

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/funeco

Commentary

Prospects and challenges for fungal metatranscriptomics of complex communities



ABSTRACT

The ability to extract and purify messenger RNA directly from plants, decomposing organic matter and soil, followed by high-throughput sequencing of the pool of expressed genes, has spawned the emerging research area of metatranscriptomics. Each metatranscriptome provides a snapshot of the composition and relative abundance of actively transcribed genes, and thus provides an assessment of the interactions between soil microorganisms and plants, and collective microbial metabolic processes in many environments. We highlight current approaches for analysis of fungal transcriptome and metatranscriptome datasets across a gradient of community complexity, and note benefits and pitfalls associated with those approaches. We discuss knowledge gaps that limit our current ability to interpret metatranscriptome datasets and suggest future research directions that will require concerted efforts within the scientific community.

Introduction

Fungi have significant ecological roles as components of complex microbial communities in many diverse environments – including soil, marine and freshwater habitats, animal and insect digestive systems, and within plant and animal hosts. Perhaps most well known for their contributions to terrestrial ecosystems, fungi are the major decomposers of plant and soil organic matter and form critical nutritional linkages with plants through mycorrhizal, pathogenic and endophytic associations. Playing central roles in nutrient and mineral availability and mobilization, fungi represent some of the most functionally diverse organisms on Earth.

Molecular approaches based on the ribosomal RNA operon or single enzyme-coding genes have been used to assess taxonomic diversity and track fungal community composition. Reverse transcriptase PCR surveys using single enzyme-encoding genes (i.e. cellobiohydrolase; [Weber et al., 2012](#)) have identified specific transcripts in fungal communities that suggest ecological roles for fungal populations. However, single-gene studies are limited by primer selectivity, inability to capture the entire community, and in scope, failing to

assess related functions or metabolic processes. Collectively surveying for all transcribed enzyme-coding genes in an environmental sample would improve our understanding of the metabolic diversity, activities, and community interactions among fungal species and in their associations with plants, bacteria, and other Eukarya.

The ability to extract and purify RNA directly from plants, decomposing organic matter, and soil, followed by high-throughput sequencing of the pool of expressed genes, is an emerging capability through which we may assess the plant–fungal interactions and fungal metabolism in soils. This approach, termed metatranscriptomics, generates a snapshot of the composition and relative abundance of actively transcribed genes at a single time point. Comparison of metatranscriptomes over time, and across environmental variables or gradients, provides information on the collective, interactive metabolism as the community responds to changing environmental conditions. Metatranscriptomic surveys have the potential to be particularly powerful when combined with parallel surveys, such as target-gene sequencing, metagenomics, meta-metabolomics and meta-proteomics. Partnered with information on soil geochemistry, plant species or nutritional status, or other experimental factors, these surveys become part of a comprehensive ecosystem study. Transcriptome and metatranscriptome surveys also facilitate discovery of novel enzymes and processes with importance to bioenergy, medical, agricultural, industrial and ecosystem climate response applications.

To efficiently survey the expressed gene pool, one must first obtain a mRNA sample from which DNA, ribosomal RNA, and impurities that inhibit downstream sample processing have been removed. This has been a significant technical challenge, especially in environmental samples where transcripts are low in abundance relative to the rRNA pool and where co-extracted inhibiting compounds are abundant. Recent advances in this area include rigorous extraction to remove RNAase enzymes and co-extracting contaminants, use of polyA-enrichment (where only the Eukarya are the target of investigation), and removal of rRNA using affinity reagents. Studies have now demonstrated successful extraction and purification of fungal mRNA from soil, plant roots, and decomposing leaf litter (see for examples, [Griffiths et al., 2000](#); [Liao et al., 2014](#); [Weber et al., 2012](#)).

Fungal transcriptomics – from single cultures to complex soil communities

Transcriptomes from single fungal cultures have provided key reference material for genome annotation and allowed the study of active fungal metabolism (for examples, [Hori et al., 2014](#); [Tisserant et al., 2012](#)). Despite these successes, there remain significant gaps in our ability to use transcriptome datasets in their entirety, and there is a significant challenge to glean biologically meaningful inferences from them. These challenges become exponentially more daunting when assessing metatranscriptomic datasets that represent pairs of organisms or more complex communities.

[Fig 1](#) illustrates a gradient of fungal transcriptome complexity from a single cultured fungus (box A) to very complex natural soil systems (C). It notes several features of current metatranscriptome studies, including approaches for analysis and key metabolic questions being addressed. The simplest case is a single fungal transcriptome ([Fig 1](#), box A). A single-culture transcriptome approach involves sequencing both the genome and the expressed genes (e.g. using RNA-seq). Transcripts are identified by mapping sequences to the annotated fungal genome. Genome annotation quality

can be variable, however, with many gene functions unidentified (hypothetical) or identified based solely on sequence homology to other genomes. This has been shown to be an exceptionally useful approach to identify metabolic pathways that are active in the presence of different substrates (See for example [Hori et al., 2014](#)). However, single-organism transcriptome datasets can be difficult to interpret because of differences in transcript turn-over rates and translational efficiencies, and the limitations of using homology-based gene assignments to correctly predict function. Interpretation is greatly improved when transcriptomes are combined with high-quality, accurately annotated genome and proteome information ([Hori et al., 2014](#)).

Metatranscriptome surveys have the potential to identify the metabolic interactions occurring between two or more organisms ([Fig 1](#), box B). An example with relevance to plant health is between a mycorrhizal fungus and its plant partner (for examples, [Liao et al., 2014](#); [Larsen et al., 2011](#)). With more organisms in the mRNA mix, the sequence datasets become more complex, but if annotated genome sequences for both fungus and plant are available, the metabolic signaling, interactions and their regulation may be determined. A notable challenge when simultaneously investigating two or

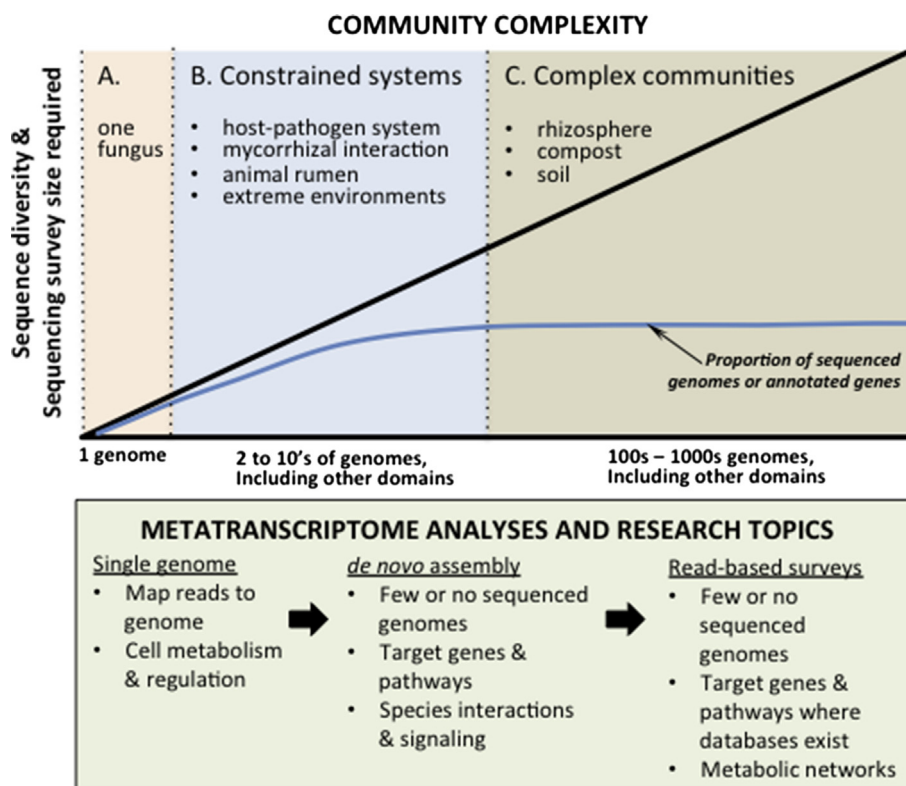


Fig 1 – The figure illustrates the impact of increasing fungal community complexity on our ability to assess and interpret metatranscriptome information. Panel A begins the continuum of complexity with transcriptome analysis of a single organism, where expressed genes may be mapped to a genome and where cell metabolism may be studied through transcriptome profiling. Panel B represents constrained systems where the focus is on a few co-habiting organisms that interact closely. For this level of metatranscriptome study, *de novo* assembly or assessment of target genes and pathways that may influence the interaction are often the focus of study. Panel C represents complex communities such as the soil, where current analyses are often limited to read-based surveys.

more organisms is that the relative abundance of their RNA may be skewed; for example fungal RNA in a mycorrhizal tissue may be under-represented compared to the plant RNA. Although a polyA-enrichment approach is suitable for plant–fungal interactions, soil- and root-inhabiting bacteria also influence this interaction. Studies that include non-eukaryotic organisms would require an rRNA depletion approach that does not rely on polyA-enrichment. Other constrained environments, such as those in ruminant and insect digestive tracts, harbor a limited number of fungal species that participate in breakdown of complex substrates (Lee et al., 2000; Brune, 2014). Because the number of resident fungi is limited in these systems, *de novo* assembly of short transcript reads into longer contiguous sequences (contigs) is possible (Qi et al., 2011). As communities become more complex, the amount of sequence data required to sufficiently sample the community gene expression increases greatly (Fig 1) and a variety of complementary analysis approaches may be needed to identify transcripts (see recent reviews for metagenomes by Desai et al., 2012; Wooley and Ye, 2010). Even in systems with only a few species, *de novo* assembly becomes more challenging and computationally intensive, and data simplification approaches may be necessary. Methods to optimize *de novo* assembly using k-mer binning and other strategies have been successfully employed in constrained bacterial-dominant communities and may, with additional computational power, be viable for depauperate eukaryotic communities (Thomas et al., 2012). As community complexity increases and sequencing coverage decreases, *de novo* assembly is less efficient for generating long contiguous sequences and is confounded by potential chimeric assembly artifacts. In these situations, approaches that do not require dataset assembly, such as mapping reads to known genomes and read-based sequence similarity searches [e.g. translated BLAST, hidden Markov models (HMMs), PFam] provide information on collective community function but are limited by genome availability, incomplete genome annotation (with many sequences remaining unidentifiable), and inability to confidently attribute a function based on a gene sequence motif.

Natural environments such as soil may contain very complex assemblages of hundreds-to-thousands of fungal species (Fig 1, box C). These communities interact with diverse bacteria and other Eukarya and collectively regulate plant carbon decomposition and nutrient cycling in terrestrial ecosystems. Fungi are the key component for initial plant material decomposition via secretion of extracellular enzymes and are thus central to nutrient cycling processes in soils and terrestrial ecosystem function. The diversity and complexity of these communities currently make *de novo* assembly of all individual genomes prohibitive, and ability to mine these datasets for interpretable functions becomes limited to read-based or k-mer based approaches. Read-based searches rely on sequence homology to known genes, and are most informative where high-quality databases containing accurately curated sequences across a wide taxonomic breadth, exist for comparison (e.g. CAZymes, www.cazy.org; Lombard et al., 2013). Unfortunately, sequence homology alone can be a poor predictor of *in situ* enzyme function, and many detected genes have no

attributed function (hypothetical genes). K-mer based strategies that bin sequences by sequence motifs of 'k' length, are useful for comparison of the metatranscriptomes of complex communities, but are more difficult to use as a binning tool when sequence coverage is lower. In examination of such complex communities, metagenomic and metatranscriptomic information is most useful when conducted in an experimental setting, with replicated, defined treatments, or across natural gradients, where direct comparisons are possible to tease apart metabolic state changes in the community (Tveit et al., 2013). Focusing the experimental system on known key components (e.g. mycorrhizal interactions, key species identified from larger ecological experiments) may also simplify the system enough to enhance ability to interpret metatranscriptome information.

Conclusions and future directions

Generation of metatranscriptomes to explore community-level metabolic interactions is a potentially rich but challenging endeavor that promises to advance our understanding of the ecological roles of fungi in many environments. Despite the technical challenges of obtaining environmental mRNA, generating metatranscriptome datasets is relatively simple compared to their analysis and interpretation. Our ability to interpret single transcriptomes and metatranscriptomic data is currently limited by the availability of high-quality, accurately annotated, and phylogenetically diverse genomes; by our restricted knowledge of fungal cell metabolism, where many mapped transcripts remain of hypothetical function and functions attributed by sequence homology may be dubious; and by the lack of curated sequence databases for specific functions. High quality, accurately annotated reference genomes are critical for all transcriptome/metatranscriptome studies, to identify transcripts in individual genomes, to provide information on gene order and neighborhoods, and to provide interpretive power for analysis of complex community datasets. Sequencing efforts such as the '1 000-fungal genomes' project (<http://1000fungalg genomes.org>) are attempting to increase the taxonomic breadth of fungal genome sequence databases and are providing coupled genomes and transcriptomes to facilitate annotation. To be most useful, the genome sequence information needs to be accompanied by gene expression and other experimental data. Such databases (e.g. <http://fungidb.org/fungidb/>) are in development.

Continued emphasis needs to be placed on the cell biology of a broader taxonomic range of fungi to facilitate unambiguous assignment of gene functions and define components of key metabolic pathways, especially for the 'hypothetical' genes. Augmenting metatranscriptome studies with advances in mass spectrometry-based metaproteome surveys will aid accurate functional assignments of transcribed genes (Mueller and Pan, 2013) and reduce our reliance on sequence homology-based approaches. Continued improvement of sequencing platforms to support longer sequence reads will also improve ability to map transcriptional reads to known functions and taxonomic groups.

As our sequence-based datasets become larger and our experiments more complex, improvements in computational speed and dataset handling are needed to facilitate comparative analyses. Overcoming the transcriptome analysis bottleneck will require a concerted effort to improve our ability to accurately assign transcript functions, identify components of key metabolic pathways, and disentangle interactions among organisms in communities. This will undoubtedly involve the creative use of available computational resources as well as the design and development of novel strategies for assigning functions. Continuing to publicly share this information through validated databases and other information resources will promote accurate interpretation of complex metatranscriptome datasets.

Fungal metabolism at the cellular and community levels is regulated at many different points - genomic content and associated local transcriptional regulatory features, post-transcriptional controls imposed on the mRNA, post-translational modification and protein cellular location, and enzyme kinetics in different biochemical environments to name a few. Metatranscriptome surveys provide a global view of community metabolism where abundant taxa and highly expressed genes are likely to be represented, but where low-abundance organisms and genes with low levels of expression may be missed. To be most informative, metatranscriptomic surveys must be combined with assessment of the identity and growth/viability of the community members, knowledge of the local biogeochemical environment, and the activities of expressed proteins.

Acknowledgment

This contribution was made possible by a workshop grant from the Los Alamos National Laboratory, and was supported in part by the U.S. Department of Energy, Office of Science, Biological and Environmental Research, under contract number 2014LANLF260. This is Los Alamos Unclassified Report LA-UR-14-28371.

REFERENCES

- Brune, A., 2014. Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* 12 (3), 168–180.
- Desai, N., Antonopoulos, D., Gilbert, J.A., Glass, E.M., Meyer, F., 2012. From genomics to metagenomics. *Current Opinion In Biotechnology* 23, 72–76.
- Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G., Bailey, M.J., 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA-and rRNA-based microbial community composition. *Applied and Environmental Microbiology* 66, 5488–5491.
- Hori, C., Ishida, T., Igarashi, K., Samejima, M., Suzuki, H., Master, E., et al., 2014. Analysis of the *Phlebiopsis gigantea* genome, transcriptome and secretome provides insight into its pioneer colonization strategies of wood. *PLOS Genetics* 10 (12), e1004759.
- Larsen, P.E., Sreedasyam, A., Trivedi, G., Podila, G.K., Cseke, L.J., Collart, F.R., 2011. Using next generation transcriptome

- sequencing to predict an ectomycorrhizal metabolome. *BMC Systems Biology* 5, 70–84.
- Lee, S.S., Ha, J.K., Cheng, K., 2000. Relative contributions of bacteria, protozoa, and fungi to *in vitro* degradation of orchard grass cell walls and their interactions. *Applied and Environmental Microbiology* 66 (9), 3807–3813.
- Liao, H.L., Chen, Y., Bruns, T.D., Peay, K.G., Taylor, J.W., Branco, S., et al., 2014. Metatranscriptomic analysis of ectomycorrhizal roots reveal genes associated with *Piloderma-Pinus* symbiosis: improved methodologies for assessing gene expression *in situ*. *Environmental Microbiology*. <http://dx.doi.org/10.1111/1462-2920.12619>.
- Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., Henrissat, B., 2013. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42, D490–D495.
- Mueller, R.S., Pan, C., 2013. Sample handling and mass spectrometry for microbial metaproteomic analyses. *Methods in Enzymology* 531, 289–303.
- Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B., et al., 2011. Snapshot of the eukaryotic gene expression in muskoxen rumen – a metatranscriptomic approach. *PLoS One* 6 (5), e20521.
- Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics – a guide from sampling to data analysis. *Microbial Informatics and Experimentation* 2, 3.
- Tisserant, E., Kohler, A., Dozolme-Seddas, P., Balestrini, R., Benabdella, K., Colard, A., et al., 2012. The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. *New Phytologist* 193, 755–769.
- Tveit, A., Schwacke, R., Svenning, M.M., Urich, T., 2013. Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *ISME Journal* 7, 299–311.
- Weber, C.F., Balasch, M.M., Gossage, Z., Porras-Alfaro, A., Kuske, C.R., 2012. Soil fungal cellobiohydrolase I gene (*cbhl*) composition and expression in a loblolly pine plantation under conditions of elevated atmospheric CO₂ and nitrogen fertilization. *Applied and Environmental Microbiology* 78, 3950–3957.
- Wooley, J.C., Ye, Y., 2010. Metagenomics: facts and artifacts, and computational challenges. *Journal of Computer Science and Technology* 25 (1), 71–81.

ARTICLE INFO

Accepted 12 December 2014

Corresponding editor: Björn D Lindahl

KEYWORDS

Environmental RNA
Fungal metabolism
Fungal transcriptome
K-mer
Metatranscriptome
Sequence databases

Cheryl R. KUSKE^{a,*}, Cedar N. HESSE^a, Jean F. CHALLACOMBE^a, Daniel CULLEN^b, Joshua R. HERR^c, Rebecca C. MUELLER^a, Adrian TSANG^d, Rytas VILGALYS^e

^aBioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

^bForest Products Laboratory, United States Department of Agriculture, University of Wisconsin, One Gifford Pinchot Drive, Madison, WI 53726, USA

^cDepartment of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA

^dCentre for Structural and Functional Genomics, Concordia University, Montreal, Quebec, H4B 1R6, Canada

^eBiological Sciences, Duke University, Durham, NC 27708, USA

*Corresponding author. Bioscience Division, M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. Tel.: +1 505 412 7934; fax: +1 505 665 3024.

E-mail address: Kuske@lanl.gov (C.R. Kuske).

1754-5048/\$ – see front matter

Elsevier Ltd and The British Mycological Society.

<http://dx.doi.org/10.1016/j.funeco.2014.12.005>