Obligate Biotrophy Features Unraveled by the Genomic Analysis of the Rust Fungi, *Melampsora larici-populina* and *Puccinia graminis* f. sp. *tritici*

Sébastien Duplessis[1+*], Christina A. Cuomo[2+*], Yao-Cheng Lin[3], Andrea Aerts[4], Emilie Tisserant[1], Claire Veneault-Fourrey[1], David L. Joly[5], Stéphane Hacquard[1], Joelle Amselem[6], Brandi Cantarel[7], Readman Chiu[8], Pedro Couthinho[7], Nicolas Feau[5#], Matthew Field[8], Pascal Frey[1], Eric Gelhaye[1], Jonathan Goldberg[2], Manfred Grabherr[2], Chinnappa Kodira[2#], Annegret Kohler[1], Ursula Kües[9], Erika Lindquist[4], Susan Lucas[4], Rohit Mago[10], Evan Mauceli[2], Emmanuelle Morin[1], Claude Murat[1], Jasmyn L. Pangilinan[4], Robert Park[11], Matthew Pearson[2], Hadi Quesneville[6], Nicolas Rouhier[1], Sharadha Sakthikumar[2], Asaf A. Salamof[4], Jeremy Schmutz[4], Benjamin Selles[1], Harris Shapiro[4], Philippe Tangay[5], Gerald A. Tuskan[12], Yves Van de Peer[3], Bernard Henrissat[7], Pierre Rouzé[3], Jeffrey G. Ellis[10], Peter N. Dodds[10], Jacqueline E. Schein[8], Shaobin Zhong[13#], Richard C. Hamelin[5], Igor V. Grigoriev[4], Les J. Szabo[14,13*], Francis Martin[1*]

[1]INRA, UMR 1136 INRA/Nancy Université, Interactions Arbres/Micro-organismes, Centre INRA de Nancy, 54280 Champenoux, France. [2]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [3]Department of Plant Systems Biology, VIB, Ghent, Belgium. [4]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA. [5]Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, 1055 du PEPS, PO Box 10380, Stn Sainte-Foy, Québec, QC, G1V 4C7 Canada. [6]INRA URGI, Route de St Cyr 78026 Versailles Cedex, France. [7]UMR 6098 CNRS-Universités Aix-Marseille I, Marseille, France. [8]Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. [9]Institute of Forest Botany, Georg-August-University, 37077 Göttingen, Germany. [10]Division of Plant Industry, Commonwealth Scientific and Industrial Research Organization, Canberra, ACT 2601, Australia. [11]University of Sydney, Plant Breeding Institute Cobbitty, PMB 11, Camden, NSW 2570, Australia. [12]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6422. [13]Department of Plant Pathology, University of Minnesota, St. Paul, MN, USA. [14]Cereal Disease Laboratory, USDA-Agricultural Research Service, St. Paul, MN, USA

**To whom correspondence may be addressed. E-mail:* To duplessi@nancy.inra.fr (SD), cuomo@broadinstitute.org (CAC), les.szabo@ars.usda.gov (LJS) & fmartin@nancy.inra.fr (FM).

9 May 2011

ACKNOWLEDGMENTS:

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Obligate Biotrophy Features Unraveled by the Genomic Analysis of the Rust Fungi, *Melampsora larici-populina* and *Puccinia graminis* f. sp. *tritici*

Sébastien Duplessis[1+*], Christina A. Cuomo[2+*], Yao-Cheng Lin[3], Andrea Aerts[4], Emilie Tisserant[1], Claire Veneault-Fourrey[1], David L. Joly[5], Stéphane Hacquard[1], Joelle Amselem[6], Brandi Cantarel[7], Readman Chiu[8], Pedro Couthinho[7], Nicolas Feau[5#], Matthew Field[8], Pascal Frey[1], Eric Gelhaye[1], Jonathan Goldberg[2], Manfred Grabherr[2], Chinnappa Kodira[2#], Annegret Kohler[1], Ursula Kües[9], Erika Lindquist[4], Susan Lucas[4], Rohit Mago[10], Evan Mauceli[2], Emmanuelle Morin[1], Claude Murat[1], Jasmyn L. Pangilinan[4], Robert Park[11], Matthew Pearson[2], Hadi Quesneville[6], Nicolas Rouhier[1], Sharadha Sakthikumar[2], Asaf A. Salamof[4], Jeremy Schmutz[4], Benjamin Selles[1], Harris Shapiro[4], Philippe Tangay[5], Gerald A. Tuskan[12], Yves Van de Peer[3], Bernard Henrissat[7], Pierre Rouzé[3], Jeffrey G. Ellis[10], Peter N. Dodds[10], Jacqueline E. Schein[8], Shaobin Zhong[13#], Richard C. Hamelin[5], Igor V. Grigoriev[4], Les J. Szabo[14,13*], Francis Martin[1*]

[1]INRA, UMR 1136 INRA/Nancy Université, Interactions Arbres/Micro-organismes, Centre INRA de Nancy, 54280 Champenoux, France. [2]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [3]Department of Plant Systems Biology, VIB, Ghent, Belgium. [4]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA. [5]Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, 1055 du PEPS, PO Box 10380, Stn Sainte-Foy, Québec, QC, G1V 4C7 Canada. [6]INRA URGI, Route de St Cyr 78026 Versailles Cedex, France. [7]UMR 6098 CNRS-Universités Aix-Marseille I, Marseille, France. [8]Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. [9]Institute of Forest Botany, Georg-August-University, 37077 Göttingen, Germany. [10]Division of Plant Industry, Commonwealth Scientific and Industrial Research Organization, Canberra, ACT 2601, Australia. [11]University of Sydney, Plant Breeding Institute Cobbitty, PMB 11, Camden, NSW 2570, Australia. [12]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6422. [13]Department of Plant Pathology, University of Minnesota, St. Paul, MN, USA. [14]Cereal Disease Laboratory, USDA-Agricultural Research Service, St. Paul, MN, USA.

#present address : for NF, UMR 1202 BIOGECO, INRA 69 Route d'Arcachon, 33612 Cestas Cedex; for CDK, 454 Life Sciences, a member of the Roche Group, Branford, CT, USA; for SZ, Department of Plant Pathology, North Dakota State University, Fargo, ND, USA.

*To whom correspondence should be addressed. E-mail: duplessi@nancy.inra.fr (SD), cuomo@broadinstitute.org (CAC), les.szabo@ars.usda.gov (LJS) & fmartin@nancy.inra.fr (FM)

+ These authors contributed equally to this work.

Authors contribution: S.D., C.A.C., L.J.S., I.V.G., G.A.T., F.M. and X designed research; S.D., C.A.C., Y.C.L., A.E., E.T., C.V.F., D.L.J., S.H., J.A., B.C., P.C., N.F., E.G., A.K., U.K., E.M., C.M., H.Q., N.R., B.S., P.T., B.H., R.C.H., L.J.S., I.V.G., F.M. and X performed research; P.R., Y.V.D.P., S.Z., R.P. contributed new reagents/analytic tools; P.F. contributed the *Melampsora larici-populina* strain that was sequenced; L.J.S. contributed the *Puccinia graminis* f.sp. *tritici* strain that was sequenced; A.A., I.V.G., S.D., and F.M. supervised the *Melampsora* genome project; C.A.C. and X supervised the *Puccinia* genome project; S.D., F.M., C.A.C., and X supervised the genome annotation teams; S.D., C.A.C., Y.C.L., A.E., E.T., C.V.F., D.L.J., S.H., J.A., B.C., P.C., N.F., E.G., A.K., U.K., E.M., C.M., H.Q., N.R., B.S., P.T., B.H., R.C.H., L.J.S., I.V.G., F.M. and X analyzed data; S.D., C.V.F., C.A.C., L.J.S., I.V.G., F.M., Y.C.L., D.L.J., S.H., N.F., P.F., A.K., U.K., C.M., N.R., P.T., B.H., R.C.H. wrote the paper.

The authors declare no conflict of interest.

This article is a direct PNAS submission.

Data deposition: Assembly accessions: AECX00000000 (for *M. larici-populina* 98AG31) and AAWC01000000 (for *Puccinia graminis* f. sp *tritici)*; Expression data in GEO: GSE23097 (for *M. larici-populina* 98AG31) and GSE25020 (for *Puccinia graminis* f. sp *tritici)*.

This article contains supporting information online.

Abstract. Rust fungi are some of the most devastating pathogens of crop plants. They are obligate biotrophs, which extract nutrients only from living plant tissues and cannot grow apart from their hosts. Their lifestyle has slowed the dissection of molecular mechanisms underlying host invasion and avoidance or suppression of plant innate immunity. We sequenced the 101–mega–base pair genome of *Melampsora larici-populina*, the causal agent of poplar leaf rust, and the 89–mega–base pair genome of *Puccinia graminis* f. sp. *tritici*, the causal agent of wheat and barley stem rust. We then compared the 16,841 predicted proteins of *M. larici-populina* to the 18,241 predicted proteins of *P. graminis* f. sp *tritici*. Genomic features related to their obligate biotrophic life-style include expanded lineage-specific gene families, a large repertoire of effector-like small secreted proteins (SSPs), impaired nitrogen and sulfur assimilation pathways, and expanded families of amino-acid, oligopeptide and hexose membrane transporters. The dramatic upregulation of transcripts coding for SSPs, secreted hydrolytic enzymes, and transporters *in planta* suggests that they play a role in host infection and nutrient acquisition. Some of these genomic hallmarks are mirrored in the genomes of other microbial eukaryotes that have independently evolved to infect plants, indicating convergent adaptation to a biotrophic existence inside plant cells.

Rust fungi (Pucciniales, Basidiomycota), is a diverse group of plant pathogens composed of more than 120 genera and 6,000 species and are one of the most economically important groups of pathogens of native and cultivated plants (1, 2). *Puccinia graminis*, the causal agent of stem rust, has caused devastating epidemics wherever wheat is grown (3) and a new highly virulent strain (Ug99) threatens wheat production worldwide (4). Similarly, major epidemics of poplar leaf rust, caused by *Melampsora* spp., is a major constraint on the development of bioenergy programs based on domesticated poplars (5) due to the lack of durable host resistance (6,7). Rust fungi are obligate biotrophic parasites with a complex life cycle that often includes two phylogeneticaly unrelated hosts (2; Fig. S1). They have evolved specialized structures, haustoria, formed within host tissue to efficiently acquire nutrients and suppress host defense responses (8). Molecular features driving adaptations to an obligate biotrophic association with plant hosts are unknown. Whether the convergent biotrophic adaptation observed in bacterial parasites (9) and other lineages of microbial eukaryotes (e.g. microsporidia) (10) has lead to functional specializations at the genome level (i.e. gene gain or loss, regulation of gene expression) remains to be determined. The recent report of the genome sequence of *Blumeria graminis*, an ascomycete biotroph pathogen responsible for Barley powdery mildew revealed a genome size expansion due to transposons proliferation concomitant with dramatic reduction in gene content, i.e. genes encoding sugar-cleaving enzymes, transporters and assimilatory enzymes for inorganic nitrate and sulfur (11). Similar gene losses were observed in the genome of the oomycete *Hyaloperonospora arabidopsidis*, a biotroph parasite infecting *Arabidopsis thaliana*, and the diversification of genes encoding RXLR-effector-like secreted proteins (12). Despite their

phylogenetic distance, these two pathogens forming haustoria seems to share striking adaptation convergences to biotrophy. To determine the genetic features underlying pathogenesis and biotrophic ability of rust pathogens, we report here the genome sequences of the rust fungi *M. larici-populina* and *P. graminis* f.sp. *tritici*.

Results and Discussion

Genome sequencing, gene family annotation and expression analysis. We have sequenced the diploid genomes of the poplar leaf rust fungus, *Melampsora larici-populina* and of the wheat stem rust fungus, *Puccinia graminis* f. sp. *tritici*, by Sanger whole-genome shotgun strategy (*SI Text*). The overall assembly sizes of the haploid genomes of *M. larici-populina* and *P. graminis* f. sp. *tritici* are 101.1 Mb and 88.6 Mb, respectively (Table S1). These genomes are much larger than the other sequenced basidiomycete genomes (13,14), but no evidence for whole-genome duplication or large scale dispersed segmental duplications was observed. The expanded size results from a massive proliferation of transposable elements (TEs), which account for nearly 45% in both assembled genomes (Fig. S4, Tables S4 and S5). Class I long-terminal-repeat (LTR) retroelements are more abundant in *M. larici-populina*, whereas class II TIR (Terminal Inverted Repeat) DNA transposons are prominent in *P. graminis* f. sp. *tritici*. Timing of TE activity using sequence divergence of extant copies suggests that a major wave of retrotransposition in the *M. larici-populina* and *P. graminis* f. sp. *tritici* lineages occurred <1 million years ago (Fig. S8).

We predicted 16,841 and 18,241 protein-coding genes in *M. larici-populina* and *P. graminis* f. sp. *tritici,* respectively. The size of these proteomes is similar to the symbiotic basidiomycete *Laccaria bicolor* (14), but strikingly larger than the corn smut fungus, *Ustilago maydis*, a pathogenic biotroph that only possesses ~6,500 proteins (15). Among the predicted proteins, only 41 and 44% in *M. larici-populina* and *P. graminis* f. sp. *tritici*, respectively, showed significant sequence similarity to documented proteins (BLASTP ≤ e-value 10^{-5}) (Fig. S9 and 10). *M. larici-populina* and *P. graminis* f. sp. *tritici* possess a large set of lineage-specific gene pairs showing high similarity levels (80-100%) (Fig. S11 and S15). To investigate protein evolution in *M. larici-populina* and *P. graminis* f. sp. *tritici*, we constructed families containing both orthologs and paralogs from a diverse set of ascomycetous and basidiomycetous fungi. The two genomes shared 3,273 orthologous Tribe-MCL families (*SI Text*) which comprised 5,709 *P. graminis* f. sp. *tritici* genes and 6,117 *M. larici-populina* genes; ~30% of the predicted proteins are lineage-specific, whereas 854 gene families were unique to these two rust fungi. Expansion of protein family sizes was prominent in both *M. larici-populina* and *P. graminis* f. sp. *tritici* (Fig 1, Fig. S13, Tables S7-9); several expanded gene families are lineage-specific, suggesting that important protein-coding innovation occurred in these lineages. Of the 4,858 *M. larici-populina* genes that have an orthologue in *P. graminis* f. sp. *tritici* (Best Reciprocal Hit, e-value ≤ 10^{-3}), very few show conservation of neighbouring orthologues (synteny) (Fig S14). This is likely due to the expansion of the TE and massive reshuffling of the genome as a result. In addition, within the rust fungi, *M. larici-populina* and *P. graminis* f. sp. *tritici* represent very divergent phylogenetic lineages (1). Marked gene family expansions also occurred in those genes coding for alpha-kinases (Fig

S21), oligopeptide membrane transporters (OPT) (Table S19), copper/zinc superoxide dismutase (Table S24), and several groups of predicted transcription factors (Fig S13).

Seventy and 54% of the predicted genes of *M. larici-populina* and *P. graminis* f. sp. *tritici*, respectively, were detected by custom microarray transcript profiling of resting and germinating urediniospores, as well as infected leaves (Fig S17). A significant proportion of the detected transcripts (18%) is differentially expressed (fold-ratio≥10.0, $P<0.05$) in infected leaves (Fig S16), whereas only ~8.0% are specifically expressed *in planta* (Fig S17). Transcripts coding for secreted peptidases and lipases, transporters of hexoses, amino-acids and oligopeptides, and carbohydrate-cleaving enzymes, such as chitin deacetylase and cutinase (Tables 1 and 2, Tables S13 and S17), are strikingly enriched (≥10-fold) *in planta*. However, the most highly upregulated transcripts *in planta* (≥100-fold) are mainly comprised of species-specific transcripts, including those coding for small secreted proteins (SSPs) (Tables S11 and S15). These *in planta*-induced, lineage-specific genes are likely involved in the specific relationship established between these rusts and their respective hosts.

Rust fungi secretomes contain candidate novel rust effectors. Microbial pathogens have evolved highly advanced mechanisms to engage their hosts in intimate contact and sabotage host immune responses by secreting effector proteins into host cells to target regulators of defense (16,17,18). Most SSPs that are specifically produced during plant infection are likely to be effectors that manipulate host cells to facilitate parasitic colonization, such as by suppressing plant innate immunity or enhancing nutrient availability (17). *In silico* gene prediction and manual annotation of SSPs in *M. larici-populina* genome identified a set of 1,184 SSPs (Table S18), of which 74% are species-specific. Homologs of known effectors from *M. lini*, such as haustorially expressed secreted proteins (HESPs) and the avirulence factors *AvrM*, *AvrL567*, *AvrP123*, *AvrP4* from the flax rust fungus *M. lini* (8,17), and the rust-transferred protein RTP1 from the bean rust pathogen (18), are present in highly upregulated *M. larici-populina* transcripts (Table 1, Tables S11-S13). At least 43% of *M. larici-populina* SSPs are expressed in infected leaves at 96 hours post infection. *P. graminis* f. sp. *tritici* contains a similar number of 1,103 SSPs, of which 85% are species-specific. In *P. graminis* f. sp. *tritici*, PGTG_17547 matches the highest number of haustorial ESTs, and is similar in sequence to a predicted secreted protein (ADA54575) from the wheat stripe rust fungus, *P. striiformis* (19). In both rust species, one protein in this group (PGTG_13212, JGI ID# 85525), is similar in sequence to a haustorially expressed protein from the flax rust pathogen, HESP-735 (17) (Tables S11 and S15). Fifty and 29 SSPs belong to the top 100 most highly transcriptionally up-regulated in infected poplar and wheat leaves compared to *M. larici-populina* and *P. graminis* f. sp. *tritici* urediniospores, respectively (Tables S11 and S15). Most upregulated SSP transcripts *in planta* were species-specific, as only 16% have an ortholog in both rust species, suggesting that these sequences are evolving at a very high rate. It remains to be determined whether upregulated SSPs are expressed in infection hyphae and/or haustoria, and whether they remain in the cell-wall, the extra-haustorial matrix, or are adressed to specific compartments of the host cell where they interact with their target proteins as showed for avirulence proteins in *M. lini* (8,17). In *M. larici-populina*, a total of 844 SSPs are organized in 91 families of 3 to 111 members (Table S18); the largest family contains a highly conserved ten-cysteine pattern (Fig. S18).

In *P. graminis* f. sp. *tritici*, a total of 1,105 SSPs are organized in 164 families of 2 to 38 members; the largest family contains a highly conserved eight-cysteine pattern. Four of these proteins show evidence of haustorial expression, suggesting they could be potential effectors.

Rust fungi Carbohydrate-Active Enzymes set. Gene families encoding host-targeted, hydrolytic enzymes acting on plant biopolymers, such as proteinases, lipases, and several sugar-cleaving enzymes (CAZymes)(20), are highly upregulated in both rust pathogen transcriptomes *in planta* (Tables 1 and 2, Tables S13 and S17), suggesting that the invading hyphae is penetrating the host cells by using these degrading enzymes. The comparison of the glycoside hydrolase (GH), glycosyltransferases, polysaccharide lyase (PL) and carbohydrate esterase (CE) of 21 sequenced fungi (Fig. 2) however revealed that *M. larici-populina* and *P. graminis* f. sp. *tritici* have a relatively smaller set of GH-encoding genes (173 and 158 members, respectively) (Table S20); similar to the basidiomycete symbiont *L. bicolor* (14), but much fewer than hemibiotrophic or necrotrophic phytopathogens (e.g., *Magnaporthe grisea*) and saprotrophs (e.g., *Neurospora crassa; Podospora anserina*). This set of CAZymes is strikingly larger than the repertoire of the biotroph *Ustilago maydis* (100 members) (15). In evolving a biotrophic lifestyle, the rust fungi have lost several secreted hydrolytic GH and PL enzymes acting on plant cell wall (PCW) polysaccharides (Fig. 2) and they are lacking the cellulose-binding CBM1 module. However, they show a moderate expansion of a few GHs cleaving plant celluloses and hemi-celluloses (e.g., GH7, GH10, GH12, GH26 and GH27) compared to the biotroph *U. maydis* or the hemibiotroph *M. grisea*. These enzymes, together with *in planta* upregulated and expanded alpha-mannosidase (GH47) and beta-1,3-glucanase (GH5) transcripts (Tables S13 and S17), may play a key role in the initial stages of host colonization, i.e. penetration of the parenchyma cells. On the other hand, induced chitin deacetylases (CE4) present in *P. graminis* f. sp. *tritici, M. larici-populina* and the symbiont *L. bicolor* (14) are likely involved in fungal cell wall remodelling and may play a role in the alteration of the fungal cell wall surface during infection to conceal the hyphae from the host (21).

Expanded rust transporters gene families are expressed during host infection. A process that is crucial to the success of rust pathogen biotrophic interactions is the acquisition of nutrients (carbohydrates and amino acids) by invading hyphae from its host plant through the haustoria (17,22,23). The repertoire of membrane transporters (Table S19) in *M. larici-populina* and *P. graminis* f. sp. *tritici* contains homologs of the hexose transporter HXT1, amino-acid transporters AAT1, AAT2 and AAT3 and H^+-ATPases from the bean rust pathogen (*Uromyces fabae*), known to be highly upregulated during the interaction with its host plant. In addition, *M. larici-populina* and *P. graminis* f. sp. *tritici* genomes display an increased genetic potential for peptide uptake with 22 and 21 oligopeptide transporter (OPT) genes, respectively, whereas other basidiomycete genomes only contain five to 16 OPT genes (Table S19). OPT genes, that are transcriptionally upregulated *in planta* (Tables S13 and S17), are likely involved in the transport of peptides released by the action of the induced proteinases (aspartic peptidase, subtilisin) expressed in infected leaf tissues. The Major Facilitator Superfamily (MFS) gene family is reduced in the *M. larici-populina* and *P. graminis* f. sp. *tritici* genomes compared to other basidiomycetes (Table S19), but many MFS transcripts are however highly expressed *in planta* including two HXT1 homologs. Consistent with *in planta* expression of *M. larici-populina* and *P. graminis* f. sp.

tritici invertase genes (Tables S13 and S17), no homolog of the sucrose transporter Srt1 recently described in *U. maydis* (22) was identified, supporting the preferential uptake of host hexoses by invading rust pathogen hyphae (23). The increased activity of membrane transporters provides the needed fuel for the high primary metabolism activity observed in the invading rust fungi (Tables S13 and S17).

Nitrate and sulfate assimilation pathways deficiencies in rust fungi. Based on the inability of rust fungi to grow *in vitro* we hypothesized that the *M. larici-populina* and *P. graminis* f. sp. *tritici* genomes may lack genes typically present in saprotrophic basidiomycetes. Major anabolic pathways of primary metabolism were manually inspected for potential deficiencies. Although the enzymes of the NH_4^+ assimilation pathway were identified, several genes involved in nitrate assimilation were lacking in both rust pathogen gene repertoires. The nitrate/nitrite porter and the nitrite reductase (NiR) are missing from the nitrate assimilation gene cluster found in other fungi (Fig. 3) (24). Genes required to perform the primary sulfate assimilation were identified in *M. larici-populina* whereas they were lacking in *P. graminis* f. sp. *tritici*. The latter fungus lacks both alpha- and beta-subunits of sulfite reductase (SiR), whereas the *M. larici-populina* beta-subunit of SiR is missing the transketolase domain present in other fungal SiRs. The apparent lack of nitrate and sulfate assimilation enzymes in both rust fungi is consistent with their obligate biotrophic life style, as they depend on reduced nitrogen (either NH4+ or amino acids) and sulfur from plant cells. These metabolic deficiencies have also been found in plant pathogens that represent two independent evolutionary lineages of obligate biotrophy in the oomycete (*H. arabidopsidis*) and ascomycete (*Blumeria graminis*) lineages (11,12).

Conclusions

The obligate biotroph status of rust fungi has limited studies to understand how they invade their hosts and avoid or suppress defense responses. The genome sequences of the poplar leaf and wheat stem rust fungi are an unparalleled opportunity to address questions related to the obligate biotrophy lifestyle. The genetic changes that brought about the evolution of obligate biotrophy from biotrophic progenitors remain obscure. Our comparisons of *M. larici-populina* and *P. graminis* f. sp. *tritici* to other saprotrophic, pathogenic and symbiotic basidiomycetes indicate that the developmental innovations in the lineages of rust fungi did not involve major changes in the ancestral repertoire of proteins with known function. On the other hand, the large set of lineage-specific, expanding gene families may provide a key source of developmental innovation and adaptation. Our analysis shows that the colonization of the host leaf, differentiation of pathogenic structures and control of the plant immune system can be associated with a large-scale invention of lineage-specific proteins. For example, the rich repertoire of candidate effector-like SSPs could underline the co-evolution and adaptation of these obligate pathogens to the plant immune system. Contrary to obligate bacterial biotrophs and microsporidian fungal parasites which oftenly undergo gene loss and genome compaction (9,10), the rust pathogen genomes are amongst the largest fungal genomes sequenced so far showing expanded gene families and massive proliferation of TEs. No massive gene loss was observed in *M. larici-populina* and *P. graminis* f. sp. *tritici*, but

irreversible deletion of genes not essential for the obligate biotrophic life-style (e.g., N and S assimilation), together with a lower set of plant cell wall polysaccharide degrading enzymes are genomic hallmarks of rust fungi and other biotrophic pathogens (11,12).

A deeper understanding of the complex array of factors, such as effector-like SSPs, affecting host–pathogen interactions and co-evolution could ensure efficient targeting of parasite-control methods in agricultural and forest ecosystems.

Material and Methods

Genome sequencing, assembly and annotation. The dikaryotic *M. larici-populina* 98AG31 and *P. graminis* f.sp. *tritici* CDL 75-36-700-3 (race SCCL) strains were sequenced by whole-genome sequencing (WGS) and were assembled into predicted 101.1Mb and 88.6 Mb genomes respectively (*SI Text*). The protein coding-genes (16,841 for *M. larici-populina* and 18,241 for *P. graminis* f.sp. *tritici*) were predicted with a combination of gene callers using ESTs produced from each rust fungus (*SI Text*). Reduced gene sets were considered to perform multigene families analyses (14,527 and 15,680 predicted genes for *M. larici-populina* and *P. graminis* f.sp. *tritici* respectively) by removing gene models presenting overlaps with specific repeats/TE fragments to avoid creation of biased gene families (*SI Text*). The *M. larici-populina* genome sequence can be accessed at http://genome.jgi-psf.org/Mellp1/Mellp1.home.html and the *P. graminis* f.sp. *tritici* genome sequence can be accessed at http://www.broadinstitute.org/annotation/genome/puccinia_group/MultiHome.html.

Microarray analysis of gene expression in urediniopsores and rust-infected plants. For both *M. larici-populina* and *P. graminis* f.sp. *tritici*, gene expression was assessed in resting and *in vitro* germinating urediniospores of the sequenced rust strains as well as in respective host plant tissues at late stages of infection using specific custom 70-mer oligoarrays (*SI Text*). Methods for RNA isolation, probes synthesis and hybridization, data capture and analysis are described in *SI Text* and data have been deposited in GEO (GSE23097 for *M. larici-populina* and GSE25020 for *P. graminis* f.sp. *tritici*).

ACKNOWLEDGEMENTS. The work conducted on *M. larici-populina* by the Joint Genome Institute of the U.S. Department of Energy is supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC02-05CH11231. This project was also funded by grants from the INRA and the Région Lorraine Council to F.M. and S.D.; a grant from Natural Resources Canada to R.C.H. The sequence of *P. graminis* f. sp. *tritici* was funded by the US National Science Foundation and and conducted by the Broad Institute Sequencing Platform. The work of Y-CL, P.R. and Y.VdP was supported by IUAP P6/25 (BioMaGNet). We thank Marie-Pierre Oudot-LeSeq for the initial *M. larici-populina* TE annotation, Benoît Hilselberger for database construction, Carine Commun and Hélène Hirzel for the annotation of the *M. larici-populina* secretome and mating-type genes, respectively, and Jerry Johnson for technical assistance.

1. Aime MC, et al. (2006) An overview of the higher level classification of Pucciniomycotina based on combined analyses of nuclear large and small subunit rDNA sequences. *Mycologia* 98:896-.

2. Cummins GB, Hiratsuka Y (2004) Illustrated genera of rust fungi, 3rd edtion (APS Press, St. Paul, 2003). pp1-240.

3. Leonard KJ, Szabo LJ (2005) Stem rust of small grains and grasses caused by *Puccinia graminis*. *Mol Plant Pathol* 6:99-111.

4. Stokstad E (2007) Deadly wheat fungus threatens world's breadbaskets. *Science* 315:1786-1787.
5. Rubin EM (2008) Genomics of cellulosic biofuels. *Nature* 454:841-845.
6. Duplessis S, Major I, Martin F, Séguin A (2009) Poplar and pathogen interactions: insights from *Populus* genome-wide analyses of resistance and defense gene families and gene expression profiling. *Crit Rev Plant Sci* 28:309-334.
7. Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of genetic and virulence diversity of *Melampsora larici*-populina populations on wild and cultivated poplar and influence of the alternate host. *Phytopathol* 96:1027-1036.
8. Dodds *et al.* (2009) Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytol* 183:993-1000.
9. Ochman H, Moran NA (2001) Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*, 292:1096-1099.
10. Corradi N, Pombert JF, Farinelli L, Didier E, Keeling PJ (2010) The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nature Com* 1:77.
11. Spanu PD et al. (2010) Genome expansion and gene loss in Powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* 330:1543-1546.
12. Baxter L et al. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549-1551.
13. Cuomo CA, Birren B (2010) The fungal genome initiative and lessons learned from genome sequencing. *Meth Enzymol* 470:833-855.
14. Martin F *et al.* (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452:88-92.
15. Kämper J *et al.* (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444:97-101.
16. Panstruga R, Dodds PN (2009) Terrific protein trafic: the mystery of effector protein delivery by filamentous plant pathogens. *Science* 324:748-750.
17. Ellis JG, Rafiqi M, Gan P, Chakrabarti A, Dodds PN (2009) Recent progress in discovery and functional analysis of effector proteins of fungal and oomycète plant pathogens. *Curr Opin Plant Biol* 12:399-405.
18. Voegele RT, Hahn M, Mendgen K (2009) The uredinales: cytology, biochemistry, and molecular biology. In *The Mycota V: Plant relationships*, H. B. Deising, Ed (Spinger, Berlin, 2009), pp69-98.
19. Godfrey D *et al.* (2010) Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics* 11:317.
20. Cantarel BL *et al.* (2009) The Carbohydrate-Active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37:D233-D238.
21. El Gueddari NE, Rauchhaus U, Moerschbacher BM, Deising HB (2002) Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. *New Phytol* 156:103-112.
22. Wahl R, Wippel K, Goos S, Kämper J, Sauer N (2010) A novel high-affinity sucrose transporter is required for virulence of the plant pathogen *Ustilago maydis*. *PLoS Biology* 8:e1000303.
23. Voegele RT, Struck C, Hahn M, Mendgen K (2001) The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. *Proc Natl Acad Sci U S A* 98:8133-8138.
24. Slot JC, Hibbett DS (2007) Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLoS One* 10:e1097.
25. Aguileta G *et al.* (2008) Assessing the performance of single-copy genes for recovering robust phylogenies. *Syst Biol* 57:613-627.
26. Tamura K *et al.* (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.

Supporting Online Material

SI Text

Figs. S1 to S23

Tables S1 to S26

References

00 December 2010; accepted 00 00 0000

Table 1. Selection of *Melampsora larici-populina* genes strongly upregulated during poplar leaf infection (see Table S13 for complete list)

		Best blast hit		Express
Mlp ID	**Function**	**Pgt ID**	**GenBank Accession #**	**96 hpi**
89465	aspartic peptidase A1, secreted	PGTG_10570	XP_001881739	44063
94889	lipase, secreted	PGTG_15782	XP_749106	27318
123524	small secreted protein, U. fabae rust-transfered protein RTP homolog	PGTG_18022	ABS86408	49354
106755	small secreted protein, active site of glycosyl hydrolase 16, GH16	no hit	no hit	25530
88574	oligopeptide transporter, OPT	PGTG_17016	XP_001394363	38726
86448	transporter, AEC (Auxin Efflux Carrier) family	PGTG_06747	XP_759229	17984
112330	a-glycosidase related to a-mannosidases, secreted, glycosyl hydrolase 47, GH47	PGTG_09507	XP_001881296	14561
36184	amino acid permease, lysine-specific permease, U. fabae PIG2 homolog	PGTG_15547	XP_001873273	10319
95696	alanine amino-transferase	PGTG_07510	XP_001837651	11018
53832	thiazole biosynthetic enzyme, U. fabae THI4 homolog	PGTG_01304	Q9UVF8	52910
39287	small secreted protein, C. ribicola Cro r I homolog	no hit	AAF87492	7916
64764	small secreted protein, M. lini hesp-376 homolog	no hit	no hit	7596
89463	subtilisin protease, secreted, peptidase subfamily S8A	PGTG_18581	XP_001877576	18072
40379	sugar transporter HXT1, MFS	PGTG_15147	XP_001874568	12387
91040	b-glycosidase, endoglucanase, glycosyl hydrolase 5, GH5	PGTG_17056	XP_001875020	7212
124202	secreted protein, M. lini AvrM-B homolog	no hit	ABB96259	3764
67013	thiamin biosynthesis enzyme Thi1p (PIG1)	PGTG_10151	ABK96768	35825
48366	carotenoid ester lipase, secreted	PGTG_13346	XP_001875752	14890
40488	chitin deacetylase, carbohydrate esterase 4, CE4	PGTG_09635	XP_774611	3704
109896	secreted protein distantly related to plant expansins	PGTG_19856	XP_771894	4998
60884	glycosyltransferase, glycosyltransferase 18, GT18	PGTG_01151	XP_001884748	3889
87910	Oligopeptide Transporter, OPT	PGTG_15138	XP_001834544	12366
39227	Zinc transporter, CDF (Cation Diffusion Facilitator) family	PGTG_14264	CAE00445	3210
25498	chitin deacetylase, carbohydrate esterase 4, CE4	PGTG_09635	XP_774611	4541
55212	small secreted protein, M. lini hesp-735 homolog	no hit	ABB96276	2221

Up-regulation in poplar infected leaves is assessed by comparing transcripts profiles to those from resting urediniospores (USp). Poplar leaves were infected by *M. larici-populina* urediniospores and left for 96 hpi Under controlled conditions. At this stage, poplar rust pathogen has formed many haustoria in planta and sporulation has not yet occured. Expression values are the means of three biological replicates for 96 hpi and USp. Based on statistical analysis of normalized fluorescence levels, a gene was considered significantly regulated if it met two criteria: (1) t-test Pvalue, 0.05 (ArrayStar, DNAStar); infected poplar leaves at 96 hpi versus urediniospores fold-change > 10. Genes were selected on the basis of homology to a function, and hypothetical proteins or genes without homology of unknown function (exception of small secreted proteins representing candidate rust pathogen effectors) were discarded. The complete list of signficantly regulated genes is detailed in Table S13. Genes presented here are a selection of in planta highly up-regulated rust genes falling in the small secreted proteins,

lipases, proteases, transporters, carbohydrate active enzymes, metabolism-related genes categories. A grey cell indicates an expression value below background expression level.

Table 2. Selection of *Puccinia graminis* f. sp*. tritici* genes strongly upregulated during wheat infection (see Table S17 for complete list)

		Best blast hit		Expression levels		Wheat/USp	
Pgt ID	**Function**	**Mlp ID**	**GenBank Accession #**	**Wheat**	**USp**	**FC**	***P*-value**
PGTG_12502	Amino acid perméase	113062	no hit	31670	68	467.2	0.004
PGTG_15174	differentiation-related protein Infp	no hit	AAD38996	23002	50	466.3	0.002
PGTG_07532	Amino acid perméase	113062	no hit	13666	47	293.8	0.005
PGTG_07938	invertase 1 precursor	44167	CAG26671	18901	70	271	3.63 E-04
PGTG_17720	Zinc finger, C2H2 type	no hit	no hit	31604	175	180.9	0.004
PGTG_16569	Multicopper oxidase, secreted	112024	BAG50320	18825	114	166.6	0.012
PGTG_15026	lipase, putative	96073	XP_001273241	21088	229	92.4	1.22 E-06
PGTG_10570	Aspartyl protease, secreted	89871	no hit	3493	46	76.1	0.04
PGTG_05667	Cu/Zn superoxide dismutase, secreted	73483	XP_002418001	10257	138	74.7	0.004
PGTG_11683	Major intrinsic protein	106246	no hit	8738	118	74.6	4.73 E-04
PGTG_19191	Serine carboxypeptidase, secreted	49959	EEY14780	6156	86	71.8	0.017
PGTG_11725	endo-1,4-beta-glucanase, secreted	47207	AAR29981	6503	100	65.3	0.038
PGTG_08842	Thiamine monophosphate synthase/TENI	63716	no hit	7343	117	63.1	7.47 E-04
PGTG_10915	Major intrinsic protein	89561	no hit	41747	686	61	0.006
PGTG_05491	MFS sugar transporter, putative	86594	XP_002480590	28494	478	59.8	0.006
PGTG_15162	endo-beta-mannanase	86044	ABR27262	6992	123	57.3	0.009
PGTG_02527	Chitin synthase N-terminal	73345	ABB70409	33954	766	44.4	8.40 E-04
PGTG_06309	plasma membrane (H+) ATPase	44104	CAA05841	10443	272	38.5	0.003
PGTG_01889	Lipase, secreted	91294	no hit	13249	348	38.2	9.55 E-04
PGTG_15889	aspartic peptidase A1	34644	XP_001880663	4880	128	38.2	0.019
PGTG_15122	chitinase	75188	CAQ51152	15175	415	36.6	2.63 E-04
PGTG_12200	MFS monocarboxylate transporter	86626	XP_001267950	1636	49	34	0.012
PGTG_15888	aspartic peptidase A1	34644	XP_001880663	2159	77	28.2	0.021
PGTG_18584	HXT1p	38418	CAC41332	8629	378	22.9	0.006

Up-regulation in infected wheat is assessed by comparing transcripts profiles to those from resting urediniospores (USp). Wheat leaves were infected by *P. graminis* f.sp. *tritici* urediniospores and left for 8 days post-inoculation under controled conditions. At this stage, wheat rust pathogen has started to sporulate and macroscopic flecking are visible. Expression values are the means of three biological replicates for 8 dpi and USp. Based on statistical analysis of normalized fluorescence levels, a gene was considered significantly regulated if it met two criteria: (1) t-test *P*-value, 0.05 (using mattes in Matlab); infected wheat at 8 dpi versus Usp fold-change > 10. Genes were selected on the basis of homology to a function, and hypothetical proteins or genes without homology of unknown function (exception of small secreted proteins representing candidate rust pathogen effectors) were discarded. The complete list of signficantly regulated genes is detailed in Table S17. Genes presented here are a selection of in planta highly up-regulated rust genes falling in the small secreted proteins, lipases, proteases, transporters, carbohydrate active enzymes, metabolism-related genes categories. A grey cell indicates an expression value below background expression level.

Legends of Figures

Fig. 1. Predicted pattern of gene families gain and loss in representative fungal genomes. The figure represents the total number of protein families in each species or node estimated by Dollo parsimony principle. The numerals on the phylogenetic tree branches show numbers of expanded (left, black), contracted (right, red) or inferred ancestral (oval) protein families along lineages by comparison to the putative pan-proteome. For each species, the number of gene families, orphan genes and the total non-TE gene number are indicated on the right.

Fig. 2. Double clustering of the carbohydrate-cleaving families (*19*) from representative fungal genomes. Top tree: the fungi named are *Aspergillus nidulans* (A_nidu), *Aspergillus niger* (A_nige), *Aspergillus oryzae* (A_oryz), *Cryptococcus neoformans* (C_neof), *Gibberella zeae* (G_zeae), *Hypocrea jecorina* (H_jeco), *Laccaria bicolor* (L_bico), *Magnaporthe grisea* (M_gris), *Malassezia globosa* (M_glob), *Melampsora larici-populina* (M_lari), *Nectria haematococca* (N_haem), *Neurospora crassa* (N_cras), *Penicillium chrysogenum* (Pe_chr), *Phanerochaete chrysosporium* (Ph_chr), *Podospora anserina* (P_anse), *Postia placenta* (P_plac), *Puccinia graminis* f. sp. *tritici* (P_gram), *Saccharomyces cerevisiae* (S_cere), *Schizosaccharomyces pombe* (S_pomb), *Tuber melanosporum* (T_mela) and *Ustilago maydis* (U_mayd). Left tree: the enzyme families are represented by their class (GH, glycoside hydrolase; PL, polysaccharide lyase) and family number according to the carbohydrate-active enzyme database (*27*). Right side: known substrate of CAZy families (most common forms in brackets): BPG, bacterial peptidoglycan; CW, cell wall; ESR, energy storage and recovery; FCW, fungal cell wall; PCW, plant cell wall; PG,protein glycosylation; U, undetermined; α-gluc, α-glucans (including starch/glycogen); α-man, α-mannan, β-glyc, β-glycans; β-1,3-gluc, β-1,3-glucan; cell, cellulose; chit, chitin/chitosan; dext, dextran; hemi, hemicelluloses; inul, inulin; N-glyc, N-glycans; N-/O-glyc, N- / O-glycans; pect, pectin; sucr, sucrose; and treh, trehalose. Abundance of the different enzymes within a family is represented by a colour scale from 0 (white) to 33 occurrences (red) per species.

Fig. 3. Structure of the nitrate assimilation cluster among Basidiomycetes. Phylogram based on the MS277 and MS456 genes (25) from eight Basidomycete fungi was obtained using the minimum evolution method implemented in MEGA4 (26), with the complete deletion option for handling alignment gaps, and with the Poisson correction model for distance computation. Bootstrap tests were conducted using 1000 replicates. Branch lengths (drawn in the horizontal dimension only) are proportional to phylogenetic distances. Description of open reading frames coding nitrate/nitrite porter (green), nitrate reductase (purple) and nitrite reductase (red) is given. Numbers indicate the start and stop codons for each ORF. Grey rectangles indicate proteins that are not functionally related to nitrate assimilation.

Figure 1

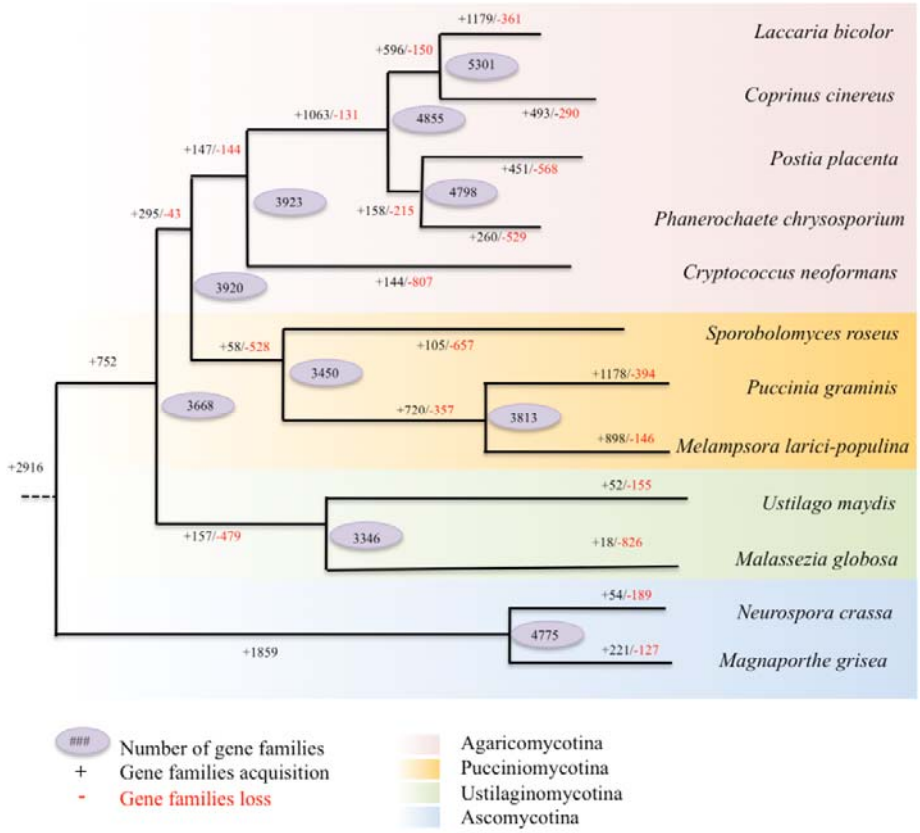

+1179/-361
+596/-150
5301
Laccaria bicolor
Coprinus cinereus
+1063/-131
4855
+493/-290
+147/-144
Postia placenta
+451/-568
4798
3923
+158/-215
+295/-43
Phanerochaete chrysosporium
+260/-529
Cryptococcus neoformans
3920
+144/-807
Sporobolomyces roseus
+105/-657
+58/-528
+752
3450
Puccinia graminis
+1178/-394
3668
+720/-357
3813
+898/-146
Melampsora larici-populina
+2916
+52/-155
Ustilago maydis
+157/-479
3346
+18/-826
Malassezia globosa
+54/-189
Neurospora crassa
4775
+1859
+221/-127
Magnaporthe grisea
Number of gene families
+ Gene families acquisition
- Gene families loss
Agaricomycotina
Pucciniomycotina
Ustilaginomycotina
Ascomycotina
Gene families
Orphans
Genes
6,119 4,184 18,983
5,504 3,437 13,367
4,681 1,507 12,399
4,529 1,806 9,963
3,260 2,168 7,146
2,898 1,387 5,529
4,597 5,906 15,680
4,565 3,644 14,527
3,243 2,087 6,507
2,538 941 4,283
4,640 3,630 9,818
4,869 4,672 12,829

Figure 2

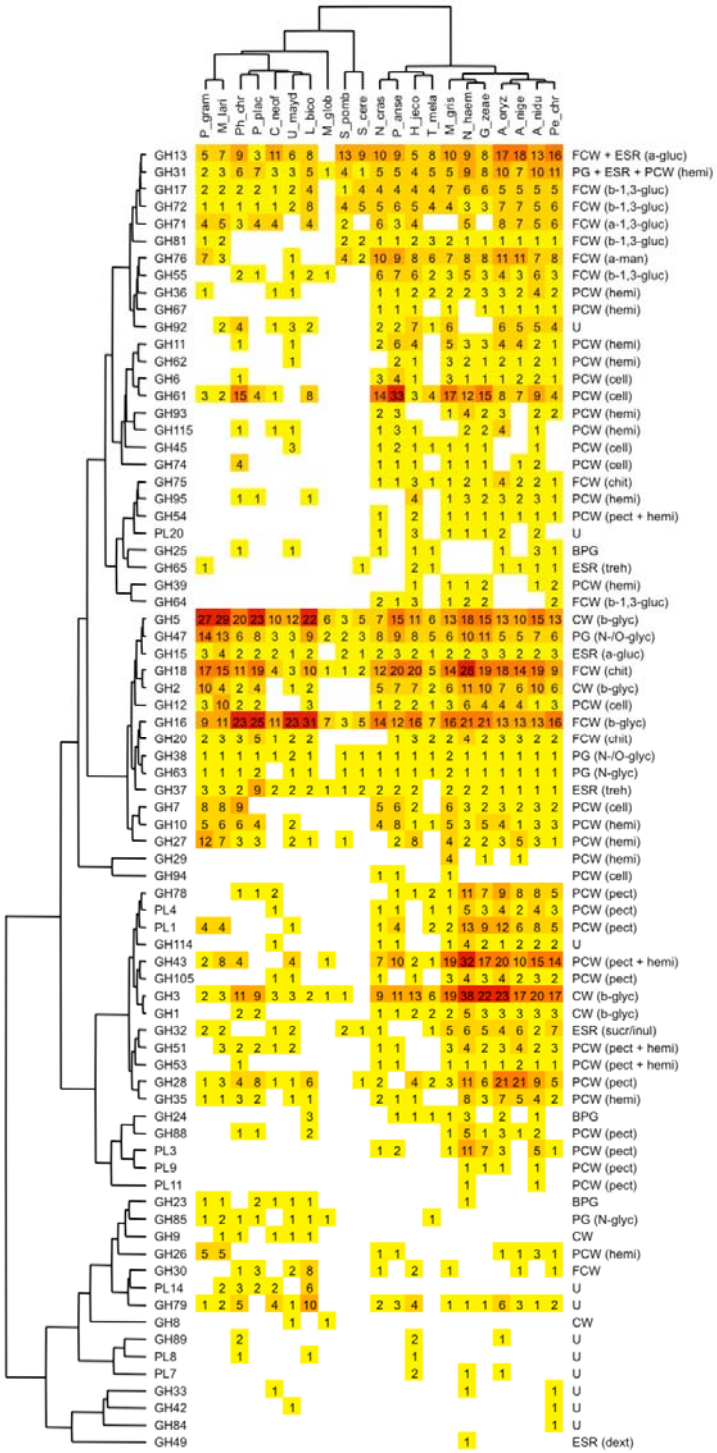
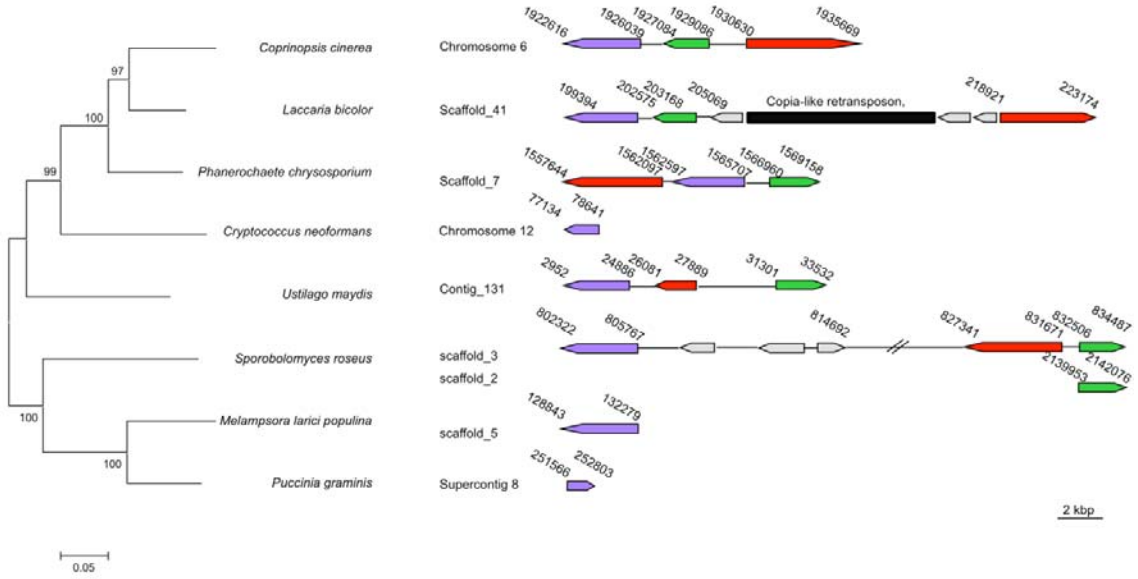

P_gram
M_tart
Ph_chr
P_plac
C_neof
U_mayd
L_bico
M_glob
S_pomb
S_cere
N_cras
P_anse
H_jeco
T_mela
M_gris
N_haem
G_zeae
A_oryz
A_nige
A_nidu
Pe_chr
GH13 FCW + ESR (a-gluc)
GH31 PG + ESR + PCW (hemi)
GH17 FCW (b-1,3-gluc)
GH72 FCW (b-1,3-gluc)
GH71 FCW (a-1,3-gluc)
GH81 FCW (b-1,3-gluc)
GH76 FCW (a-man)
GH55 FCW (b-1,3-gluc)
GH36 PCW (hemi)
GH67 PCW (hemi)
GH92 U
GH11 PCW (hemi)
GH62 PCW (hemi)
GH6 PCW (cell)
GH61 PCW (cell)
GH93 PCW (hemi)
GH115 PCW (hemi)
GH45 PCW (cell)
GH74 PCW (cell)
GH75 FCW (chit)
GH95 PCW (hemi)
GH54 PCW (pect + hemi)
PL20 U
GH25 BPG
GH65 ESR (treh)
GH39 PCW (hemi)
GH64 FCW (b-1,3-gluc)
GH5 CW (b-glyc)
GH47 PG (N-/O-glyc)
GH15 ESR (a-gluc)
GH18 FCW (chit)
GH2 CW (b-glyc)
GH12 PCW (cell)
GH16 FCW (b-glyc)
GH20 FCW (chit)
GH38 PG (N-/O-glyc)
GH63 PG (N-glyc)
GH37 ESR (treh)
GH7 PCW (cell)
GH10 PCW (hemi)
GH27 PCW (hemi)
GH29 PCW (hemi)
GH94 PCW (cell)
GH78 PCW (pect)
PL4 PCW (pect)
PL1 PCW (pect)
GH114 U
GH43 PCW (pect + hemi)
GH105 PCW (pect)
GH3 CW (b-glyc)
GH1 CW (b-glyc)
GH32 ESR (sucr/inul)
GH51 PCW (pect + hemi)
GH53 PCW (pect + hemi)
GH28 PCW (pect)
GH35 PCW (hemi)
GH24 BPG
GH88 PCW (pect)
PL3 PCW (pect)
PL9 PCW (pect)
PL11 PCW (pect)
GH23 BPG
GH85 PG (N-glyc)
GH9 CW
GH26 PCW (hemi)
GH30 FCW
PL14 U
GH79 U
GH8 CW
GH89 U
PL8 U
PL7 U
GH33 U
GH42 U
GH84 U
GH49 ESR (dext)

Figure 3


Coprinopsis cinerea
Laccaria bicolor
Phanerochaete chrysosporium
Cryptococcus neoformans
Ustilago maydis
Sporobolomyces roseus
Melampsora larici populina
Puccinia graminis
97
100
99
100
100
0.05
Chromosome 6
1922616
1926039
1927084
1929086
1930630
1935669
Scaffold_41
199394
202575
203168
205069
Copia-like retransposon,
218921
223174
Scaffold_7
1557644
1562097
1562597
1565707
1566960
1569158
Chromosome 12
77134
78641
Contig_131
2952
24886
26081
27889
31301
33532
scaffold_3
scaffold_2
802322
805767
814692
827341
831671
832506
834487
2139953
2142076
scaffold_5
128843
132279
Supercontig 8
251566
252803
2 kbp