Transcription factor Xpp1 is a switch between primary and secondary fungal metabolism

Christian Derntl[a], Bernhard Kluger[b], Christoph Bueschl[b], Rainer Schuhmacher[b], Robert L. Mach[a], and Astrid R. Mach-Aigner[a,1]

[a]Research Area Biochemical Technology, Institute of Chemical Engineering, TU Wien, A-1060 Vienna, Austria; and [b]Center for Analytical Chemistry, Department of Agrobiotechnology, University of Natural Resources and Life Sciences, A-3430 Tulln, Austria

Fungi can produce a wide range of chemical compounds via secondary metabolism. These compounds are of major interest because of their (potential) application in medicine and biotechnology and as a potential source for new therapeutic agents and drug leads. However, under laboratory conditions, most secondary metabolism genes remain silent. This circumstance is an obstacle for the production of known metabolites and the discovery of new secondary metabolites. In this study, we describe the dual role of the transcription factor Xylanase promoter binding protein 1 (Xpp1) in the regulation of both primary and secondary metabolism of *Trichoderma reesei*. Xpp1 was previously described as a repressor of xylanases. Here, we provide data from an RNA-sequencing analysis suggesting that Xpp1 is an activator of primary metabolism. This finding is supported by our results from a Biolog assay determining the carbon source assimilation behavior of an *xpp1* deletion strain. Furthermore, the role of Xpp1 as a repressor of secondary metabolism is shown by gene expression analyses of polyketide synthases and the determination of the secondary metabolites of *xpp1* deletion and overexpression strains using an untargeted metabolomics approach. The deletion of Xpp1 resulted in the enhanced secretion of secondary metabolites in terms of diversity and quantity. Homologs of Xpp1 are found among a broad range of fungi, including the biocontrol agent *Trichoderma atroviride*, the plant pathogens *Fusarium graminearum* and *Colletotrichum graminicola*, the model organism *Neurospora crassa*, the human pathogen *Sporothrix schenckii*, and the ergot fungus *Claviceps purpurea*.

transcription factor | secondary metabolism | low molecular compounds | fungi | gene regulation

Fungi are prominent producers of a broad variety of so-called secondary metabolites (1, 2). They are highly variable in structure and effects but share the following common feature: they are produced via the secondary metabolism (3, 4). Whereas primary metabolites are shared between all living cells, secondary metabolites are highly diverse and frequently produced by a limited number of species or cell types. Fungi use secondary metabolites for different purposes [e.g., protection against predation (5) or harsh environments (6), communication (7), competition and toxicity against bacteria (8) and other fungi (9), and pathogenicity (10)]. Some fungal secondary metabolites have toxic characteristics, such as aflatoxins, fumonisins, trichothecenes, fusarins, zearalenone, and ergot alkaloids (11–16), whereas other secondary metabolites are of major interest because of their potential application in the treatment of infectious diseases [e.g., antibiotics (17)] or cancer [e.g., immunosuppressants (18)] and as a potential source for novel therapeutic agents and drug leads (19). Interestingly, some of these natural products have already been used by ancient human populations (20). Numerous new compounds have been identified within the last decade that are now applied by the biotechnology and pharmaceutical industry (4). However, a vast number of compounds still awaits discovery (1, 21–23). A hindrance to the discovery of new products is the fact that a majority of secondary metabolite biosynthesis genes remain silent under standard laboratory conditions (1, 21). A strategy to overcome this problem is the exploitation of pleiotropic regulators of the secondary metabolism.

One well-studied example is the nuclear protein LaeA, which was originally described as a regulator of secondary metabolism in *Aspergillus* sp. (24). In additional studies, orthologs of LaeA were also shown to globally regulate secondary metabolism in other fungi (25–28). Other than regulation on the transcriptional level, a class of 4-phosphopantetheinyl transferases was found to be necessary for the activation of polyketide synthases (PKSs) and nonribosomal peptide synthases (NRPSs) in eukaryotes (29) on the posttranslational level. PKSs and NRPSs are large multidomain enzymes responsible for the synthesis of polyketides [e.g., norsolorinic acid, an intermediate in the biosynthesis pathway of aflatoxins (30)] or nonribosomal peptides [e.g., L-ergopeptam, an intermediate in the biosynthesis pathway of peptide ergot alkaloids (31)]. In this study, we describe the identification of a transcription factor that acts as a switch between primary and secondary metabolism. The deletion of the previously described xylanase repressor Xylanase promoter binding protein 1 (Xpp1) (32, 33) results in a decline of the primary metabolism and the up-regulation of secondary metabolism in terms of compound diversity and quantity in the saprotrophic ascomycete *Trichoderma reesei* [teleomorph; *Hypocrea jecorina* (34)]. To study the extent of the Xpp1 regulon in detail, we performed RNA-sequencing (RNA-Seq) analysis. Global carbon source assimilation analysis was used to understand its impact on the primary metabolism. To study the biological relevance of Xpp1 on the formation of secondary metabolites, the relative transcript levels of PKS-encoding genes were determined, and the metabolite profiles of a *T. reesei xpp1* deletion strain and an *xpp1* overexpression strain were compared with their parent strains using a stable isotopic labeling-assisted untargeted metabolomics approach. Based

Significance

Fungi produce a vast number of different chemical compounds via secondary metabolism. These compounds are of great interest because of their potential applicability in medicine, pharmacy, and biotechnology. In addition, a number of such compounds are toxins that potentially represent severe threats to human and animal health. However, under standard cultivation conditions, fungal secondary metabolism remains largely inactive. Here, we show that the deletion of the regulator Xylanase promoter binding protein 1 (Xpp1) results in the production of significantly more secondary metabolites in terms of both number and concentration. Because homologs of Xpp1 exist in fungi with numerous bioactive secondary metabolites, our results can lead to the discovery of secondary metabolites.

on the biological relevance of Xpp1 for secondary metabolite synthesis, its presence in other fungi was investigated in silico, and conserved motifs were identified.

Results

The Absence of Xpp1 Leads to the Earlier and Enhanced Secretion of Sorbicillin-Related Yellow Pigments.

The industrially used, cellulase-producing fungus *T. reesei* secretes a typical yellow pigment on a range of carbon sources (35–37). The pigment is a mixture of different sorbicillin derivatives, of which biosynthesis results from a gene cluster that is present in a range of not closely related ascomycetes (38, 39). However, in a previous study, we cultivated an *xpp1* deletion strain and its parent strain on carboxymethylcellulose (CMC) to determine the expression levels of cellulases (33). Throughout growth, we observed that cultures of the *xpp1* deletion strain turned yellow earlier and that coloring was more pronounced compared with the parent strain. For this study, we cultivated these strains together with an *xpp1* overexpression strain (33) on CMC, lactose, and D-glucose media. The absence of Xpp1 enhances the secretion of the yellow pigment on all three carbon sources (Fig. 1 *A–C*). In contrast, the overexpression of *xpp1* reduced the accumulation of yellow pigment compared with the parent strain on lactose and D-glucose (Fig. 1 *B* and *C*). On CMC, the parent strain itself barely produced any yellow pigment (Fig. 1*A*). Notably, on D-glucose, the secretion of the yellow pigment was delayed in the overexpression strain compared with the other two strains (Fig. 1*C*). We also observed these differences on D-glucose plates (Fig. 1*D*).

To learn whether Xpp1 regulates the expression of the sorbicillin cluster genes, we compared the expression of Yellow pigment regulator 1 (Ypr1), the main regulator of the cluster (39), in the *xpp1* deletion strain with its parent strain. The elevated *ypr1* transcript levels in the *xpp1* deletion strain (Fig. 1*E*) point toward an indirect regulation of the cluster by Xpp1 via regulation of Ypr1. Notably, the putative Xpp1 binding motif (an indirect AGAA repeat overlapping with the palindrome TCTAGA) is present in the *ypr1* upstream regulatory region (at −1,394 to −1,382). Because the yellow pigments are products of PKS activity and therefore, secondary metabolism, we sought to investigate the potential regulatory role of Xpp1 on secondary metabolism.

Xpp1 Affects the Expression of Genes of Primary and Secondary Metabolism.

To gain insight into how Xpp1 influences the fungal transcriptome, we performed RNA-Seq analysis. For this purpose, the *xpp1* deletion and parent strains were grown for 48 h on CMC, on which the two strains grow similarly (33). The processing of individual samples (Fig. S1*A*) was equally successful (50,209,455–57,345,532 reads without a significant difference between the two strains). The sequences of the reads were mapped to the reference genome of *T. reesei* (40) (genome.jgi.doe.gov/Trire2/Trire2.home.html) with coverage of 93.1–93.7%. In total, 9,129 unique transcripts were detected. A clustering of the samples based on the respective number of the unique reads showed a clear difference between the two strains (Fig. S1*B*). Next, differential gene expression analysis was performed. Genes were considered to be differentially expressed between the two strains when the average reads of the corresponding transcripts differed with an adjusted P value < 0.01 (41). We found 995 differentially expressed genes (DEGs) comparing the *xpp1* deletion and the parent strains. The number of up- and down-regulated genes and the extent of different
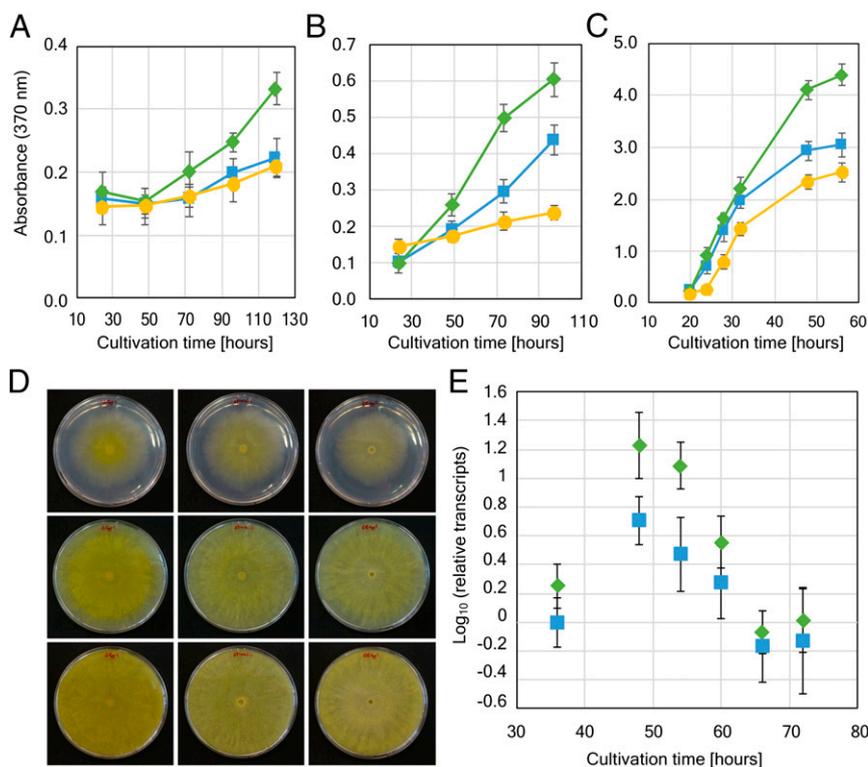
Fig. 1. Xpp1 influences the secretion of sorbicillin-related yellow pigments in *T. reesei*. The *xpp1* deletion strain (green diamonds), the *xpp1* overexpression strain (yellow circles), and the parent strain (blue squares) were grown in (*A*) CMC, (*B*) lactose, and (*C*) D-glucose. Samples of the supernatant were taken at indicated time points, and absorbance at 370 nm was photometrically measured. Error bars indicate SDs from three independently grown cultures. (*D, Left*) The *xpp1* deletion, (*D, Right*) the *xpp1* overexpression, and (*D, Center*) the parent strains were grown on D-glucose. Pictures were taken at (*D, Top*) 48, (*D, Middle*) 72, and (*D, Bottom*) 96 h. (*E*) The *xpp1* deletion (green diamonds) and the parent (blue squares) strains were grown in CMC. Samples were taken at 36, 48, 52, 60, 66, and 72 h. Relative transcript levels of *ypr1* were determined by qPCR, normalized by using the reference genes *sar1* and *act*, and correlated to the reference sample (parent strain, 36 h). Error bars indicate SDs from three independently grown cultures.

expression were similar—490 genes were up-regulated, and 505 genes down-regulated, with medians of the $\log_2$ fold changes of 1.478 and −1.272, respectively.

Next, we categorized the DEG according to their eukaryotic clusters of orthologous groups (KOGs) assignment as published on genome.jgi.doe.gov/Trire2/Trire2.home.html. KOGs represent the basic functional groups of genes. Because not all genes could be assigned to a KOG, the overview provided in Table S1 contains a total number of 669 DEGs. To estimate which functional group of genes is most prevalently influenced by Xpp1, we tested which KOG has an overrepresented number of DEG. First, the expected portion of genes was calculated as the percentage of the genes assigned to a specific KOG of the total number of all KOG-assigned genes (i.e., 6,836 according to genome.jgi.doe.gov/Trire2/Trire2.home.html). Analogously, the obtained portion of DEGs was calculated as the percentages of DEGs assigned to a specific KOG of the total number of KOG-assigned DEGs (i.e., 669 genes). Second, we compared the obtained percentages with the expected percentages for each KOG. We found a higher proportion of DEGs than expected in the KOG "metabolism" (Fig. 2 and Table S1). Precisely, a higher portion of DEGs than expected was found in six of its nine KOG classes, including the KOG class "secondary metabolites biosynthesis, transport, and catabolism" (Fig. 2 and Table S1). The other five classes could be considered to represent the primary metabolism. A detailed analysis of the DEGs of each class drew our attention particularly to the KOG class "carbohydrate transport and metabolism." Nearly all up-regulated DEGs within this class were genes encoding transporters (Table S2). The other DEGs in this class (encoding for, e.g., hydrolases and enzymes involved in glycolysis and the TCA) were down-regulated, suggesting a suppression of the primary metabolism.

Xpp1 Supports Fungal Growth. Because the KOG overrepresentation analysis pointed toward an influence of Xpp1 on primary metabolism, we were interested in its effect on the carbon assimilation behavior of *T. reesei*. Therefore, we compared the *xpp1* deletion strain with its parent strain in a Biolog assay. Only those carbon sources on which both strains grew better than on water (control) were included in the analysis. Throughout growth, the *xpp1* deletion strain accumulated equal or less biomass than the parent strain on all tested carbon sources (Fig. 3A). We performed a Wilcoxon test on the data pairs, namely the mean values of the two strains for each carbon source, and found them to be significantly different (P value < 0.001) for both tested time points (48 and 72 h). Next, we plotted the data pairs in scatter graphs for both time points and calculated the respective trend lines (Fig. 3 B and C). The statistical significance of the differences of the data pairs and the high R^2 values of the trend lines indicate that the impaired growth behavior of the *xpp1* deletion strain is an inherent property of the strain. Accordingly, Xpp1 seems to promote growth in *T. reesei*. This observed phenotype is concordant with the RNA-Seq results, which had already suggested an attenuation of primary metabolism in the absence of Xpp1.

Xpp1 Regulates the Gene Expression of PKSs. The KOG overrepresentation analysis based on the RNA-Seq data pointed to a regulatory influence of Xpp1 on secondary metabolism. However, the difference between the percentage of obtained DEGs (4.48%) and the expected percentage (3.83%) in this KOG class was relatively small (Fig. 2 and Table S1). In the course of the detailed analysis of the DEGs, we realized that the genes encoding PKS were assigned to the KOG class "lipid transport and metabolism." Consequently, the calculated difference between the obtained and expected percentages for the secondary metabolism class would have been higher if the PKS-encoding genes were assigned to the secondary metabolite class. Surprisingly, the genes encoding the two PKSs from the sorbicillin cluster (39) (protein IDs 73621 and 73618) were not included in the list of DEGs because of their P values. Table 1 separately lists the results obtained from our RNA-Seq analysis for all PKS-encoding genes. We decided to investigate the potential regulatory role of Xpp1 on the expression of PKS-encoding genes in
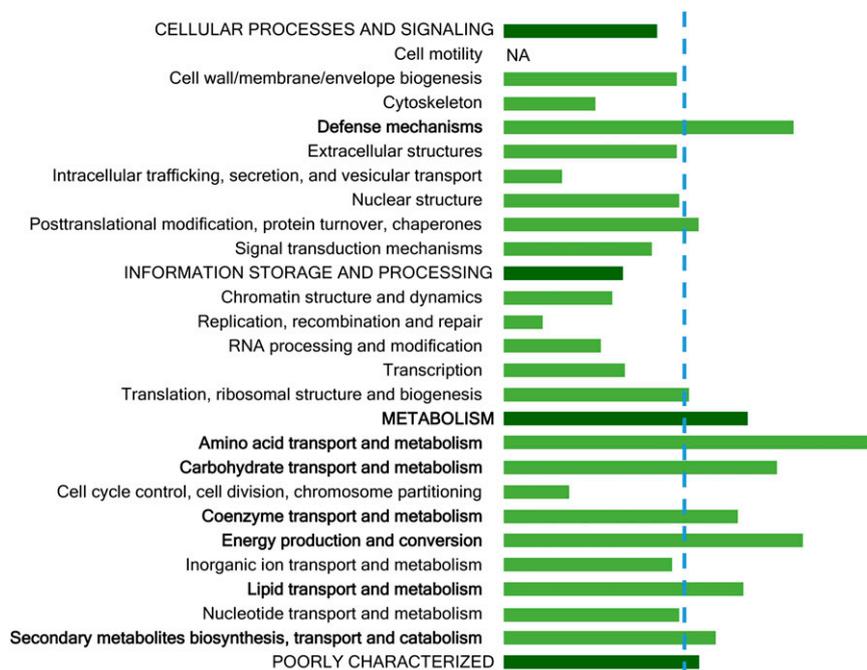
Fig. 2. KOG overrepresentation analysis The *xpp1* deletion and the parent strains were grown on CMC for 48 h and subjected to an RNA-Seq analysis. The obtained DEGs were assigned to their KOGs. The obtained number of DEGs was normalized to the expected number of DEGs (blue dashed line; according to the normal distribution of genes) for each KOG (capitalized letters and dark green bars) and KOG class (light green bars). Bold indicates a KOG or KOG class with a higher number of DEGs than could be expected. NA, not applicable.

Fig. 3. Influence of Xpp1 on the growth behavior of *T. reesei*. The *xpp1* deletion (green bars) and the parent (blue bars) strains were subjected to a Biolog assay and grown at 30 °C in darkness. Absorbance at 750 nm was measured as an indicator for biomass accumulation. (*A*) Absorbance was measured after 72 h on a representative selection of carbon sources. The selection accounts for all types of carbon sources used in the assay and covers the entire value range. Error bars indicate SDs from three independently performed cultivations in one assay. Mean values of both strains after (*B*) 48 and (*C*) 72 h were blotted against each other (*x* axis, parent strain; *y* axis, *xpp1* deletion strain), and linear trend lines were inserted with a forced intercept of (0,0). Equations of the trend lines and corresponding coefficients of determination are shown.

more detail. To this end, we measured the transcript levels of the most strongly regulated PKS-encoding genes (as indicated in Table 1) by quantitative PCR (qPCR) analysis in the parent and the *xpp1* deletion strain grown on CMC. With this method, we found the strongest differences for the genes encoding the two PKSs from the sorbicillin cluster, namely protein IDs 73621 and 73618 (Fig. 4 *A* and *B*). This result matches the elevated transcript levels of the gene encoding their main regulator, *ypr1* (Fig. 1*E*). No difference could be detected for the protein ID 65116 (Fig. 4*C*). The results for the protein IDs 65172, 60118, and 81964 were unclear (Fig. 4 *D*–*F*). The expression of protein ID 65172 was up-regulated earlier and thereafter, also down-regulated earlier in the *xpp1* deletion strain (Fig. 4*D*). Protein ID 60118 generally appeared to be increased in expression at later time points (Fig. 4*E*). The expression of protein ID 81964 followed an oscillating expression pattern, without clear differences in overall expression strength between

the two strains (Fig. 4*F*). However, these results prompted us to investigate the influence of Xpp1 on the phenotype of *T. reesei* in the context of the secretion of secondary metabolites.

Absence of Xpp1 Enhances the Secretion of Low-Molecular Weight Compounds. The KOG overrepresentation analysis and the transcript levels of some PKS-encoding genes pointed toward a regulatory influence of Xpp1 on secondary metabolism. Therefore, we sought to investigate to what extent deletion of Xpp1 results in the secretion of low-molecular weight compounds (LMCs) that could be used as indicators for secondary metabolism. We previously established an untargeted metabolomics workflow based on liquid chromatography–high-resolution MS (LC-HRMS) for the preferential detection of secondary metabolites in the supernatants of fungal cultures. The stable isotope labeling approach is based on parallel cultivation on native and highly ^{13}C-enriched, uniformly

Table 1. Influence of Xpp1 on expression of PKSs in *T. reesei*

Protein ID	Clade	Mean reads	Log$_2$ (fold change)	Adjusted *P* value
65172*	Reducing clade I: lovastatin/citrinin diketide	28.1	2.06	0.00028
65891	Reducing clade I: lovastatin/citrinin diketide	169.5	−0.34	0.33321
105804	Nonreducing fungal clade III	4.0	0.18	0.79373
82208†	Nonreducing fungal clade I	877.5	0.77	0.03328
59482	Reducing clade III: t-toxin	484.5	0.79	0.04114
60118*	Reducing clade I: lovastatin/citrinin diketide	1,429.5	−1.37	1.1E-09
65116*	Reducing clade IV: fumonisins	30.3	−2.22	6.6E-06
81964*	Nonreducing fungal clade I–II	445.5	−3.02	1.3E-12
106272	Reducing clade I: lovastatin/citrinin diketide	451.1	−0.37	0.14051
73621‡	Nonreducing fungal clade III	6,491.1	1.98	NA
73618§	Reducing clade I: lovastatin/citrinin diketide	6,966.4	2.24	NA

NA, not applicable.
*PKSs that are DEGs according to the RNA-Seq analysis.
†Termed PKS4; essential for conidial pigmentation (35).
‡Homolog of *P. chrysogenum* SorB, which is involved in sorbicillin biosynthesis (38, 39).
§Homolog of *P. chrysogenum* SorA, which is essential for sorbicillin biosynthesis (38, 39).

labeled D-glucose (U-^{13}C$_6$ D-glucose) as sole carbon source in a small-scale setup (42). In a preexperiment, the *T. reesei* strains were grown under these modified cultivation conditions. Whereas the strains grew slower than under generally applied conditions (compare Fig. S2*A* with ref. 33), the reduced growth of the *xpp1*

deletion compared with the WT strain also was observed here (Fig. S2*A*). Again, we observed more pronounced secretion of the yellow pigment in the *xpp1* deletion strain in contrast to reduced amounts of pigment in the supernatant of the *xpp1* overexpression strain (Fig. S2*B*). The expression levels of the gene coding for PKS
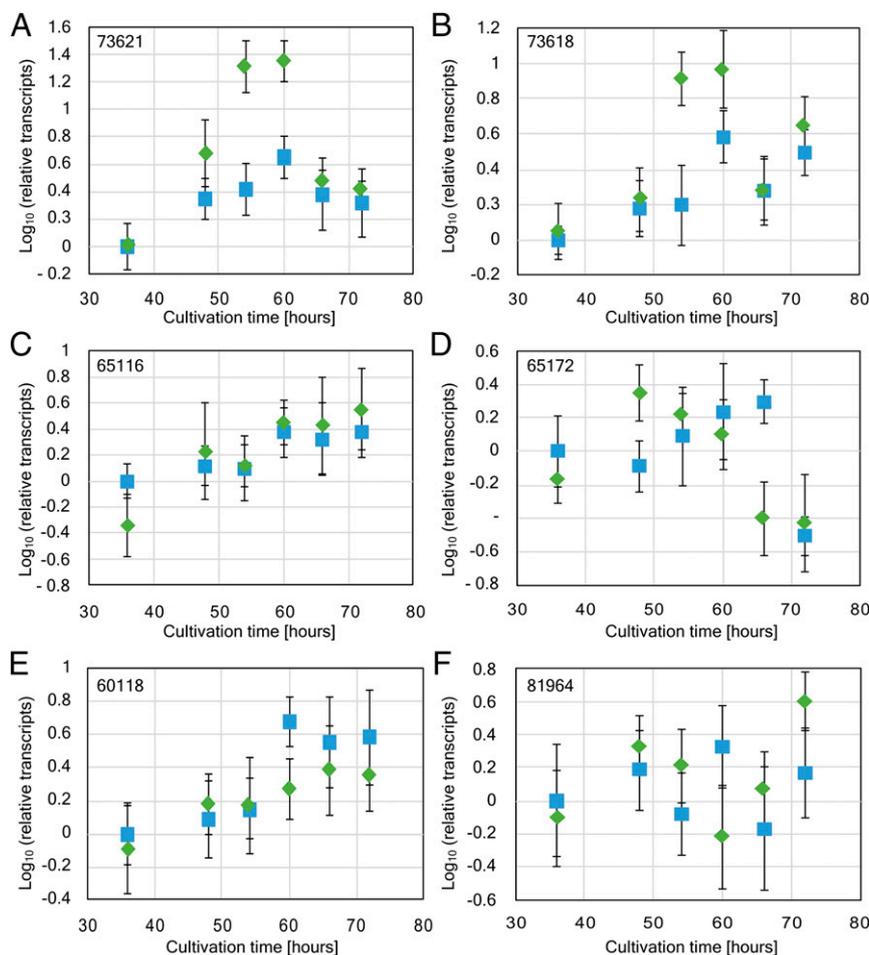

Fig. 4. Xpp1 influences the expression of PKSs in *T. reesei*. The *xpp1* deletion (green diamonds) and the parent (blue squares) strains were grown in CMC. Samples were taken at 36, 48, 52, 60, 66, and 72 h. Relative transcript levels of the genes encoding PKSs with the protein IDs (*A*) 73621, (*B*) 73618, (*C*) 65116, (*D*) 65172, (*E*) 60118, and (*F*) 81964 were determined by qPCR, normalized by using the reference genes *sar1* and *act*, and correlated to the reference sample (parent strain, 36 h). Error bars indicate SDs from three independently grown cultures.

73621 from the sorbicillin cluster were increased in the *xpp1* deletion compared with the parent strain (Fig. S2*C*) as previously observed on CMC (Fig. 4*A*). Interestingly, the difference in expression between the *xpp1* deletion and the parent strains for the second PKS-encoding gene (protein ID 65172) was more pronounced under these conditions than on CMC (compare Fig. 4*D* with Fig. S2*D*). However, we did not observe any differences between the *xpp1*

overexpression and parent strain, except for a clear delay of the expression peak of protein ID 73621 (Fig. S2 *C* and *D*). In summary, we conclude that the small-scale setup and the modified medium were suitable for our purposes.

For metabolite profiling, the *xpp1* deletion, *xpp1* overexpression, and the parent strains were grown in parallel on native and U-^{13}C$_6$–labeled D-glucose. The constituents of culture
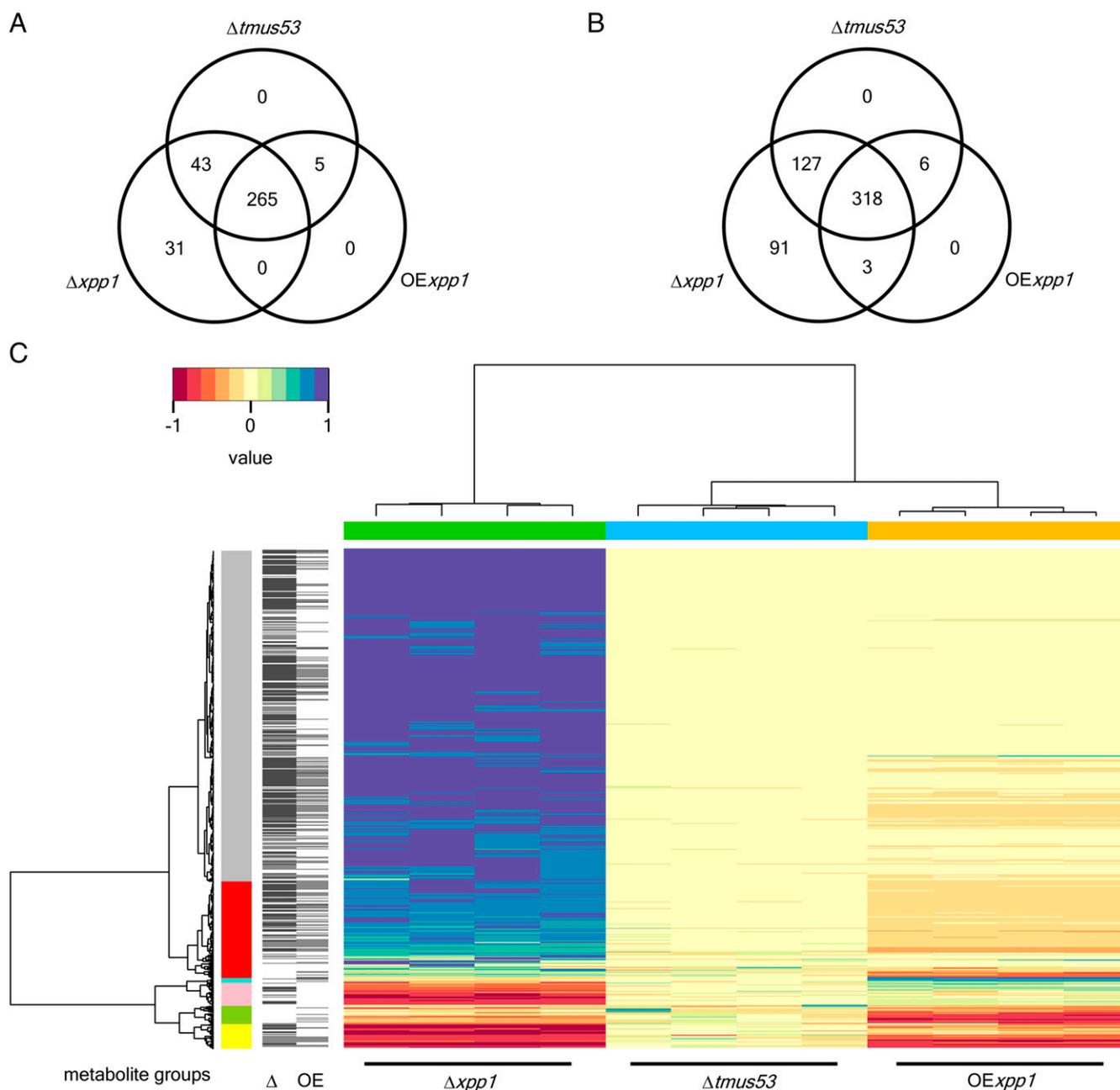

Fig. 5. Xpp1 influences the secretion of LMCs in *T. reesei*. The *xpp1* deletion (Δ*xpp1*), the *xpp1* overexpression (OE*xpp1*), and the parent (Δ*tmus53*) strains were grown in modified minimal medium containing only D-glucose (unlabeled or U-^{13}C$_6$ labeled) as the sole carbon source in four biological replicates. Secreted compounds of three strains were analyzed using a UHPLC system coupled to an LTQ Orbitrap XL. The Venn diagrams depict the numbers of different compounds secreted by three strains at (*A*) 72 and (*B*) 96 h. (*C*) The results obtained for the 96-h cultures were subjected to a hierarchical analysis and are represented in the heat map normalized to the levels of the parent strain Δ*tmus53*. Six different metabolite groups based on the relative abundances in three strains are indicated: gray, increase in metabolite levels in Δ*xpp1*; red, increase in metabolite levels in Δ*xpp1* and decrease in OE*xpp1*; turquoise, increase in metabolite levels in OE*xpp1*; pink, decrease in metabolite levels in Δ*xpp1*; green, decrease in metabolite levels in OE*xpp1*; and yellow, metabolites that tended to be less abundant in both mutant strains (Δ*xpp1* and OE*xpp1*) compared with the parent strain. The results of *t* tests between Δ*tmus53*/Δ*xpp1* (Δ) and Δ*tmus53*/OE*xpp1* (OE) are indicated next to the clustering (gray: significant difference, *P* value < 0.05, fold change > 2; dark gray: highly significant difference, *P* value < 0.00009, fold change > 2).

supernatants were then separated on a reverse-phase C18 HPLC column and recorded by high-resolution MS in full-scan mode. Subsequently, the resulting LC-HRMS raw data were evaluated using the latest version of the in house-developed MetExtract algorithm (42, 43). Screening for corresponding mass peaks of native and uniformly ^{13}C-labeled metabolites facilitated the detection of only true *T. reesei* strain-derived metabolites. We detected total numbers of 344 and 545 metabolites at 72 and 96 h, respectively. According to the chosen chromatographic conditions (i.e., reverse-phase C18) and observed retention behavior of the detected metabolites (~90% of the compounds exhibited retention times >10 min), the chosen approach predominantly captures secondary fungal metabolites (44). In contrast, LMCs of the central metabolism, such as non-aromatic amino acids, sugars, sugar phosphates, small alcohols, or organic acids, are not retained on the used C18 HPLC column and therefore, are not contained among the measured fungal metabolites. Accordingly, a series of the detected LMCs, including sorbicillin and sorbicillinol, was annotated to secondary metabolites by searching m/z values and corresponding numbers of carbon atoms per metabolite ion against the Antibase (45) database for matching entries of the genus *Trichoderma* (*SI Text*). As shown in the Venn diagrams (Fig. 5 *A* and *B*), a large number of the detected compounds was exclusively present in the supernatants of the *xpp1* deletion strain at both time points (i.e., 31 after 72 h and 91 after 96 h). In contrast, many compounds were not detected in the *xpp1* overexpression strain (i.e., 43 and 127, respectively) (Fig. 5 *A* and *B*). Next, we performed hierarchical cluster analysis and heat map analysis of the results obtained at 96 h (Fig. 5*C*). Our results indicated a clear separation of three strains, and the metabolite dendrogram in the heat map yielded six distinct metabolite clusters. Increased amounts of almost all compounds were detected in the *xpp1* deletion strain, and smaller amounts were in the *xpp1* overexpression compared with the parent strain (Fig. 5*C*). In addition, univariate *t* tests of the *xpp1* deletion and the parent strain were performed for all 545 detected metabolites. In total, 168 metabolites exhibited significant differences (*P* value ≤ 0.05, fold change ≥ 2), and 320 metabolites exhibited highly significant differences (*P* value ≤ 0.00009, fold change ≥ 2). We conclude that Xpp1 represses the secretion of LMCs in terms of number and concentration levels. Comparing the number of carbon atoms per metabolite of the detected LMCs, we observe a higher relative abundance of large compounds in the *xpp1* deletion strain (Fig. S3). Because large compounds are more likely secondary metabolites, these findings strongly support that Xpp1 is a regulator of secondary metabolism.

Homologs of Xpp1 Are Found in a Broad Range of Ascomycetes. Based on these promising findings, we were interested whether homologs of Xpp1 are found in other fungi. To this end, we performed a BLAST analysis and found homologs in a broad range of ascomycetes. For the identification of conserved domains and motifs, we performed a conserved domain search and a multiple alignment of all homologs obtained from the BLAST analysis with a reasonable similarity (i.e., chosen cutoff of total scores above 175 bits) and sufficient sequence information (Fig. 6, black). Within the sequences of all aligned Xpp1 homologs, six highly conserved motifs were found (Fig. S4 *A*, Roman numbers and *B*, double underlined). Additionally, a glycine- and proline-rich stretch (Fig. S4*B*, waved underlined) is present in most homologs directly after the conserved motif III—although notably, not very pronounced in *Trichoderma* sp. Furthermore, two semi-conserved motifs (not conserved throughout all tested sequences) were identified in the multiple alignment (Fig. S4 *A*, Arabic numbers and *B*, single underlined). *T. reesei* Xpp1 contains a basic helix–loop–helix domain at R396–K462 (32) (Fig. S4*B*). The motifs IV and V compose the DNA binding domain (Fig. S4*B*). Notably, the motifs I, 1, II, 2, and III might constitute a conserved domain. In the light of the results for Xpp1 of *T. reesei*, the elucidation of the role of its homologs is warranted.

Discussion

In a recent publication, we could link the production of the typical yellow pigment to a secondary metabolism gene cluster containing the two PKSs 73618 and 73621. The gene cluster is also present in a series of not closely related ascomycetes, including *Penicillium chrysogenum*. Both studies, the former on *P. chrysogenum* and our recent study on *T. reesei*, showed that this cluster is responsible for the production of sorbicillin. Notably, here we found again a number of metabolites in the supernatant of *T. reesei*, two of which are annotated as sorbicillin and sorbicillinol (*SI Text*).

In the KOG overrepresentation analysis, only one class that does not belong to the KOG metabolism exhibited an overrepresented number of DEGs, namely the class "defense mechanisms" (Fig. 2 and Table S1). This result is caused by the fact that specific DEGs identified during this study (encoding the proteins with IDs 119805, 123475, 123976, 109748, and 105342)—next to a series of other genes—are annotated as "von Willebrand factor and related coagulation proteins" (according to genome.jgi.doe. gov/Trire2/Trire2.home.html). We could not confirm this annotation regarding the mentioned DEGs by performing manual in silico analyses (conserved domain search and BLAST analysis). Without these DEGs, the KOG class defense mechanisms would not have exhibited an overrepresented number of DEGs. Therefore, we did not include defense mechanisms in the subsequent design of the study and the experiments.

The results obtained from RNA-Seq analysis, Biolog assay, and LMC screening indicate that Xpp1 is both a positive regulator of primary metabolism and a repressor of secondary metabolism. It seems to be involved in a central switch mechanism. To test whether Xpp1 might act on secondary metabolism indirectly (e.g., via known regulators of secondary metabolism), we examined the results from the RNA-Seq analysis for Lae1 and the two 4-phosphopantetheinyl transferases, protein IDs 56081 and 48788. They were not differentially expressed in the *xpp1* deletion strain, with log$_2$ fold changes of 0.265, −0.097, and 0.626 and adjusted *P* values of 0.539, 0.880, and 0.120, respectively.

As mentioned in the Introduction, Xpp1 was originally identified and described as a repressor of xylanase expression (32, 33). Therein, Xpp1 was reported to regulate the expression of xylanolytic enzymes at late time points during cultivation on xylan. In this study, we showed a regulatory influence of Xpp1 on the expression of a total of 995 genes involved in primary metabolism and secondary metabolism. This finding shifts the role of Xpp1 from a narrow-range regulator to a putative wide-domain regulator. This assumption is supported by the fact that 28 transcription factors are regulated by Xpp1 according to our RNA-Seq analysis. The latter finding suggests that Xpp1 influences the expression of a large number of genes (also) by regulating a set of narrow-range regulators. This type of action was, for example, shown in the case of the sorbicillin cluster genes, which are indirectly controlled by Xpp1 via regulation of *ypr1* expression.

Previously, the expression of Xpp1 itself was shown to depend on growth rate (33). Taken together, we propose that the general role of Xpp1 is to direct the flow of material and energy toward the accumulation of biomass. Xpp1 might facilitate this by (*i*) delivering a feedback signal on the expression of nutrient-degrading enzymes when sufficient levels have been reached (33) and (*ii*) negatively controlling secondary metabolism and balancing it with growth. In this regard, it needs to be considered that Xpp1 might be involved in the regulation of additional biological processes.

Materials and Methods

Fungal Strains and Cultivation Conditions. *T. reesei* QM6aΔ*tmus53* (ATCC 13631), the *xpp1* deletion QM6aΔ*tmus53*Δ*xpp1* (QM6aΔ*xpp1*) (33), and the *xpp1* overexpression (QM6aOE*xpp1*) (33) strains were maintained on malt extract (MEX) agar at 30 °C. Hygromycin B was added when applicable to a final concentration of 113 U/mL.

Fig. 6. Phylogenetic tree of Xpp1 (unrooted). Relative similarities of Xpp1 homologs in the given organisms were calculated using the COBALT for multiple protein sequences and visualized in the cladogram, which was generated using the Fast Minimum Evolution algorithm based on the Grishin distance model. Homologs of Xpp1 in the organisms written in gray were omitted in the motif search, because their sequences either were only partially available (depicted by asterisks) or differed too strongly from their own close homologs.

For growth comparison on plates, *T. reesei* was pregrown on Mandels–Andreotti (MA) medium (46) plates containing 1% glycerol at 30 °C. Equal pieces of overgrown agar were used to inoculate MA medium plates containing 50 mM D-glucose. Plates were incubated at 30 °C in darkness. Pictures were taken from the bottom of the plates.

For cultivation on CMC, *T. reesei* was grown in 60 mL MA medium containing 1% CMC (Carl Roth GmbH) in 1-L Erlenmeyer flasks stationary at 30 °C. For cultivation on D-glucose and lactose, *T. reesei* was grown in MA medium containing 50 mM D-glucose or 27.5 mM lactose at 30 °C on a rotary shaker at 180 rpm. Mycelia and supernatants were separated by filtration through Miracloth (EMD Millipore). Mycelia were stored in liquid nitrogen.

For small-scale cultivation, *T. reesei* was grown in 24-well plates stationary at 30 °C in 1.5 mL modified MA medium containing either 50 mM native D-glucose or U-^{13}C$_6$–labeled D-glucose with a degree of enrichment of 99% (Euriso-Top GmbH). The phosphate citrate buffer was replaced by a 0.1 M sodium phosphate buffer, and the medium was sterilized by filtration.

Supernatants were quenched with 30% (vol/vol) acetonitrile and centrifuged at 20,000 × g for 10 min at 4 °C.

Transcript Analysis by RNA-Seq. Illumina RNA sequencing and differential gene expression analysis were performed by Microsynth using their standardized analysis pipeline. RNA was isolated using the RNeasy Plant Mini Kit (Qiagen), and the quality of the RNA samples was controlled using an Agilent 2100 Bioanalyzer. RNA libraries were prepared using a TruSeq Stranded mRNA Sample Prep Kit including poly(A) enrichment (Illumina) and quantified with an Illumina Library Quantification Kit (KAPA Biosystems). The pooled libraries were sequenced on a NextSeq500 instrument (Illumina) with read lengths of 75 nt. For bioinformatic analysis, the sequence reads were mapped onto the reference genome of *T. reesei* (genome.jgi.doe.gov/Trire2/Trire2.home.html) using the TopHat and Bowtie 2 software (47, 48) before reads were counted using HTSeq (49). The counted transcripts were normalized and subjected to statistical analysis through the DESeq2 software package (50). The mapped reads were provided as bam files and visualized using the Integrative Genomics Viewer (software.broadinstitute.org/software/igv/). Both strains were analyzed in three independent biological replicates.

KOG-Related Analysis of DEG. The DEGs obtained from the RNA-Seq analysis were assigned to a KOG. To determine which KOGs are more prevalently affected by the deletion of *xpp1*, the obtained number of DEGs assigned to a certain KOG was compared with the annotated number of genes assigned to this KOG. The obtained portion of DEGs (obtained percentage) was calculated as the percentage of the DEGs of a certain KOG of the total number of DEGs that could be assigned to a KOG (i.e., 669 genes). Analogously, the expected portion of DEGs (expected percentage) was calculated as the percentage of the genes of a certain KOG of the total number of all KOG-assigned genes (i.e., 6,836; according to genome.jgi.doe.gov/Trire2/Trire2.home.html). For graphical visualization, the obtained percentage was normalized to the expected percentage.

Biolog Microarray Technique. The global carbon assimilation profiles were evaluated by using Biolog FF MicroPlate (Biolog, Inc.) following a previously described protocol (51), with minor modifications as follows: the inoculum was prepared from cultures on MEX plates incubated at 30 °C; mycelial growth was measured after 18, 24, 30, 36, 42, 48, 66, 72, and 90 h using biological triplicates.

Transcript Analysis by qPCR. Between 0.01 and 0.03 g harvested mycelia were homogenized in 1 mL peqGOLD TriFast DNA/RNA/Protein Purification System Reagent (PEQLAB Biotechnologie) using a FastPrep FP120 BIO101 Thermo-Savant Cell Disrupter (Qbiogene). RNA was isolated according to the manufacturer's protocols, and the concentrations were measured using the NanoDrop 1000 (Thermo Scientific). Synthesis of cDNA from mRNA was performed using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer's protocols.

qPCRs were performed in a Mastercycler Ep Realplex 2.2 System (Eppendorf). All reactions were performed in triplicate. The amplification mixture (final volume of 25 μL) contained 12.5 μL 2× iQ SYBR Green Mix (Bio-Rad Laboratories), 100 nM forward and reverse primers, and 2.5 μL cDNA (diluted 1:100). Primer sequences are provided in Table S3. Cycling conditions and control reactions were performed as previously described (52). Calculations using *sar1* and *act* as reference genes were performed as previously published (52).

Screening for LMCs. Equal volumes of all quenched U-^{13}C–labeled supernatants (three strains in four biological replicates each) were pooled together. Quenched unlabeled supernatants (three strains in four biological replicates each) were individually mixed at a ratio of 1:1 (vol/vol) with the U-^{13}C–labeled pool. All samples were analyzed as described previously (53) using an ultra-high-performance liquid chromatography (UHPLC) System (Accela; Thermo Fisher Scientific) coupled to an LTQ Orbitrap XL (Thermo Fisher Scientific) with

an electrospray ionization interface in positive ionization (ESI) mode. A reversed-phase XBridge C18 150 × 2.1-mm i.d., 3.5-μm Particle Size Analytical Column (Waters) preceded by a C18 4 × 3-mm i.d. Security Cartridge (Phenomenex) were used with MeOH and water, both containing 0.1% formic acid as solvents. The flow rate was maintained at 250 μL/min using a linear gradient program. The chromatographic method held the initial mobile-phase composition (10% B) constant for 2 min followed by a linear gradient to 100% B within 30 min. This final condition was held for 5 min followed by 8 min of column reequilibration at 10% B. The ESI interface was operated with the following settings: sheath gas: 60 arbitrary units; auxiliary gas: 15 arbitrary units; sweep gas: 5 arbitrary units; capillary voltage: 4 kV; and capillary temperature: 300 °C. The orbitrap mass analyzer was operated in full-scan mode in a scan range from *m/z* 100 to 1,000, with a resolving power setting of 60,000 FWHM (at *m/z* 400).

Before data processing, LC-HRMS raw data were centroided and converted to the mzXML format with the Proteowizard Toolbox (54) (version 3.0.8789). Subsequently, the data files were processed with an updated version of MetExtract (43). In brief, each MS scan was inspected for the isotope patterns derived from both native and uniformly ^{13}C-labeled metabolite ions as described by Bueschl et al. (42). The minimum intensity threshold for mono-isotopic ^{12}C and uniformly ^{13}C-labeled derived MS signals was set to 5,000 in at least three scans, the maximum allowed *m/z* deviation of related *m/z* values was set to 2.5 ppm, and a maximum isotopolog abundance error of ±20% was used. Chromatographic peaks of MS signal pairs were detected in the extracted-ion chromatograms of the native and the U-^{13}C–labeled metabolite isotopologs with the algorithm by Du et al. (55). Two such matching chromatographic peaks representing a native and a U-^{13}C–labeled ion had to show a minimum correlation of 0.7. Different ions of a metabolite formed during ionization of the metabolite were convoluted using their similar retention time (±10 scans) and a minimum correlation coefficient of 0.85. All detected feature pairs were bracketed over the LC-HRMS files into a data matrix, which was used for statistical analysis.

For statistical analysis, the intensity ratios of the relative monoisotopic unlabeled and the fully ^{13}C-labeled feature abundances (metabolome-wide internal standardization) were used. For univariate significance testing, a global *P* value threshold of 0.05 was used, which was reduced to 0.00009411 after multiple testing correction with the Šidák method. For multivariate analysis, missing values were imputed by zero, and the abundances of the detected metabolites were range-scaled (56) and centered relative to the mean value of the WT replicates. For hierarchical cluster analysis and heat map analysis, squared Euclidean distance and ward linkage were used.

In Silico Analysis of Xpp1 and Its Orthologs. The amino acids sequence of *T. reesei* Xpp1 was used as the query for a BLAST analysis (57) at blast.ncbi.nlm.nih.gov/Blast.cgi using the blastp algorithm searching the nonredundant protein sequences database. Relative distance of the orthologous proteins with a total score above 175 bits was calculated using the Constraint-Based Multiple Protein Alignment Tool (COBALT) (58) at www.st-va.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi and visualized in a tree using the Fast Minimum Evolution algorithm (59) based on the Grishin distance model (60). Conserved motifs were determined manually based on the multiple alignment. Conserved domain searches were performed against the CDD v3.14 database (61) at www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.

1. Wiemann P, Keller NP (2014) Strategies for mining fungal natural products. *J Ind Microbiol Biotechnol* 41(2):301–313.
2. Bérdy J (2005) Bioactive microbial metabolites. *J Antibiot (Tokyo)* 58(1):1–26.
3. Brakhage AA (2013) Regulation of fungal secondary metabolism. *Nat Rev Microbiol* 11(1):21–32.
4. Hoffmeister D, Keller NP (2007) Natural products of filamentous fungi: Enzymes, genes, and their regulation. *Nat Prod Rep* 24(2):393–416.
5. Rohlfs M, Churchill AC (2011) Fungal secondary metabolites as modulators of interactions with insects and other arthropods. *Fungal Genet Biol* 48(1):23–34.
6. Eisenman HC, Casadevall A (2012) Synthesis and assembly of fungal melanin. *Appl Microbiol Biotechnol* 93(3):931–940.
7. Fu SF, et al. (2015) Indole-3-acetic acid: A widespread physiological code in interactions of fungi with other organisms. *Plant Signal Behav* 10(8):e1048052.
8. Netzker T, et al. (2015) Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Front Microbiol* 6:299.
9. Almassi F, Ghisalberti EL, Narbey MJ, Sivasithamparam K (1991) New antibiotics from strains of *Trichoderma harzianum*. *J Nat Prod* 54(2):396–402.
10. Scharf DH, Heinekamp T, Brakhage AA (2014) Human and plant fungal pathogens: The role of secondary metabolites. *PLoS Pathog* 10(1):e1003859.
11. Abrar M, et al. (2013) Aflatoxins: Biosynthesis, occurrence, toxicity, and remedies. *Crit Rev Food Sci Nutr* 53(8):862–874.
12. Nesic K, Ivanovic S, Nesic V (2014) Fusarial toxins: Secondary metabolites of *Fusarium* fungi. *Rev Environ Contam Toxicol* 228:101–120.
13. Streit E, et al. (2012) Current situation of mycotoxin contamination and co-occurrence in animal feed–focus on Europe. *Toxins (Basel)* 4(10):788–809.

14. Marin S, Ramos AJ, Cano-Sancho G, Sanchis V (2013) Mycotoxins: Occurrence, toxicology, and exposure assessment. *Food Chem Toxicol* 60:218–237.
15. Schardl CL, Panaccione DG, Tudzynski P (2006) Ergot alkaloids–biology and molecular biology. *Alkaloids Chem Biol* 63:45–86.
16. Ward TJ, Bielawski JP, Kistler HC, Sullivan E, O'Donnell K (2002) Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. *Proc Natl Acad Sci USA* 99(14):9278–9283.
17. Ozcengiz G, Demain AL (2013) Recent advances in the biosynthesis of penicillins, cephalosporins and clavams and its regulation. *Biotechnol Adv* 31(2):287–311.
18. Rüegger A, et al. (1976) [Cyclosporin A, a peptide metabolite from Trichoderma polysporum (Link ex Pers.) Rifai, with a remarkable immunosuppressive activity]. *Helv Chim Acta* 59(4):1075–1092.
19. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 75(3):311–335.
20. Bassett EJ, Keith MS, Armelagos GJ, Martin DL, Villanueva AR (1980) Tetracycline-labeled human bone from ancient Sudanese Nubia (A.D. 350). *Science* 209(4464):1532–1534.
21. Brakhage AA, Schroeckh V (2011) Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal Genet Biol* 48(1):15–22.
22. Chávez R, Fierro F, García-Rico RO, Vaca I (2015) Filamentous fungi from extreme environments as a promising source of novel bioactive secondary metabolites. *Front Microbiol* 6:903.
23. Andersen MR, et al. (2013) Accurate prediction of secondary metabolite gene clusters in filamentous fungi. *Proc Natl Acad Sci USA* 110(1):E99–E107.
24. Bok JW, Keller NP (2004) LaeA, a regulator of secondary metabolism in *Aspergillus* spp. *Eukaryot Cell* 3(2):527–535.
25. Butchko RA, Brown DW, Busman M, Tudzynski B, Wiemann P (2012) Lae1 regulates expression of multiple secondary metabolite gene clusters in Fusarium verticillioides. *Fungal Genet Biol* 49(8):602–612.
26. Karimi-Aghcheh R, et al. (2013) Functional analyses of *Trichoderma reesei* LAE1 reveal conserved and contrasting roles of this regulator. *G3 (Bethesda)* 3(2):369–378.
27. Bi Q, Wu D, Zhu X, Gillian Turgeon B (2013) Cochliobolus heterostrophus Llm1 - a Lae1-like methyltransferase regulates T-toxin production, virulence, and development. *Fungal Genet Biol* 51:21–33.
28. Kim HK, et al. (2013) Functional roles of FgLaeA in controlling secondary metabolism, sexual development, and virulence in *Fusarium graminearum*. *PLoS One* 8(7):e68441.
29. Lambalot RH, et al. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* 3(11):923–936.
30. Watanabe CM, Wilson D, Linz JE, Townsend CA (1996) Demonstration of the catalytic roles and evidence for the physical association of type I fatty acid synthases and a polyketide synthase in the biosynthesis of aflatoxin B1. *Chem Biol* 3(6):463–469.
31. Ortel I, Keller U (2009) Combinatorial assembly of simple and complex D-lysergic acid alkaloid peptide classes in the ergot fungus *Claviceps purpurea*. *J Biol Chem* 284(11):6650–6660.
32. Mach-Aigner AR, Grosstessner-Hain K, Poças-Fonseca MJ, Mechtler K, Mach RL (2010) From an electrophoretic mobility shift assay to isolated transcription factors: A fast genomic-proteomic approach. *BMC Genomics* 11:644.
33. Derntl C, Rassinger A, Srebotnik E, Mach RL, Mach-Aigner AR (2015) Xpp1 regulates the expression of xylanases, but not of cellulases in Trichoderma reesei. *Biotechnol Biofuels* 8:112.
34. Kuhls K, et al. (1996) Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. *Proc Natl Acad Sci USA* 93(15):7755–7760.
35. Atanasova L, Knox BP, Kubicek CP, Druzhinina IS, Baker SE (2013) The polyketide synthase gene *pks4* of *Trichoderma reesei* provides pigmentation and stress resistance. *Eukaryot Cell* 12(11):1499–1508.
36. Seiboth B, et al. (2012) The putative protein methyltransferase LAE1 controls cellulase gene expression in *Trichoderma reesei*. *Mol Microbiol* 84(6):1150–1164.
37. Nakari-Setälä T, Aro N, Kalkkinen N, Alatalo E, Penttilä M (1996) Genetic and biochemical characterization of the *Trichoderma reesei* hydrophobin HFBI. *Eur J Biochem* 235(1-2):248–255.

38. Salo O, et al. (2016) Identification of a polyketide synthase involved in sorbicillin biosynthesis by *Penicillium chrysogenum*. *Appl Environ Microbiol* 82(13):3971–3978.
39. Derntl C, Rassinger A, Srebotnik E, Mach RL, Mach-Aigner AR (2016) Identification of the Main Regulator Responsible for Synthesis of the Typical Yellow Pigment Produced by *Trichoderma reesei*. *Appl Environ Microbiol* 82(20):6247–6257.
40. Martinez D, et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol* 26(5):553–560.
41. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57(1):289–300.
42. Bueschl C, et al. (2014) A novel stable isotope labelling assisted workflow for improved untargeted LC-HRMS based metabolomics research. *Metabolomics* 10(4):754–769.
43. Bueschl C, et al. (2012) MetExtract: A new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research. *Bioinformatics* 28(5):736–738.
44. Kluger B, Lehner S, Schuhmacher R (2015) Metabolomics and secondary metabolite profiling of filamentous fungi. *Biosynthesis and Molecular Genetics of Fungal Secondary Metabolites*, eds Zeilinger S, Martín J-F, García-Estrada C (Springer, New York), Vol 2, pp 81–101.
45. Laatsch H (2012) *AntiBase 2012 Upgrade: The Natural Compound Identifier* (Wiley-VCH, Weinheim, Germany).
46. Mandels M (1985) Applications of cellulases. *Biochem Soc Trans* 13(2):414–416.
47. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
48. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
49. Anders S, Pyl PT, Huber W (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.
50. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
51. Druzhinina IS, Schmoll M, Seiboth B, Kubicek CP (2006) Global carbon utilization profiles of wild-type, mutant, and transformant strains of *Hypocrea jecorina*. *Appl Environ Microbiol* 72(3):2126–2133.
52. Steiger MG, Mach RL, Mach-Aigner AR (2010) An accurate normalization strategy for RT-qPCR in *Hypocrea jecorina* (*Trichoderma reesei*). *J Biotechnol* 145(1):30–37.
53. Kluger B, et al. (2013) Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat. *Anal Bioanal Chem* 405(15):5031–5036.
54. Kessner D, Chambers M, Burke R, Agus D, Mallick P (2008) ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics* 24(21):2534–2536.
55. Du P, Kibbe WA, Lin SM (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* 22(17):2059–2065.
56. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ (2006) Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics* 7:142.
57. Johnson M, et al. (2008) NCBI BLAST: A better web interface. *Nucleic Acids Res* 36(Web Server issue):W5–W9.
58. Papadopoulos JS, Agarwala R (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23(9):1073–1079.
59. Desper R, Gascuel O (2004) Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol Biol Evol* 21(3):587–598.
60. Grishin NV (1995) Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *J Mol Evol* 41(5):675–679.
61. Marchler-Bauer A, et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue):D222–D226.