Fungal metabolite analysis in genomics and phenomics
Michael C Jewett, Gerald Hofmann and Jens Nielsen

Metabolomics consists of strategies to quantitatively identify cellular metabolites and to understand how trafficking of these biochemical messengers through the metabolic network influences phenotype. The application of metabolomics to fungi has been strongly pursued because these organisms are widely used for the production of chemicals, are well known for their diverse metabolic landscape and serve as excellent eukaryotic model organisms for studying metabolism and systems biology. Within the context of fungal systems, recent progress has been made in the development of analytical tools and mathematical strategies used in metabolite analysis that have enhanced our ability to crack the code underpinning the cellular inventory, regulatory schemes and communication mechanisms that dictate cellular function. Metabolomics has played a key role in functional genomics and strain classification.

Introduction
Being major players in food and pharmaceutical biotechnology, intense interest surrounds the study of fungal kingdom members, such as filamentous fungi and yeasts. The enormous biodiversity within the Mycota has resulted in their application as model organisms for the production of fuels, chemicals, food ingredients, pharmaceuticals and enzymes. Equally important is the central role that fungi occupy as model systems for basic research. For example, Saccharomyces cerevisiae, one of the most well established eukaryotic model organisms [1], has played and continues to play a crucial role in the development of numerous techniques that simultaneously detect multiple signals at the molecular level (e.g. DNA microarrays) for elucidating general rules and descriptions about living cells. Finally, the relevance of fungi in food and feed spoilage (primarily through mycotoxin production) and their pathogenicity has also elevated the importance of fungal research.

To understand, characterize and exploit fungi more completely, research efforts have sought to build a more detailed understanding of fungal metabolism. Metabolism captures the most salient traits of fungi and, in particular, the exquisite chemical diversity of their metabolites and the utilization of eukaryotic cellular features to ensure operation of dedicated pathways in different compartments. By offering a window to core attributes responsible for diverse phenotypes, metabolome analysis capitalizes on the information content obtained from primary and secondary metabolism to elucidate fundamental principles that describe the relationship between genotype and phenotype (Figure 1).

Genome sequencing and annotation identifies the inventory of parts that make up the cell. A parts list, however, void of contextual, integrative and quantitative information describing active elements, lacks the power to completely decipher the mechanisms of cellular function [2]. As the intermediates of biochemical reactions, metabolites link together a web of complex interactions, cellular pathways, molecular participants (e.g. DNA, mRNA, proteins) and environmental stimuli. They act as a spoken language, broadcasting signals from the genetic architecture and the environment through the metabolic network to help achieve the functional objectives of the cell. The integrated nature of metabolic networks, predominantly in primary metabolism, has been underscored by genome-scale metabolic models [3–7]. First, we observe that more than 70% of all metabolites participate in more than just two reactions (Figure 2a) [4]. Second, cellular biochemical reactions usually involve more than one substrate and one product (Figure 2b) [4]. Third, the average path length to get from any metabolite to any other metabolite is approximately three [5,8*]. Clearly, the maze-like nature of metabolism reveals that even small perturbations in metabolite concentrations are likely to impact the overall functional operation of the network [8*]. This coordinated structural organization highlights the significance of metabolite quantification in achieving a systems-level understanding.

In addition to finding utility in systems biology applications, metabolite analyses have also proven effective for identification of bioactive molecules with potential for therapeutic application and fungal taxonomy. Here, black-box approaches, in which there is no need for gene sequence information, can be taken to explore the enormous metabolic chemodiversity present in fungi.
Although this can generally be applied to all parts of metabolism, secondary metabolic products are often more interesting for such purposes because of their specificity, uniqueness and high diversity.

This review highlights recent advances in fungal metabolomics. Because the genome provides the code of cellular parts, which helps to define the theoretical boundary of the metabolic network, we begin with a brief update on fungal genomics. Next, we concentrate on new methods for quantitative metabolite measurement used in functional genomics and strain classification. Finally, we underscore the need for improvements in standardization of how we generate and report our data. Despite our focus on fungal systems, many of the technological advances have been mirrored by or originated from similar reports in plants.

**Genomics**

To map changes in metabolites back to gene sequence, an accurate genomic blueprint is required. The era of fungal genomics started in 1996 with the publication of the complete genome sequence of *S. cerevisiae* [9]. Since then, the number of available fungal genome sequences has increased dramatically. Although only 12 complete fungal genomic sequences have been published, more than 51 fungal genomes can be accessed using BLAST searches at the National Center for Biotechnology Information (NCBI) internet website (http://www.ncbi.nlm.nih.gov/). Because no comprehensive repository for fungal genomes is available, Table 1 provides a list of URL links that can be consulted for up-to-date information on the status of fungal genome sequencing and for access to various genome data. Although genome sequences might ultimately play an essential role in increasing the information content from metabolomics studies, the intrinsically complex relationship between genes and metabolites has made interpretations of these data difficult.

**Metabolomics**

We define a metabolite as any chemical compound of the cell that is not genetically encoded and is a substrate, intermediate or product of metabolism. Metabolome analysis [10] seeks to identify and quantify the entire collection of intracellular and extracellular metabolites. Conceptually, there are two basic approaches used in metabolomics (Figure 3) [11**]. Mainly exploited for classification, metabolite profiling strategies investigate
qualitative scanning of all detectable metabolites observed by a selected analytical technique. Here, the pattern of known and unknown metabolites (or spectra from chromatography or mass spectrometry) is used to find discriminatory elements via high-throughput detection followed by data deconvolution methods [12,13]. Metabolite profiling comprises metabolic fingerprinting, which covers the endometabolome (intracellular metabolites), and metabolic footprinting, which covers the exometabolome (metabolites secreted into the growth media or extracellular fluid). The other general method used in metabolomics is target analysis. Here, absolute, or at least semi-, quantification and unambiguous detection of predefined metabolites are achieved. Although target analysis has been historically reserved for interrogating relatively small numbers of metabolites (e.g. <20), new developments enable quantitative analysis of more expanded metabolome coverage [14**.15.16].

A variety of analytical platforms has been utilized for metabolite detection [11**]. Although most quantitative strategies couple a separation technique (e.g. capillary electrophoresis [CE], liquid chromatography [LC] or gas chromatography [GC]) with MS- or NMR-based detection, it is not uncommon to make use only of direct infusion MS for metabolite profiling. From a practical standpoint, our inability to quantitatively extract and detect highly diverse families of metabolites in their original state over a large dynamic range with a single or even limited set of analytical techniques makes the analysis of the complete set of all metabolites impossible. Thus, metabolomics is more appropriately used to describe an ‘area of science rather than an analytical approach’ [17].

Metabolomics in functional genomics

A fundamental premise of metabolomics is that acquiring a snapshot of the metabolic composition will provide insight into the governing functional and regulatory behavior that connects gene sequence to gene function. Here, we highlight several platforms that have emerged to guide systems-level phenomics by exploring metabolite levels and flow through the primary metabolic scaffold.

Elucidating a metabolic image of central carbon metabolism has provided insights for linking normal anabolic and catabolic trafficking with other branches of metabolism. For example, the use of LC–MS [18] has been exploited to map metabolic activity and flexibility through dynamic analysis of intracellular metabolites during the yeast cell-cycle [19] and to determine the effect of culture age on metabolite pools [20].

Even though quantification of biomolecules involved in central metabolism offers many insights into key nodes of metabolism, other applications have also laid the foundation for target analysis of metabolic hubs that lie one step beyond central metabolism. To quantify metabolites containing an amino or carboxylic acid group, Villas-Boas et al. [14] applied a sensitive GC–MS method coupled to a statistical data-mining strategy for the integrated analysis of clearly identified and quantified intracellular and extracellular metabolites (~60) in S. cerevisiae. By isolating statistically significant differences among metabolite levels from four biological conditions, they observed discriminatory metabolic features that hinted at the potential for future integration with comparative omic analyses. Highlighting the generality of this method, Panagiotou et al. [21] have utilized this analytical approach to determine the influence of aerobic and anaerobic cultivation conditions on the metabolic state of Fusarium oxysporum.

Equally important in guiding a systems-level understanding of the overlapping layers of global regulation and network flexibility are efforts to measure the flow of material through central metabolism experimentally. Characterization of cellular metabolic operation is achieved by using 13C-labeled substrates followed by determination of characteristic metabolite patterns, which can indicate directional flow [22]. The most general approach uses proteinogenic amino acid analysis to infer labeling patterns and flux distributions. However, the
application of rapid sampling and quenching has recently been applied to analyze intracellular metabolites directly from *S. cerevisiae*, without being impeded by the high metabolic turnover rates [23]. This approach generates direct data without inference, but caution must be exercised because of the rapid dynamics of exchange between metabolites and amino acids incorporated into cellular proteins [24].

High-throughput efforts for comparative flux analysis offer an unprecedented view of the rigidity, flexibility and performance of metabolic networks. For example, Blank et al. [25] considered flux data from over 30 mutants of *S. cerevisiae* to investigate potentially flexible fitness reactions during growth on glucose. Combining these measurements with mathematical modeling revealed that metabolic network robustness to single gene knockouts was principally a result of genetic redundancy (duplicate genes) with alternative pathways (redirection of carbon flow) having less importance. This approach was taken further in a larger systematic flux analysis of 137 null mutants of *Bacillus subtilis* (selected from all major functional categories) on its preferred substrate [26**]. In the previous report, this strategy enabled identification of fundamental design principles of *in vivo* network operation. A key feature is the manifestation of rigid distribution patterns, which are ‘largely independent of the rate and yield of biomass formation’. Specifically, they observed that network operation in *B. subtilis* is not solely operating to optimize growth, but also invests cellular resources in anticipation of changing environmental stress. The above cases represent powerful strategies to uncover the structure and function of the interplay between genetic regulatory networks and phenotype.

Relying less on connections between genetic sequence and metabolites, metabolite profiling and target analysis have also been effectively used to classify the phenotype of silent and unknown mutations [27–29]. Weckwerth et al. [28] demonstrated the application of target analysis, using GC–time-of-flight (TOF)–MS for quantification of more than 1000 metabolites, to characterize the features responsible for a silent plant phenotype. Exploiting statistical tools, metabolic correlations were determined between identified metabolites (e.g. trehalose-erythritol) and used to reveal network maps, which suggested hypotheses for the impact of an exact phenotype on carbohydrate and amino acid metabolism.

Hierarchical metabolomics, developed in plants, has also proved to be well suited to guide targeted analysis of metabolism [30**]. Catchpole et al. used comprehensive metabolome coverage of conventional and genetically modified (GM) potato crops to reveal that, apart from anticipated engineered differences, metabolic compositions were comparable among several types of cultivars. First, they applied metabolic fingerprinting of potato tuber extracts to classify several potato genotypes. Second, target analysis of defined and specific classes of metabolites, using LC–MS and GC–TOF–MS, was exploited to identify specific fructans responsible for the global classifications. Finally, data analysis tools were applied to remove the influence of anticipated differences in the GM crops and show that the GM and conventional crops were within the variation observed from investigating several unmodified metabolic phenotypes. Hierarchical analysis provides a rapid and relatively inexpensive screen for many functional genomics and screening applications.

**Metabolomics in strain classification**

Whereas exploiting results obtained with metabolome analysis in functional genomics is often limited by our inability to unravel highly interconnected networks of molecular constituents, the huge metabolic chemodiversity within the Mycota kingdom has become an indispensable tool for classification and identification of fungi. The secondary metabolism of fungi stands second only to that of plants, and secondary metabolites, which are generally secreted, represent a highly interesting subset of the fungal exometabolome, which has been exploited for fungal taxonomy.

Detection and quantification of mycotoxins are the major focal points for characterization studies (see [31] for an overview of mycotoxins). The increase in public awareness over the safety of food and feed during the past few years has led to the establishment of many new laws and guidelines with respect to mycotoxins [32]. For the latest developments in analysis and detection methods, we refer the reader to an exceptional review detailing this topic [33]. A key development is that unconventional biosensor methods, such as electronic nose or tongue technology, which typically rely on metabolite profiling, have a strong potential to mature into key techniques for the detection of mycotoxins and toxigenic fungi [34].

Use of metabolite profiling in fungal classification and comparative analysis is more commonly associated with chromatographic and mass spectrometric techniques [29,35,36*,37]. Although it is relatively simple to generate an enormous amount of metabolite data, processing and extracting useful information is often problematic. To address this limitation, reference databases and data deconvolution methods are crucial. For example, Nielsen and Smidsgaard [38] have collected high-performance liquid chromatography (HPLC)–UV data of 474 mycotoxins. By covering the biodiversity of secondary metabolites from fungi, reference libraries provide an invaluable tool for the identification of known mycotoxins, as well as for the discovery of potentially new compounds within fungal metabolome datasets.

Sophisticated data-mining strategies, such as a new algorithm named X-hitting that was developed to identify
Advances in internal standardization are also paving the future targeted analysis strategies of several metabolite analysis. It is organism specific, it provides a benchmark for biological story (level of metabolites). Even though their several extraction strategies appear to result in the same amino acids, an important outcome of this work was that was principally directed toward organic and non-organic extraction protocols. Although the explicit detection tool their focus on comparative profiling from six different the yeast preparation on targeted metabolite analysis using GC–MS for with limited degradation and losses. Recently, Villas-Boas consistent and boasts excellent recovery characteristics access, it is important that the extraction procedure is imaging the metabolic inventory of the cell depends on structure and compartments) make metabolome cover-

Challenges in sample preparation and standardization

Despite several examples establishing the utility of metabolomics as a tool for functional analysis and classification in fungi, ensuring unbiased and robust quantification of a large number of metabolites is still a major challenge. The main obstacles to this objective, sample preparation and standardization, have recently received attention.

Relative to the other molecular participants in the cell (e.g. genes, mRNAs and proteins), metabolites typically exist on a considerably shorter time-scale (by more than an order of magnitude). Thus, to obtain an accurate picture of the metabolic state of the cell, rapid inactivation of biological activity is needed to prevent enzymatic exchange and turnover of metabolites. In addition to short time-scales, the chemical diversity of metabolite classes and the physical barriers of the cell (e.g. cell structure and compartments) make metabolome coverage, particularly for the endometabolome, an issue. As imaging the metabolic inventory of the cell depends on access, it is important that the extraction procedure is consistent and boasts excellent recovery characteristics with limited degradation and losses. Recently, Villas-Boas and colleagues explored the impact of sample preparation on targeted metabolite analysis using GC–MS for the yeast S. cerevisiae. One substantial contribution was their focus on comparative profiling from six different extraction protocols. Although the explicit detection tool was principally directed toward organic and non-organic amino acids, an important outcome of this work was that several extraction strategies appear to result in the same biological story (level of metabolites). Even though their analysis is organism specific, it provides a benchmark for future targeted analysis strategies of several metabolite classes in fungi.

Advances in internal standardization are also paving the way for more robust metabolite measurements. Heijnen and co-workers have recently created a platform for quantitative metabolite analysis, which is independent of ion suppression effects, metabolite modification/degradation during extraction and variations in instrument response. This elegant method holds significant promise for unifying quantitative metabolome analysis. The foundation for this strategy, initially proposed by Mashago et al., is based on the generation of a stable-isotope-labeled metabolome from 13C-saturated microbial cultivations. Labeled metabolite libraries enable correction for extraction recoveries and provide standardization strategies for quantitative measurements. To address several drawbacks from the original methodology, Wu et al. extended this concept through a more generally applicable scheme. Although less universal because limited to nitrogen-containing metabolites, 15N-saturated cultivations have also shown a strong potential to impact metabolome study standardization.

Developments to establish guidelines for reporting metabolomic data and to create public and open-access of mass spectral identification libraries from MS data are also important. These coordinated efforts are expected to make an immediate impact on minimizing variability between researchers, leading to more-accurate biological insight.

Conclusions and future perspectives

Recent progress in the field of metabolite analysis can be attributed to two major driving forces: the need for identification and quantification of natural compounds from complex matrices and the reorientation of biology towards systems analysis. Together, these factors require not only the collection of comprehensive datasets from different molecular levels, but also their integration to ultimately allow the understanding of the phenotype on the basis of the genome sequence and the environment. Because of their integrative nature, metabolome data represent an important milestone on the road towards this goal. The combination of metabolite profiling and targeted analysis through hierarchical metabolomics, as demonstrated by Catchpole et al., and strategies to detect and quantify hundreds of metabolites in one-shot hold significant promise. Despite foreseeable improvements in our ability to quantitatively measure more metabolites using standardized methods and progress in mathematical approaches to identify statistically relevant features, the limiting step for the utilization of metabolome data will be in our ability to develop appropriate frameworks to integrate and map data from multiple cellular levels.

Update

Recent work demonstrated the separation and detection of more than 40 sugars and sugar derivatives from a GC–MS platform. This platform expands the toolbox available for targeted metabolome analysis.

A comprehensive review by Larsen et al. was published, describing not only the importance of fungal natural products and the necessity of intelligent screening for the discovery of new drugs with the help of chemotaxonomy, but also the techniques applied for their analysis.

Three fungal genomes have been published that give insight into the genome evolution of the interesting
genus Aspergillus spp. and show the high diversity of secondary metabolites in these fungi.

Acknowledgements

MCJ, GH, and JN are most grateful to the National Science Foundation International Research Fellowship Program, the EU EuroBloodSubstitutes project (www.eurobloodsubstitutes.com) and the Danish Research Council for Technology and Production Sciences for supporting their work. The authors thank K Patil, AP Oliveira and J Højer-Pedersen for their helpful comments and suggestions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


This paper describes a novel algorithm for exploring metabolic regulation by inference from mRNA global expression and genome-scale metabolic models. By using topological network structure defined from stoichiometry, this approach is useful in uncovering co-regulated structures in the metabolic network.


A seminal review on the use of MS in metabolomics, offering excellent comparisons of sample preparation methods (quenching and extraction) and suggested applications for specific types of MS analyses (e.g. GC-MS, LC-MS, CE-MS).


This work demonstrates an experimental methodology and software platform for the rapid diagnosis of intracellular amino and non-amino organic acids using GC-MS. Of particular interest are discussions relative to normalization. The robustness and sensitivity of this approach offer significant promise for future studies that attempt to integrate metabolism to study or exploit biological complexity.


By comparing the flux distributions of 137 null mutants from Bacillus subtilis, the authors demonstrate fundamental principles of metabolic network rigidity and robustness.


The authors detail a case study revealing the complementary nature of hierarchical metabolomics with physiological data, comparing GM and conventional potato crops. By utilizing metabolite profiling and...
targeted analysis, this work features one of the most complete analyses of metabolic content ever reported. It demonstrates the importance of coupling sensitive analysis techniques with data-mining strategies. Moreover, it illustrates the potential for metabolomics applications in food biotechnology.


37. This chapter highlights the importance of standardizing analytical procedures and the utilization of databases for the identification and dereplication of fungal secondary metabolites.


