



**HAL**  
open science

## Comparative Transcriptome Analysis Shows Conserved Metabolic Regulation during Production of Secondary Metabolites in Filamentous Fungi.

Jens Christian Nielsen, Sylvain Prigent, Sietske Grijseels, Mhairi Workman, Boyang Ji, Jens Nielsen

### ► To cite this version:

Jens Christian Nielsen, Sylvain Prigent, Sietske Grijseels, Mhairi Workman, Boyang Ji, et al.. Comparative Transcriptome Analysis Shows Conserved Metabolic Regulation during Production of Secondary Metabolites in Filamentous Fungi.. *mSystems*, 2019, 4 (2), pp.e00012-19. 10.1128/mSystems.00012-19. hal-02622105

**HAL Id: hal-02622105**

**<https://hal.inrae.fr/hal-02622105>**

Submitted on 26 May 2020

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

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



Distributed under a Creative Commons Attribution 4.0 International License



# Comparative Transcriptome Analysis Shows Conserved Metabolic Regulation during Production of Secondary Metabolites in Filamentous Fungi

Jens Christian Nielsen,<sup>a</sup> Sylvain Prigent,<sup>a\*</sup> Sietske Grijseels,<sup>b</sup> Mhairi Workman,<sup>b</sup> Boyang Ji,<sup>a</sup> Jens Nielsen<sup>a,c</sup>

<sup>a</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

<sup>b</sup>Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark

<sup>c</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

**ABSTRACT** Filamentous fungi possess great potential as sources of medicinal bioactive compounds, such as antibiotics, but efficient production is hampered by a limited understanding of how their metabolism is regulated. We investigated the metabolism of six secondary metabolite-producing fungi of the *Penicillium* genus during nutrient depletion in the stationary phase of batch fermentations and assessed conserved metabolic responses across species using genome-wide transcriptional profiling. A coexpression analysis revealed that expression of biosynthetic genes correlates with expression of genes associated with pathways responsible for the generation of precursor metabolites for secondary metabolism. Our results highlight the main metabolic routes for the supply of precursors for secondary metabolism and suggest that the regulation of fungal metabolism is tailored to meet the demands for secondary metabolite production. These findings can aid in identifying fungal species that are optimized for the production of specific secondary metabolites and in designing metabolic engineering strategies to develop high-yielding fungal cell factories for production of secondary metabolites.

**IMPORTANCE** Secondary metabolites are a major source of pharmaceuticals, especially antibiotics. However, the development of efficient processes of production of secondary metabolites has proved troublesome due to a limited understanding of the metabolic regulations governing secondary metabolism. By analyzing the conservation in gene expression across secondary metabolite-producing fungal species, we identified a metabolic signature that links primary and secondary metabolism and that demonstrates that fungal metabolism is tailored for the efficient production of secondary metabolites. The insight that we provide can be used to develop high-yielding fungal cell factories that are optimized for the production of specific secondary metabolites of pharmaceutical interest.

**KEYWORDS** comparative transcriptomics, cell factories, filamentous fungi, secondary metabolism

Filamentous fungi are economically and ecologically important microorganisms and serve diverse applications in industrial biotechnology. Some of the key industrial processes utilizing these organisms include the production of pharmaceuticals (1), bulk chemicals (2), and enzymes (3) and the manufacture of fermented food products (4). One important characteristic of many filamentous fungi is their ability to secrete bioactive compounds, called secondary metabolites, which are renowned for their pharmaceutical properties, e.g., as antibiotics, but also for their toxic effects, which are major health concerns when they contaminate food and feed (5). The fungal genus

**Citation** Nielsen JC, Prigent S, Grijseels S, Workman M, Ji B, Nielsen J. 2019. Comparative transcriptome analysis shows conserved metabolic regulation during production of secondary metabolites in filamentous fungi. *mSystems* 4:e00012-19. <https://doi.org/10.1128/mSystems.00012-19>.

**Editor** Matthew F. Traxler, University of California, Berkeley

**Copyright** © 2019 Nielsen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Jens Nielsen, [nielsenj@chalmers.se](mailto:nielsenj@chalmers.se).

\* Present address: Sylvain Prigent, UMR 1332 BFP, INRA, Université de Bordeaux, Villenave d'Ornon, France.

 Comparative transcriptome analysis links primary and secondary metabolism in filamentous fungi

**Received** 20 January 2019

**Accepted** 16 March 2019

**Published** 16 April 2019

*Penicillium* is especially well-known for the production of many secondary metabolites (6, 7).

The large-scale industrial production of secondary metabolites is one of the greatest success stories of industrial biotechnology. The yield and titers of penicillin have been improved by orders of magnitude through decades of random mutagenesis and selection of *Penicillium chrysogenum* mutants (8, 9). Conversely, only limited success has been achieved through the use of rational metabolic engineering strategies to improve secondary metabolite production in fungi (10, 11). A major explanation for this is the limited understanding of the metabolic processes that govern the production of secondary metabolites in filamentous fungi.

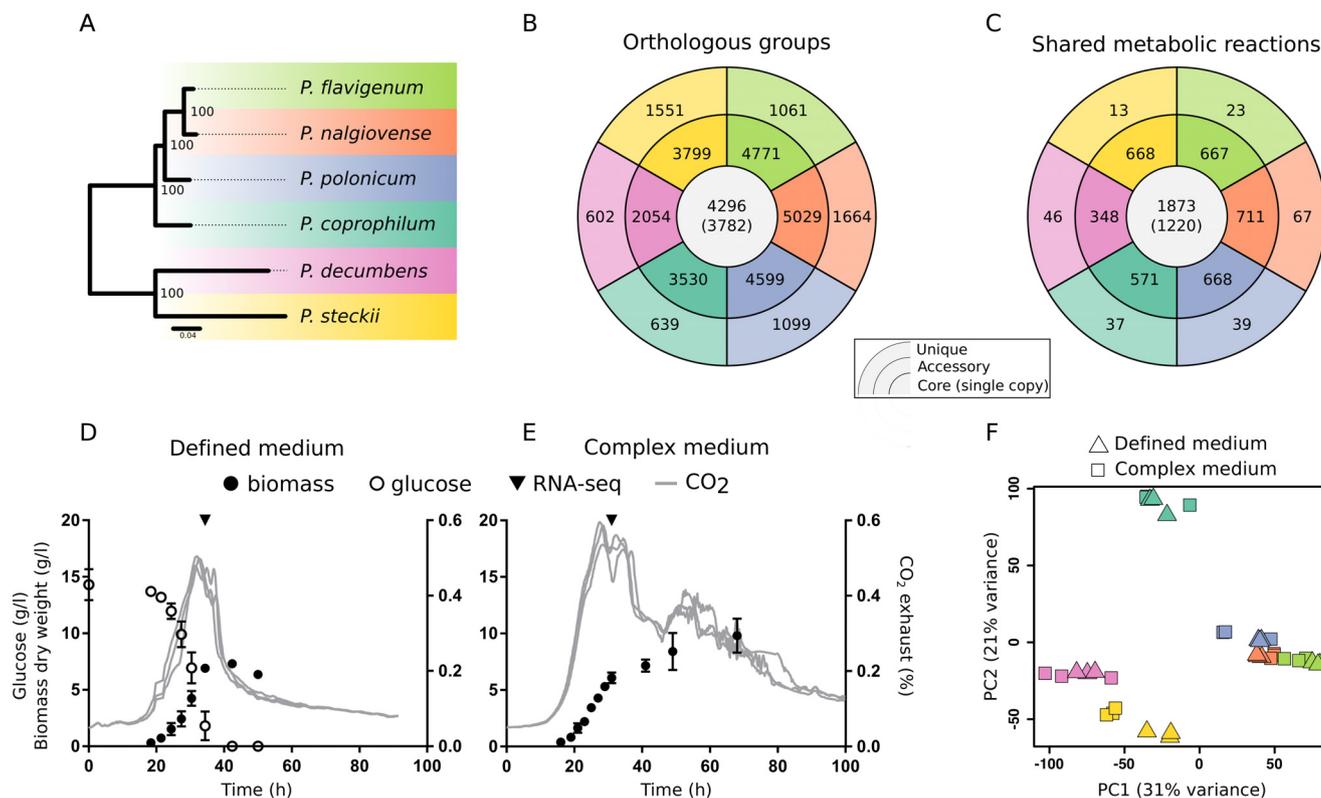
Secondary metabolism is strongly connected to primary metabolism, in the sense that precursors and cofactors for secondary metabolites are derived from processes in the central carbon metabolism. The two main classes of secondary metabolites in fungi are polyketides (PKs), which are derived from short-chain acyl coenzyme A (acyl-CoA) units and which are synthesized by polyketide synthases (PKS), and nonribosomal peptides (NRPs), which are derived from amino acids and which are synthesized by nonribosomal peptide synthetases (NRPSs). Additionally, secondary metabolite pathways use cofactors, such as ATP and reducing power in the form of NADPH derived from energy metabolism. The link between the generation of secondary metabolite precursors and secondary metabolites is, however, poorly understood. Since secondary metabolites often are induced under suboptimal growth conditions, e.g., after depletion of a nutrient source (12), precursor units might not be synthesized *de novo* through glycolytic catabolism of carbon sources but, rather, might be synthesized through the degradation of stored macromolecules. The degradation of fatty acids and branched-chain amino acids (BCAAs) has been suggested to contribute to the acetyl-CoA supply for certain PKs in *Aspergillus* species (13), but a more comprehensive overview is needed.

Here, we analyzed secondary metabolism at the transcriptional level during stationary phase of batch fermentation experiments of secondary metabolite-producing *Penicillium* species (7, 14). The aim was to elucidate the link between primary and secondary metabolism and to improve the understanding of how the metabolism of native secondary metabolite producers is wired for the efficient production of secondary metabolites. We conducted a comparative transcriptome analysis of six *Penicillium* species cultivated in a defined medium (DM) and a complex medium (CM) with the objective to identify orthologous protein groups representing conserved metabolic features in these species.

## RESULTS

**Experimental setup for comparative transcriptomics of *Penicillium*.** Six *Penicillium* species (*P. flavigenum*, *P. nalgiovense*, *P. polonicum*, *P. coprophilum*, *P. decumbens*, and *P. steckii*) whose genomes were recently sequenced were subject to a comparative transcriptome analysis in order to assess evolutionarily conserved patterns of expression, with a specific emphasis on secondary metabolite biosynthesis. The species were selected because they represent the phylogenetic diversity within the *Penicillium* genus (15) (Fig. 1A) and because studies have highlighted their prolific capabilities for the biosynthesis of secondary metabolites (6, 14), as well as a genomic potential for additional secondary metabolite production (7).

For all species, protein-coding genes were clustered into orthologous groups (OGs). Among the 4,296 genes representing the core genome, 3,782 genes were present only in a single copy in the genomes (Fig. 1B), and these formed the basis for a phylogenetic assessment (Fig. 1A). Our previously published genome-scale metabolic models of the *Penicillium* species (16) served as an annotation of metabolic genes as defined in the MetaCyc database (17). We found that 582 single-copy OGs were part of the core metabolism and that 1,220 metabolic reactions were present in all metabolic networks (Fig. 1C). These core reactions were significantly depleted for reactions involved in secondary metabolism (adjusted  $P = 5e-9$ , hypergeometric test), in particular, alkaloid



**FIG 1** Overview of study and comparative analysis of *Penicillium* species. (A) Phylogenetic tree of the species investigated in this study. The colors in the tree serve as a legend for the entire figure. (B and C) Numbers of genes (B) and reactions (C) within each of the species that are shared by all species (core), that are shared by a subset of species (accessory), or that are specific to one species (unique). (D and E) Example growth profile of *Penicillium flavigenum* in defined medium (D) and complex medium (E), indicating the time point of the RNA-seq sample (generated based on previously published data [14]). (F) Principal-component analysis of the six *Penicillium* species based on the normalized expression of single-copy core genes.

and terpenoid biosynthesis. Conversely, all other parts of metabolism (as defined in the MetaCyc database), except for inorganic nutrient metabolism, were significantly enriched in the core genome fraction (adjusted  $P < 0.05$ , hypergeometric test).

All species were cultivated in batch cultures in two different media: one defined medium (DM) for *Penicillium* containing glucose and ammonium and one industrially relevant complex medium (CM) based on yeast extract, sucrose, and nitrate. Since the aim was to investigate secondary metabolism, which is often induced under suboptimal growth conditions (12, 18), samples for transcriptome analysis were collected in the stationary phase several hours after  $\text{CO}_2$  production had peaked, indicating that the preferred carbon source had been depleted from the two carbon-limited media (Fig. 1D and E; see also Fig. S1 in the supplemental material).

To assess the global differences in gene expression between the species, we performed a principal-component analysis (PCA) based on single-copy core genes (Fig. 1F). The clustering of the samples in the PCA largely reflected the phylogeny of the species (Fig. 1A), indicating that the regulation of the core genes is evolutionarily related across species, in agreement with previous observations in different yeast species (19). In contrast, the medium composition had a minor impact on the clustering of the samples. Since batch effects could influence the comparison of expression levels across species, we tested the robustness of the PCA by evaluating multiple different strategies for normalization of the data, but in all cases, similar clustering patterns were observed (Fig. S2).

**Transcriptional landscape of *Penicillium*.** Differentially expressed genes (DEGs) in CM relative to DM were identified in each species (adjusted  $P < 0.05$ ). A great variation in the number of DEGs per species was observed, ranging from 327 upregulated genes and 331 downregulated genes in *P. decumbens* to 2,363 upregulated genes and 1,885



low-affinity ammonium transporter which has proven to be upregulated under nitrogen starvation and induced by nitrate (21). No shared DEGs were downregulated in all species.

To further investigate which functions were affected across species, a gene set analysis was conducted, and significantly enriched gene sets were defined as either up- or downregulated. This allowed us to identify processes affected across the species (Fig. 2B; Table S1).

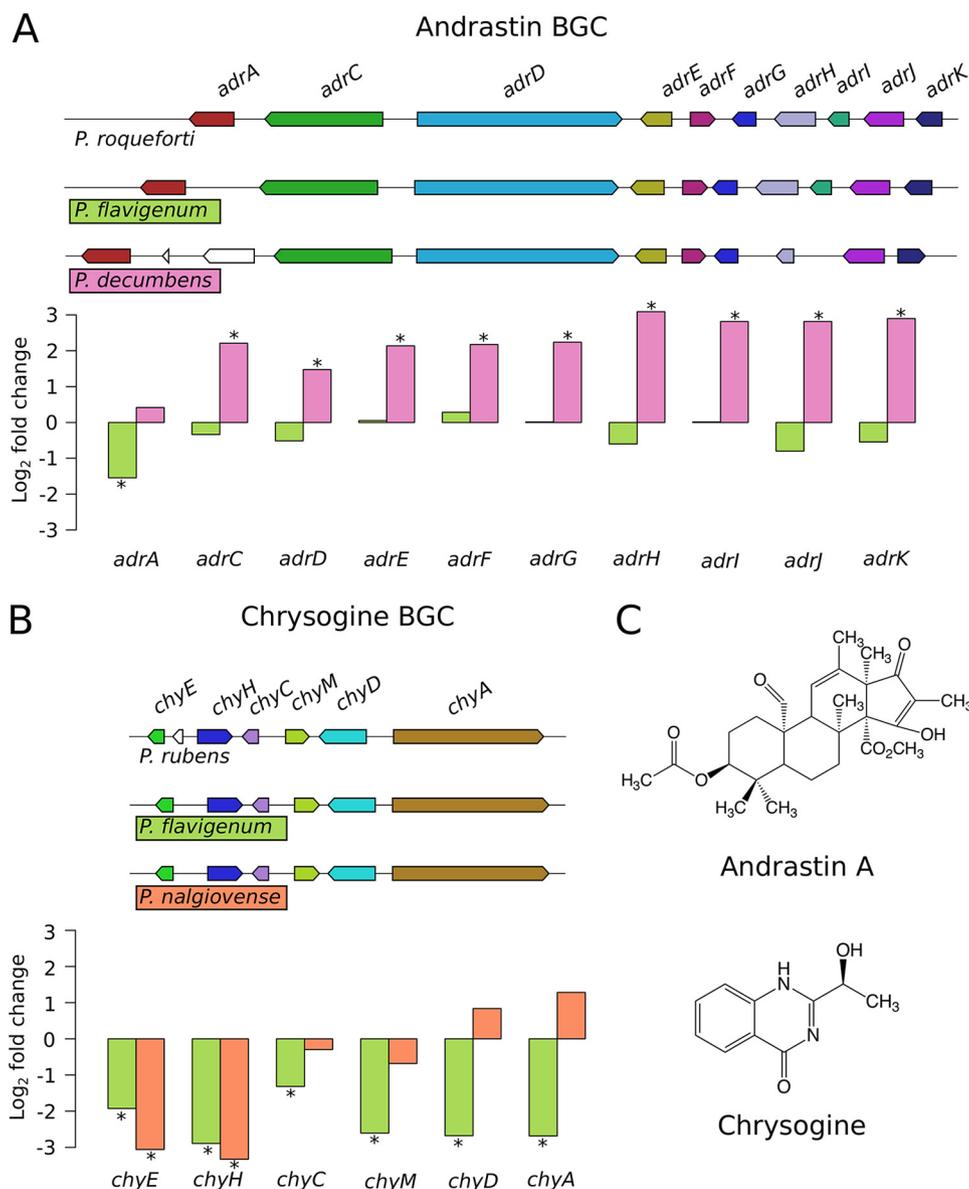
Only one process was conserved among the six species, namely, nitrate assimilation in the GO annotation, which was upregulated in CM, in agreement with the difference in the nitrogen source between the two media. Ammonium transport was upregulated in five species, based on GO terms, and similarly, ammonia assimilation and nitrate reduction were upregulated in four species, based on MetaCyc pathways. Additional upregulated biological processes included the siroheme and porphyrin biosynthesis GO terms in four species. No processes were significantly downregulated in all species, but in five of the species, the GO term copper ion transport and the MetaCyc degradation pathways of the branched-chain amino acids (BCAAs) leucine, isoleucine, and valine were downregulated. In four species, the MetaCyc pathways thiamine and 2-methylcitrate cycle (2MCC) were downregulated as well.

Previous studies have shown that *Aspergillus* species respond to hypoxia by upregulating BCAA metabolism and mitochondrial processes. The latter is seen as an increase in the content of metals, such as iron and copper cofactors, that are used in heme and porphyrin (22). Since we observed similar upregulated pathways, hypoxia might be a common driver of some of the metabolic changes observed in the species. In contrast, nitrate and ammonium assimilation as well as copper transport may be unrelated to hypoxia and, rather, may be a specific response highlighting the difference in nutrient sources and availability between the two media.

In general, few processes were statistically significantly differentially regulated in the majority of the species, indicating that despite the species being within the same genus and, hence, phylogenetically related, many of the responses were species specific rather than general. As a consequence of this, one should be careful when extrapolating regulatory information between species in a diverse genus, such as *Penicillium*.

**Expression and clustering of secondary metabolism in *Penicillium*.** Genome mining of the six *Penicillium* genomes revealed a total of 311 encoded secondary metabolite biosynthetic gene clusters (BGCs), and these were grouped into 42 gene cluster families (GCFs), consisting of similar BGCs (Fig. S4). Among these GCFs, 32 contained PKS, NRPS, NRPS-like, or PKS-NRPS backbone genes. Our recent annotation of *Penicillium* BGCs was updated using the same approach described previously (7) and allowed us to link six of the GCFs to a metabolic product. Among the 32 GCFs, 17 were differentially expressed in at least one species (as determined based on the expression level of the backbone genes), and these were approximately equally distributed between up- and downregulation (Fig. 2C).

We correlated the expression levels of the annotated BGCs with the secondary metabolites detected in the fermentation media (Fig. S5). For four of the annotated BGCs, we detected the corresponding compound under at least one of the fermentation conditions. This included (i) andrastins for *P. decumbens*, (ii) chrysogines for *P. flavigenum* and *P. nalgiovense*, (iii) roquefortine/meleagrins intermediates for *P. coprophilum* and *P. flavigenum*, and (iv) fungisporin for *P. coprophilum*, *P. nalgiovense*, and *P. flavigenum* (14). We further looked into the expression of genes involved in previously characterized BGCs encoding andrastin biosynthesis in *P. roqueforti* (23) and chrysogine biosynthesis in *P. rubens* (24) (Fig. 3). All genes in the andrastin BGC in *P. decumbens* were upregulated, while chrysogine genes in *P. flavigenum* were downregulated. In the chrysogine BGC in *P. nalgiovense*, some genes were upregulated, while others were downregulated. Specifically, the genes *chyE* and *chyH* were significantly downregulated, while *chyA* and *chyD* were upregulated (Fig. 3). ChyA and ChyD catalyze the two first steps in the pathway, while ChyE and ChyH are thought to catalyze later steps (24).



**FIG 3** Expression of biosynthetic gene clusters (BGCs) in *Penicillium*. Andrastin BGC (A), chrysoyine BGC (B), and chemical structures of the end products of the pathways (C) are shown. Bar plots represent the log<sub>2</sub> fold change in expression levels in complex medium (CM) relative to defined medium (DM), and the colors of the bars correspond to the species. Asterisks denote genes that were significantly differentially expressed.

One explanation for observing these differences in expression could thus be a temporal transcriptional control based on when the individual enzymes are needed in the pathway cascade. It should be noted here that the RNA-seq data are indicative only for flux through the pathway and thus cannot be translated directly to increased production levels of the final compound. Quantitative metabolomics would be required to correlate gene expression with production levels and pathway fluxes.

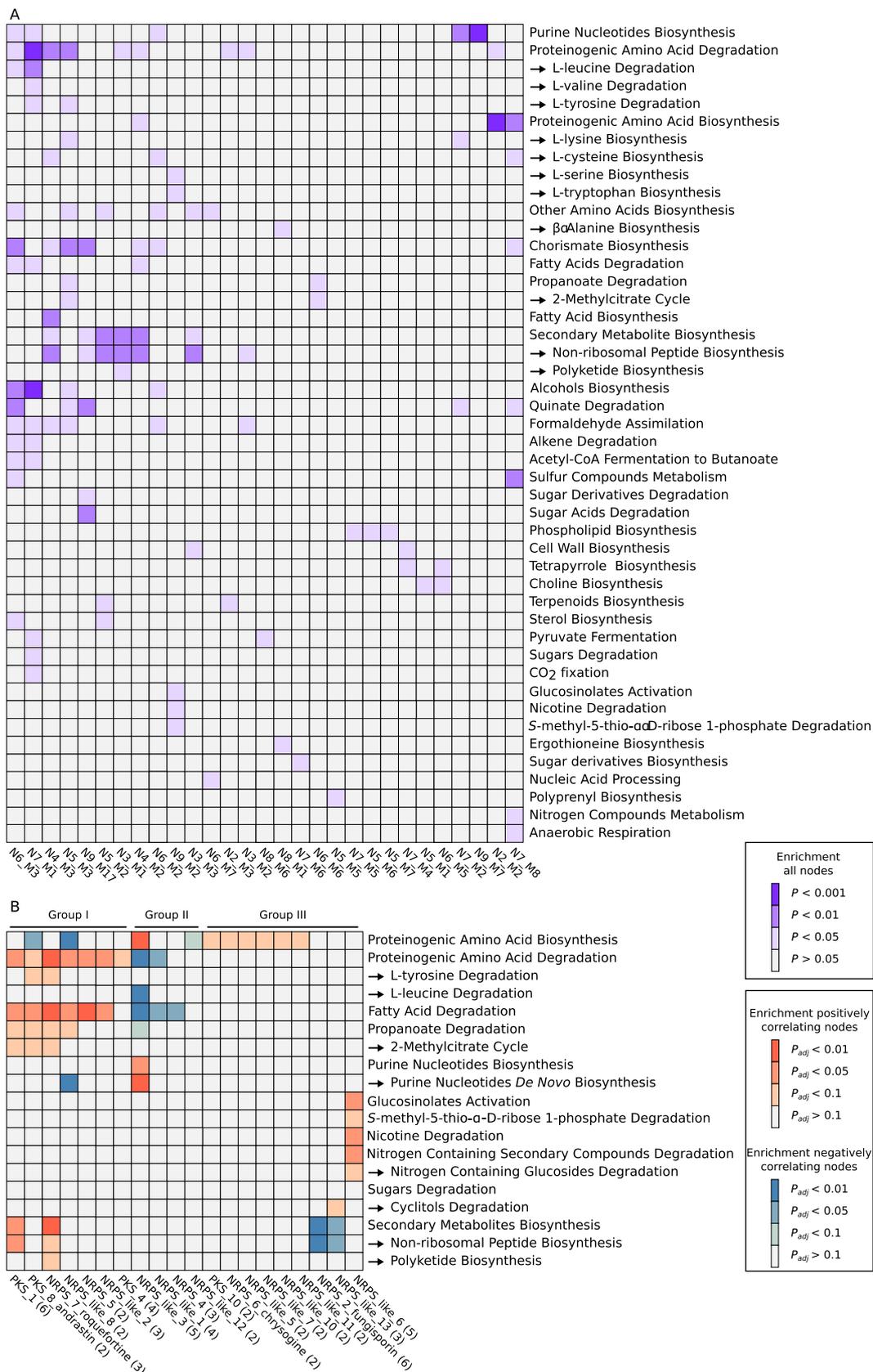
**Coexpression modules in *Penicillium*.** In order to identify gene modules (groups of coexpressed genes) general to the *Penicillium* genus, the Pearson correlation coefficient (PCC) among all 3,782 single-copy core OGs as well as the 33 backbone genes identified in the 32 GCFs present in at least two species was computed. The correlations among these 3,815 OGs constituted a coexpression network from which nine coexpression subnetworks were extracted by removing connections between OGs based on absolute PCC cutoffs in the range of 0.1 to 0.9, which were denoted N1 to N9, respectively

(Fig. S6). For each coexpression subnetwork, we identified highly connected gene modules using the ClusterONE algorithm (25), which allows for identification of overlapping modules, in agreement with the biological context where one gene product can have activities in multiple pathways. For larger networks, fewer but larger modules were found (Fig. S7), in agreement with previous observations (26), and a total of 59 significant modules were detected ( $P < 0.1$ ). To assess the overlap between these modules, the Jaccard index was computed based on module nodes (OGs), and this highlighted that many OGs were shared among multiple modules and, hence, represented different levels of resolution (Fig. S8). We removed highly redundant modules that shared the majority of nodes, and this reduced the total number of modules for the downstream analysis to 54.

The 54 nonredundant coexpression modules were tested for enrichment of metabolic pathways using the MetaCyc annotation of metabolism, and a total of 29 modules were significantly enriched for at least one pathway (Fig. 4A). The most frequently occurring pathway was proteinogenic amino acid degradation, which was enriched in nine modules, and this included, in particular, degradation of leucine, valine, and tyrosine. Conversely, only three modules were enriched for amino acid biosynthesis genes, indicating that degradation processes were more predominant in the fermentations at the time of sampling. Biosynthesis of chorismate, a precursor of aromatic amino acids, was enriched in seven modules but did not overlap tryptophan biosynthesis, for which it is a precursor, thus indicating that chorismate might be used to a large extent for the synthesis of non-amino-acid metabolites, such as tetrahydrofolate, in fungi. There was no clear distinction between modules enriched for either biosynthetic or degradative pathways. For example, the network N7\_M1 was enriched for both the degradation of amino acids and the biosynthesis of purine nucleotides (Fig. 4A).

Six modules were enriched for secondary metabolite biosynthesis. In three of these modules, amino acid degradation was also enriched, indicating a connection between the biosynthesis of secondary metabolites and the degradation of amino acids. In three of the modules, chorismate biosynthesis was enriched, which may be related to the biosynthesis of alkaloids derived from chorismate or to certain NRPs consisting of chorismate-derived amino acid building blocks.

**Expression of primary and secondary metabolism is tightly linked.** To further investigate interactions between primary and secondary metabolism, the connections of all secondary metabolite backbone genes were investigated in each of the 54 coexpression modules. The connections were divided into positive and negative correlations, and an enrichment analysis was conducted. All evaluated backbone genes were found in at least one of the coexpression modules, and among these, 20 were statistically enriched for MetaCyc pathways (Fig. 4B). Interestingly, the enriched pathways proved largely to revolve around the same pathways that were enriched in the coexpression modules as a whole (Fig. 4A) as well as some of the processes that were enriched across species in the gene set analysis (Fig. 2B). These pathways included degradation of amino acids, in particular, leucine and tyrosine; fatty acid degradation; and propanoate degradation via the 2-methylcitrate cycle (2MCC). Further, it was observed that, based on the correlated pathways, the backbone genes could be divided into three groups (Fig. 4B). Group I contained PKS and NRPS genes that were positively correlated with degradation processes, in particular, amino acid and fatty acid degradation, while they were negatively correlated with the biosynthesis of amino acids and purine nucleotides. Conversely, group II contained NRPS genes showing a reciprocal pattern of correlation, with a positive correlation to biosynthesis and a negative correlation to degradation pathways. Interestingly, the only biosynthesis process that was positively correlated to genes in group I was secondary metabolite biosynthesis. Group III contained PKS and NRPS genes that correlated with either high-level processes, such as amino acid biosynthesis, or specific pathways, such as cyclitol degra-



**FIG 4** Enriched pathways in coexpression modules. (A) Pathway enrichment among orthologous groups in each coexpression module. Modules are denoted as NX\_MY, where X is the network number, and Y is the module number. (B) Pathway enrichment among orthologous groups connected to secondary metabolite backbone genes in the coexpression modules.

(Continued on next page)

dition. This group might represent secondary metabolite pathways that are activated only under very specific conditions.

We further investigated the pathways that were enriched in groups I and II to gain information about the specific correlations at the gene/reaction level (Fig. 5). This confirmed the sharp division between the backbone genes in group I and II as either positively or negatively correlated to degradation or biosynthesis. The products of the investigated degradation pathways proved largely to revolve around the mitochondrial acetyl-CoA pool, except for 2MCC, which is used for degradation of propionate, a toxic by-product of valine and isoleucine degradation. Purine biosynthesis was positively correlated with the three backbone genes of group II that were negatively correlated with degradation pathways, indicating that these backbone genes might be expressed under conditions where nucleotides are needed for growth. We further investigated other acetyl-CoA-generating reactions and observed that the genes in group II that were correlated with biosynthesis pathways were also correlated with one of the subunits in the pyruvate dehydrogenase complex. On the other hand, ATP-citrate lyase (ACL) did not show any clear correlation to any of the backbone genes. This suggests that pyruvate dehydrogenase activity, but not ACL activity, might contribute to the acetyl-CoA supply for secondary metabolism under growth conditions (Fig. 5).

## DISCUSSION

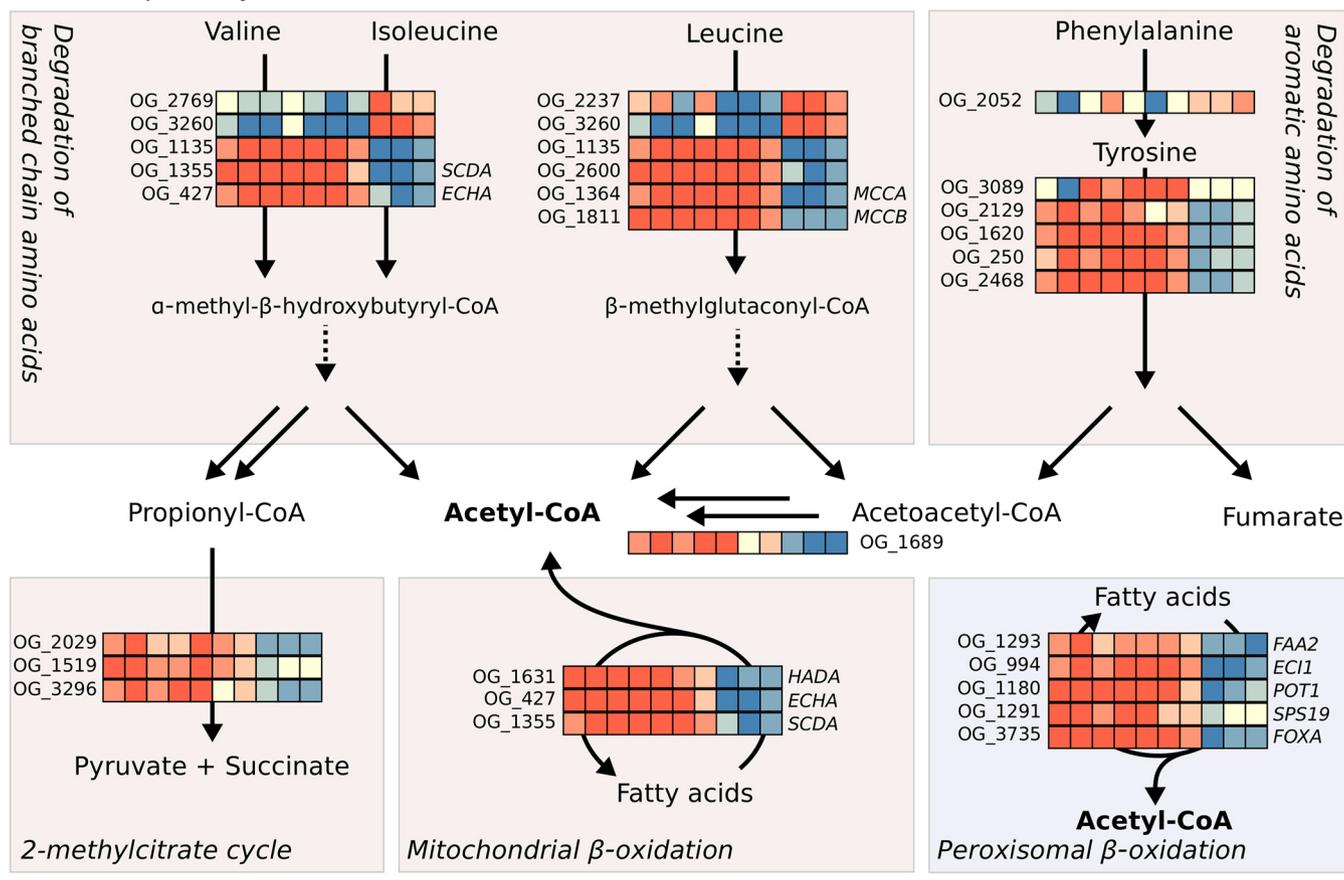
In this study, we have conducted a comparative transcriptome analysis of six *Penicillium* species in the stationary phase of bioreactor batch fermentations in a defined medium and a complex medium (14). The diversity among the species and the conditions under which they were cultivated were evident from the gene expression profiles, since a gene set analysis showed that only a few cellular processes were enriched across species. A coexpression network was constructed and allowed us to identify modules of genes that showed similarity in expression patterns. This enabled the identification of a metabolic signature which links central carbon metabolism to the production of secondary metabolites.

We found seven conserved secondary metabolite backbone genes (group I) with a strong correlation to the genes in mitochondrial and peroxisomal  $\beta$ -oxidation pathways. Previous studies have implied a fundamental role of both of these pathways in secondary metabolism. Disruption of the mitochondrial and peroxisomal  $\beta$ -oxidation pathways individually has been shown to reduce the production of the PK sterigmatocystin in *A. nidulans* (27). Further, mitochondrial  $\beta$ -oxidation is known to share an enoyl-CoA hydratase, *ECHA*, and an acyl-CoA dehydrogenase, *SCDA*, with the degradation pathway of the BCAAs isoleucine and valine (28, 29). Our data demonstrate a strong correlation between secondary metabolite backbone genes and genes encoding the degradation of all three BCAAs, including *ECHA* and *SCDA*, thus showing a close interplay between mitochondrial degradation of fatty acids and BCAAs. Both of these pathways yield acetyl-CoA, a precursor of PKs, as well as propionyl-CoA, and indicate that fatty acid and BCAA degradation might be the main sources of acetyl-CoA for PK biosynthesis under nutrient-limited conditions, such as in the stationary phase in DM, where the carbon source has been depleted. This is in agreement with previous metabolomics studies showing how disruption of the global transcriptional regulator of secondary metabolism, *veA*, leads to decreased PK biosynthesis as well as fatty acid and BCAA degradation in *Aspergillus parasiticus* (13). In addition to these pathways, we found that expression of genes involved in tyrosine degradation correlated with expression of secondary metabolite backbone genes, and this pathway has, to the best of our knowledge, not been linked to the precursor supply for secondary metabolism

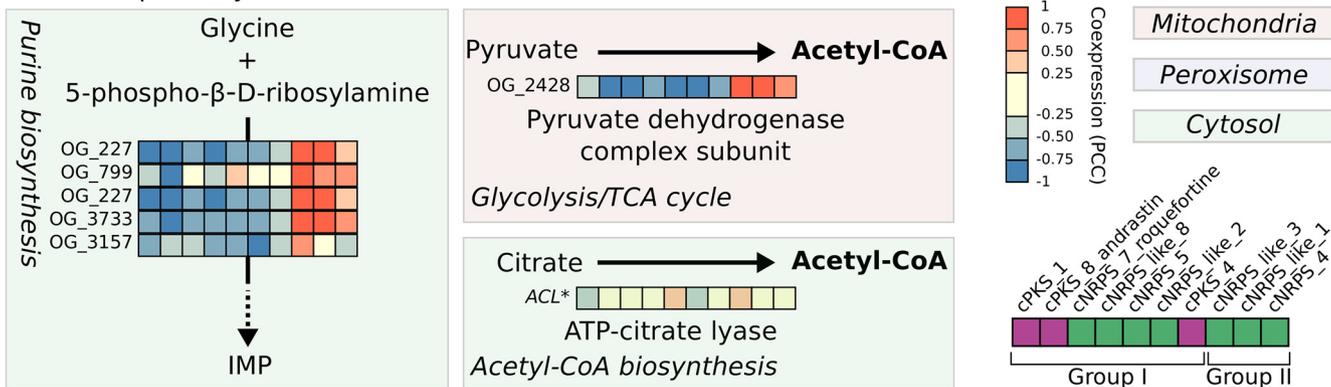
### FIG 4 Legend (Continued)

Pathways were based on the third level of the MetaCyc annotation, while subpathways are indicated with arrows (the fourth level of the MetaCyc annotation). BGCs are denoted based on class, followed by an underscore and an identification number (e.g., PKS\_1). Numbers in parentheses indicate how many of the six investigated species contain the BGC in their genomes.  $P_{adj}$ , adjusted *P* value.

Catabolic pathways



Anabolic pathways



**FIG 5** Main metabolic pathways correlating with secondary metabolite backbone genes. Heat maps represent backbone genes in the columns (as specified in the figure key) and orthologous groups (OG) responsible for the catalysis of the reactions in the pathways as rows. The figure key indicates PKs in purple and NRPSs in green. Coexpression with genes marked with an asterisk was calculated for only five of the species. TCA, tricarboxylic acid.

before. Intermediates from the degradation pathways of BCAAs could also possibly be utilized as building blocks for some PKs, as seen in bacteria (30, 31), as could propionyl-CoA (32). Propionyl-CoA is toxic and can be degraded via the 2MCC pathway, which was observed among the correlated pathways as well. Disruption of this pathway has been shown to have marked negative effects on the production of sterigmatocystin in *A. nidulans*, possibly because propionyl-CoA can block the active site of PKs (33), thus highlighting the importance of detoxification of propionyl-CoA during PK production.

Among the seven backbone genes of group I that were seen to correlate with the above-mentioned acetyl-CoA-generating processes, only three were PKs, while four

were NRPSs (Fig. 5). The correlated pathways explain well how acyl-CoA units can be derived for PK biosynthesis during nutrient limitation but not why expression of NRPSs, which utilize amino acids as building blocks, was correlated as well. We did not observe any NRPS genes of group I to be correlated with amino acid biosynthetic pathways, so amino acids for NRPs are possibly derived from the breakdown of existing proteins instead of the *de novo* synthesis of amino acids. The directionality of this, i.e., if macromolecular breakdown leads to secondary metabolite formation or the other way around, cannot be determined from our data and would be interesting to investigate as a parameter for induction of secondary metabolite formation in an industrial setting. If, however, amino acids are synthesized *de novo*, the products from the degradation of BCAAs, i.e., acetyl-CoA, glutamate, and NADH, would be a favorable starting point for the synthesis of many amino acids. It could be speculated that upon induction of secondary metabolism, many PKs and NRPs are produced at the same time, and this could explain why we observed that NRPS genes also correlate with the degradation processes generating acetyl-CoA.

The three backbone genes of group II were anticorrelated with the acetyl-CoA-yielding degradation pathways and thus might encode secondary metabolites which are induced under nutrient excess. This is supported by the fact that they were correlated to nucleic acid and amino acid biosynthetic genes, which are active under growth conditions. During nutrient excess, pyruvate is produced through glycolysis and can then be converted into acetyl-CoA via the pyruvate decarboxylase complex. Our data suggest that the secondary metabolites that are produced during nutrient excess might use the acetyl-CoA generated via pyruvate decarboxylase activity as a building block (Fig. 5).

It is intriguing to speculate whether certain regulatory elements determine the observed grouping of backbone genes. We were not able to identify any conserved promoter motifs in the genes of BGCs defined in the same groups. This is likely due to the hierarchical structure of regulatory networks (34), resulting in BGC-specific transcription factor activation.

Taken together, our results suggest that the metabolism of filamentous fungi is tailored for the biosynthesis of secondary metabolites. During nutrient limitation, filamentous fungi direct metabolic flux through degradation pathways that generate the necessary precursor metabolites for the biosynthesis of secondary metabolites. We identified a metabolic signature that highlights the main pathways for precursor supply for secondary metabolite biosynthesis and found that expression of these pathways correlates with expression of certain PKS and NRPS genes. We further found that the precursor-generating degradation pathways were enriched in several coexpression modules, suggesting that major parts of metabolism are concerned with generating supplies for secondary metabolite biosynthesis. Even though many metabolic changes are also concerned with adapting the physiology to environmental changes, our data indicate that metabolism in filamentous fungi is tailored to meet the demands of secondary metabolite production.

As mentioned earlier in the Discussion, our data align well with observations on the metabolic regulation of secondary metabolism in *Aspergillus* species, which is why our findings likely can be extrapolated to other fungi as well. Recent findings for *Penicillium* (7, 16) and *Aspergillus* (35, 36) have shown a high genomic diversity and the adaptation of secondary metabolism to natural environments, but the supply of precursors for secondary metabolism might, on the contrary, be highly conserved. Our findings on the regulation of precursor-supplying pathways can thus aid in the design of metabolic engineering strategies to optimize the precursor supply for secondary metabolites in fungi, e.g., by overexpression of the precursor-generating pathways identified in this study. Although future experimental validation is necessary to fully map precursor-supplying pathways for secondary metabolites in fungi and to get a more detailed description of how metabolism is shaped toward secondary metabolite production, our findings provide a starting point in understanding how to manipulate metabolism for more efficient production of secondary metabolites.

## MATERIALS AND METHODS

**Cultivation conditions.** Samples for RNA-seq analysis were collected from cultivation experiments previously described (14). Briefly, six different *Penicillium* species (*P. coprophilum* [IBT 31321], *P. nalgiovense* [IBT 13039], *P. polonicum* [IBT 4502], *P. decumbens* [IBT11843], *P. flavigenum* [IBT 14082], and *P. steckii* [IBT 24891]) were cultivated in controlled bioreactors in a defined medium (DM) for *Penicillium* and in a complex medium (CM). Glucose was quantified by high-performance liquid chromatography, and CO<sub>2</sub> concentrations were determined mass spectrometrically. Biomass samples for RNA-seq were withdrawn in the stationary phase, rinsed with ice-cold water through a Miracloth, and snap-frozen in liquid nitrogen until further analysis. Samples were collected from fermentations carried out in biological triplicate. All strains used in this study are available from the culture collection (IBT) at the Technical University of Denmark.

**Transcriptome analysis.** Cells from frozen biomass samples were disrupted using a TissueLyser LT disrupter (Qiagen), and total RNA was extracted using an RNeasy minikit (catalog no. 74104; Qiagen). DNA libraries for sequencing were prepared from the total RNA using Illumina's TruSeq protocol and sequenced using an Illumina 2500 machine, yielding 99 nucleotide paired-end reads with an average insert size of 600 nucleotides. Raw RNA-seq reads were mapped to the individual *Penicillium* genomes using a documented work flow (37) based on the TopHat2 (version 2.0.9) (38) and HTSeq (39) programs. Both programs were run with default parameters. Gene-level statistics for CM versus DM were calculated using the DESeq2 program with default parameters, and differentially expressed genes were identified based on an adjusted *P* value cutoff of 0.05. For all downstream analyses, log-transformed expression levels were used.

**Determination of orthology and phylogeny.** The genome sequences of the six species were downloaded from NCBI, orthologous protein groups were identified using the orthoMCL algorithm (40) with default parameters, and these orthologous groups were used to define the core genome and the pangenome. Based on single-copy core genes, a concatenated maximum likelihood phylogenetic tree was reconstructed by using the RAxML program (41) and by applying a work flow previously described (7).

**Gene set enrichment analysis.** A gene set analysis was conducted based on a MetaCyc annotation (17) that was retrieved from previously published genome-scale metabolic models of the *Penicillium* species (16), and GO terms were annotated using the InterProScan (version 5.7-48) program (42). For the MetaCyc pathways, PKs and NRPSs identified using the antiSMASH program (43) were added to the annotation of secondary metabolism. Gene set enrichment analysis of the DEGs was performed using the R package PIANO (44). Significant gene sets were identified based on a Benjamini-Hochberg-corrected *P* value cutoff of 0.05. Other enrichment analyses were conducted using the hypergeometric test implementation in R (phyper function).

**Identification and clustering of biosynthetic gene clusters.** BGCs were identified in the genomes using antiSMASH (version 4.0.0rc1 for fungi) (43), and the detected BGCs were clustered into gene cluster families (GCFs) using the BIG-SCAPE program (<https://git.wageningenur.nl/medema-group/BiG-SCAPE>). We tested various network clusterings in BIG-SCAPE, and finally, a cutoff of 0.6 for the overall score was selected on the basis of having similarity to previous clustering of *Penicillium* BGCs (7). Within GCFs, orthologous genes were identified using the MultiGeneBlast program (45), based on having 25% coverage and 30% identity.

**Coexpression network analysis.** Gene expression was correlated using the Pearson correlation coefficient (PCC). In addition, the single-copy orthologous genes were correlated with backbone genes (PKs, NRPSs, NRPS-like, or PKS-NRPSs) present in at least two different species, as identified in the BGC clustering. The pairwise correlations of the expression of all 3,815 single-copy core genes constituted a weighted coexpression network with orthologous groups as nodes and PCC values as edges. This coexpression network was divided into nine subnetworks by applying different cutoffs for the PCC values (0.1 to 0.9), and the corresponding networks were denoted N1 to N9, respectively. For each of these subnetworks, correlation coefficients were converted to absolute values and normalized to distribute between 0 and 1 (minimum/maximum normalization). Highly connected clusters of genes, referred to as modules, were detected in the subnetworks using the ClusterONE algorithm (25). A total of 56 significant modules were identified using a *P* value cutoff of 0.1 for a *t* test assessing the connectivity within a module versus outside a module.

**Data availability.** The data sets supporting the conclusions of this article are available in NCBI's Gene Expression Omnibus (46) repository under accession number [GSE106983](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106983).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00012-19>.

**FIG S1**, PDF file, 0.2 MB.

**FIG S2**, PDF file, 0.02 MB.

**FIG S3**, PDF file, 0.1 MB.

**FIG S4**, PDF file, 0.02 MB.

**FIG S5**, PDF file, 0.02 MB.

**FIG S6**, PDF file, 0.02 MB.

**FIG S7**, PDF file, 0.1 MB.

**FIG S8**, PDF file, 0.04 MB.

**TABLE S1**, XLSX file, 0.3 MB.

## ACKNOWLEDGMENTS

This work was supported by the European Commission Marie Curie Initial Training Network Quantfung (FP7-People-2013-ITN, grant 607332). We also acknowledge funding from the Novo Nordisk Foundation and the Knut and Alice Wallenberg Foundation.

## REFERENCES

- Barrios-González J, Miranda RU. 2010. Biotechnological production and applications of statins. *Appl Microbiol Biotechnol* 85:869–883. <https://doi.org/10.1007/s00253-009-2239-6>.
- Okabe M, Lies D, Kanamasa S, Park EY. 2009. Biotechnological production of itaconic acid and its biosynthesis in *Aspergillus terreus*. *Appl Microbiol Biotechnol* 84:597–606. <https://doi.org/10.1007/s00253-009-2132-3>.
- Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Barbote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol* 26:553–560. <https://doi.org/10.1038/nbt1403>.
- Ropars J, Rodriguez de la Vega RC, Lopez-Villavicencio M, Gouzy J, Sallet E, Dumas E, Lacoste S, Debuchy R, Dupont J, Branca A, Giraud T. 2015. Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Curr Biol* 25:2562–2569. <https://doi.org/10.1016/j.cub.2015.08.025>.
- Keller NP, Turner G, Bennett JW. 2005. Fungal secondary metabolism—from biochemistry to genomics. *Nat Rev Microbiol* 3:937–947. <https://doi.org/10.1038/nrmicro1286>.
- Frisvad JC, Smedsgaard J, Larsen TO, Samson RA. 2004. Mycotoxins, drugs and other extrolites produced by species in *Penicillium* subgenus *Penicillium*. *Stud Mycol* 49:201–241.
- Nielsen JC, Grijseels S, Prigent S, Ji B, Dainat J, Nielsen KF, Frisvad JC, Workman M, Nielsen J. 2017. Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nat Microbiol* 2:17044. <https://doi.org/10.1038/nmicrobiol.2017.44>.
- van den Berg MA, Albang R, Albermann K, Badger JH, Daran J-M, Driessen AJM, Garcia-Estrada C, Fedorova ND, Harris DM, Heijne WHM, Joardar V, Kiel J, Kovalchuk A, Martin JF, Nierman WC, Nijland JG, Pronk JT, Roubos JA, van der Kleij IJ, van Peij N, Veenhuis M, von Döhren H, Wagner C, Wortman J, Bovenberg R. 2008. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat Biotechnol* 26:1161–1168. <https://doi.org/10.1038/nbt.1498>.
- Thykaer J, Nielsen J. 2003. Metabolic engineering of beta-lactam production. *Metab Eng* 5:56–69. [https://doi.org/10.1016/S1096-7176\(03\)00003-X](https://doi.org/10.1016/S1096-7176(03)00003-X).
- Mattern DJ, Valiante V, Unkles SE, Brakhage AA. 2015. Synthetic biology of fungal natural products. *Front Microbiol* 6:775. <https://doi.org/10.3389/fmicb.2015.00775>.
- Nielsen JC, Nielsen J. 2017. Development of fungal cell factories for the production of secondary metabolites: linking genomics and metabolism. *Synth Syst Biotechnol* 2:5–12. <https://doi.org/10.1016/j.synbio.2017.02.002>.
- Brakhage AA. 2013. Regulation of fungal secondary metabolism. *Nat Rev Microbiol* 11:21–32. <https://doi.org/10.1038/nrmicro2916>.
- Roze LV, Chanda A, Laivenieks M, Beaudry RM, Artymovich KA, Koptina AV, Awad DW, Valeeva D, Jones AD, Linz JE. 2010. Volatile profiling reveals intracellular metabolic changes in *Aspergillus parasiticus*: veA regulates branched chain amino acid and ethanol metabolism. *BMC Biochem* 11:33. <https://doi.org/10.1186/1471-2091-11-33>.
- Grijseels S, Nielsen JC, Nielsen J, Larsen TO, Frisvad JC, Nielsen KF, Frandsen RJN, Workman M. 2017. Physiological characterization of secondary metabolite producing *Penicillium* cell factories. *Fungal Biol Biotechnol* 4:8. <https://doi.org/10.1186/s40694-017-0036-z>.
- Visagie CM, Houbraken J, Frisvad JC, Hong S-B, Klaassen CHW, Perrone G, Seifert KA, Varga J, Yaguchi T, Samson RA. 2014. Identification and nomenclature of the genus *Penicillium*. *Stud Mycol* 78:343–371. <https://doi.org/10.1016/j.simyco.2014.09.001>.
- Prigent S, Nielsen JC, Frisvad JC, Nielsen J. 2018. Reconstruction of 24 *Penicillium* genome-scale metabolic models shows diversity based on their secondary metabolism. *Biotechnol Bioeng* 115:2604–2612. <https://doi.org/10.1002/bit.26739>.
- Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Weerasinghe D, Zhang P, Karp PD. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 42:D459–D471. <https://doi.org/10.1093/nar/gkt1103>.
- Brakhage AA. 1998. Molecular regulation of beta-lactam biosynthesis in filamentous fungi. *Microbiol Mol Biol Rev* 62:547–585.
- Thompson DA, Roy S, Chan M, Styczynski MP, Pffiffer J, French C, Socha A, Thielke A, Napolitano S, Muller P, Kellis M, Konieczka JH, Wapinski I, Regev A. 2013. Evolutionary principles of modular gene regulation in yeasts. *Elife* 2:e00603. <https://doi.org/10.7554/eLife.00603>.
- Terabayashi Y, Shimizu M, Kitazume T, Masuo S, Fujii T, Takaya N. 2012. Conserved and specific responses to hypoxia in *Aspergillus oryzae* and *Aspergillus nidulans* determined by comparative transcriptomics. *Appl Microbiol Biotechnol* 93:305–317. <https://doi.org/10.1007/s00253-011-3767-4>.
- Schinko T, Berger H, Lee W, Gallmetzer A, Pirker K, Pachlinger R, Buchner I, Reichenauer T, Güldener U, Strauss J. 2010. Transcriptome analysis of nitrate assimilation in *Aspergillus nidulans* reveals connections to nitric oxide metabolism. *Mol Microbiol* 78:720–738. <https://doi.org/10.1111/j.1365-2958.2010.07363.x>.
- Vödisch M, Scherlach K, Winkler R, Hertweck C, Braun H-P, Roth M, Haas H, Werner ER, Brakhage AA, Knimemeyer O. 2011. Analysis of the *Aspergillus fumigatus* proteome reveals metabolic changes and the activation of the psurotin A biosynthesis gene cluster in response to hypoxia. *J Proteome Res* 10:2508–2524. <https://doi.org/10.1021/pr1012812>.
- Rojas-Aedo JF, Gil-Durán C, Del-Cid A, Valdés N, Álamos P, Vaca I, García-Rico RO, Levicán G, Tello M, Chávez R. 2017. The biosynthetic gene cluster for andrastin A in *Penicillium roqueforti*. *Front Microbiol* 8:813. <https://doi.org/10.3389/fmicb.2017.00813>.
- Viggiano A, Salo O, Ali H, Szymanski W, Lankhorst PP, Nygård Y, Bovenberg RAL, Driessen A. 2017. Pathway for the biosynthesis of the pigment chrysoerythrin by *Penicillium chrysogenum*. *Appl Environ Microbiol* 84:02246-17. <https://doi.org/10.1128/AEM.02246-17>.
- Nepusz T, Yu H, Paccanaro A. 2012. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* 9:471–472. <https://doi.org/10.1038/nmeth.1938>.
- Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. 2017. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* 29:944–959. <https://doi.org/10.1105/tpc.17.00009>.
- Maggio-Hall LA, Wilson RA, Keller NP. 2005. Fundamental contribution of beta-oxidation to polyketide mycotoxin production in planta. *Mol Plant Microbe Interact* 18:783–793. <https://doi.org/10.1094/MPMI-18-0783>.
- Maggio-Hall LA, Keller NP. 2004. Mitochondrial  $\beta$ -oxidation in *Aspergillus nidulans*. *Mol Microbiol* 54:1173–1185. <https://doi.org/10.1111/j.1365-2958.2004.04340.x>.
- Maggio-Hall LA, Lyne P, Wolff JA, Keller NP. 2008. A single acyl-CoA dehydrogenase is required for catabolism of isoleucine, valine and short-chain fatty acids in *Aspergillus nidulans*. *Fungal Genet Biol* 45:180–189. <https://doi.org/10.1016/j.fgb.2007.06.004>.
- Denoya CD, Fedechko RW, Hafner EW, McArthur HAI, Morgenstern MR, Skinner DD, Stutzman-Engwall K, Wax RG, Wernau WC. 1995. A second branched-chain alpha-keto acid dehydrogenase gene cluster (bkdFGH)

- from *Streptomyces avermitilis*: its relationship to avermectin biosynthesis and the construction of a *bkdF* mutant suitable for the production of novel antiparasitic avermectins. *J Bacteriol* 177:3504–3511. <https://doi.org/10.1128/jb.177.12.3504-3511.1995>.
31. Stirrett K, Denoya C, Westpheling J. 2009. Branched-chain amino acid catabolism provides precursors for the type II polyketide antibiotic, actinorhodin, via pathways that are nutrient dependent. *J Ind Microbiol Biotechnol* 36:129–137. <https://doi.org/10.1007/s10295-008-0480-0>.
  32. Chan YA, Podevels AM, Kevany BM, Thomas MG. 2009. Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* 26:90–114. <https://doi.org/10.1039/B801658P>.
  33. Zhang YQ, Brock M, Keller NP. 2004. Connection of propionyl-CoA metabolism to polyketide biosynthesis in *Aspergillus nidulans*. *Genetics* 168:785–794. <https://doi.org/10.1534/genetics.104.027540>.
  34. Yu H, Gerstein M. 2006. Genomic analysis of the hierarchical structure of regulatory networks. *Proc Natl Acad Sci U S A* 103:14724–14731. <https://doi.org/10.1073/pnas.0508637103>.
  35. de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, Anderluh G, Asadollahi M, Askin M, Barry K, Battaglia E, Bayram Ö, Benocci T, Braus-Stromeyer SA, Caldana C, Cánovas D, Cerqueira GC, Chen F, Chen W, Choi C, Clum A, dos Santos RAC, Damásio ADL, Diallynas G, Emri T, Fekete E, Flipphi M, Freyberg S, Gallo A, Gournas C, Habgood R, Hainaut M, Harispe ML, Henriessat B, Hildén KS, Hope R, Hossain A, Karabika E, Karaffa L, Karányi Z, Kraševc N, Kuo A, Kusch H, LaButti K, Legendijk EL, Lapidus A, Levasseur A, Lindquist E, Lipzen A, Logrieco AF, et al. 2017. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol* 18:28. <https://doi.org/10.1186/s13059-017-1151-0>.
  36. Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, Hoof JB, Brandl J, Salamov A, Riley R, Gladden JM, Phatale P, Nielsen MT, Lyhne EK, Kogle ME, Strasser K, McDonnell E, Barry K, Clum A, Chen C, Labutti K, Haridas S, Nolan M, Sandor L, Kuo A, Lipzen A, Hainaut M, Drula E, Tsang A, Magnuson JK, Henriessat B, Wiebenga A, Simmons BA, Mäkelä MR, Vries RP, De Grigoriev IV, Mortensen UH, Baker SE, Andersen MR. 2018. Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section *Nigri*. *Nat Genet* 50:1688–1695. <https://doi.org/10.1038/s41588-018-0246-1>.
  37. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–578. <https://doi.org/10.1038/nprot.2012.016>.
  38. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
  39. Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
  40. Li L, Stoeckert CJJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <https://doi.org/10.1101/gr.1224503>.
  41. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
  42. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
  43. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.
  44. Våremo L, Nielsen J, Nookaew I. 2013. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* 41:4378–4391. <https://doi.org/10.1093/nar/gkt111>.
  45. Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol* 30:1218–1223. <https://doi.org/10.1093/molbev/mst025>.
  46. Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210. <https://doi.org/10.1093/nar/30.1.207>.