
Meetings

Fungal networks made of humans: UNITE, FESIN, and frontiers in fungal ecology

UNITE/FESIN meeting, Dragør, Denmark, September 2007

The recent bloom of fungal community ecology can be credited in large part to the development of molecular tools for identification of fungi (Koljalg *et al.*, 2005). These tools have been largely self-assembled, as individual researchers have borrowed techniques from other fields such as molecular systematics and medical diagnostics and applied them to particular studies. As a result, the development of the field has had some resemblance to an unplanned boomtown; growth has been breathtakingly rapid but weak on development of coordinated infrastructure. To address these challenges, researchers have started to assemble into larger networks. Two such groups, the Fungal Environmental Sampling and Informatics Network (FESIN; <http://www.bio.utk.edu/fesin/title.htm>) and the User-friendly Nordic ITS Ectomycorrhizal database group (UNITE; <http://unite.ut.ee/>) held a joint meeting in September in Dragør, Denmark to struggle with the problems of building and maintaining infrastructure for the burgeoning field of fungal ecology.

'The relationship between biodiversity and ecosystem function remains a "holy grail" issue in ecosystem science ...'

UNITE and FESIN

UNITE led the way 5 yr ago by focusing on the problem of unreliable names and undersampled ectomycorrhizal taxa in GenBank and creating a sequence database for the identification of ectomycorrhizal fungi. FESIN is a newly formed group whose broad goals are to make fungi accessible to ecologists by (1) coordinating the development of rapid identification and

information retrieval tools, (2) fostering interactions between mycologists and ecologists, and (3) supporting educational initiatives in fungal ecology. The UNITE/FESIN meeting was coordinated between the two groups because of the strong overlap in goals and interests. Both organizations are funded as research coordination networks, which means that they have funds to organize meetings, but not to do any actual research. It was realized by both groups that, to transcend this constraint, the best outcome of the Dragør meeting would be to spearhead community research proposals. To that end, the first part of the meeting had a strong focus on methods, so that current tools, and their limitations, were in the forefront of participants' minds.

Microarrays

Microarrays provided the axis for much discussion regarding the development of new methods in fungal ecology. Three different types of array were discussed. Gary Andersen (Lawrence Berkeley National Laboratory, CA, USA) presented a talk on development of oligonucleotide arrays for fungal identification. His group had previously developed such tools for prokaryotic taxa using Affymetrix chips (Wilson *et al.*, 2002; DeSantis *et al.*, 2005), and they are now directing their efforts to a fungal chip based on nuclear internal transcribed spacer (ITS) region and large subunit rRNA (LSU) sequences. The advantages of an Affymetrix array approach are manifold. Firstly, chips have a capacity of 500 000 probes in a 1.28-cm² array, which allows multiple probes for each operational taxonomic unit (OTU) as well as mismatch probes to test for specific hybridization. Secondly, a single hybridization experiment can be used for each biological replicate and will simultaneously identify tens of thousands of taxa from mixed assemblages. Thirdly, results from array experiments yield presence/absence determinations for organisms differing in abundance by roughly five orders of magnitude; this method is currently much more sensitive at detecting low-abundance targets than sequencing of clone pools (although see pyrosequencing below). In addition, semi-quantitative changes in abundance can be measured between hybridizations (i.e. samples and replicates). The cost per array is estimated to be \$200 USD, but this is after substantial development cost, and assumes one has access to the equipment needed to read the chips.

Overall, the UNITE/FESIN group saw the Affymetrix chips as an incredibly useful but untapped tool for analyzing fungal communities at large scales, and they discussed ways that the community might be involved in developing a consortium to lower the price and to test the chip across multiple habitats.

The main disadvantage with the identification array is that one cannot identify taxa that are not already known and sequenced. For this reason, it became obvious to the group that one of the best community-wide actions would be to ensure that Andersen's group (Lawrence Berkeley National Laboratory, CA, USA) has access to all available data, so that the chip would be widely useful across diverse study areas.

Functional gene arrays

Chris Schadt (Oakridge National Laboratory, TN, USA) presented an overview of 'functional gene arrays' (FGA) which target protein-coding genes important for ecosystem function. The relationship between biodiversity and ecosystem function remains a 'holy grail' issue in ecosystem science, and FGAs offer a rapid way to compare functional capacity across systems and conditions. In their current form, these arrays are printed 'in-house' on glass slides, and probes are relatively long and therefore tolerate more mismatches than the short oligonucleotide probes used in Affymetrix arrays. Environmentally extracted RNA is used to probe the arrays. This approach provides a view of functions that are expressed in the microbial community, and when coupled with random amplification methods it is sensitive enough to detect activities in even minute samples (Gao *et al.*, 2007). As with current identification arrays, the functional genes are primarily drawn from prokaryotic organisms (Wu *et al.*, 2001; He *et al.*, 2007), with the exception of genes for cellulose and lignin degradation which are heavily drawn from the fungi. However, the potential for extending the coverage of fungal functions is increasing as more fungal genome sequences become available.

Oligonucleotide fingerprinting

James Borneman (University of California Riverside, USA) discussed his oligonucleotide fingerprinting approach (Valinsky *et al.*, 2002). This is essentially a reverse array in which environmental clones are spotted and probed with a battery of oligonucleotide probes. He uses the approach to compare conditions, such as suppressive and nonsuppressive soils, and to pick out key organisms that are enriched by particular environmental conditions. Unlike the other arrays, this method is used for finding 'a needle in a haystack' rather than to enumerate all members of the haystack. It also allows one to identify previously unknown taxa because the probes are only used to screen clones, which are later sequenced if they prove interesting in their patterns of occurrence. The main limiting factor is that it is work-intensive, and the number of clones one can screen is still fairly modest (approx. 9600) – at least in relation to hyperdiverse systems such as soil. To work around this problem, Borneman is exploring the use of 'colonies', which are tiny colonies amplified from single molecules on acrylamide-coated slides (Mitra & Church, 1999). This would increase the screening capacity to over a million colonies per slide.

Sequencing and the future

Advances in sequencing technology are poised to increase the amount of fungal sequence by orders of magnitude in the next few years. It is clear that much of this increase will be from environmental samples – that is to say, sequences that are retrieved directly from complex substrates containing multiple unidentified fungi. Two examples were reported at the meeting. Lee Taylor (University of Alaska–Fairbanks, USA) discussed his work in boreal forest at Bonanza Creek Long-term Ecological Research (LTER) site. His group teamed up with the genome sequencing facility of the Broad Institute at Massachusetts Institute of Technology to produce over 70 000 sequences that included both the ITS and LSU regions. Ari Jumpponen (University of Kansas, USA) made it clear in his talk on pyrosequencing that even 70 000 sequences will be a small number in the near future. Although still costly, pyrosequencing avoids most of the PCR biases, eliminates cloning entirely, and yields megabases of sequence in a single run (Ronaghi *et al.*, 1996; Ronaghi *et al.*, 1998). Length of sequence reads has been a limitation, but improvements in chemistry and protocols are predicted to substantially increase the read length in the near future (Mashayekhi & Ronaghi, 2007). With sequence capacity increasing at such a rapid rate it seems possible that array approaches might be quickly replaced by direct sequencing, although the trade-off between resolution and replication will likely persist until the costs of sequencing drop substantially. Furthermore, it is clear that sequencing capacity has temporarily outpaced bioinformatics approaches; in both Taylor's and Jumpponen's work, one of the main bottlenecks has been analyzing the huge volume of sequences.

Fungal genomics projects represent another huge source of new fungal sequences that will revolutionize fungal ecology. Francis Martin (French National Institute for Agricultural Research (INRA)-Nancy, France) presented an update on these projects focused on taxa that might be of special interest to ecologists (e.g. *Laccaria*, *Tuber* and *Melampsora*). These data will have at least three applications in ecology: (1) elucidating gene expression and signaling between symbiotic partners; (2) inferring the functional capacity of sequenced taxa based on gene content; and (3) deriving population and phylogenetic markers from sequenced taxa and expanding their use to related taxa.

Although the methodological talks painted a bright future for fungal ecology, they also underlined the huge need on the bioinformatics side of the field. Simple tools that allow sequences to be used for identification such as BLASTN, galaxieblast (Nilsson *et al.*, 2004), and emerencia (Nilsson *et al.*, 2005) all require significant human interpretation to avoid inaccurate classification. This is largely because GenBank is filled with inaccurately identified sequences (Bridge *et al.*, 2004; Nilsson *et al.*, 2006) and because BLAST hits are strongly affected by sequence length. The option of expanding third-party databases such as the UNITE database is certainly available, but the permanence of these is not guaranteed and the vast majority

of data will always be in GenBank. Furthermore, it is almost certain that the bulk of future sequences will be environmental samples and thus will not be tied directly to an identified specimen. For these reasons, meeting participants were united in their call for third-party annotation of GenBank sequences (something that is not currently allowed), and for using community-based annotation in a concerted effort to clean up the problems in GenBank. This is a critical need for the field in the near term; existing data need to be corrected and reliable automated classification systems need to be developed before the onslaught of new data makes these goals much more difficult.

Global collaborations

With these methodological advances in mind, further discussion centered on large-scale community ecology projects that would apply such tools at continental or global scales and would involve collaborations among laboratories across Europe and North America. The basic idea was that by scaling up we could achieve a view of fungal systems that is unobtainable by individual research groups. This is likely to be an ongoing discussion over the next several years and there is no reason that it should be limited to development of only one such project, as there are many pressing questions that would benefit from community-wide efforts. The only danger in higher level coordination is that it can stifle innovation of individual researchers: as more resources are funneled into large community-wide projects, less may be available for independent studies. This danger may be avoided by integrating fungal ecology more fully into microbial and ecosystems studies so that it becomes a necessary part of many fields.

A list of participants, individual presentations, and more details on group discussions are available at the FESIN website (<http://www.bio.utk.edu/fesin/title.htm>). All interested are encouraged to attend the first public meeting of FESIN at the annual meeting of the Ecological Society of America in Milwaukee, Wisconsin in August 2008.

**Thomas D. Bruns^{*1}, A. Elizabeth Arnold² and
Karen W. Hughes³**

¹Plant and Microbial Biology, University of California
Berkeley, Berkeley, CA 97720-3102, USA;

²Plant Sciences, University of Arizona,
Tucson, AZ 85721, USA;

³Ecology and Evolutionary Biology, University of Tennessee,
Knoxville, TN 37996-1610, USA

(*Author for correspondence:

tel +1 510 642 7987; fax +1 510 642 4995;
email pogon@berkeley.edu)

References

- Bridge PD, Roberts PJ, Spooner BM, Panchal G. 2004. On the unreliability of published DNA sequences. *New Phytologist* 161: 15–17.
- DeSantis TZ, Stone CE, Murray SR, Moberg JP, Andersen GL. 2005. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. *FEMS Microbiology Letters* 245: 271–278.
- Gao HC, Yang ZMK, Gentry TJ, Wu LY, Schadt CW, Zhou JZ. 2007. Microarray-based analysis of microbial community RNAs by whole-community RNA amplification. *Applied and Environmental Microbiology* 73: 563–571.
- He ZL, Gentry TJ, Schadt CW, Wu LY, Liebich J, Chong SC, Huang ZJ, Wu WM, Gu BH, Jardine P *et al.* 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *Isme Journal* 1: 67–77.
- Koljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjöller R, Larsson E *et al.* 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* 166: 1063–1068.
- Mashayekhi F, Ronaghi M. 2007. Analysis of read length limiting factors in pyrosequencing chemistry. *Analytical Biochemistry* 363: 275–287.
- Mitra RD, Church GM. 1999. *In situ* localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Research* 27: e34.
- Nilsson RH, Kristiansson E, Ryberg M, Larsson KH. 2005. Approaching the taxonomic affiliation of unidentified sequences in public databases – an example from the mycorrhizal fungi. *BMC Bioinformatics* 6: 178.
- Nilsson RH, Larsson K-H, Ursing BM. 2004. galaxie – CGI scripts for sequence identification through automated phylogenetic analysis. *Bioinformatics* 20: 1447–1452.
- Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Koljalg U. 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *Plos Biology* 1: e59.
- Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* 242: 84–89.
- Ronaghi M, Uhlen M, Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281: 363–365.
- Valinsky L, Della Vedova G, Jiang T, Borneman J. 2002. Oligonucleotide fingerprinting of rRNA genes for analysis of fungal community composition. *Applied and Environmental Microbiology* 68: 5999–6004.
- Wilson KH, Wilson WJ, Radosevich JL, DeSantis TZ, Viswanathan VS, Kuczmarski TA, Andersen GL. 2002. High-density microarray of small-subunit ribosomal DNA probes. *Applied and Environmental Microbiology* 68: 2535–2541.
- Wu LY, Thompson DK, Li GS, Hurt RA, Tiedje JM, Zhou JZ. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Applied and Environmental Microbiology* 67: 5780–5790.

Key words: ectomycorrhiza, FESIN, fungal ecology, GenBank, microarray, sequence, UNITE.