* 37:67–93.
- Suh SO, Sugiyama J. 1993. Phylogeny among the basidiomycetous yeasts inferred from small subunit ribosomal DNA sequence. *J Gen Microbiol.* 139:1595–1598.
- Tanaka S, et al. 2014. A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *Elife* 3:e01355.
- Thines M, Kamoun S. 2010. Oomycete-plant coevolution: recent advances and future prospects. *Curr Opin Plant Biol.* 13:427–433.
- Tian M, et al. 2007. A *Phytophthora infestans* cystatin-like protein targets a novel tomato papain-like apoplastic protease. *Plant Physiol.* 143:364–377.
- Trapnell C, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 28:511–515.
- Tyler BM, et al. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261–1266.
- Wahl R, Wipfel K, Goos S, Kamper J, Sauer N. 2010. A novel high-affinity sucrose transporter is required for virulence of the plant pathogen *Ustilago maydis*. *PLoS Biol.* 8:e1000303.
- Wallace AR. 1858. On the tendency of varieties to depart indefinitely from the original type. *J Proc Linn Soc Lond Zool.* 3:53–62.
- Waskiewicz A, et al. 2010. Formation of fumonisins and other secondary metabolites by *Fusarium oxysporum* and *F. proliferatum*: a comparative study. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess.* 27:608–615.
- Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859–1875.
- Xu J, et al. 2007. Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci U S A.* 104:18730–18735.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Yoshida K, et al. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* 2:e00731.
- Yu J, Bhatnagar D, Cleveland TE. 2004. Completed sequence of aflatoxin pathway gene cluster in *Aspergillus parasiticus*. *FEBS Lett.* 564:126–130.
- Yu J, Chang PK, et al. 2004. Clustered pathway genes in aflatoxin biosynthesis. *Appl Environ Microbiol.* 70:1253–1262.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22:2472–2479.

Associate editor: Sarah Schaack