

LARGE-SCALE BIOLOGY ARTICLE

The Transition from a Phytopathogenic Smut Ancestor to an Anamorphic Biocontrol Agent Deciphered by Comparative Whole-Genome Analysis ^{WJOPEN}

François Lefebvre,^a David L. Joly,^b Caroline Labbé,^a Beate Teichmann,^a Rob Linning,^b François Belzile,^a Guus Bakkeren,^b and Richard R. Bélanger^{a,1}

^aDépartement de Phytologie, Université Laval, Quebec G1V 0A6, Canada

^bAgriculture and Agri-Food Canada, Pacific Agri-Food Research Centre, Summerland V0H 1Z0, Canada

ORCID ID: 0000-0002-3646-024X (F.L.).

***Pseudozyma flocculosa* is related to the model plant pathogen *Ustilago maydis* yet is not a phytopathogen but rather a biocontrol agent of powdery mildews; this relationship makes it unique for the study of the evolution of plant pathogenicity factors. The *P. flocculosa* genome of ~23 Mb includes 6877 predicted protein coding genes. Genome features, including hallmarks of pathogenicity, are very similar in *P. flocculosa* and *U. maydis*, *Sporisorium reilianum*, and *Ustilago hordei*. Furthermore, *P. flocculosa*, a strict anamorph, revealed conserved and seemingly intact mating-type and meiosis loci typical of Ustilaginales. By contrast, we observed the loss of a specific subset of candidate secreted effector proteins reported to influence virulence in *U. maydis* as the singular divergence that could explain its nonpathogenic nature. These results suggest that *P. flocculosa* could have once been a virulent smut fungus that lost the specific effectors necessary for host compatibility. Interestingly, the biocontrol agent appears to have acquired genes encoding secreted proteins not found in the compared Ustilaginales, including necrosis-inducing-*Phytophthora*-protein- and Lysin-motif- containing proteins believed to have direct relevance to its lifestyle. The genome sequence should contribute to new insights into the subtle genetic differences that can lead to drastic changes in fungal pathogen lifestyles.**

INTRODUCTION

The Ustilaginomycetes, commonly known as smut fungi, comprise important plant pathogens that have been extensively studied for their pathogenic determinants. The smut fungi studied share common features essential for pathogenicity, such as mating, and the different mating types intrinsic to smut fungi have been reviewed in detail (Bakkeren et al., 2008). Indeed, as part of their sexual life cycle, haploid cells of compatible mating type must fuse and generate dikaryotic hyphae in order to infect and penetrate the host. Once inside the plant, these smut fungi establish a biotrophic interaction with their host in which infected cells remain alive throughout the disease cycle.

Over the years, the *Ustilago maydis* interaction with maize (*Zea mays*) has become the model pathosystem to investigate factors essential for the establishment of the biotrophic interaction typical of Ustilaginomycetes. In 2006, the generation and analysis of the genome sequence of *U. maydis* has provided unprecedented insight into novel genes that play key roles during pathogenesis (Kämper et al., 2006). Among these is a distinctive set of genes

that code for secreted proteins referred to as effector proteins (or effectors), of which many have unknown function. However, some have since been reported to be essential for infection and several seem to interfere with plant defense responses, thus facilitating infection by the pathogen (Brefort et al., 2009; Doehlemann et al., 2009, 2011).

First described in prokaryotes, effectors have now been found in all important groups of plant pathogens, including powdery mildews, rusts, and oomycetes. In the latter group, many effectors share a common conserved Arg-X-Leu-Arg (RXLR) motif that plays an important role in translocation of the effectors into the plant cell (Whisson et al., 2007). In the case of *U. maydis*, the secreted effectors lack such a motif, but many are arranged in clusters and are upregulated upon recognition of the host plant, invasion, and in tumor tissue. Through cluster deletion analysis, the importance in pathogenicity of these novel effectors was established, the first such demonstration for a eukaryotic pathogen (Kämper et al., 2006; Schirawski et al., 2010).

Comparison of the *U. maydis* genome to genomes of other sequenced basidiomycetes, including *Laccaria bicolor*, *Cryptococcus neoformans*, *Phanerochaete chrysosporium*, and *Malassezia globosa*, revealed great variability in terms of genome size, synteny, structure, amount of repetitive elements, and number and homology of effector proteins. These observations reinforced the relevance of using phylogenetically related species to improve our understanding into the evolution of pathogenic factors through comparative genomic approaches. Additionally, the recent genome annotation of *Sporisorium reilianum* (Schirawski et al., 2010)

¹ Address correspondence to richard.belanger@fsaa.ulaval.ca.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Richard R. Bélanger (richard.belanger@fsaa.ulaval.ca).

^{WJ} Online version contains Web-only data.

^{OPEN} Articles may be viewed online without a subscription.

www.plantcell.org/cgi/doi/10.1105/tpc.113.113969

and *Ustilago hordei* (Laurie et al., 2012) has provided an opportunity to perform comparative analyses that highlight common and divergent features among these closely related pathogens. Of particular interest was the finding that all three fungi share many conserved effectors but also possess sets of species-specific effectors. These results suggest that common effectors are essential for basic plant pathogenicity, while species-specific effectors may be responsible for divergent interaction strategies, targeting different host tissues and molecules and thereby dictating host specificity and pathogen development.

In earlier classification, all smuts were ecologically characterized by their ability to infect plants and shared a similar life cycle with a yeast-like haploid phase and a parasitic dikaryophase, culminating in the production of numerous powdery black teliospores, hence, their common name (Begerow et al., 2006). However, a number of anamorphic fungi lacking sexual development, initially placed in deuteromycetous taxa, were found to be morphologically and phylogenetically related to the Ustilaginales. In order to integrate these anamorphs into the general phylogenetic system of Ustilaginomycetes, Begerow et al. (2000) analyzed and compared diagnostic ribosomal DNA sequences of teleomorphic and anamorphic species of Ustilaginomycetes. Their analyses confirmed that species of *Pseudozyma* and Ustilaginales parasitizing grasses form a monophyletic group. *Pseudozyma* species thus represent the sole known members of the Ustilaginales that cannot parasitize plants. On the basis of these findings, it appears that *Pseudozyma* spp are either anamorphic vestiges or descendants of plant pathogenic smut species that have lost their phytoparasitic and/or sexual capacity. It was recently proposed that comparative genomics, in the context of understanding the evolution of biotrophic pathogens, should target phylogenetically relevant organisms that are not pathogens (Kemen and Jones, 2012). As such, *Pseudozyma* spp constitute a rare and unique model to investigate the determinants that differentiate them from pathogenic Ustilaginales.

Pseudozyma spp have received little attention in the literature. One notable exception is *Pseudozyma flocculosa* (Traquair, Shaw, and Jarvis) Boekhout and Traquair, a species that has been studied for its ability to antagonize powdery mildews (Figure 1). *P. flocculosa* was discovered in 1987 and originally identified as *Sporothrix flocculosa*, an ascomycetous yeast (Traquair et al., 1988). Because of its potential as a biofungicide, it was extensively characterized for its properties, mode of action, secondary metabolites, and environmental fate, before being registered by the Environmental Protection Agency and Pest Management Regulatory Agency in the US and Canada, respectively (Bélangier et al., 2012). *P. flocculosa* has no known sexual stage and is considered a leaf epiphyte although the sole report of its isolation specified it was found in association with clover powdery mildew. This particularity may explain its rarity in nature and the scarcity of known isolates, with only six reported but most of them being isolated from the same strain following biocontrol trials (Avis et al., 2001). However, its reclassification among Ustilaginomycetes in 2000 (Begerow et al., 2000) launched unexpected opportunities to better characterize *P. flocculosa*. At first, the notion that a biocontrol agent was related to smut fungi, including the well-studied model species *U. maydis*, appeared biologically incongruous. Nevertheless, many observations have since confirmed its link

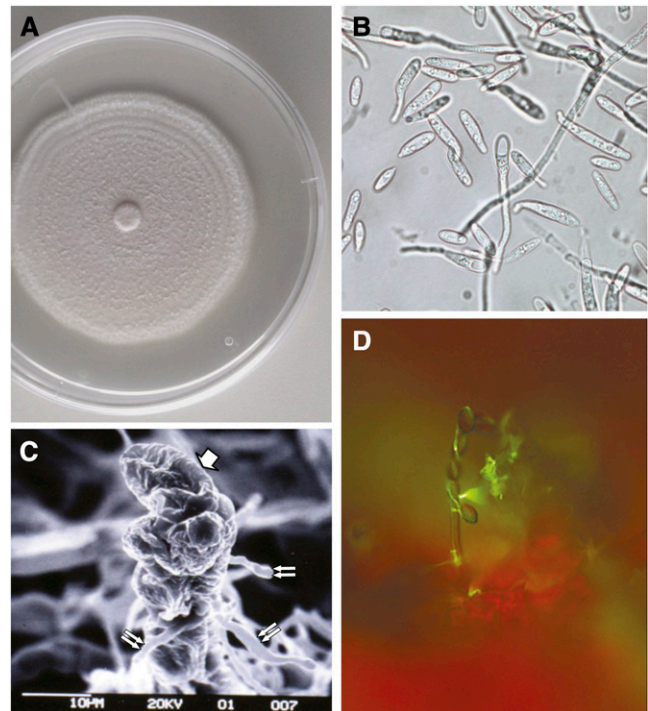


Figure 1. Different Life Forms of the Biocontrol Agent *P. flocculosa*.

- (A) *P. flocculosa* growing on potato dextrose agar medium.
- (B) Sporidia in 24-h-old culture of *P. flocculosa*.
- (C) Scanning electron micrograph of collapsed cucumber powdery mildew conidia (*Podosphaera xanthii*; wide arrow) antagonized by *P. flocculosa* hyphae (double arrow).
- (D) Green fluorescent protein-transformed *P. flocculosa* coiling around conidia of cucumber powdery mildew.

to *U. maydis*. Cheng et al. (2001) were the first to bring forth biological evidence by observing that the fungus was only amenable to genetic transformation when using promoters from *U. maydis* to drive expression of a selectable marker. This was followed by the discovery of an unusual glycolipid produced by *P. flocculosa* that was nearly identical to ustilagic acid, a rare metabolite identified in culture filtrates of *U. maydis* in 1951 (Lemieux et al., 1951; Mimee et al., 2005). Taking advantage of the annotated genome of *U. maydis* and based on specific chemical characteristics of the molecule, Böcker's group was able to find a cluster of 10 genes responsible for the synthesis of ustilagic acid in *U. maydis* (Teichmann et al., 2007). As expected, a similar cluster, containing homologous genes and responsible for the synthesis of the glycolipid, was recently reported in *P. flocculosa* (Teichmann et al., 2011b).

Here, we report the generation and analysis, including annotation, of the complete genome of *P. flocculosa*. Comparison to the genomes of phytopathogenic Ustilaginales (*U. maydis*, *S. reilianum*, and *U. hordei*) highlighted similarities and differences. Particular emphasis was placed on reported hallmarks of pathogenicity and biocontrol activity, such as mating factors, cell wall-degrading enzymes (CWDEs), secondary metabolites, and effector proteins in an effort to identify the conserved and/or lost elements that account for the divergent lifestyle of *P. flocculosa*.

RESULTS

Genome Statistics

The genome of *P. flocculosa* was sequenced using the Roche 454 sequencing platform. Three sequencing runs yielded 522 Mb of whole-genome sequencing data and 240 Mb of paired-end data to generate an average genome coverage of 28 X (Table 1). The assembly yielded 1583 contigs from which 1187 were oriented and ordered into 37 scaffolds to which three contigs larger than 2 kb were added. The N50 value shows that 50% of the bases that compose the assembly are contained in scaffolds larger than 920 kb, the longest scaffold being 1.975 Mb long.

Of the 40 genome scaffolds, seven are considered to represent complete chromosomes since telomeric motifs (TTAGGG) are present at both ends and 17 present such a motif at one end. The total assembly of the *P. flocculosa* genome is slightly bigger in size than that of the available phytopathogenic Ustilaginales, but the number of protein coding genes is remarkably similar (Table 1). Accordingly, a lower proportion of coding sequences is found in the genome, correlating with lower gene density. Regarding

structural annotation, more than 80% of the gene calls are supported by homology to known proteins and manually curated transcriptomic data, while 8.2 and 6.8% are supported by homology to known proteins or transcriptomic data alone. Less than 3% of the genes are based solely on ab initio predictions.

The most important differences between the genome structures of *P. flocculosa* and the three Ustilaginales were found in the proportion of guanine and cytosine (GC) residues and in the structure of genes. First, *P. flocculosa* has a substantially higher GC content than the three phytopathogenic Ustilaginales spp, and this proportion exceeded 66% in the coding regions (Table 1). All codon positions are richer in GC with the highest proportion reaching 82.3% at the third position. Second, genes identified in *P. flocculosa* contained an average of 4 times more introns than *U. maydis* (Table 1). Consequently, *P. flocculosa* contains roughly three times fewer genes without any intron, though the size of the introns is well conserved across all four species. Rapid analysis of intron position within genes showed that introns of all three phytopathogenic Ustilaginales are depleted toward the 3' end of genes, while introns of *P. flocculosa* are more equally distributed across genes (see Supplemental Figures 1 and 2 online).

Table 1. Comparative Genome Statistics of *P. flocculosa*, *U. maydis*, *U. hordei*, and *S. reilianum*

	<i>P. flocculosa</i>	<i>U. maydis</i>	<i>U. hordei</i>	<i>S. reilianum</i>
Assembly Statistics				
Total contig length (Mb)	23.2	19.7	20.7	18.2
Total scaffold length (Mb)	23.4	19.7	21.2	18.5
Average base coverage	28x	10x	20x	20x
N50 contig (kb)	38.6	127.4	48.7	50.3
N50 scaffold (kb)	919.9	885.1	307.7	772.4
GC content (%)	65.1	54.0	52.0	59.7
Assembled	65.2	54.0	52.2	59.7
Coding	66.3	56.3	54.3	62.6
Noncoding	63.7	50.5	49.0	54.3
Coding sequences				
Percentage coding (%)	54.3	61.1	57.3	64.7
Average gene size (bp)	2,097	1,836	1,782	1,858
Average gene density (genes/Mb)	291.5	344.5	335.4	360.7
Protein coding genes	6,877	6,787	7,111	6,673
Homology-supported gene calls ^a	6,239 (90.5%)	6,475 (95.4%)	6,443 (90.6%)	6,450 (96.7%)
RNA-seq-supported gene calls ^b	6,142 (89.3%)	n.a. ^c	n.a	n.a
RNA-seq- and homology-supported predictions ^{a,b}	5,673 (82.3%)	n.a.	n.a	n.a
Ab initio only predictions	183 (2.7%)	312 (4.6%)	668 (9.4%)	223 (3.3%)
Exons	19,318	9,778	10,995	9,777
Average exon size	658	1,230	1,103	1,222
Exons/gene	2.8	1.44	1.55	1.47
tRNA genes	176	111	110	96
Noncoding sequences				
Introns	12,427	2,991	3,884	3,104
Introns/gene	1.8	0.44	0.55	0.47
Intron-free genes	1,565 (22.7%)	4,910 (72.3%)	4,821 (67.8%)	4,732 (70.9%)
Average intron length (base)	141	146	141	146
Average intergenic length (bp)	1,273	1,061	1,098	919
Transposable elements and repeats (%)	3.4	2.0	10.0	0.8

^aHomology-supported gene calls are based on BLASTP similarity search to the nonredundant GenBank database (cutoffs: e-value 1e-05, bits 80).

^bRNA-sequencing-supported gene calls are based on tBLASTN similarity search to database of retrieved transcript sequences after manual curation (cutoffs: e-value 1e-30, alignment on >70% of protein).

^cn.a., not available.

The number of transposable elements and simple repeats found in the *P. flocculosa* genome is similar to that reported in *U. maydis* (Table 1; see Supplemental Tables 1 and 2 online). It is 3 times lower than in *U. hordei*, in which species-specific transposable element families seem to have expanded over time (Laurie et al., 2012). Repetitive elements do not cluster in specific areas of the genome but appear to be randomly scattered (data not shown). Some of them bear signatures specific to some of the elements in the other Ustilaginales, such as the long terminal repeat (LTR) retrotransposon and LINE-type elements. Repeat-induced point-like mutation signatures as described for *U. hordei* could not be found in *P. flocculosa*, a situation similar to the one reported in *U. maydis* and *S. reilianum* (see Supplemental Figure 3 online).

Genome Synteny

Genome synteny evaluated by genome-to-genome alignments is an efficient way to assess the level of conservation in gene order and content. It allows the visualization of translocation, inversion, and gene loss events that occurred during evolution from a common ancestor. In the case of *P. flocculosa*, its level of genome synteny with *U. maydis* is much lower than that observed between *U. maydis* and *S. reilianum* (Figure 2; Schirawski et al., 2010). However, many smaller *P. flocculosa* genome regions do share a common content and gene orientation with genome regions in *U. maydis*. Figure 2 is visually similar to mesosynteny described by Hane et al. (2011). These observations are further supported by OrthoCluster analysis (see Supplemental Table 3 online). The lowest levels of synteny conservation are obtained when comparing any of the three phytopathogenic species with *P. flocculosa*.

Gene Families

In order to group genes with similar functions regardless of whether that function is documented or not, we used BLASTP (Altschul et al., 1990) to perform an all-against-all similarity search with the 27,448 genes from the four organisms and TribeMCL (Enright et al., 2002) to generate clusters of genes based on their similarity. All genes were grouped into 5841 clusters, further referred to as gene families.

A total of 602 genes (2.5%) were grouped in species-specific families, while the bulk of genes (24,105, and 87.8%) were grouped in core families present in all four organisms (Figure 3A; see Supplemental Data Set 1 online). Nearly 91% (6238) of *P. flocculosa* genes are part of the core gene families, while only 4.0% (275) are specific to this biocontrol agent, the vast majority (>90%) lacking known functions. Based on gene families, *P. flocculosa* has the least diversified genome among the four species, with 5288 gene families (Figure 3A).

The data generated in this exhaustive similarity search were used to derive with a high level of confidence and using 306 core gene families, the phylogenetic relationship between the four species and *M. globosa* (Figure 4). Based on the consensus tree, *P. flocculosa* is indeed more closely related to Ustilaginales than *Malassezia* but does branch outside of the three pathogens.

Mating, Dimorphic Switch, and Meiosis

P. flocculosa is not known to be a plant pathogen, and no sexual stage has ever been discovered; the biocontrol activity toward mildews is achieved by the haploid form (Bélanger and Deacon, 1996). Therefore, a particular emphasis was placed on examining the presence or absence in *P. flocculosa* of genes typical of pathogenic activity in Ustilaginales. Since successful mating between haploid yeast-like cells is essential to trigger a switch to dikaryotic hyphal growth and pathogenicity in the Ustilaginales, genes involved in these processes are of great importance. Complete mating-type genes constituting the Ustilaginales *a* and *b* loci were localized in the *P. flocculosa* genome assembly, on scaffolds 4 and 10, respectively. The *b* locus was similarly organized in two divergently transcribed, homeodomain-containing heterodimeric transcription factor open reading frames, *bE1* and *bW1*. The *a* locus housed a pheromone receptor gene, *pra1*, and two mating pheromone genes, *mfa1.1* and *mfa1.2*, arranged in an orientation and spacing similar to the *S. reilianum* *a3* locus (Schirawski et al., 2005) (see Supplemental Table 4 online). A phylogenetic analysis based on amino acid sequences of Pra (STE3 homolog) revealed a similar relationship between *P. flocculosa* and other Ustilaginales species (Figure 5) as derived from the previous comparison of the 306 core gene families (Figure 4). Despite a close phylogenetic relationship, there was no indication *P. flocculosa* was responsive to mating interactions with various *U. maydis* strains and mating types (see Supplemental Figure 4 online). Second, all genes coding for the proteins involved in the mitogen-activated protein kinase and cAMP signaling cascades downstream of the G protein-coupled pheromone receptor and leading to crucial morphological and physiological modifications in pathogenic Ustilaginales were present and mostly highly conserved in *P. flocculosa* (see Supplemental Data Set 2 online). In *U. maydis*, meiosis occurs after the mating of compatible haploid mating types and in planta development of diploid teliospores (Saville et al., 2012). All genes reported to be involved in meiosis and sporulation were found in the genome of *P. flocculosa* (see Supplemental Data Set 3 online).

Components necessary for transcriptional gene silencing and chromatin remodeling were also found in *P. flocculosa* (see Supplemental Table 4 online). These genes are missing from the *U. maydis* genome but are present in *U. hordei* and *S. reilianum* (Laurie et al., 2012).

Cell Wall-Degrading and Carbohydrate-Active Enzymes

CWDEs play important roles in plant pathogenicity (Walton, 1994) and biocontrol activity (Vinale et al., 2008). Analysis of CWDEs in the *P. flocculosa*, *U. maydis*, *U. hordei*, and *S. reilianum* genomes revealed that all four organisms possess fairly similar sets of genes (Figures 3C and 3D).

However, when comparing individual carbohydrate-active enzyme (CAZy) families, *P. flocculosa* revealed interesting differences with respect to possible functions in fungal cell wall degradation, specifically dealing with chitosanase and chitinase activities (see Supplemental Data Set 4 online). For example, the glycoside hydrolase family 18 (GH18) associated with fungal cell wall degradation (chitinase) clearly differentiates *P. flocculosa* from the phytopathogenic Ustilaginales. First, *P. flocculosa* possesses seven

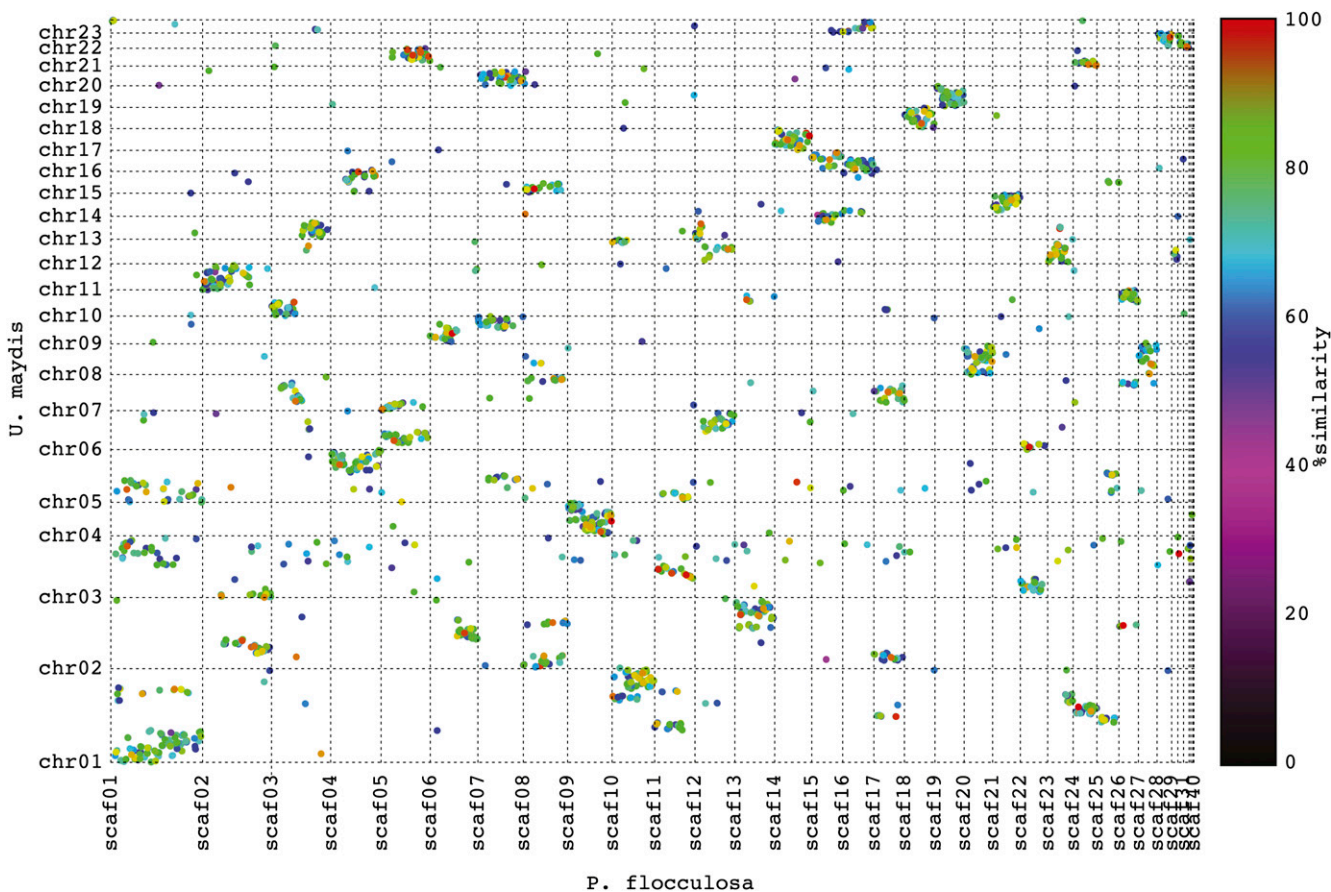


Figure 2. Alignment of *P. flocculosa* and *U. maydis* Genome Assemblies.

Regions of the genomes were compared on the basis of their amino acid sequences generated from six-frame translations. The graph was plotted using the promoter program from MUMmer 3.0 with the following options: -mum (promer) and -color (mummerplot). Colors refer to the percent similarity of each alignment as shown by the color chart on the right.

genes coding for members of the GH18 family compared with only four in the phytopathogenic species. Second, in *P. flocculosa* only, two of those GH18 chitinase-encoding genes are each positioned right next to a gene coding for a protein bearing five (pf01239) or seven (pf06228) lysin motifs (LysM, CAZy CBM50), known for their affinity for chitin, and to a gene encoding a protein harboring a calcium binding EF-hand motif. Both of the latter genes also encode secreted proteins and neither have homologs in the three smuts.

Regarding plant CWDEs, no apparent differences could be found between gene sets. However, a motif that was associated with pectin degradation by Amselem et al. (2011) led to the identification of two *P. flocculosa*-specific putative lipases with a consensus Gly-Asp-Ser-Leu (GDSL) motif and a secretion signal (pf01013 and pf03087). Both have a carbohydrate esterase family 16 motif, but, surprisingly, they share a very low level of similarity with one another and are not homologous to any protein from any of the three plant pathogenic species.

Biosynthesis of Secondary Metabolites

Secreted secondary metabolites also play an important role in pathogenicity and biocontrol. Using the JCVI Secondary Metabolite

Unique Regions Finder server (SMURF) (Khaldi et al., 2010), we identified pivotal genes associated with the synthesis of secondary metabolites, also referred to as backbone genes, and predicted their presence in a broader cluster of genes involved in biosynthesis.

P. flocculosa, *U. hordei*, and *S. reilianum* species had a nearly identical number of SMURF clusters and backbone genes, while *U. maydis* had approximately 50% more in each category (Table 2). All backbone genes have at least one homolog in all other species except for two *U. maydis* dimethylallyltryptophan synthases (um05796 and um06508). Based on homology between their gene content, 13 SMURF-predicted clusters could be grouped into six groups, two of which contain a predicted cluster in *P. flocculosa* (see Supplemental Table 5 online). Not surprisingly, one group linked the flocculosin gene cluster with the ustilagic acid gene cluster.

Secretome

Based on the method described by Mueller et al. (2008), 547 secreted proteins were identified in *P. flocculosa*. Of this number, 345 could not be assigned an enzymatic function and were therefore considered as candidate secreted effector proteins (CSEPs). In

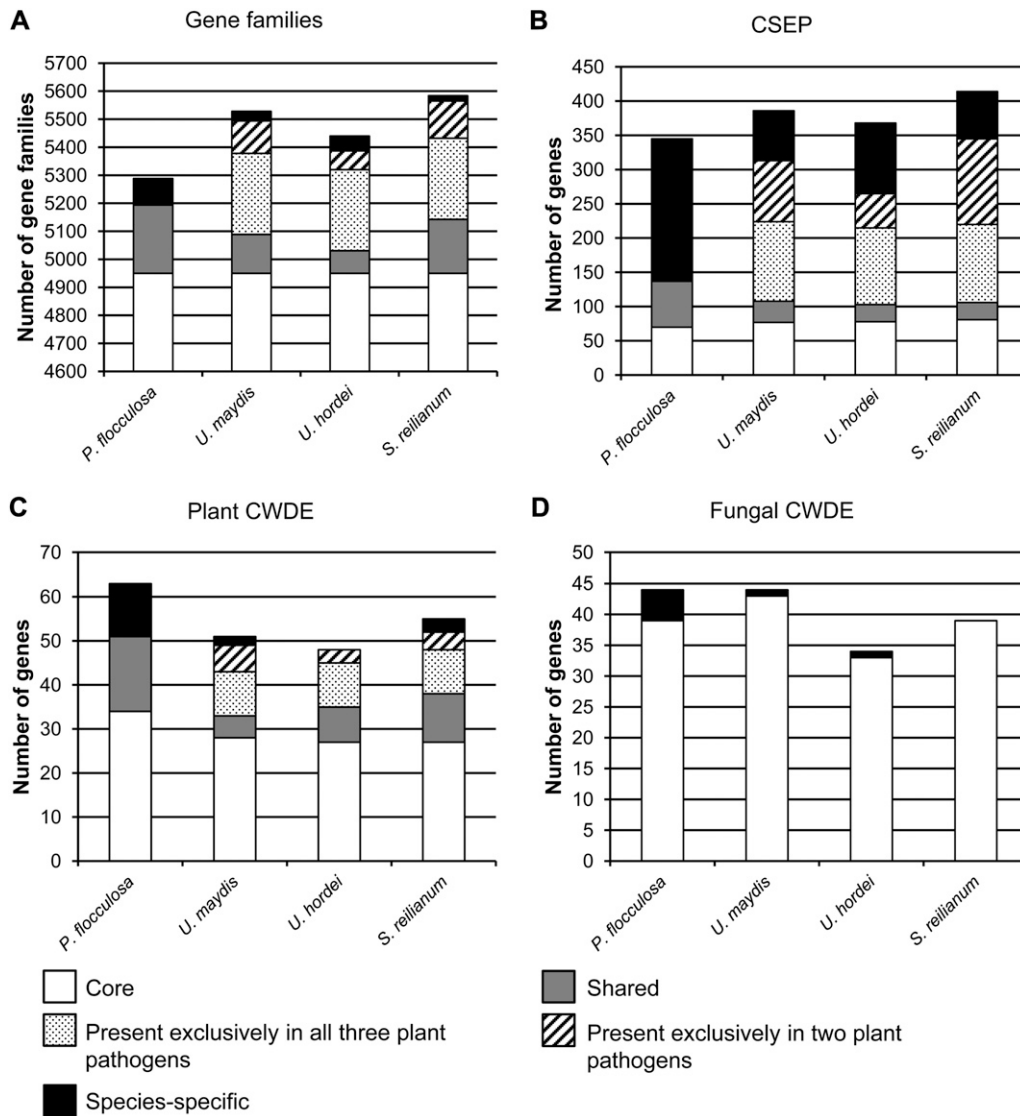


Figure 3. Composition of Four Gene Categories Based on Gene Families in Four Ustilaginomycete Fungi.

Four types of gene families were defined: “Core” refers to gene families composed of genes present in all four species. “Shared” represents gene families composed of genes found in two or three species including *P. flocculosa*. The two other types of gene families refer to genes families “present exclusively in two plant pathogens” or “present exclusively in all three plant pathogens.” “Species-specific” refers to gene families composed of genes found in only one species.

(A) Graph of gene families reports frequencies of all gene families present in species.

(B) to (D) Otherwise, genes were classified according to the category (CSEP, plant CWDEs, and fungal CWDE), the species, and the type of gene family to which they belong. For specific gene product categories, types of gene families were defined exclusively on the basis of genes belonging to the targeted category and being present in the family.

comparison to *U. maydis*, the genome of *P. flocculosa* has nearly the same number of predicted secreted proteins overall but ~10% less identified CSEPs (554 secreted proteins and 386 CSEPs for *U. maydis* according to Mueller et al., 2008) (see Supplemental Data Set 5 online). Clustering is also similar in *P. flocculosa* with 112 (20%) secreted proteins being clustered into 21 clusters versus 137 and 139 (28%) into 22 clusters for *U. maydis* and *S. reilianum*, respectively. In comparison, only seven clusters grouping 30 genes (7%) could be found in *U. hordei*.

When focusing on potential effectors, 60% (209 genes) of the identified *P. flocculosa* CSEPs are distributed over its 138 species-specific gene families (Figure 3B). This proportion is just 21% in the plant pathogens, on average. The closer relatedness among the CSEP sets in the plant pathogens is also reflected in the number of gene families they share. On average, 75% of gene families containing CSEPs are shared in the plant pathogenic species, whereas this number is only 43% between *P. flocculosa* and the other plant pathogens. Surprisingly, a NPP1

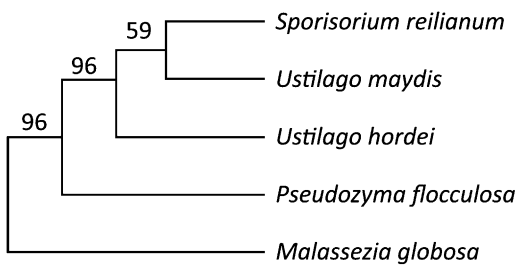


Figure 4. Molecular Phylogeny of Five Sequenced Ustilaginomycota.

Consensus phylogenetic tree constructed from 306,000 trees (306 gene families \times 1000 bootstrap replicates) estimated by the maximum likelihood method under the constraint of a molecular clock and the JTT matrix-based model (Jones et al., 1992). Each tree was estimated from one bootstrap replicate generated from the alignment of amino acid sequences from a gene family containing only one member per species. Families had to group genes encoding an ortholog involved in Metabolism or Cellular Process as defined by KEGG (Kanehisa and Goto, 2000). Trees were rooted using the midpoint criterion. Branch lengths are unscaled.

(for necrosis-inducing *Phytophthora* protein) domain was identified in two *P. flocculosa* CSEPs (pf04735 and pf04834), a feature not found in the smuts.

Orthology to *U. maydis* Genes Proven to Affect Virulence

From the Pathogen-Host Interaction database (PHI-base) (Winnenburg et al., 2006), we extracted 97 genes that had been biologically characterized as playing a role in the ability of *U. maydis* to infect maize or affect the level of virulence after infection. Using this *U. maydis* set, we then verified the presence of orthologs in the other three organisms and grouped them into four categories based on the type of response they provoked in the *U. maydis*-maize interaction (loss of pathogenicity versus increased or reduced virulence) and on whether they are secreted (secreted versus nonsecreted) (Figure 6; see Supplemental Data Set 6 online). Roughly half of the genes analyzed in this manner are conserved and represent equal numbers in each of three categories. Conserved genes are, for the most part, genes coding for nonsecreted components of signal transduction pathways, chitin synthases, and proteins with nucleic acid binding motifs. For example, components of the cAMP and mitogen-activated protein kinase signaling cascades downstream of the pheromone receptor are all conserved. Of the four gene categories created, the one represented by secreted proteins influencing the level of virulence clearly showed the highest level of variation among the species (Figure 6). Most notably, *P. flocculosa* contained only two genes sharing homology with the 51 found in *U. maydis*, a striking difference considering the high conservation of genes in all other categories studied.

DISCUSSION

Comparative genomics has been extensively exploited to unravel specific virulence determinants and their evolution among the phytopathogenic Ustilaginomycetes *U. maydis*, *S. reilianum*, and

U. hordei (Kämper et al., 2006; Schirawski et al., 2010; Laurie et al., 2012). In this work, the genome release of the closely related fungus *P. flocculosa* constitutes a new step in the study of Ustilaginomycete fungi evolution and phytopathogenicity-associated traits. In addition, because *P. flocculosa* cannot parasitize plants but is a powerful antagonist of powdery mildews (Jarvis et al., 1989; Hajlaoui and Bélanger, 1993; Clément-Mathieu et al., 2008), it represents a unique model to assess the presence and role of virulence factors inherent to pathogenic activity against both plants and fungal pathogens.

In order to generate high-confidence gene structures for a maximum of protein-coding genes, RNA sequencing was performed based on four previously defined growth phases of *P. flocculosa* (Hammami et al., 2008). The resulting contigs were aligned against the *P. flocculosa* genome along with homologous predicted protein sequences from the closely related Ustilaginales species already sequenced. This determined, with accuracy, gene models for more than 90% of the predicted genes composing the *P. flocculosa* genome.

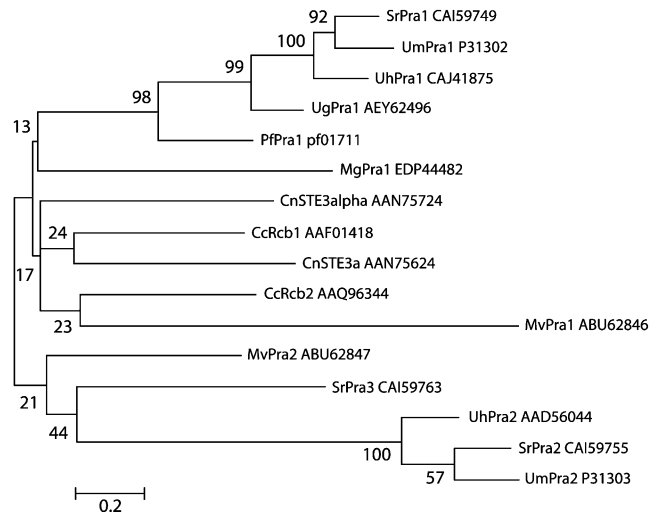


Figure 5. Molecular Phylogeny of Basidiomycete Pheromone Receptor Protein Sequences.

Maximum likelihood method based on the JTT matrix-based model (Jones et al., 1992) and evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011). Pheromone receptor sequences were C-terminally truncated to exclude the cytoplasmic tail and to optimize the alignment and were from the following species (protein name, accession number, number of amino acids): *Coprinopsis cinerea* (CcRcb1, GenBank AAF01418, 293) and (CcRcb2, GenBank AAQ96344, 298); *Cryptococcus neoformans* (CnSte3a, GenBank AAN75624, 295) and (CnSte3alpha, GenBank AAN75724, 295); *Malassezia globosa* (MgPra1, GenBank EDP44482, 298); *Microbotryum violaceum* (MvPra1, GenBank ABU62846, 300) and (MvPra2, GenBank ABU62847, 299); *Pseudozyma flocculosa* (PfPra1, pf01711, 279); *Sporisorium reilianum* (SrPra1, GenBank CAI59749, 296), (SrPra2, GenBank CAI59755, 299), and (SrPra3, GenBank CAI59763, 303); *Ustanciosporium gigantosporum* (UgPra1, GenBank AEY62496, 279); *Ustilago hordei* (UhPra1, GenBank CAJ41875, 296) and (UhPra2, GenBank AAD56044, 300); *U. maydis* (UmPra1, GenBank P31302, 296) and (UmPra2, GenBank P31303, 298). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed.

Table 2. Secondary Metabolite Backbone Genes Found and Clusters Predicted in Four Ustilaginomycete Fungi

Species	PKS ^a	PKS-Like	NRPS ^b	NRPS-Like	DMAT ^c	Total	No. of Predicted Clusters
<i>P. flocculosa</i>	4	1	1	4	0	10	8
<i>U. maydis</i>	3	2	3	6	2	16	12
<i>U. hordei</i>	2	1	3	4	0	10	7
<i>S. reilianum</i>	2	1	2	4	0	11	9

^aPolyketide synthase.

^bNonribosomal peptide synthetase.

^cDimethylallyl Trp synthase.

Although *P. flocculosa* has never been reported to infect plants, its genome sports several conserved features similar to those in the genomes of the three phytopathogenic Ustilaginales. The genome was comparable in terms of size and gene number, and many regions revealed mesosynteny to the *U. maydis* genome. Also, the repeat analysis showed a low proportion of TEs and repetitive sequences similar to the fairly streamlined genomes of *U. maydis* and *S. reilianum*. Of the four, *U. hordei* seems to stand out with over 10% of TEs and repeats. We did not find evidence of a repeat-induced point mutation mechanism as in *U. hordei*, but, as in *S. reilianum*, too few available homologous repeats made a comparative analysis difficult to assess mutation rates (Laurie et al., 2012). In general, repeat content in genomes of fungi seems to vary widely and is much higher in the biotrophic rusts and mildews (Spanu, 2012). Even though TEs are now seen as genome modifiers creating variation in gene structure and expression, streamlined genomes having few TEs and repeats may be a result of combinations of mating strategies, sex, and recombination potential (Laurie et al., 2013). In contrast with *U. maydis*, *P. flocculosa* does not have a high rate of homologous recombination (Teichmann et al., 2011a), a feature it shares with *U. hordei*.

A striking feature unique to *P. flocculosa* is the high number of introns per gene. It is unclear if this is the result of intron gain in the case of *P. flocculosa* or intron loss in phytopathogenic Ustilaginales. Interestingly, the three Ustilaginales species revealed a biased intron distribution toward the 5' end of genes, a characteristic common to intron-poor species, possibly associated with a mechanism of reverse transcriptase-mediated intron loss (RTMIL) (Mourier and Jeffares, 2003). Alternatively, the 5' bias observed in intron-poor species could be linked to selection pressure favoring the conservation of the 5'-proximal intron having a functional role (Nielsen et al., 2004; Jeffares et al., 2006) even though different studies showed lineage-specific variability in the pattern of intron gain and loss and contradicted the thesis of a simple RTMIL mechanism (Sharpton et al., 2008). However, considering that evolution of eukaryotes is characterized more by intron loss than gain, Roy and Gilbert (2006) suggested that if RTMIL occurred in some species, then the rate of paralogous recombination will directly influence the rate of intron loss. If this holds true for Ustilaginomycetes, since recombination events are favored during meiosis in fungi, it would explain why species with a sexual cycle possess fewer introns than others. The possible lack of a sexual cycle in *P. flocculosa* could also explain the state of mesosynteny observed between its genome and that of *U. maydis*. As reported for *Magnaporthe grisea*, a fungus

that exclusively propagates asexually in nature, genome rearrangements are suspected to be maintained as long as vegetative fitness is conserved (Dean et al., 2005).

Phylogenetic reconstruction based on more than 300 core gene families supports the relatively close proximity of *P. flocculosa* to phytopathogenic Ustilaginales. However, its exclusion from a common branch with the three smut fungi may indeed suggest that it has adopted a divergent lifestyle after the speciation event from a common ancestor. When classified by Begerow et al. (2006), *P. flocculosa* was described as an anamorph lacking a sexual stage. One hypothesis and possible differentiating factor with pathogenic Ustilaginales is that loss of mating-type genes accounts for the inability of *P. flocculosa* to infect plants. However, mating-type genes, constituting the pheromone-receptor module and the homeodomain-containing heterodimeric transcription factors, are present in its genome. Although we currently cannot confirm functionality based on compatibility assays

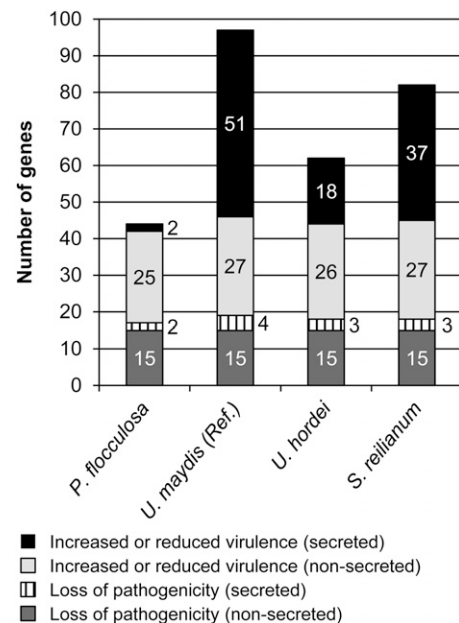


Figure 6. Conservation of *U. maydis* Genes Proven to Affect Virulence.

For each of the reported 97 *U. maydis* genes, presence of an ortholog in *P. flocculosa*, *U. hordei*, or *S. reilianum* was assessed. Genes were grouped according to the type of response observed and the secretion status of the coded protein.

with *U. maydis* (see Supplemental Figure 4 online), initial RNA sequencing analysis revealed that the receptor and both pheromone genes are transcribed (data not shown). The presence in the *P. flocculosa* genome of all other genes shown to be necessary for the initiation of mating, maintenance of the dikaryotic pathogenic state, and meiosis in *U. maydis* suggests that *P. flocculosa* is either capable of mating and sex under certain circumstances or has recently lost this capability. One would expect that selection pressure would alter the presence or function of mating genes in strict anamorphs, but recent genome sequence analyses have shown that fungal species reproducing asexually in nature possessed all genes involved in sexual reproduction and were even capable of mating under in vitro conditions (Xu et al., 2007; O'Gorman et al., 2009; Böhm et al., 2013). Interestingly, *P. flocculosa* does not require mating to unleash its pathogenic potential onto mildew fungi, although this mechanism of epiphytic competition is likely different from the host infection requirement of the smuts.

The *P. flocculosa* *a* and *b* loci were localized in the current genome assembly to scaffolds 4 and 10, respectively. Scaffold 4 (1.1 Mb) reveals patchy synteny with interrupted stretches to 600-kb regions in *U. maydis* chromosome 5 and *S. reilianum* chromosome 20, and to a much shorter region in *U. hordei* scaffold 5.00019, all of which house the homologous *pra* and *mfa* genes (Laurie et al., 2012). Incidentally, the *P. flocculosa pra* and two *mfa* genes are arranged similarly as at the *S. reilianum a3* locus (Schirawski et al., 2005). The *P. flocculosa b* locus localizes to scaffold 10, which, over a length of ~920 kb, matches homologs present in a 830-kb region of chromosome 1 in both *U. maydis* and *S. reilianum*; only a very short 9-kb stretch containing the *bE* and *bW* genes matches *U. hordei* scaffold 5.00019. Interestingly, genes in between the homologs on chromosome 1 locate to various other chromosomes but which are often the same in both *U. maydis* and *S. reilianum*. Overall, this indicates an arrangement closer to the tetrapolar mating-type organization of the *U. maydis*/*S. reilianum* lineage. The presence of two mating pheromone genes, also found at the *S. reilianum a3* locus, could indicate that this arrangement with multiple pheromone genes is ancestral; the structure of the mating-type gene-containing chromosomes in *S. reilianum* also pointed to this fungus being closer to an ancestral lineage (Laurie et al., 2012).

In addition to the presence of pivotal genes known to trigger the drastic morphological and physiological changes that enables the smut fungi to infect their host, *P. flocculosa* has also conserved all other typical plant pathogenicity traits found in smuts. This finding indicates that *P. flocculosa* likely has evolved from a plant pathogenic ancestor common to all four species.

In previous studies of fungal genomes, the presence and number of CAZy have been associated with the specific lifestyle of the fungus under study. For instance, the strict biotrophic pathogen *Blumeria graminis* f. sp. *hordei* contains very few CAZy capable of plant cell wall degradation (Spanu et al., 2010). Other biotrophs, such as *Puccinia graminis* f. sp. *tritici* (Duplessis et al., 2011) and *U. maydis* (Kämper et al., 2006), are also characterized by reduced enzyme systems capable of degrading plant cell walls but do possess predicted cellulases and xylanases. Analysis of CAZy in *P. flocculosa* revealed a similar set of CAZy, including plant CWDEs as found in plant pathogenic Ustilaginales. It is

unclear why it would have conserved these enzymes given that they are not deemed necessary for its epiphytic lifestyle on the leaf surface (Bélanger and Avis, 2002) nor for its parasitic activity on powdery mildews.

Ever since the establishment of the phylogenetic link between *P. flocculosa* and *U. maydis*, their shared property for production of secondary metabolites, namely, unusual glycolipids, has been an intriguing biological feature. It was established that both organisms contain an elaborate cluster of 10 genes regulating the synthesis of highly similar glycolipids termed flocculosin (*P. flocculosa*) and ustilagic acid (*U. maydis*) (Teichmann et al., 2007, 2011a, 2011b). Owing to the antifungal property of flocculosin, it was originally hypothesized that it played a determinant role in the biocontrol activity of *P. flocculosa*. This hypothesis came under scrutiny when *U. maydis* was tested and found to be unable to antagonize powdery mildews under specific assay conditions in spite of producing antifungal ustilagic acids and when genes involved in flocculosin synthesis were observed to be repressed in spores that came into contact with powdery mildew (Hammami et al., 2011). While these glycolipids are believed to contribute to the fitness of both organisms, possibly in fending off competitors, their true ecological role remains to be determined.

Following the release of the first fungal genomes and the development of reliable bioinformatic tools to predict secretion signals in protein sequences, emphasis has been placed on the study of effector proteins as determinants of pathogenicity (Torto et al., 2003). In a variety of plant pathogens, including smut fungi, powdery mildews, rusts, and oomycetes, effectors have been found to affect virulence, suppress plant defense responses, dictate host specificity, and/or to maintain a biotrophic interaction. In this study, 345 CSEPs were identified in *P. flocculosa*, a number similar to those found in the plant pathogenic Ustilaginales. However, the single most striking difference between *P. flocculosa* and the three smut fungi was found in terms of gene homology between CSEPs (Figure 5). Borrowing from the wealth of information available from sequenced Ustilaginales genomes and functional analyses in *U. maydis*, it appeared that orthologs for nearly all *U. maydis* genes (51 out of 55) encoding secreted proteins deemed to influence pathogenicity and virulence are absent in *P. flocculosa*. Considering that *P. flocculosa* has conserved all the necessary elements to mate and that it shows a high level of conservation for all other pathogenicity-related genes (e.g., CWDEs and biosynthesis of secondary metabolites), this result strongly suggests that loss of effectors represent the single most determinant factor that relates to the nonpathogenic lifestyle of *P. flocculosa*. At the same time, the presence of many orthologs in both *U. hordei* (21) and *S. reilianum* (40) strengthens the role that those proteins play in the establishment of a compatible pathogenic interaction in their respective host plants.

Allowing for this absence of homologs to a large number of virulence-promoting effectors in the phytopathogenic Ustilaginales, *P. flocculosa* seems to have evolved its own set of secreted proteins, the role of which is unknown at present. While not a phytopathogen, a feature shared with the other described *Pseudozyma* spp (Avis et al., 2001), *P. flocculosa* possesses nonetheless the unique ability to antagonize powdery mildews. Based on our results, this interaction between a biocontrol agent

and a fungal pathogen might also be dictated by effector proteins. For example, the secretome of *P. flocculosa* includes two NPP1-containing proteins that were not found in plant pathogenic Ustilaginales. In basidiomycetes, such proteins, labeled necrosis and ethylene-inducing peptide 1-like proteins, have only been identified in *Moniliophthora perniciosa*, the causal agent of witches'-broom disease of cacao. These proteins are absent from all other sequenced basidiomycetes, including Ustilaginomycetes (Kämper et al., 2006; Schirawski et al., 2010; Laurie et al., 2012). Horizontal gene transfer has been suggested as one mechanism to account for the phylogenetic distribution of NPP1-containing proteins (Gijzen and Nürnberger, 2006). However, neither the GC content nor the codon bias supports this hypothesis. These proteins were recently determined to exhibit structural similarities to cytolytic toxins produced by marine organisms (actinoporins), which forms transmembrane pores (Ottmann et al., 2009). Interestingly, the rapid collapse of powdery mildew colonies by *P. flocculosa* has always been associated with an alteration of the plasma membrane and cytoplasmic leaking (Hajlaoui and Bélanger, 1991; Hajlaoui et al., 1994; Mimeo et al., 2009). Thus, NPP1-containing proteins could be key elements explaining the antagonism of *P. flocculosa* toward powdery mildews. For now, the paucity of data detailing the interaction has prevented the validation of this hypothesis, a challenge that this new set of data will help to address.

Other genes from species-specific gene families revealed here might explain how *P. flocculosa* acquired the potential to antagonize powdery mildews. For instance, the discovery of two divergent GDSL lipases/esterases, containing a carbohydrate esterase family 16 motif that is exclusive to *P. flocculosa*, may be of relevance to its activity as an epiphytic competitor. GDSL lipases are multifunctional enzymes with a broad range of substrate specificities (Akoh et al., 2004). Oh et al. (2005) reported that a GDSL produced by *Arabidopsis thaliana* had direct antifungal activity against *Alternaria brassicicola*. Of particular interest, the effect of the protein on fungal spores was reminiscent of the effect observed on powdery mildew spores when antagonized by *P. flocculosa* (Hajlaoui et al., 1992; Hajlaoui and Bélanger, 1993; Mimeo et al., 2009). This distinctive feature may therefore be linked to the biocontrol activity of *P. flocculosa*.

Another important observation differentiating *P. flocculosa* from the plant pathogens was the identification of two loci similar to others found in mycoparasitic *Trichoderma* species (Kubicek et al., 2011). These loci contain a gene encoding a subgroup C GH18 chitinase adjacent to another gene on the opposite strand encoding a LysM protein. Interestingly, LysM protein TAL6 was shown to inhibit spore germination of *Trichoderma atroviride* while having no effect on *Aspergillus niger* or *Neurospora crassa* (Seidl-Seiboth et al., 2013). Since *tal6* is mainly expressed in mature hyphae, the authors suggested that TAL6 could have a self-regulatory role in fungal growth and development. TAL6 could also act to protect the fungus against self-degradation by its own TAC6 during mycoparasitism. A similar protective function for LysM against wheat (*Triticum aestivum*) chitinases was described during infection by *Mycosphaerella graminicola* (Marshall et al., 2011). A role for chitinases in the biocontrol activity of *P. flocculosa* has never been established (Bélanger et al., 2012), but this

finding suggests a feature shared by mycoparasites that warrants further investigation.

In conclusion, this work highlighted how *P. flocculosa* represents a unique model for comparative genomics because it provides insights into both the virulence determinants of fungal plant pathogens and the elusive properties of biocontrol agents. It can thus serve as a powerful tool to advance concerted efforts toward plant protection.

METHODS

Sequencing, Assembly, and Alignment of Genomes

Genomic DNA was extracted from *Pseudozyma flocculosa* (DAOM 196992) according to the protocol of the Qiagen DNeasy plant mini kit (Venlo) for fungal cells. Two sequencing runs were performed by GenomeQuebec (McGill University, Montreal, Canada) and yielded 522 Mb data from a whole-genome shotgun library and 43 Mb data from a 2.5-kb paired-end library, respectively. A third run was performed at the Institut de Biologie Integrative et des Systèmes (Université Laval, Quebec City, Canada) and yielded 197 Mb data from 4.5-kb paired-end library. Sequences obtained represent an average coverage of 28 times the expected 25-Mb genome. All runs used the Roche 454 Life Sciences technology of pyrosequencing. Sequencing information was assembled using Roche proprietary software Newbler v2.6. MUMmer v.3.0 was used with promoter program and `-mum` parameter for the alignment of genomes. Metrics for genome synteny showed in Supplemental Table 3 online were generated using OrthoCluster (version 2013/02/14; Zeng et al., 2008). Genomes from plant pathogenic species were retrieved from the Munich information center for protein sequences (<ftp://ftpmips.gsf.de>).

Structural Annotation

Structural annotation of the *P. flocculosa* genome relied on several sources of information. First and foremost, the sequencing of the transcriptome was used as the most reliable source of information. The data were obtained from an experiment specifically designed to yield an exhaustive bank of transcripts. Briefly, *P. flocculosa* was cultured in yeast malt peptone dextrose broth and harvested at four different times (4, 8, 16, and 30 h) spanning the different growth phases of the fungus (Hammami et al., 2008). Total RNA from the four samples was extracted using the RNeasy mini kit from Qiagen. Library construction was done by Genome British Columbia at the Michael Smith Genome Sciences Centre. Only the poly(A)⁺ fraction was used, and cDNA were synthesized from random hexamers to cover entire transcripts. Both ends of fragments from the four libraries were sequenced on a single lane using the Illumina HiSeq sequencer and yielded 2×192 M 100-bp sequences.

Sequences from RNA sequencing were de novo assembled using CLC Genomics Workbench v5.0.1 (CLC bio, Aarhus, Denmark) and Trinity assembler v2012-01-25 (Grabherr et al., 2011). All assembled contigs, mapped to the genome with GMAP v2007-09-28 (Wu and Watanabe, 2005), were selected based on reads mapped to the genome by TopHat v1.3.3 (Trapnell et al., 2009), alignment of protein sequences from *Ustilago maydis* and *Sporisorium reilianum* by Exonerate v2.2.0 (Slater and Birney, 2005), and a provisory gene prediction by AUGUSTUS v.2.5.5 (Stanke et al., 2004) based on training information from *U. maydis* genome annotation. All this information was visualized with GBrowse v2.0 (Stein et al., 2002). Following manual curation, 6458 highly reliable coding sequences were retrieved. A subset of 400 handpicked annotations was then used to train ab initio gene predictors AUGUSTUS and SNAP v2006-07-28 (Korf, 2004). The non-trainable gene predictor GeneMark-ES v2.3e (Lomsadze et al., 2005) was also used as a supplementary source of evidence.

Annotation files from transcript sequences processed by PASA (Haas et al., 2003), protein sequences of *U. maydis* and *S. reilianum* aligned with Exonerate, and predictions from ab initio gene predictors were fed to EvidenceModeler (Haas et al., 2008) to produce an initial annotation that was further improved by manual curating. tRNA gene predictions were obtained using tRNAScan-SE v1.3.1 (Lowe and Eddy, 1997) with default settings for eukaryotes.

Repetitive Sequences

A custom library of repeat sequences was created from de novo calling of repeats by RepeatModeler (<http://repeatmasker.org>) on the target genome, repeats obtained via RepeatModeler run performed on *Ustilago hordei*, repeats identified from the *U. hordei* MAT-1 locus, and the entire RepeatMasker library of repeats. RepeatMasker (<http://repeatmasker.org>) was run with the custom library on the assembled 40 scaffolds and on the contigs that were not included in the assembly (contigs 1191 and higher). Analysis of repeat induced point mutations was done using RIPCAL v1.0 (Hane and Oliver, 2008) on sequences retrieved from the analysis of repetitive sequences and coding sequences.

Functional Annotation

Gene naming was first established on the basis of a BLASTP v2.2.27 (Altschul et al., 1990) similarity search against the Swiss-Prot database (Boeckmann et al., 2003). At this point, only sequences with 70% identity and 70% coverage were assigned a name with the prefix "probable." Other names were obtained via the Blast2GO tool v2.6.2 (Conesa et al., 2005) using default settings and were assigned the prefix "putative." Gene functions were also predicted using InterProScan v4.8 (database v38.0) (Quevillon et al., 2005).

Annotation of CAZy was performed using the dbCAN Web server (Yin et al., 2012). Original output was downloaded and parsed with custom python script to retrieve only matches the following parameters: e-value < 10⁻³ if alignment length ≤ 80 amino acids, otherwise e-value < 10⁻⁵; covered fraction > 0.3. Association of CAZy modules with substrates specificity followed the classification used by Amselem et al. (2011).

A search for genes involved in the biosynthesis of secondary metabolites was performed using JCVI Secondary Metabolite Unique Regions Finder Web server (SMURF) (Khaldi et al., 2010). The server first identifies what is called backbone genes for which the coded protein harbors a domain related to polyketide synthases (PKS and PKS-Like), nonribosomal peptide synthetase (NRPS and NRPS-Like), or dimethylallyl Trp synthase (DMAT). From the located backbone gene, a region is identified as grouping the other genes usually associated to the biosynthesis and transport of secondary metabolites. Predictions of backbone genes are highly reliable. Otherwise, prediction of biosynthetic clusters is useful but results are approximate. Overall, both information outputs give a good portrait of global genetic resources putatively involved in the biosynthesis of secondary metabolites.

Annotation of secreted and CSEPs was accomplished according to the method described by Mueller et al. (2008). Secreted proteins were selected based upon SignalP v3.0 (Bendtsen et al. 2004) D-value and Dmax cutoffs, TargetP v1.1 (Emanuelsson et al., 2000) predicted location of protein, TMHMM v2.0 (Krogh et al., 2001) predicted number of transmembrane domains and position according to cleavage site, and finally, correlation to LocDB or PotLocDB ProtComp v9.0 (<http://www.softberry.com>) databases. Based on InterProScan-assigned domains, proteins lacking enzymatic functions were classified as candidate effectors (CSEP).

A list of all *P. flocculosa* genes and annotations is presented in Supplemental Data Set 7 online.

Clusters of Genes Encoding Secreted Proteins

Identification of physically clustered secreted proteins was based on a more permissive version of the definition given by Kämper et al. (2006).

Here, clusters are defined as groups of three directly neighboring genes that are predicted to encode secreted proteins or groups of more than three genes with no more than three genes encoding nonsecreted gene products in between subgroups of at least two neighboring genes encoding secreted products.

Phylogeny Analysis

A first phylogenetic tree was constructed using 306 gene families that were generated using Tribe-MCL (Enright et al., 2002) and the protein sets from *P. flocculosa*, *Ustilago maydis*, *U. hordei*, *Sporisorium reilianum*, and *Malassezia globosa*. Gene families had to group only one gene from every species and to have a positive hit to a member of the Metabolism or Cellular Process categories of the KEGG database (Kanehisa and Goto, 2000) (see Supplemental Data Set 8 online). Alignments were performed using Clustal omega 1.1.0 with default parameters. Alignments were edited for optimization (see Supplemental Data Set 9 online). For each gene family, 1000 replicates were generated using the bootstrap method. For each replicate, a phylogenetic tree was estimated with the maximum likelihood method for protein amino acid sequences under the constraint of a molecular clock and the Jones-Taylor-Thornton model by the program Promk from the Phylip package version 3.69 (<http://evolution.genetics.washington.edu/phylip/>) using default parameters. A consensus tree was constructed by a majority rule (extended) on the basis of the topologies of the 306,000 trees that were considered as rooted (midpoint) by the program Consense (Phylip). Each branch of the consensus tree shows the percentage of the 306,000 trees that support it.

For the phylogenetic tree of basidiomycete pheromone receptor protein sequences, the maximum likelihood method based on the JTT matrix-based model (Jones et al., 1992) and evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011). Pheromone receptor sequences were C-terminally truncated to exclude the cytoplasmic tail and to optimize the alignment (see Supplemental Data Set 10 online). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in <50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Initial trees for the heuristic search were obtained automatically as follows: When the number of common sites was <100 or less than one-fourth of the total number of sites, the maximum parsimony method was used; otherwise, the BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 270 positions in the final data set.

Gene Families, Orthologs, and Characterized Genes for Virulence and Pathogenicity

All 27,462 sequences from *P. flocculosa*, *U. maydis*, *U. hordei*, and *S. reilianum* were searched for similarity, one against the other, using the BLASTP algorithm. Tribe-MCL (Enright et al., 2002) was used to group sequences into gene families on the basis of the expectation value (e-value) of all hits. The inflation value was set to 2.1 as by Kämper et al. (2006). Each family was assigned to a category based on whether it contained genes from all four organisms (core), genes from two to three organisms including *P. flocculosa* (shared), genes from the three plant pathogens (present exclusively in all three plant pathogens), genes from two plant pathogens (present exclusively in two plant pathogens), and genes from only one species (species-specific). For the analysis of CSEP, non-CSEP genes were not considered in the assignment of categories.

Orthology was established on the basis of bidirectional best hits from the results of BLASTP all-against-all similarity search with a cutoff of

1×10^{-6} . For homology, BLASTP local alignment had to cover at least 50% of query sequence and show 25% identity.

Information concerning effects of *U. maydis* gene deletion on pathogenicity and virulence was retrieved from PHI-base (Winnenburg et al., 2006).

Mating Assays

The mating test was done as described by Müller et al. (1999) by mixing *U. maydis* strain FB1 (*a1* and *b1*) or MB215 (*a2* and *b13*) with *P. flocculosa* to observe the formation of fluffy white colonies indicative of dikaryotic growth following successful mating. Using overnight cultures of each strain with an equal OD₆₀₀ of 0.5, 5 μ L of the strains were mixed with each other on a potato dextrose agar plate supplemented with 1% (w/v) activated charcoal. The assays were conducted for 48 h at room temperature.

Accession Numbers

This Whole Genome Shotgun project has been deposited at GenBank/ DDBJ/EMBL under accession number AOUS00000000. The version described in this article is the first version, AOUS01000000.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Number of Introns in Four Ustilaginomycotina Fungi According to Relative Position in Genes.

Supplemental Figure 2. Relative Frequency of Introns According to Relative Position in Genes.

Supplemental Figure 3. Analysis of Dinucleotide Frequency in *P. flocculosa*.

Supplemental Figure 4. Mating Assays between *P. flocculosa* and *U. maydis*.

Supplemental Table 1. Inventory of Repetitive Sequences in the *P. flocculosa* Genome Assembly.

Supplemental Table 2. Inventory of Repetitive Sequences in Non-assembled *P. flocculosa* Sequences.

Supplemental Table 3. OrthoCluster Analysis.

Supplemental Table 4. Subset of Gene Correspondence for Important Biological Processes.

Supplemental Table 5. Conserved SMURF Secondary Metabolite Gene Clusters.

Supplemental Data Set 1. Genes and Gene Families Distribution.

Supplemental Data Set 2. Inventory of Genes Related to Mating and Pathogenicity.

Supplemental Data Set 3. Inventory of Genes Related to Meiosis and Other Functions.

Supplemental Data Set 4. Gene Counts for CAZy Families.

Supplemental Data Set 5. CSEP IDs.

Supplemental Data Set 6. Characterized Genes for Pathogenicity and Virulence in *U. maydis*.

Supplemental Data Set 7. Gene Set.

Supplemental Data Set 8. Genes for Phylogeny.

Supplemental Data Set 9. Amino Acid Sequence Alignments of the 306 Gene Families Used for Phylogeny Reconstruction (Figure 4).

Supplemental Data Set 10. Amino Acid Sequence Alignment of Pheromone Receptor Proteins Used for Phylogeny Reconstruction (Figure 5).

ACKNOWLEDGMENTS

We thank Michael Bölker for supplying *U. maydis* strains used in this study, Marie Vasse and Anaïs Pitarch for technical assistance, and Jérôme Laroche and Brian Boyle from the Institut de Biologie Intégrative et des Systèmes at Laval University for help and support in bioinformatics. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, by grants from Fonds de Recherche du Québec–Nature et Technologies, and Natural Sciences and Engineering Research Council of Canada to F.L. and the Canada Research Chairs Program to R.R.B.

AUTHOR CONTRIBUTIONS

F.L. carried out genomic analyses (genome and transcriptome assemblies, gene structure prediction and curation, functional annotation, and comparative analyses), participated in the interpretation of data, submitted sequences to GenBank, and drafted the article. D.L.J. verified presence of pathogenicity, mating and meiosis genes, improved annotation of secreted proteins, analyzed genomic data, and participated in the interpretation of results and drafting of the article. C.L. participated to genome annotation and interpretation of data. B.T. conducted mating assays experiments. R.L. carried out bioinformatic analysis of repetitive elements. F.B. participated in the interpretation of data and drafting of the article. G.G.B. carried out analysis of mating-type genes and phylogeny and participated in the interpretation of data and drafting of the article. R.R.B. participated in genome annotation, interpretation of data, and drafting of the article. All authors have read and approved the final article.

Received May 24, 2013; revised May 24, 2013; accepted June 12, 2013; published June 25, 2013.

REFERENCES

- Akoh, C.C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H., and Shaw, J.-F. (2004). GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43**: 534–552.
- Amselem, J., et al. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* **7**: e1002230.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Avis, T.J., Caron, S.J., Boekhout, T., Hamelin, R.C., and Bélanger, R.R. (2001). Molecular and physiological analysis of the powdery mildew antagonist *Pseudozyma flocculosa* and related fungi. *Phytopathology* **91**: 249–254.
- Bakkeren, G., Kämper, J., and Schirawski, J. (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. *Fungal Genet. Biol.* **45** (suppl. 1): S15–S21.
- Begerow, D., Bauer, R., and Boekhout, T. (2000). Phylogenetic placements of ustilaginomycetous anamorphs as deduced from nuclear LSU rDNA sequences. *Mycol. Res.* **104**: 53–60.
- Begerow, D., Stoll, M., and Bauer, R. (2006). A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data. *Mycologia* **98**: 906–916.

- Bélanger, R.R., and Avis, T.J.** (2002). Ecological processes and interactions occurring in leaf surface fungi. In *Phyllosphere Microbiology*, S.E. Lindow, E.I. Hecht-Poinar, and V.J. Elliott, eds (St. Paul, MN: The American Phytopathological Society), pp. 193–208.
- Bélanger, R.R., and Deacon, J.W.** (1996). Interaction specificity of the biocontrol agent *Sporothrix flocculosa*: A video microscopy study. *Phytopathology* **86**: 1317–1323.
- Bélanger, R.R., Labbé, C., Lefebvre, F., and Teichmann, B.** (2012). Mode of action of biocontrol agents: All that glitters is not gold. *Can. J. Plant Pathol.* **34**: 469–478.
- Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S.** (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**: 783–795.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M.** (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**: 365–370.
- Böhm, J., Hoff, B., O'Gorman, C.M., Wolfers, S., Klux, V., Binger, D., Zadra, I., Kümsteiner, H., Pöggeler, S., Dyer, P.S., and Kück, U.** (2013). Sexual reproduction and mating-type-mediated strain development in the penicillin-producing fungus *Penicillium chrysogenum*. *Proc. Natl. Acad. Sci. USA* **110**: 1476–1481.
- Brefort, T., Doehlemann, G., Mendoza-Mendoza, A., Reissmann, S., Djamei, A., and Kahmann, R.** (2009). *Ustilago maydis* as a pathogen. *Annu. Rev. Phytopathol.* **47**: 423–445.
- Cheng, Y.L., Belzile, F., Tanguay, P., Bernier, L., and Bélanger, R.R.** (2001). Establishment of a gene transfer system for *Pseudozyma flocculosa*, an antagonistic fungus of powdery mildew fungi. *Mol. Genet. Genomics* **266**: 96–102.
- Clément-Mathieu, G., Chain, F., Marchand, G., and Bélanger, R.R.** (2008). Leaf and powdery mildew colonization by glycolipid-producing *Pseudozyma* species. *Fungal Ecol.* **1**: 69–77.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M.** (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.
- Dean, R.A., et al.** (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980–986.
- Doehlemann, G., Reissmann, S., Assmann, D., Fleckenstein, M., and 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.
- Doehlemann, G., van der Linde, K., Assmann, D., Schwambach, D., Hof, A., Mohanty, A., Jackson, D., and Kahmann, R.** (2009). Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. *PLoS Pathog.* **5**: e1000290.
- Duplessis, S., et al.** (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. USA* **108**: 9166–9171.
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G.** (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**: 1005–1016.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A.** (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**: 1575–1584.
- Gijzen, M., and Nürnberger, T.** (2006). Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry* **67**: 1800–1807.
- Grabherr, M.G., et al.** (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**: 644–652.
- Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., Salzberg, S.L., and White, O.** (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**: 5654–5666.
- Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R.** (2008). Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**: R7.
- Hajlaoui, M.R., and Bélanger, R.R.** (1991). Comparative effects of temperature and humidity on the activity of three potential antagonists of rose powdery mildew. *Eur. J. Plant Pathol.* **97**: 203–208.
- Hajlaoui, M.R., and Bélanger, R.R.** (1993). Antagonism of the yeast-like phyllophane fungus *Sporothrix flocculosa* against *Erysiphe graminis* var *tritici*. *Biocontrol Sci. Technol.* **3**: 427–434.
- Hajlaoui, M.R., Benhamou, N., and Bélanger, R.R.** (1992). Cytochemical study of the antagonistic activity of *Sporothrix flocculosa* on rose powdery mildew, *Sphaerotheca pannosa* var. *rosae*. *Cytology and Histology*. **82**: 583–589.
- Hajlaoui, M.R., Traquair, J.A., Jarvis, W.R., and Bélanger, R.R.** (1994). Antifungal activity of extracellular metabolites produced by *Sporothrix flocculosa*. *Biocontrol Sci. Technol.* **4**: 229–237.
- Hammami, W., Castro, C.Q., Rémus-Borel, W., Labbé, C., and Bélanger, R.R.** (2011). Ecological basis of the interaction between *Pseudozyma flocculosa* and powdery mildew fungi. *Appl. Environ. Microbiol.* **77**: 926–933.
- Hammami, W., Labbé, C., Chain, F., Mimee, B., and Bélanger, R.R.** (2008). Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa*. *Appl. Microbiol. Biotechnol.* **80**: 307–315.
- Hane, J.K., and Oliver, R.P.** (2008). RIPCAL: A tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics* **9**: 478.
- Hane, J.K., Rouxel, T., Howlett, B.J., Kema, G.H., Goodwin, S.B., and Oliver, R.P.** (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* **12**: R45.
- Jarvis, W.R., Shaw, L.A., and Traquair, J.A.** (1989). Factors affecting antagonism of cucumber powdery mildew by *Stephanoascus flocculosus* and *S. rugulosus*. *Mycol. Res.* **92**: 162–165.
- Jeffares, D.C., Mourier, T., and Penny, D.** (2006). The biology of intron gain and loss. *Trends Genet.* **22**: 16–22.
- Jones, D.T., Taylor, W.R., and Thornton, J.M.** (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.
- Kämper, J., et al.** (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**: 97–101.
- Kanehisa, M., and Goto, S.** (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**: 27–30.
- Kemen, E., and Jones, J.D.** (2012). Obligate biotroph parasitism: Can we link genomes to lifestyles? *Trends Plant Sci.* **17**: 448–457.
- Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K. H., and Fedorova, N.D.** (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* **47**: 736–741.
- Korf, I.** (2004). Gene finding in novel genomes. *BMC Bioinformatics* **5**: 59.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.** (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**: 567–580.
- Kubicek, C.P., et al.** (2011). Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biol.* **12**: R40.
- Laurie, J.D., Ali, S., Linning, R., Mannhaupt, G., Wong, P., Güldener, U., Münsterkötter, M., Moore, R., Kahmann, R., Bakkeren, G., and Schirawski, J.** (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, J.D., Linning, R., Wong, P., and Bakkeren, G.** (2013). Do TE activity and counteracting genome defenses, RNAi and methylation, shape the sex lives of smut fungi? *Plant Signal. Behav.* **8**: e23853.

- Lemieux, R.U., Thorn, J.A., Brice, C., and Haskins, R.H. (1951). Biochemistry of the ustilaginales. II. Isolation and partial characterization of ustilagic acid. *Can. J. Chem.* **29**: 409–414.
- Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **33**: 6494–6506.
- Lowe T.M. and Eddy S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.
- Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K.E., Thomma, B.P.H.J., and Rudd, J.J. (2011). Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiol.* **156**: 756–769.
- Mimee, B., Labbé, C., Pelletier, R., and Bélanger, R.R. (2005). Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa*. *Antimicrob. Agents Chemother.* **49**: 1597–1599.
- Mimee, B., Pelletier, R., and Bélanger, R.R. (2009). In vitro antibacterial activity and antifungal mode of action of flocculosin, a membrane-active cellobiose lipid. *J. Appl. Microbiol.* **107**: 989–996.
- Mourier, T., and Jeffares, D.C. (2003). Eukaryotic intron loss. *Science* **300**: 1393.
- Mueller, O., Kahmann, R., Aguilar, G., Trejo-Aguilar, B., Wu, A., and de Vries, R.P. (2008). The secretome of the maize pathogen *Ustilago maydis*. *Fungal Genet. Biol.* **45** (Suppl 1): S63–S70.
- Müller, P., Aichinger, C., Feldbrügge, M., and Kahmann, R. (1999). The MAP kinase kpp2 regulates mating and pathogenic development in *Ustilago maydis*. *Mol. Microbiol.* **34**: 1007–1017.
- Nielsen, C.B., Friedman, B., Birren, B., Burge, C.B., and Galagan, J.E. (2004). Patterns of intron gain and loss in fungi. *PLoS Biol.* **2**: e422.
- O’Gorman, C.M., Fuller, H.T., and Dyer, P.S. (2009). Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Nature* **457**: 471–474.
- Oh, I.S., Park, A.R., Bae, M.S., Kwon, S.J., Kim, Y.S., Lee, J.E., Kang, N.Y., Lee, S., Cheong, H., and Park, O.K. (2005). Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternaria brassicicola*. *Plant Cell* **17**: 2832–2847.
- Ottmann, C., Luberaeki, B., Kufner, I., Koch, W., Brunner, F., Weyand, M., Mattinen, L., Pirhonen, M., Anderluh, G., Seitz, H.U., Nürnberger, T., and Oecking, C. (2009). A common toxin fold mediates microbial attack and plant defense. *Proc. Natl. Acad. Sci. USA* **106**: 10359–10364.
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: Protein domains identifier. *Nucleic Acids Res.* **33** (Web Server issue): W116–W120.
- Roy, S.W., and Gilbert, W. (2006). The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nat. Rev. Genet.* **7**: 211–221.
- Saville, B.J., Donaldson, M.E., and Doyle, C.E. (2012). Investigating host induced meiosis in a fungal plant pathogen. In *Meiosis: Molecular Mechanisms and Cytogenetic Diversity*, A. Swan, ed (New York: InTech), pp. 411–460.
- Schirawski, J., Heinze, B., Wagenknecht, M., and Kahmann, R. (2005). Mating type loci of *Sporisorium reilianum*: Novel pattern with three a and multiple b specificities. *Eukaryot. Cell* **4**: 1317–1327.
- Schirawski, J., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* **330**: 1546–1548.
- Seidl-Seiboth, V., Zach, S., Frischmann, A., Spadiut, O., Dietzsch, C., Herwig, C., Ruth, C., Rodler, A., Jungbauer, A., and Kubicek, C.P. (2013). Spore germination of *Trichoderma atroviride* is inhibited by its LysM protein TAL6. *FEBS J.* **280**: 1226–1236.
- Sharpton, T.J., Neafsey, D.E., Galagan, J.E., and Taylor, J.W. (2008). Mechanisms of intron gain and loss in *Cryptococcus*. *Genome Biol.* **9**: R24.
- Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31.
- Spanu, P.D. (2012). The genomics of obligate (and nonobligate) biotrophs. *Annu. Rev. Phytopathol.* **50**: 91–109.
- Spanu, P.D., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543–1546.
- Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32** (Web Server issue): W309–W312.
- Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., and Lewis, S. (2002). The generic genome browser: A building block for a model organism system database. *Genome Res.* **12**: 1599–1610.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**: 2731–2739.
- Teichmann, B., Labbé, C., Lefebvre, F., Bötker, M., Linne, U., and Bélanger, R.R. (2011b). Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa*. *Mol. Microbiol.* **79**: 1483–1495.
- Teichmann, B., Lefebvre, F., Labbé, C., Bötker, M., Linne, U., and Bélanger, R.R. (2011a). Beta hydroxylation of glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-dependent β -hydroxylase. *Appl. Environ. Microbiol.* **77**: 7823–7829.
- Teichmann, B., Linne, U., Hewald, S., Marahel, M.A., and Bötker, M. (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Mol. Microbiol.* **66**: 525–533.
- Torto, T.A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N.A., van West, P., and Kamoun, S. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res.* **13**: 1675–1685.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.
- Traquair, J.A., Shaw, L.A., and Jarvis, W.R. (1988). New species of *Stephanoascus* with *Sporothrix* anamorphs. *Can. J. Bot.* **66**: 926–933.
- Vinale, F., Sivasithamparam, K., Ghisalberti, E.L., Marra, R., Woo, S.L., and Lorito, M. (2008). *Trichoderma*–plant–pathogen interactions. *Soil Biol. Biochem.* **40**: 1–10.
- Walton, J.D. (1994). Deconstructing the cell wall. *Plant Physiol.* **104**: 1113–1118.
- Whisson, S.C., et al. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450**: 115–118.
- Winnenburg, R., Baldwin, T.K., Urban, M., Rawlings, C., Köhler, J., and Hammond-Kosack, K.E. (2006). PHI-base: A new database for pathogen host interactions. *Nucleic Acids Res.* **34** (Database issue): D459–D464.
- Wu, T.D., and Watanabe, C.K. (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. USA* **104**: 18730–18735.
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40** (Web Server issue): W445–W451.
- Zeng, X., Nesbitt, M.J., Pei, J., Wang, K., Vergara, I.A., and Chen, N. (2008). OrthoCluster: A new tool for mining synteny blocks and applications in comparative genomics. In *Proceedings of the 11th International Conference on Extending Database Technology*, J. Teubner, ed (New York: Association for Computing Machinery), pp. 656–667.