

Letter

Sequencing the fungal tree of life

Terrestrial ecosystems host a complex array of interacting communities, with thousands of species of animals, plants, fungi and bacteria. In soils, this complex web of life is responsible for the cycling of carbon (C), for water and nutrients, for soil quality and for plant nutrition and health. To predict future changes of these threatened ecosystems and to fully grasp the biological and chemical workings of these complex interactions, one must not only regard organisms as individuals but also as members of a larger community, considering the interplay and communication between individuals within these entangled populations, that is, their extended phenotype (Whitham *et al.*, 2008). One emerging model for such studies is the interaction between soil-borne fungi and plant communities.

Fungi are one of the largest and most diverse kingdoms of eukaryotes and function as important biological components of all terrestrial ecosystems. They are central to the global C cycle, constitute the major group of plant pathogens in managed and natural ecosystems, serve as symbionts with heterotrophic and autotrophic organisms alike, and play an integral and growing role in the development and production of renewable bio-based fuels and chemicals. The success and importance of fungi to life on Earth are directly attributable to the remarkable diversity of enzymes and metabolites that they produce, which afford them a broad range of nutritional modes and grant them access to an amazing breadth of C sources and ecological niches.

Understanding how saprotrophic, symbiotic and pathogenic fungi achieve their lifestyle is crucial for understanding their ecological functions and their subsequent impact on the fate of plant communities. The inconspicuous nature of soil fungi, the inaccessibility of their habitats and our inability to culture many of them have made them difficult to study. Advances in large-scale DNA barcoding surveys have circumvented some of these limitations and allowed us to determine the composition and the dynamics of several fungal soil communities, including mycorrhizal fungi (Buée *et al.*, 2009; Öpik *et al.*, 2009; Jumpponen *et al.*, 2010). Several hundred species are active in soils, and ongoing metagenomics and metatranscriptomics studies will uncover the functions encoded in their genomes as well as their expressed transcripts (Martin & Martin, 2010). Many of the fungi whose genomes have been sequenced are residents of

soil and plants (Martinez *et al.*, 2004, 2009; Martin *et al.*, 2008, 2010; Ohm *et al.*, 2010; Spanu *et al.*, 2010; Stajich *et al.*, 2010) and these sequences will provide baseline genomic information that enables scientists to explore the genomes and functions of thousands of soil fungal species that cannot be cultured and sequenced directly. Unfortunately, reference fungal genomes sequenced to date, and those in progress, show significant bias towards fungi of medical importance (Cuomo & Birren, 2010). The availability of genome sequences from ecologically and taxonomically diverse fungi not only would allow ongoing research on those species, but enhances the value of other sequences through comparative studies of gene evolution, genome structure, metabolic and regulatory pathways, and saprotrophism/symbiosis/pathogenesis lifestyles. Recently, the US Department of Energy Joint Genome Institute (JGI) launched the Fungal Genomics Program (FGP), aimed at exploration of fungal diversity for energy and environmental sciences and applications through the scale-up of sequencing and analysis (Grigoriev *et al.*, 2011).

Fungal genomics for energy and environment

The Genomic Encyclopedia of Fungi of the JGI FGP (<http://www.jgi.doe.gov/fungi>) targets fungal genomes in three areas: plant health, biorefinery and fungal diversity. Plant health depends on interactions of the plant with both fungal mutualists and parasites. While on the surface, the effects of both on a host are orthogonal; they may share common characteristics to escape the plant defense system and acquire host nutrients (e.g. the role of effector-like secreted proteins in *Laccaria bicolor* and in *Melampsora larici-populina*), live within the host (e.g. the semibiotroph *Mycosphaerella graminicola*) and balance between saprobic and parasitic (or mutualistic) lifestyles (e.g. *Heterobasidion annosum*). Comparative genomics should lead us to understand these mechanisms critical for sustainable growth of forest trees and bioenergy feedstocks, such as poplars, pines and eucalypts. Biorefinery includes industrial hosts such as *Trichoderma reesei* and *Aspergillus niger*. Novel 'modules' or 'parts' – metabolic processes, enzymes and regulatory elements with desired properties from other fungi – will be plugged into these established hosts to accelerate the development of new bioprocesses that efficiently convert biomass into biofuels and renewable chemicals. Finally, exploration of fungal ecological and phylogenetic diversity should bring us more useful 'modules' or 'parts', such as thermostable enzymes or thermotolerant hosts from

extreme environments, as well as provide clues on fungal evolution. Dramatic improvements in genome-sequencing technologies enable us to go beyond sequencing a single reference genome, allowing 'resequencing' of several accessions (e.g. resequencing of a dozen strains of the ectomycorrhizal model species *L. bicolor*). Through comparative analysis of gene expression, genomic variation across the diversity of fungi and metagenomics of complex systems, such as soil, we will reach a higher level of understanding of fungal lifestyles, their interactions with plants and their evolution, laying a strong foundation upon which to build practical applications for energy and environment.

Sequencing saprotrophic Agaricomycotina

Diverse fungi obtain nutrition from primary plant tissues, exudates or phloem sap, but only a relatively small number are able to efficiently degrade wood. The majority of the wood-decaying fungi are in the Agaricomycotina (Basidiomycota), which also includes mycorrhizal and pathogenic forms. Within the Agaricomycotina, a polyphyletic assemblage, known as white-rot fungi, have the unique ability to depolymerize and mineralize the recalcitrant lignin in order to gain access to cell wall carbohydrates as C and energy sources. Another assemblage, the so-called brown-rot fungi, rapidly depolymerize cellulose but do not extensively degrade the lignin, which remains *in situ* as a modified polymeric residue that is highly resistant to further microbial decay. Both types of wood-decay fungi are common inhabitants of forest ecosystems and play an important, if not pivotal, role in the C cycle.

The genomes of three wood-decaying Agaricomycotina – *Phanerochaete chrysosporium*, *Postia placenta* and *Schizophyllum commune* – have been published (Martinez *et al.*, 2004, 2009; Ohm *et al.*, 2010). *P. chrysosporium* produces a white rot whereas *P. placenta* is a brown-rot species. They are both members of the Polyporales, but were found to have radically different decay chemistries. For example, *Phanerochaete* has 15 genes encoding class II fungal peroxidases, which function in lignin degradation, whereas *Postia* has only one low-redox-potential peroxidase, of unknown function. In addition, *Postia* lacks exocellobiohydrolases and cellulose-binding modules, which were previously thought to be present in all cellulolytic fungi. By contrast, *S. commune* has one of the largest sets of cellulolytic enzymes (Ohm *et al.*, 2010). These findings, from three closely related taxa, suggest that there has been considerable diversification in decay mechanisms in Agaricomycotina, warranting a broad sampling of this clade.

The JGI initiative has targeted 30 wood-decay fungi representing 12 orders. Fourteen first-tier species (nine white rot and five brown rot), are designated for sequencing within the first year. Five of the first-tier species represent major clades of Agaricomycotina that include widespread wood-

decay species for which there are currently no published genomes or (to our knowledge) active genome projects (i.e. Auriculariales, Dacrymycetales, Hymenochaetales and Corticiales). The second-tier species further broaden the phylogenetic diversity of wood-decay genomes (including the as-yet unsampled Atheliales, Amylocorticiales, Jaapiales and Phallomycetidae). Comparative analyses of these annotated genomes will also integrate gene-expression profiles determined by high-throughput Illumina RNA-Seq and by protein mass spectrometry from 12 fungal species. The genome sequence is already available for five species (*P. chrysosporium*, *P. placenta*, *Pleurotus ostreatus*, *Serpula lacrymans* and *H. annosum*) and, for the other seven (*Gloeophyllum trabeum*, *Fistulina hepatica*, *Fomitiporia mediterranea*, *Dacryopinax spathularia*, *Fomitopsis pinicola*, *Auricularia auricular-judae* and *Punctularia strigosozonata*), the genome sequence will be produced in the saprotrophic Agaricomycotina project, as already described. Comparative whole-transcriptome analyses of different decayer types and clades will facilitate the long-range understanding of their biological lifestyles and lignocellulolytic strategies.

Sequencing mycorrhizal genomes

In soils of most forests, hundreds of species of ectomycorrhizal fungi establish mutualistic symbioses with lateral roots of trees (Buée *et al.*, 2009; Martin & Nehls, 2009). Genomic-based analyses of the ectomycorrhizal symbiosis in environmental settings are in their infancy (Courty *et al.*, 2008), but research in environmental genomics is rapidly developing and is crucial for understanding the role of the mycorrhizal symbiosis in nutrient cycling and plant health. One key step for the future of mycorrhizal research is the production of additional sequenced genomes, both for a variety of mycorrhizal fungi and also for their plant hosts. To date, the genomes of two symbionts – the basidiomycete *L. bicolor* (Martin *et al.*, 2008) and the ascomycete *Tuber melanosporum* (Martin *et al.*, 2010) – have been sequenced. Based on their symbiosis-related gene networks, evolution of the ectomycorrhizal lifestyle appears to be quite divergent (Plett & Martin, 2011). To better understand the differences between symbiotic clades and types of symbiosis, the JGI initiative has targeted 25 mycorrhizal fungi from different orders (Plett & Martin, 2011). Genomic DNA from *Amanita muscaria*, *Cenococcum geophilum*, *Hebeloma cylindrosporum*, *L. amethystina*, *Oidiodendron maius*, *Piloderma croceum*, *Paxillus involutus*, *Pisolithus microcarpus* and *P. tinctorius* is currently being sequenced using next-generation sequencing platforms. Sequencing of *Boletus edulis*, *Cantharellus cibarius*, *Coltricia cinnamomea*, *Cortinarius glaucopus*, *Gymnomyces xanthosporus*, *Lactarius quietus*, *Meliniomyces bicolor*, *Paxillus rubicundulus*, *Ramaria formosa*, *Rhizoscyphus ericeae*, *Scleroderma citrinum*, *Suillus luteus*, *Sebacina vermifera*, *Tomentella sublilacina*, *Tricholoma*

matsutake, *Tulasnella calospora* and *Terfezia boudieri* will soon follow. The sequenced mycorrhizal species were selected for their ability to promote plant growth and health, their phylogenetic novelty, their ability to establish different types of mycorrhizal symbiosis (ectomycorrhizas, ericoid and orchid endomycorrhizas) and their ecological niche and host specificity. The comparative genomics of the available *L. bicolor* and *T. melanosporum* genomes (Martin *et al.*, 2008, 2010), and the mycorrhizal genomes that are currently sequenced, will provide new insights that include but are not limited to:

- a better understanding of the complex interactions between trees and mycorrhizal symbionts;
- comparative genomics with the other economically important saprobic and pathogenic fungi;
- bioinformatics identification of important symbiosis-related genes; and
- molecular markers for investigating adaptation signatures of symbiotic fungi in various ecosystems and environmental conditions.

The fact that mycorrhizal fungi appear to be independently derived from multiple saprobic lineages means that genomic data will provide independent assessments of the genetic underpinnings of mycorrhizal competence (Plett & Martin, 2011).

Exploring the fungal tree of life

The fungal tree of life (FTOL) serves as the foundation for comparative fungal biology and is the focus of considerable phylogenetic research (Blackwell *et al.*, 2006). The early diverging lineages of the FTOL comprise species that possess a smooth posterior flagellum. Flagellated fungi were previously classified in Chytridiomycota, but are now recognized as members of multiple paraphyletic lineages. The remaining early diverging lineages of fungi include the zygomycetous fungi and the arbuscular mycorrhizal fungi, which are characterized by aseptate filamentous growth and lack of sporocarp formation. They were classified in Zygomycota, but like Chytridiomycota may comprise multiple phylogenetic lineages. Dikarya is the most derived clade of fungi and consist of Ascomycota (e.g. yeasts and molds) and Basidiomycota (e.g. mushrooms and shelf-fungi).

Our increased understanding of the FTOL has resulted in refinement of concepts of morphological homologies, evolution of ecologies and physiologies, and a more accurate assessment of previously unappreciated phylogenetic diversity. Importantly, the FTOL informs our selection of taxa for genomic sequencing. As already stressed, the sequencing of Dikarya is focused on lineages of current economic and medicinal importance (Cuomo & Birren, 2010). While these data have advanced insights into fungal biology, a more complete understanding of fungi and their potential application and benefit to human society is limited by

current gaps in our sampling of genomes throughout the FTOL.

As an initial step to advance a more complete genomic coverage of the FTOL, JGI FGP's Genomic Encyclopedia of Fungi project will sequence 10 target species from unsampled lineages of the FTOL. These species collectively represent over 1 billion years of eukaryotic evolution and occupy diverse ecological niches (e.g. insect endosymbionts and osmophilic fungi) and exhibit novel metabolic capabilities (e.g. detoxification of noxious plant compounds and colonization of extreme environments). Fungi have been used for the benefit of humans for over 5000 yr, from fermentation to antibiotics, yet we have only harnessed a fraction of their metabolic potential. Sequencing these genomes will significantly improve our understanding of early fungal evolution through phylogenomic analyses and it will result in a more complete assessment of the genomic and metabolic diversity of fungi that is needed to more fully apply them to the human endeavor.

Outlook

Genomic analyses have so far been restricted to a limited set of ecologically relevant fungal species. Within the next 2 yr, we will see a massive shift towards the inclusion of the environmental and evolutionary perspectives in genome-sequencing initiatives. Environmental genomics of the communities of soil-borne fungi will lead to a better understanding of the biological and ecological roles of these eukaryotic microbes (Kyrpides, 2009). Several fungal species selected in the present JGI projects are amongst the most abundant species found in large-scale 454 pyrosequencing of soil fungal ribosomal DNA and are known to interact with plants (Buée *et al.*, 2009; Jumpponen *et al.*, 2010). The sequenced genomes from across the tree of life will therefore serve as anchors for sorting through thousands of metagenomic repertoires and categorizing them. This will undoubtedly lead to a better understanding of plant-microbe interactions.

Acknowledgements

We would like to thank the members of our scientific communities for their outstanding support. F.M. is supported by the European Commission within the Project ENERGYPOPLAR (FP7-211917), the US Department of Energy – Oak Ridge National Laboratory Scientific Focus Area for Genomics Foundational Sciences (Project Plant-Microbe Interfaces). A.P. is supported by grants AGL2008-05608 and the Bioetanol-Euroinnova project. This work was conducted by the US Department of Energy Joint Genome Institute, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231 (I.V.G.).

F. Martin^{1*}, D. Cullen², D. Hibbett³, A. Pisabarro⁴,
J. W. Spatafora⁵, S. E. Baker^{6,7} and I. V. Grigoriev^{7*}

¹UMR INRA/UHP 1136, Interactions Arbres/Micro-Organismes, INRA-Nancy, 54280 Champenoux, France; ²USDA, Forest Products Laboratory, Madison, WI, USA; ³Clark University, Worcester, MA, USA; ⁴Department of Agrarian Production, Public University of Navarre, 31006 Pamplona, Spain; ⁵Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA; ⁶Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, Richland, WA 99352, USA; ⁷US DOE Joint Genome Institute, Walnut Creek, CA, USA

(*Authors for correspondence: F. Martin tel +33 383 39 40 80; email fmartin@nancy.inra.fr; I.V. Grigoriev email ivgrigoriev@lbl.gov)

References

- Blackwell M, Hibbett DS, Taylor JW, Spatafora JW. 2006. Research coordination networks: a phylogeny for Kingdom Fungi (Deep Hypha). *Mycologia* 98: 829–837.
- Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S, Martin F. 2009. 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist* 184: 449–456.
- Courty PE, Poletto M, Duchaussoy F, Buée M, Garbaye J, Martin F. 2008. Gene transcription in *Lactarius quietus*–*Quercus petraea* ectomycorrhizas from a forest soil. *Applied Environmental Microbiology* 74: 6598–6605.
- Cuomo CA, Birren BW. 2010. The fungal genome initiative and lessons learned from genome sequencing. *Methods in Enzymology* 470: 833–855.
- Grigoriev IV, Cullen D, Hibbett D, Goodwin SB, Jeffries TW, Kubicek CP, Kuske C, Magnuson J, Martin F, Spatafora J *et al.* 2011. Fueling the future with fungal genomics. *Mycology*, in press.
- Jumpponen A, Jones KL, Mattox JD, Yaeger C. 2010. Massively parallel 454-sequencing of fungal communities in *Quercus* spp. ectomycorrhizas indicates seasonal dynamics in urban and rural sites. *Molecular Ecology* 19: 41–53.
- Kyrpides NC. 2009. Fifteen years of microbial genomics: meeting the challenges and fulfilling the dream. *Nature Biotechnology* 27: 627–632.
- Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V *et al.* 2008. The genome sequence of the basidiomycete fungus *Laccaria bicolor* provides insights into the mycorrhizal symbiosis. *Nature* 452: 88–92.
- Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R *et al.* 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464: 1033–1038.
- Martin F, Martin N. 2010. From galactic archeology to soil metagenomics – surfing on massive data streams. *New Phytologist* 185: 343–347.
- Martin F, Nehls O. 2009. Harnessing ectomycorrhizal genomics for ecological insights. *Current Opinion in Plant Biology* 12: 508–515.
- Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, Ferreira P, Ruiz-Duenas FJ, Martinez AT, Kersten P *et al.* 2009. Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proceeding of the National Academy of Sciences, USA* 106: 1954–1959.
- Martinez D, Larrondo LF, Putnam N, Sollewijn Gelpke MD, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F *et al.* 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology* 22: 695–700.
- Ohm R, de Jong J, Lugones L, Aerts A, Kothe E, Stajich J, de Vries R, Record E, Levasseur A, Baker S *et al.* 2010. Genome sequence of the model mushroom *Schizophyllum commune*. *Nature Biotechnology* 28: 957–963.
- Öpik M, Metsis M, Daniell TJ, Zobel M, Moora M. 2009. Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. *New Phytologist* 184: 424–437.
- Plett J, Martin F. 2011. Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. *Trends in Genetics* 27: 14–22.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, van Themaat EVL, Brown JKM, Butcher SA, Gurr Sarah J. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* 330: 1543–1546.
- Stajich J, Wilke S, Ahren D, Au C, Birren B, Borodovsky M, Burns C, Canback B, Casselton L, Cheng C *et al.* 2010. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proceedings of the National Academy of Sciences, USA* 107: 11889–11894.
- Whitham TG, DiFazio SP, Schweitzer JA, Shuster SM, Allan GJ, Bailey JK, Woolbright SA. 2008. Extending genomics to natural communities and ecosystems. *Science* 320: 492–495.

Key words: comparative genomics, DNA sequencing, energy and environment, fungal genomics, fungal tree of life, meta-genomics (–transcriptomics), mycorrhiza, saprotrophs.