SUMMARY – FESIN/UNITE Joint Meeting

Rapid molecular identification technologies and their development for assessing fungal community structure and dynamics

Dragør, Denmark
Sept. 13-15, 2007

Members of the NSF-supported FESIN research coordination network met with members of the UNITE network to discuss methods and technology development for rapid assessment of fungal community structure and dynamics. The meeting was held at Dragør Badehotel, near Copenhagen, Denmark. Thirty-three participants were involved, representing 12 countries and 23 institutions.

The meeting involved both formal presentations as well as extensive formal and informal discussion centered on (1) current challenges in identifying fungi in the environment; (2) current tools for those purposes; (3) principles that should guide the development of new tools; (4) special opportunities for joint papers and community-wide policy changes; and (5) progress toward funding opportunities in fungal ecology.

Many presentations are available online through the FESIN website. The notes below summarize the main points of each presentation as well as the discussion that followed each talk. Please send comments, corrections, and suggestions for these notes to Betsy Arnold at Arnold@ag.arizona.edu.

THURSDAY, SEPT. 13

1. Betsy Arnold: Introduction to FESIN.
Betsy highlighted the origins, goals, and specific meeting plans for FESIN, the Fungal Environmental Sampling Informatics Network – an NSF-funded research coordination network in fungal ecology. She highlighted the success of Deep Hypha (fungal systematics RCN, led by Meredith Blackwell, Joey Spatafora, and John Taylor), and the rationale for FESIN. FESIN's goals are to coordinate and facilitate information retrieval for fungi in environmental settings, and to expand the field of fungal ecology through educational innovation. Formal meetings through 2011 will alternate between the Mycological Society of America and the Ecological Society of America. Our first formal meeting in the US will be held with ESA in Milwaukee, WI in 2008, and will involve a symposium-style meeting and a workshop focusing on nucleic acid-based identification of fungi in ecological settings: current limitations and future directions. Future meetings will focus on connecting sequence data with the ecology of fungal species (MSA 2009), –omics in fungal ecology (ESA 2010), and cultivating a new generation of fungal ecologists (MSA 2011). FESIN is open to all interested participants.

Discussion. David Hibbett noted that it's important to remember that ecologists are not just consumers of taxonomy: ecology can push mycology in new and useful directions.

This presentation is available online.

2. Urmas Kõljalg: Introduction to UNITE.
Urmas introduced UNITE (User-friendly Nordic ITS Ecomycorrhiza database) and the UNITE group, of which the database is the best-known product. UNITE grew from seed money obtained in Höör in 2001, followed by a 2002 planning meeting in Tartu. Funding from 2003-2005 was obtained through Nor-Fa, and was extended until present. Activities include twice-annual meetings, as well as courses in the identification of ectomycorrhizal fungi.

Problems that motivated the project included (1) ID of ectomycorrhizae through RFLP not always reliable (2) general dissatisfaction with 'holes' and 'garbage' in GenBank; (3) disconnects and redundancy among overlapping in-house databases in many disparate labs; (4) limited collaboration between ecology/taxonomy. Many of these issues are in concert with those inspiring aspects of FESIN.

From the experience of the UNITE group, any database should have an array of desirable features: (1) rigorous quality control of sequences; (2) only vouchered and expert-identified specimens flagged as reliable; (3) more useful annotation than in GenBank; (4) inclusion of in-house sequences that could be searched but not taken (locked); (5) alternative ID tools beyond BLAST. Urmas further noted that by limiting the scope of the project to

EM from northern Europe, the project was tractable. The current UNITE database holds 2511 sequences, which represent 1046 species (>90% of EM spp. in North Europe).

Looking ahead, UNITE will seek funding in 2008, with a continued focus on fungal ecology – but with a relaxation of the geographic, taxonomic, functional, and gene restrictions previously in place for the database. Urmas notes that the database, which was developed for ecologists, could grow to better serve the needs of taxonomists – and should benefit from maintenance/curation by taxonomists.

Discussion. Jean-Marc Moncalvo noted that the Consortium for the Barcoding of Life (CBOL) in Guelph is still only accepting COI; population of a database with other genes is an open discussion. Significant problems with ITS remain, but perhaps by approaching Genome Canada/etc. we may have a long-term way to maintain the UNITE data. Betsy Arnold asked about long-term provisions; Urmas Köljalg responded that the database lives at Tartu and can be maintained there for a period if funding is not obtained. Rytas Vilgalys asked about the proportion of GenBank seq data that could be at the UNITE standard, but this was not known. Infrastructure is a major bonus of GenBank, but the limited fields/filtering/second-party annotation are major issues. Tom Bruns suggested an open letter to GenBank to argue for 3rd party annotations. Rytas seconded the letter concept.

3. Gary Andersen: **High-density oligonucleotide microarrays for microbial community characterization.** Gary illustrated the principles of high-density oligo microarray approaches to characteriziiing microbial communities through the example of bacteria, focusing on 16S data. Drawing from over 200,000 available sequences, Greengenes was used to select long sequences and to chimera-check them, and provided a means to update holdings every 1-2 weeks. A database of high-quality sequences was used to make an array.

Gary described the Affymetrix Phylochip, which allows a massively parallel study effort: 500,000 probes in a 1.28 cm2 array. The chip allows identification of multiple species in a mixed population, and importantly, incorporates multiple probes for each OTU as well as central position mismatches to ensure specific hybridization. Each chip = one biological replicate, and Gary's group recommends three replicates per use. Work by Todd DeSantis is leading to the development of a chip with 2.5 million probes.

Gary gave an overview of their work examining bacterial communities in air samples in two Texas cities. He discussed the importance and relevance of comparing clone data with chip data – for example, the apparent presence of *Nitrospira* (according to the chip) was then confirmed by cloning with specific primers. In response to a question from James Borneman, Gary stated that the sensitivity range of the chip spans five orders of magnitude.

In fungi, 18s is not variable enough for this work, leading Gary to focus on ITS and the D1/D2 region of 28S. Gary highlighted several issues: (1) high quality sequences are needed; (2) many published sequences have ITS or D1/D2, but not both. Gary noted that he and Todd DeSantis had begun work on a fungal chip.

Discussion. Tom asked whether ITS is problematic for probe design: can software developed for 18S be used for ITS? Gary said that Todd has developed software to address this. Jean-Marc pointed out the issues with databases that have only ITS2, for example, or otherwise lack anchor points. Gary noted that the issue with alignment – traditionally very problematic for ITS -- may be less critical in the future than it is now thanks to alignment-free methods. James asked what we really want as a community; Tom says that we want species- or near-species-level matches. A discussion of ITS limitations ensued, along with general feeling that ITS+D1/D2 is the best way to go for now.

The group then focused on ITS sequences in GenBank, noting that there are ca. 67,000 ITS sequences for fungi, but that without curation, their use is limited. Lee Taylor noted that most of those sequences aren't environmental samples, which was surprising to some. Rytas pointed out the problem of environmental samples clustering with environmental samples. All were reminded that even well-curated environmental samples can be very useful. Gary noted that environmental sequences are important for helping to test the array.

Prototype microarrays, including the Nimblegen prototype, were briefly discussed, and the point was raised that DNA or RNA would be ok for such a tool, depending on the questions.

In response to questions from the group, Gary indicated a cost of about $150 per microarray (up to $400), plus $50 for enzymes, etc. About a month is needed to print an array, and six months are needed to design it. ABI arrays are expensive up front, but not too expensive thereafter.

4. Chris Schadt: **Functional gene arrays.**
Chris suggested that understanding fungal communities depends on our ability to identify, functionally characterize, and quantify these organisms. He outlined triumphs and pitfalls of microarray approaches, referring to probe specificity issues given the large number of unknown fungi, but also the ability to quantifying relative abundance of fungi thanks to the linear response of signal intensity to DNA abundance.

He outlined three general methods: phylogenetic oligonucleotide arrays – POAs – as per Gary and Todd's work – which detect and identify based on phylogenetic signature genes; functional gene arrays – FGAs – which detect genes associated with various processes of environmental interest; and community genome arrays (CGAs). His talked focused on FGAs, highlighting the work of Wu et al. 2001 (AEM), including novel programs for probe design, data mining, etc. (e.g., GeneDownloader; Commoligo).

Array work at Oak Ridge involves in-house printers that top out at about 24,500 probes/array, with a one-year lifetime of the array after spotting. The cost runs about $250,000 up front; slides cost about $10; and an additional $50-60 is needed for labeling with enzymes, etc. One special target lies in capturing low-abundance samples, and limiting bias in amplification. Chris talked about optimized methods such as phi 29 polymerase random amp DNA. This presentation is available online.

5. James Borneman: Oligonucleotide reverse probing arrays.
James discussed suppressive soils (soils that suppress nematode infections), and his approach to trying to understand the underlying communities and species that might confer that suppressive effect. He described his 'inverted array' approach, which uses an oligonucleotide fingerprinting method for rRNA genes (OFRG), and highlighted the many advantages of this method. One important advantage is that OFRG can identify rRNA gene sequences that haven't been previously described, and cloned sequences are available for downstream use. Also, up to 3kb fragments can be analyzed. Disadvantages are that it is labor-intensive, the array is not reusable, the method yields fingerprints rather than nucleotide sequences, and there's a limit of 9,600 clones/array. James then discussed his work with 'polonies' – colonies of PCR amplicons derived from a single molecule of DNA. This puts library construction and array building into one process, allowing millions of genes to be analyzed on a microscope slide. He concluded with a discussion of PRISE – which allows design of sequence-selective PCR primers.

This presentation is available online.

6. Henrik Nilsson: Advances in automated sequence identification.
Henrik provided a compelling plot of the accumulation of sequences, and unidentified/insufficiently identified sequences in GenBank, as well as an array of data regarding problems with these sequences: 10-21% of species names are compromised for fungi, for example; 12% have IUPAC ambiguities; 40% of entries are not up to date. Only 18% have explicit voucher data; explicit geographical data (at any level) are given only for 46%; explicit names for collectors or determinants are available for only 2%. Henrik then gave an overview of the UNITE approach, including modified BLAST searches. He described emerencia, which lists, for all fungal ITS, the best non-self BLAST match, and is updated weekly, and offers a unique email service to indicate the recovery of a better match.

This presentation is available online. His paper on taxonomic reliability of GenBank sequences is available on the FESIN web site under "GenBank Issues."

FRIDAY, SEPT. 14

7. Ari Jumpponen: Pyrosequencing.
Ari presented an overview of pyrosequencing, starting with the history of the method and company ownership of the technology. A flash demo of the method is available at www.biotagebio.com. Massively parallel sequencing uses beads with ssDNA template (one strand per bead), emulsion PCR, and subsequent

movement of ‘microreactors’ (bubble with PCR reagents) into picotiter wells (ca. 400k wells/plate). These are packed with packing beads and reagents, and then pyro-sequenced.

The pros of this method are that it is expedient; high-throughput (>200k); and is predicted to yield sequence length of >200bp (500bp) in the near future. Plates can be partitioned to yield multiple replicates per pate, or replicates can be tagged with different primers and sorted out later. But the process is expensive, the outputs currently result in short reads, and the bead-distribution method is not always successful. Reagents are costly ($9000), bioinformatics are fairly cumbersome, and at present there is a high error rate (0.0098). That said, the method was generally viewed as having great potential, especially in years to come. Ari advertised several potential project ideas, which were received with enthusiasm.

Discussion. James asked about the length of reads; Ari says that within a year, 800 bp may be possible (more likely 400-500bp). Tom asked about the sampling intensity for saturation; Ari talked about some biases in the method (PCR biases, etc.), and clarified that only the amplicons, not the soil fungi in his research example, were sampled to saturation. Tom indicated that connecting LTER sites with a method like this is a good idea.

This presentation is available online.

8. Francis Martin: Ecological applications of fungal genomics.
Francis discussed the linkages between whole genome shotgun sequencing to the ecogenomics of ectomycorrhizal symbiosis, and highlighted parts of a larger project looking at the sequence of the *Populus* community genome (including *Laccaria, Glomus, Tuber, Melampsora, Pseudomonas fluorescens*). He noted that comparative genomics for fungi of different lifestyles allows us to ask ecologically relevant questions – for example, what makes a good EM fungus?

Focusing on *Laccaria*, which is being worked on by a consortium of ca. 50 people, he highlighted aspects of the genome, including lineage-specific multigene families; the *Laccaria* secretome; and *Laccaria* as a saprotroph. The genome, for example, is organized into 10 chromosomes, with an overall genome size of ca. 65Mbp. Of this, ca. 37% is coding; there are 19,128 putative proteins; 14,464 of these have known homologs in other species, but ca. 25% are unique to *L. bicolor*. To date, 6515 genes have been confirmed by EST; at least 17,000 genes have been supported by NimbleGen oligoarrays. A large proportion have no homolog in *Phanerochaete, Coprinopsis*, and other fungi. The transcriptome of *Laccaria* is quite different for mycelia, mycorrhiza, and fruiting bodies.

Laccaria appears to have undergone gene loss for the arsenal of tools needed to degrade plant cell walls – but it retains a high capacity for degrading non-plant oligo- and polysaccharides. In particular, it has a very striking ability to degrade plant and non-plant proteins (a large set of secreted proteases). Symbiosis genes families were specifically amplified in *Laccaria* – including several mycorrhiza-induced proteins (small cys-rich proteins, sensor proteins, and receptors, gpcrs, gtpases). Genes for litter degradation are present in *Laccaria* as well but the number of cellulose and lignin degrading genes is reduced relative to saprobes. Francis highlighted a new project – genome sequencing of black truffle (*Tuber melanospora*), and noted that genomes can be used to select genes for other uses – e.g, population genetics using functional markers.

Discussion. Tom noted that the genomics revolution is rolling ahead. In an ecologically focused group, he said there are many interfaces. Are there things that a group like this could complement or facilitate Martin’s kind of work? Could we accelerate things in any way with our contributions? Rytas noted that ecologists are still counting species, and raised the point of gene sets in the environment – which ones? Chris Schadt seconded the idea that genomes could be helpful to us in this way. Francis noted that *L. bicolor* is not so common in Europe, but other conspecifics are, creating some interesting opportunities to examine related species. Rytas asked about oxidases and lignin degradation, followed by Tom’s question about whether lignin-degrading genes are also low (yes). David notes that this talk reminds us to be careful about the assumptions we make regarding ecological roles of fungi – many have multifaceted roles. That led to a discussion of bacterial degradation work by G. Thorne; bacterial degradation may be very common as a nutritional mode. Francis wonders if this could be tied to interactions between *Laccaria* and *Pseudomonas.*

9. David Hibbett: **Deep Hypha and AFToL 1/AFToL 2.**
David provided an overview of the achievements of AFToL I, which include the generation of 5000 sequences representing 7 genes and covering 2087 species, as well as the record of 41 subcellular characters from 30 species. As of submission of the AFToL 2 grant, 49 articles were in print and 14 in press/in review; now there

are over 60 articles now from the project. AFToL was a multi-laboratory collaboration made possible by broad support from the taxonomy community. Consequently, project leaders had to work through communication issues among labs, the obligation to provide tangible benefits to the community, and the obligation to work transparently and share data – and also had to balance the interests of individual students, postdocs, and labs vs. the AFTOL consortium and the broader mycological community.

Along these lines, David highlighted some things that worked, including the data release policy, which was very open and involved a web-accessible database. The data-sharing approach was non-traditional behavior for systematists. Also, the AFToL policy was one of inclusive authorship, and this was quite effective. Tom asked about potential conflicts of interest given that inclusive policy: who can review papers or grants? David said that NSF understands the new policy of multi-authored publications, and what constitutes a conflict of interest is thus less stringent now. Additional successes include the classification paper (Hibbett et al.), and the phylogeny issue of *Mycologia* [Mycologia 98(6) 2006. A phylogeny for kingdom fungi] was great. Tree of Life web pages were promised, but most haven't yet been posted; David discussed the need to motivate the preparation of ToL pages, but said that the ways in which one can motivate people to complete these pages is limited.

The new AFToL (2) spans 10 labs and is centered on phylogenomics, with a significant outreach component. The group has identified 80 putative loci – with the goal of selecting 25 genes for 160 species – and will explore additional methods instead of traditional cloning/sequencing/etc. David highlighted Hal, which data-mines from genome sequences, and he discussed mor, which provides automated phylogenetic analysis, etc., pioneered by Henrik Nilsson in David's lab.

Discussion. Ari asked whether a large FASTA file could be uploaded and then be assembled into clusters by mor; David said that mor can work with GenBank, but that Ari's scenario represents a long-term goal. David then described relationships between Wasabi and Mesquite. He noted that six postdoc positions are currently available. Karen Hughes asked about laccase genes for phylogenetic analysis; David says that they are not using laccase genes and instead are focusing on single-copy genes. Jean-Marc asked about bioinformatics; David said that bioinformatics will be led by François Lutzoni and Frank Kauff, while David's own lab will start with data first and then transition to bioinformatics.

This presentation is available online.

10. Urmas Kõljalg and Karl-Henrik Larsson: UNITE's future plans.
Urmas reminded the group that UNITE is now open to sequences from other regions, and other ecological modes. He explored the importance of linking environmental sequences to names, and used a study of *Tomentella* and *Thelephora* to illustrate one approach. In that case, sequences were added to a large, existing alignment, and various cutoffs were used to determine species boundaries. He noted that for many fungi, there may never be sequences from fruitbodies, but the sequences are still important for capturing the diversity of the group.

Todd asked about linking soil data to sequences, and searching by soil characteristics. David noted that ontologies for ecological descriptors could be a major contribution of FESIN, and this was agreed to as a long-term goal. Rytas added air, oceans, and other ecological milieux to the list. Todd noted that there are difficulties in determining ecological parameters for bacteria. David noted that FESIN can play a large role in standardization, perhaps mirroring the international agreements on soil characteristic terminology, etc. Lee asked about uploading of large datasets to UNITE, and the ability of the database to harbor large datasets; and the importance of both was acknowledged. Jean Marc asked about inclusion of chromatograms; Urmas mentioned that the 'pictures' field in the UNITE database can be used to show those.

11. Ursula Eberhardt: The barcoding scene
Ursula gave an historical overview of the barcoding movement, and highlighted the occurrence of six major meetings on fungal barcoding in 2007. She noted that there is variation in appreciation of barcoding: although the perspective on barcoding was very positive at the Canadian meeting, but other meetings have featured less enthusiasm. Ursula gave an overview of BOLI – the Barcode of Life Initiative, organized by the Consortium for the Barcoding of Life (CBOL) – and reminded all of the functional definition of barcoding: 'DNA barcoding is about using short, standardized DNA sequence from a previously agreed upon position of the genome as a diagnostic tool for the identification of species.' While COI (coxI) was the original focus, COI is no longer the only locus; this was adopted early on as a way to sell the idea. Several members of the meeting mentioned the All Fungi Barcoding meeting, held in May of 2007 (Virginia); mycological community members who were there

focused on ITS data and voted to make ITS the fungal barcode, but it's not clear where that recommendation has ended up (there is no mention of ITS data on the various barcoding websites).

Ursula then highlighted iBOL, the new international initiative, which seeks $150 million through contributions from various countries and is part of the drive to sequence (barcode) one million specimens. It was not clear who would be doing all of that barcoding. Chris noted that we all barcode by default, so why should someone be in charge per se?

Ursula described the Consortium of European Taxonomic Facilities, and EDIT -- the European Distributed Institute of Taxonomy. Collaborations across Europe can be challenging due to language barriers and tradition, but new collaborations such as these are promising. EDIT is linked with the barcoding movement, but is not presently engaged in that, per se, nor offering funding for that work.

Discussion. The topic of barcoding is a sensitive one in several ways, and discussion was fairly passionate. Gary noted that capacity for sequencing can be high, but that the cost for collecting one million specimens is very high – and there is no obvious source of funds to do the collecting or curating of the organisms. Rytas noted that CBOL proposed that the Consortium would be more rigorous than those microbial people who were already doing 'barcoding' by default. Rytas feels that there's too much promised by the barcoding administration – and concluded that scientifically, the movement is flawed. Jean-Marc noted that the funding can be helpful to non-barcoding endeavors, including systematics. David asked what the relationship is between barcoding and the Encyclopedia of Life; no one seemed to know. Also, to Ursula's knowledge, prokaryotic microbes are not included in the barcoding movement.

12. Lee Taylor: Sequencing the boreal forest.
Lee described his work examining soil clone libraries from boreal forest in Alaska. He introduced the history of Long Term Ecological Research sites, and gave an introduction to Bonanza Creek. He noted that T-RFLPs are problematic, and described their efforts with sequencing ITS and the first part of the 28S. They've sequenced about 70,000 ITS+LSU clones. He presented data from 99 soil cores (2000 clones); they found about 400 OTUs in these samples, which came from a 150X150m plot. When OTUs were based on 97% ITS sequence similarity, >50% of OTUs were found only once. Lee discussed limitations of presence/absence data, citing Chao and Colwell (who stated that presence/absence comparisons are biased toward finding marked differences in communities that might not be that different). Ari suggested that T-tests for taxon frequency to compare communities might be helpful.

Lee addressed several questions that are important for ecologists doing this sort of work. For example, are fungal PCR clone libraries biased? Lee's group looked at the Topo-TA approach: samples have to have TA-overhangs to be captured, but certain sequences are less likely to get the overhangs. Also, some cloned inserts cannot be recovered because they are lethal to bacterial hosts. that's not likely, he suggested that it was important to posit as a possibility. His group tested Lucigen blunt cloning, with a wide insert range; this does not require the TA overhang, and prevents transcription of the insert. They examined 300 clones generated from each method and found that rank abundance data were consistent – and that the same major taxonomic groups were represented by each method. His group also established artificial communities of fungi from different phyla and attempted to recover them using tagged markers and additional primers, and found consistency with the actual community.

Discussion. Karen asked about above-ground ground-truthi of the clone libraries. Lee mentioned that the correspondence between fruitbody results and below-ground cloning was low. Chris asked about the Broad Institute, which provides the sequencing, and Lee described his relationship with them. Tom asked about an update on the status of the other 68,000 clones or so; Lee says things are chaotic, reflecting some real limitations on the bioinformatics side. Gary suggested that their pipeline might be helpful. Tim asked about the singleton sequences – how ecologically relevant might they be? This remains to be evaluated. Lee mentioned that he prepared some specific primers and has found a variety of 'long-branch taxa;' they seem to be there, but their ecological significance has yet to be clarified.

This presentation is available online.

SYNTHESIS DISCUSSIONS AND BREAK-OUT SESSIONS

Ian Alexander helped to synthesize the major topics that had emerged from the presentations and group discussions, and with Tom, assigned leaders for break-out sessions focusing on each topic:

- Affymetrix arrays: How can the FESIN/UNITE communities contribute to enhancing and validating that effort? Gary Andersen and Lee Taylor led this group.

- Thresholds for delimiting ITS based 'species' – is it possible to have standards? Working group with recommendations? Ursula pointed out that several barcoding papers have come out with discussions based on various loci. However, ITS data for fungi haven't been dealt with. Urmas and Karen led this group.

- GenBank: curation, historical annotation; third party annotation; integration/incorporation into the UNITE database (etc.). What could we as a group suggest, or do, to help improve the GenBank situation? David Hibbett and Henrik Nilsson led this group.

- Nomenclature, ontology, and curating new environmental sequences: redundancy, artifacts, etc.cThis prompted some discussion and disagreement about the utility of this topic. Tom reminded us of the presence of this topic on a future meeting agenda, and it was left for future discussion.

- What can the FESIN/UNITE community bring to the functional analysis of fungal communities? Interactions with the genomics community? Possible projects? This was met with much enthusiasm, and discussion was led by Chris Schadt and Francis Martin.

- What are the big questions in fungal ecology for the next five years? This group was led by Rytas Vilgalys and James Borneman.

Discussion of each major topic:

1. Affymetrix arrays
Gary outlined the rationale for focusing on ITS, D1/D2 of the 28S, and possibly a portion of the 18S. He mentioned that private sequence data – as well as those from major initiatives (e.g., AFToL) or collections (CBS) are welcome for this effort, including environmental and named sequences. In order to avoid scooping anyone, a method or protocol for datasharing will be important. Probe design, bioinformatics tools, and potential applications in food, medicine, etc. were described – for example, Phylotrack. A call for sequences, published in *Inoculum*, was suggested, as was a mention in the *New Phytologist* summary from the present meeting. Mentions of this work could be made in the ESA newsletter, as well as that of the BMS and BES.

2. ITS-based taxa
Urmas presented the group's feeling that species can't be based on ITS only, but instead, on biology *sensu lato*. We can't use single threshold values for estimating species boundaries. As taxonomists, we need to think about a functional ITS taxonomy for ecologists. How can we do this in the context of knowing about the limitations of ITS and, for that matter, any single-locus, nonphylogenetic approach? Urmas outlined automated identification tools. Given what's available, both in terms of sequences and taxonomy, and highlighted the importance of pre-made alignments, prepared by taxonomists, into which ecologists can fit their sequences. This approach, while appealing, worried some in terms of its feasibility. It's also problematic to estimate percent sequence similarity from the resulting phylogeny -- most phylogenies that aren't densely sampled, and our samples may inform the relationships there – and perhaps dissolve the species concepts that we think we see. Betsy noted that identification and enumeration are different, but intertwined, goals.

3. GenBank
David stated that (1) fungal ecologists need to know that their top BLAST hits are correctly identified; (2) data need to be permanent; (3) garbage needs to be cleaned up: many sequences are mis-identified; others lack vouchers; and data are hard to filter by various criteria – it's hard to find the right field. Potential solutions to the issues raised by the discussion include (1) an effort to force/cajole/persuade Genbank to allow third-party annotation; this would help, but it's not the whole solution. Who will annotate all those records? How will we know what sequences are updated? The process could add noise – and how feasible is it, given that right now people don't update according to publications for their own sequences?

Several group members suggested that as a community, we could help set standards for future submission: standards for field use, strain number, accession number, etc. could be helpful and cheap fixes. Similarly, we could expand the link-out capability, linking to databases like UNITE and MycoBank. Could link-out specificity (so you know where you'll go when you click) be improved by a community lobbying of GenBank? Along these lines, there's a new barcode identifier in GenBank, but it's controlled by CBOL and presently will only include COI. Could we as a community have an arrangement with CBOL, or with GenBank, as a means to get some of the attributes we want? It was noted that no single database or tool can provide all of the kinds of information needed to inform the user about that sequence.

Although validated, rigorously identified sequences are the goal, it was suggested that correcting/updating records in GenBank may be the wrong direction; instead, a separate database, like UNITE, could be a 'clean' supplement to GenBank, and would draw from GenBank (which would sesrve as the long-term depository). For high-quality validation, David noted that AFToL networks can help – but that the community would need easy tools to validate records. It is important that such databases include multiple, validated sequences per taxon. We may want to include type or anchor sequences, and variation in taxonomic concepts/species names should be accommodated: names should not be based on a single, restrictive recognition criterion.

Following discussion, the group concluded that there really are two issues: one of Genbank itself, and the other on the basis of how we can improve and better use the data. We can put guidelines on that, perhaps. Tom thinks we could solve the problem of who gets to annotate.

4. Functional analysis of fungal communities

Chris presented thoughts on moving functional gene approaches into fungal ecology. He focused on strategies for informing and lobbying genome sequence efforts to help define target taxa, in particular by using ecological and evolutionary rationales. Project proposals would be another product that could grow out of our group, including building datasets to start to design conserved primer sets to look for target genes and processes. Rytas asked if it would be possible to generate a list of the most interesting genes as of now, and this was judged to be possible in a rough sense.

The group generally felt that this area is an important frontier – and moving genomes into the field will be exciting. Different life stages of fungi and plants are especially tractable and exciting for controlled, comparative work. Lee discussed the potential to propose fungi for genome sequencing at the Broad Institute.

SATURDAY, SEPT. 15

5. Big questions in fungal ecology

Rytas gave a report of the ecology group's perspective. Major questions identified by the group included the following: (1) How can we link ecological diversity with phylogenetic- and species diversity? (2) Can we develop predictive ecology – e.g., making predictions based on genomics, and testing those empirically at the species and especially community level? (3) Can we address the rules of community assembly for multi-lineage communities, including stability, remediation, temporal and spatial context-dependency, etc.? The group suggested that diversity surveys are really important, but probably need to be packaged differently to be funded. Large-scale projects, and taking advantage of existing infrastructure (e.g., long-term manipulative experiments, or the 50ha tropical forest dynamics plots or LTERs) will be useful. A series of hypotheses that could be addressed with new tools/perspectives: the everything is everywhere hypothesis; the nature of latitudinal and altitudinal gradients, the relationship of diversity to stability, etc. A series of useful themes might include microbial/fungal ecology in the context of the biodiversity crisis, planetary/soil/human health, etc.

It was proposed that Meredith Blackwell and Cathie Aime attend the NEON meeting at LSU in February to represent FESIN. The group felt that we should target other society meetings to ensure that we're present. Tom expressed some enthusiasm about the Tilman plots: to what degree would changes in plant diversity be coupled with fungal surveys? Rytas brought up all-taxon surveys of national parks as an example of ways to package diversity surveys. Gary noted that the Sloan foundation might represent a useful source of funds. Rytas noted that Sloan funded the barcode meeting. Meredith noted that a colleague at LSU has a fish project funded by Sloan; Meredith will look into the contact. Gary reviews grants for Sloan and offered to help; he and Meredith with follow up. It was noted that astrobiology seems to have lost funding; that became, at least at NSF, the MIP/MO program(s).

Tom suggested a change to the 'everything is everywhere' to the topic of beta diversity; rather than going after a microbial strawman, we could instead consider many theories of beta diversity, including those currently applied to plant communities.

Tom asked about actionable items; Betsy reported on the idea ITS synthesis paper. Tim suggested other loci as well. Rytas asked about Tom's review paper and metadatasets. Lee suggested that to get at community assembly, a phylogenetically hierarchical approach would be useful. Lee conveyed some confusion about the interplay of meta-analysis and future proposals. Tom noted that most beta-diversity work to date is based on the species level; what about other phylogenetic levels of organization? David raised the point that ranks are problematic and will lead to criticism; having chronologically defined clades, based on phylogenetic structure, to make relevant comparisons would be powerful, and it's possible that a dated tree could come from AFToL II. Tom noted that such a tree would represent a good relationship between ecology and systematics. David is familiar with that literature – relaxed clock dating approaches, etc., and may be a resource in the future. Lee will share a few papers on the topic of beta diversity in animals and plants.

FINAL SESSION

Tom addressed the issue of the FESIN steering committee, which will organize future meetings. The FESIN PIs will be on it; Tom targeted several participants at this meeting, and recruited Jeri Parrent (among other potential participants). The final decisions will be made in the next few weeks. We'd like to have a UNITE presence, but funding for that is difficult to obtain.

Tom then addressed actionable items, and assigned leaders/core people for each:

1. Generating a report (full/long version) for the FESIN website and perhaps, in a distilled form, for society newsletters (*Inoculum*, ESA, etc.): Betsy, Tom, Karen.
2. Generating a report (short version) as a forum piece for *New Phytologist*: Betsy, Tom, Karen.
3. Preparing a letter about third-party annotation, and seeking support from large societies, etc., as a prelude to publishing/submitting (e.g., to *Science*): Martin Bidartondo.
4. Broadcasting a call for unpublished sequences, for use in array development: Meredith and Gary.
5. Preparing a prioritized list of fungi for genome sequencing would be important for ecological studies, and criteria for such choices: Chris.
6. Preparing an ecological and systematics 'perspective' paper on ITS limitations/uses: Tom.
7. Preparing a list of known problem groups for which typical primers fail – e.g., many *Tulasnella*, *Cantherellus*. Karen and Urmas.
8. Generating a list of available tools for ITS work, and linking those to the FESIN website: Henrik.

Discussion then turned to proposal development:
Links with NEON were discussed. Specific funding opportunities also were pondered: David mentioned Biotic Inventories, and Betsy mentioned the SGER program. Ian mentioned fungal biogeography as a great focal point for this group. Ari reminded everyone of the need for a theoretical context and hypothesis testing for NSF proposals. Chris suggested a project similar to the RDP currently in place for bacteria. David described the potential to tap into 'cyberinfrastructure' at NSF. Ian mentioned fungal biogeography as a great focal point for this group. Rytas mentioned the modularity of a proposal for comparative biogeography that could allow development of complementary proposals in other countries. Standardized data collection methods, and clear ways of analysis, nomenclature, etc., are all desirable.

Concluding this discussion, it seemed that the general topics for proposals include (1a) Affymetrix array development and (1b) testing; (2) comparative biogeography; (3) database development; (4) GenBank-centered database/improvement initiatives.

The last point led to discussion of the changes that have to be made in GenBank -- or not. The group's general feeling was that GenBank is the deep, long-term depository; we can't duplicate those benefits, nor should we try. But we can try to improve it, and to mine it.

Karen presented an overview of funding opportunities in the US. These included the FIBR program through NSF, and the potential to find support through the National Science Foundation's International Programs (as supplements to existing grants). The USDA has several programs – Plant Biosecurity NRI, and Microbial Observatories/MIP program. She highlighted the DOE's genomics to life initiatives, with special focus on

bioenergy. JGI and Rockefeller may help support genome sequencing; the Mellon foundation may be useful for biogeography; Homeland Security might be supportive (through ARPA) of biomonitoring. Human health issues could be a basis for approaching the Bill and Melinda Gates foundation. Lee brought up NIH – natural reservoirs of pathogens.

Ian Alexander presented an overview of funding opportunities in the EU. The European Science Foundation is growing, says Urmas; the European Research Council appears to have funds for large collaborative efforts. This may be promising. Soren showed and overview of the European budget allocations; climate change has been behind many of the calls from the EU. Previous projects on plants and soils have rarely, if ever, taken into account the fungi. Most money in health; much less (4.5X less) has been allocated for work in the environment.

Final thoughts
-Tom invited UNITE members to consider attending ESA, and the FESIN workshop there.
-Ideas for a chip were briefly discussed: ITS Fast; My*E*cology, ChipSuper, MycoChip, MyChip, FunChip, MyFunChip
-Branching out to additional societies is important, including ISME, BMS, ASM, APS.