

# Abstracts

## Academy colloquium 'Fungal Phylogenomics'

Royal Netherlands Academy of Arts and Sciences

&

Centraalbureau voor Schimmelcultures

Amsterdam, the Netherlands

May 11-12, 2004

organized by Teun Boekhout, Centraalbureau voor Schimmelcultures  
(CBS-KNAW), Utrecht, the Netherlands



Koninklijke  
Nederlandse  
Akademie van  
Wetenschappen



Netherlands Genomics Initiative

NWO ALW



sponsored by:

KNAW, CBS, NWO-ALW and Netherlands Genomics Initiative

# Nonparametric and parametric approaches for inferring evolutionary processes in fungal populations

*I. Carbone*

North Carolina State University, Department of Plant Pathology, Raleigh, NC 27695, USA,  
ignazio\_carbone@ncsu.edu

In recent years, rapid advances in DNA sequencing technology have stimulated population geneticists to develop a plethora of new approaches for making inferences on population processes from DNA sequence variation. Among these are tools for estimating population mutation and migration rates, recombination rates, and selection. These approaches build on Wright-Fisher population genetic models by incorporating probability-based coalescent methods, which take full advantage of the data and its inherent stochastic properties. In practice, there are three main limitations with using these coalescent-based methods: 1) they make strict assumptions that must be verified *a priori*, 2) their parameter-rich nature makes the estimation of all model parameters very complex and computationally intensive, and 3) they are distributed as console applications written in C and require the user to navigate through console menus or specify complex command line arguments. The challenge will be to develop software tools that will eliminate the requirement for using command line and that can integrate a wide array of approaches for analyzing population genetic data, based on both traditional summary-statistic methods and the newer coalescent-based population genetic models. I will describe a workbench tool that we have developed to address these issues and show how this tool can be used to infer evolutionary processes in fungal populations.

# Bayesian Markov Chain Monte Carlo versus traditional methods of phylogenetic reconstruction

Michael Weiß

Spezielle Botanik und Mykologie, University of Tübingen, Auf der Morgenstelle 1, D-72076 Tübingen, Germany, michael.weiss@uni-tuebingen.de

Probabilistic methods have become more and more important in the reconstruction of phylogenetic trees from molecular data and meanwhile play a key role in this field. Since Felsenstein's milestone publication (1981) it is possible to effectively compute the likelihood of a phylogenetic tree, i.e. the probability that the given alignment has evolved under this tree. Likelihood values are most often used as an optimality criterion to estimate a phylogenetic tree from an alignment (maximum likelihood method).

However, up to now there is no analytical way to calculate another conditional probability that has a much more relevant biological interpretation: the probability that a given tree or topology truly reflects the phylogenetic relationships of the involved sequences given the alignment (the *a posteriori* probability that the given tree is the true tree). These *a posteriori* probabilities have only become accessible after the introduction of Markov Chain Monte Carlo methods (MCMC), i.e. the application of stochastic processes, in phylogenetics (e.g. Yang & Rannala 1997; Mau, Newton & Larget 1999; Li, Pearl & Doss 2000). Since the implementation of MCMC in freely available software (e.g. Huelsenbeck & Ronquist 2001) this approach has received some popularity in molecular phylogenetic reconstruction.

I give an introduction to main principles of this method, review the recent discussion in the literature and compare MCMC with other methods of phylogenetic inference in applications to own data.

## References:

- Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17, 368–376.
- Huelsenbeck, J. P. & Ronquist, F. R. (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Li, S., Pearl, D. K. & Doss, H. (2000) Phylogenetic tree construction using Markov Chain Monte Carlo. *Journal of the American Statistical Association* 95, 493–508.
- Mau, B., Newton, M. & Larget, B. (1999) Bayesian phylogenetic inference via Markov Chain Monte Carlo methods. *Biometrics* 55, 1–12.
- Yang, Z. & Rannala, B. (1997) Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology and Evolution* 14, 717–724.

## Fungal phylogenetics: from species recognition to adaptation

*John W. Taylor*

Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102, USA, jtaylor@socrates.berkeley.edu

Phylogenetics has provided a means of studying the pattern of evolution, and now the results of these studies provide a framework for studying the process of evolution. Regarding pattern, phylogenetic species recognition has been shown to find more species than morphological species recognition, but most such studies could not include biological species recognition because fungi are difficult to mate in the laboratory. We are using outbreeding *Neurospora* species to compare species that are genetically-differentiated or genetically-isolated (as recognized by phylogenetic comparison of several gene genealogies) with species that are reproductively-isolated (as recognized by mating tests). We find that almost the same species are recognized by phylogenetic and biological methods, which indicates that phylogenetic species recognition can be applied to the many fungi for which biological species recognition is difficult or impossible.

Regarding process, our study of biological species recognition has shown that hybrid matings are more successful when the parents are geographically distant. This observation, coupled with our inability to find any true hybrids among natural *Neurospora* individuals, suggests that reinforcement is operating in *Neurospora* and that genes can be found that reduce hybrid fecundity. These genes would contribute to the process that keeps species distinct following divergence. We are using quantitative trait locus analysis to attempt to find genomic regions that correlate with variation in fecundity in matings between individuals from genetically distinct clades within one *Neurospora* species and between *Neurospora* species. Adaptation is another process that contributes to species differentiation and we aim to characterize adaptation to environmental variation within a species. The keys will be to study a confirmed single species that ranges across different environments and to concentrate on traits that contribute to fitness and for which genetic control can be assigned. *Neurospora discreta* will be our model species because it thrives in different environments over a tremendous latitudinal gradient.

To understand genetic relationships among individuals in the diverse *N. discreta* clade, we are applying phylogenetic species recognition as we have done for the other outbreeding *Neurospora* species. We are assessing fitness by growth rate and sporulation and we are profiling transcription in the hope of finding genes responsible for adaptation.

## **Assembling and communicating the fungal branch of the tree of life: toward an automated phylogenetic taxonomy of fungi**

*David S. Hibbet*

Biology Department, Clark University, Worcester, MA 01610, USA,  
dhibbett@blackntan.clarku.edu

The goal of fungal systematics is to create classifications that are both complete (meaning that they contain all species of fungi) and phylogenetically accurate. This endeavor is made difficult by the immense diversity of fungi, which includes many cryptic and undescribed taxa, as well as the usual challenges of phylogenetic inference. Using molecular characters, fungal systematists are making great progress toward resolving the major branches of fungal phylogeny, as well as the fine branching of many terminal clades. This work is continuous; new phylogenetic analyses appear in every issue of the major mycological journals, and the database of molecular sequences expands daily. Changes in taxonomy are not keeping pace with the rapid growth of phylogenetic knowledge, however. The current taxonomic system is characterized by episodic progress, with long delays between the publication of phylogenetic analyses (which may themselves be out of date when they appear in print), and their translation into new classifications.

A model for automated phylogenetic taxonomy of fungi will be presented, using the homobasidiomycetes as an example. The system involves two major components: 1) automated phylogenetic analysis, and 2) automated tree-based classification. The first component has been implemented in a computer program called "mor", which was written in PERL by H. Nilsson and M. Snyder, based on a prototype written in BioPYTHON by M. Shonfeld. mor automatically retrieves, aligns, and analyzes nuclear large subunit ribosomal DNA (nuc-lsu rDNA) sequences from homobasidiomycetes. The nuc-lsu rDNA was chosen because it is the most commonly sampled gene for phylogenetic studies above the species level in homobasidiomycetes, and can be aligned (albeit with ambiguities) across the homobasidiomycetes. However, the nuc-lsu rDNA alone has been shown to do a poor job of resolving the deeper nodes within the homobasidiomycetes. Consequently, the analyses in mor are performed with a backbone monophyly constraint that reflects phylogenetic hypotheses derived from prior multi-locus studies. As of this writing, mor has generated phylogenetic trees with 2074 sequences, which is 62% of the 3350 nuc-lsu rDNA sequences of homobasidiomycetes in GenBank. Most of the other sequences were automatically excluded because they are less than 800 bp, which was established as a minimum length criterion.

The second component, automated tree-based classification, is currently in development. Phylogenetic taxon definitions will play a central role in this aspect of the system, because they are stable in the face of rearrangements in phylogenetic trees, although the composition of the taxa they specify may change. For example, a taxon "X" could be defined as "the least inclusive clade that contains species C and F". Given a tree (A(B(C(D(E(FG)))))), X would include C, D, E, F, and G. If a new analysis, incorporating new data, suggested (A(B(D(C(H(E(FG))))))) then X would include C, E, F, G, and H. Unlike Linnaean taxon definitions, phylogenetic taxon definitions are amenable to algorithmic representation, opening the door to automated taxonomy and other phyloinformatics applications.

## **Molecular approaches for estimating fungal environmental diversity: Probing fungal diversity using DNA sequences as the sampling unit**

*R Vilgalys*

Department of Biology, Box 90338, 139 Biological Sciences Bldg, Duke University, Durham NC 27708, USA, fungi@duke.edu

Little is known about microbial communities in forest soils. We applied culture-independent methods developed for bacteria to assess fungal community diversity in soil samples collected from Piedmont

forest soils in the southeastern USA. Identification of DNA sequences using standard bioinformatics tools (BLAST) and phylogenetic analyses reveals a highly diverse community of eukaryotic microorganisms that is dominated by fungi, but which also includes protistan, chlorophyte and even metazoan lineages. Samples representing litter, O, A, and B soil horizons from mixed hardwood/pine forests were investigated by direct DNA extraction and cloning of DNA libraries for both variable (ITS) and conserved (18S) regions within the rDNA genes. ITS sequences were generated and sorted using BLAST searches; ITS sequences with > 80% similarity to known fungal groups were further identified using phylogenetic methods. Comparison of ITS sequences against taxonomic reference databases revealed a high degree of species diversity. Ectomycorrhizal basidiomycetes were found to comprise up to 40% of the total eukaryotic diversity in these forest soils.

As part of a larger study examining the influence of different land use histories on biological complexity within soils, we also used molecular-based approaches to assess the microbial community

structure from replicated recovered communities that include cultivated fields, pastures, loblolly pine stands, and remnant hardwood stands. Significant changes in fungal community assemblages

can be observed across different habitats, with hardwood and pine forests showing a much greater preponderance of ectomycorrhizal species than cultivated fields and pastured grasslands.

These studies demonstrate the potential of DNA-based studies for complete-taxon inventories, however, their general utility can sometimes be limited by technical problems (chimeric sequences, PCR-bias, etc.), and also by the limited availability of complete reference databases for many important soil taxa.

## Non-conventional yeasts are becoming new model organisms

*Jure Piskur*

BioCentrum-DTU, Building 301, Technical University of Denmark, DK-2800 Lyngby, Denmark,  
jp@biocentrum.dtu.dk

The sequence of over a dozen yeast genomes has recently been completed and several additional sequencing projects are in progress, providing the basis to resolve several complex biological questions. For example, comparative genomics provides a tool to understand the origin of yeasts, and has already pointed out several crucial molecular events in yeast evolutionary history. In addition, the availability of the sequence data from so far rarely studied yeasts species has, because of their interesting traits, already provided the background for development of novel model organisms.

Yeasts belonging to the *Saccharomyces* complex have a number of unique characters not found in other yeast genera. For example, *Saccharomyces sensu stricto* yeasts, as well as other *Saccharomyces* members, primarily degrade hexoses only to the C<sub>3</sub> and C<sub>2</sub> compounds pyruvate and ethanol, even in the presence of oxygen. This phenomenon, also known as Crabtree effect, relies on saturation of the oxidative metabolism, overflow at pyruvate branch-point and a “glucose repression” circuit that represses the respiratory part in the presence of glucose. While a majority of yeasts cannot grow in the absence of oxygen (aerobic yeasts), a majority of the *Saccharomyces* complex yeasts can also survive without any oxygen and they can generate mitochondrial petite mutants. An interesting evolutionary question is when and how did the progenitor of *Saccharomyces* yeasts develop these basic characters and what were the molecular mechanisms operating during this yeast’s evolutionary history? Apparently, two molecular mechanisms, whole-genome duplication and horizontal gene transfer, have been proposed, to play a major role in the evolutionary history of the *Saccharomyces* complex yeasts.

A very powerful approach to understand evolution is to study yeast species with intermediate phenotypes regarding *S. cerevisiae*. One recently sequenced species, *S. kluyveri*, is a facultative anaerobe but petite-negative yeast, and is now used in comparison with *S. cerevisiae* to understand glucose repression. *S. kluyveri* has also become an attractive model to study metabolism of nucleic acids precursors, specially the pyrimidine catabolic pathway, which is missing in other yeasts. Catabolism of pyrimidines is very important to understand the fate of several chemotherapeutic drugs in cancer patients. Several *S. kluyveri* genes and the corresponding enzymes, including their 3D structures, have recently been analysed in detail and now serve as a general model to understand this pathway in other organisms including humans.

Species belonging to another yeast genus, *Brettanomyces/Dekkera*, which is from the phylogenetic point of view not closely related to the *Saccharomyces* complex yeasts, are also petite-positive. At least some of them can grow anaerobically and are also good ethanol producers. Apparently, this group has in parallel developed similar phenotypic traits as found in *S. cerevisiae*. However, so far the corresponding genetic background of *Brettanomyces/Dekkera* is not known, and therefore these yeasts represent a challenging research subject.

## Fungal Phylogenomics, understanding the fungal way of life

Eiko K. Kuramae<sup>1</sup>., Vincent Robert<sup>1</sup> V., Berend Snel<sup>2</sup>, and Teun Boekhout<sup>1</sup>

<sup>1</sup>Centraalbureau voor Schimmelcultures – CBS, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands; <sup>2</sup>Centre for Molecular and Biomolecular Informatics – CMBI, University of Nijmegen, The Netherlands, kuramae@cbs.knaw.nl

For the first time, fungal evolutionary hypotheses can be based on the analysis of full genome sequences. Before this fungal phylogeny was based entirely on morphological criteria, and more recently on partial macromolecular sequence analyses. This has led to the conclusion that no single marker molecule can resolve fully the fungal tree of life or the genealogy within major fungal lineages. Also phylogenetic analyses of different genes do not always converge on the same tree.

The clusters of eukaryotic orthologous groups (KOGs) of proteins from complete fungal genomes were used for phylogeny analysis. The analyses were carried out based on the presence or absence of KOGs, as well as by supertree approach. To efficiently handle the large amount of information present in full genome sequences, we developed a Correlation Super Tree (CST) approach in order to estimate the phylogenetic signal present among the different groups of orthologs, followed by concatenation of the sequences from those orthologs showing the same evolutionary fate and a phylogenomic analysis. The resulting phylogenetic tree clearly supported the classification of the endomycetous yeasts (viz. *Saccharomyces* spp.) and the Euascomycete *Neurospora crassa* as a sister groups in a single clade, whereas *Schizosacharomyces pombe* and *Cryptococcus neoformans* seems to represent more basal fungal lineages.

The distribution of KOGs among *Cr. neoformans*, *N. crassa*, *S. cerevisiae* and *Sch. pombe* in 24 different functional categories was analysed. The distribution of the genes per KOG over the functional categories demonstrated that the genome of *C. neoformans* was more similar to that of the filamentous fungus *N. crassa*, whereas the yeasts *S. cerevisiae* and *Sch. pombe*, seem different. The *C. neoformans* and *N. crassa* genomes have more genes in the KOGs related to lipid metabolism, defense mechanisms, secondary metabolites biosynthesis, transport and catabolism, energy production and conversion; carbohydrate transport and metabolism, inorganic transport and metabolism than *S. cerevisiae* and *Sch. pombe*. Therefore, it seems that *Cr. neoformans* has the gene machinery of filamentous fungus, which may not be surprising if the dimorphic character of this pathogen is taken into account.

## Fungal Evolution, a mitochondrial perspective

Charles E. Bullerwell, Jessica Leigh, Joyce E. Longcore and B. Franz Lang

Département de Biochimie, Université de Montréal, 2900 Boul. Edouard-Montpetit, Montréal (Québec), H3T 1J4, Canada, Franz.Lang@Umontreal.ca

Over the past several years, the number of available, completely sequenced mitochondrial genomes from fungi has increased from just three to over two dozen sequences, including representatives of the four principle divisions of this kingdom: Ascomycota, Basidiomycota, Zygomycota and Chytridiomycota. This wealth of data from a wide range of diverse fungi has allowed a more complete understanding of the organization and content of their mitochondrial genomes. In addition, mtDNA-encoded protein sequences have proven invaluable for molecular phylogenetics, elucidating the phylogeny of the fungi and their relationship to other eukaryotes. In fact, mitochondrial data have proven invaluable in phylogenetic reconstruction in general, and the resolution of fungal phylogeny using these data is unprecedented (e.g., (Forget et al., 2002; Lang et al., 2002; Bullerwell et al., 2003a; Bullerwell et al., 2003b; Leigh et al., 2003).

These analyses show that the Fungi clearly form a monophyletic group, as do the Holozoa (Metazoa plus protists along the animal lineage). A highly supported monophyletic superset of these two groups (termed Ophisthokonts, i.e., Fungi plus Metazoa) is also recovered. Bootstrap support for the internal branches among the Fungi is high, and the four fungal divisions are clearly defined, with the Ascomycota and Basidiomycota appearing as a monophyletic group, the Zygomycota branching prior to the ascomycete-basidiomycete divergence, and the Chytridiomycota branching at the base of the Fungi. Although this topology is in accordance with most of the established taxonomy, certain peculiarities remain. First, there is little support for the monophyly of Chytridiomycota including *Allomyces macrogynus*. However, previously reported support for chytridiomycete paraphyly has dropped considerably with the advent of more sophisticated methods of phylogenetic analysis, and with the availability of data from additional species. Another problematic issue is the exact placement of the *Schizosaccharomyces* genus within Ascomycota.

In light of this phylogenetic framework, the comparison of fungal mitochondrial genome sequences has allowed for an appreciation of how mitochondrial genomes have evolved in terms of gene content, gene order, gene organization, gene expression and genome conformation. These advances will help us to better understand fungal biology, and therefore some of our most important eukaryotic model organisms.

### References:

- Bullerwell CE, Forget L, Lang BF. 2003a. Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences. *Nucleic Acids Res* 31:1614-1623.
- Bullerwell CE, Leigh J, Seif E, Longcore JE, Lang BF. 2003b. Evolution of the fungi and their mitochondrial genomes. In: Arora D, Khachatourians GG, eds. *Applied Mycology and Biotechnology*. Amsterdam: Elsevier Science. pp 133-160.
- Forget L, Ustinova J, Wang Z, Huss VA, Lang BF. 2002. *Hyaloraphidium curvatum*: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Mol Biol Evol* 19:310-319.
- Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G. 2002. The closest unicellular relatives of animals. *Curr Biol* 12:1773-1778.
- Leigh J, Seif E, Rodriguez N, Jacob Y, Lang BF. 2003. Fungal evolution meets fungal genomics. In: Arora D, ed. *Handbook of Fungal Biotechnology*. New York: Marcel Dekker Inc. pp 145-161.

# Comparative genomics of Hemiascomycete yeasts and mechanisms of genome dynamics

*Bernard Dujon and the Génolevures Consortium*

Institut Pasteur, Unité Genetique Moleculaire des Levures, 25, rue du Docteur Roux, 75724 Paris Cedex 15, France, bdujon@pasteur.fr

Eukaryotic molecular evolution proceeds through a succession of distinct events that leave only distorted and superimposed traces in living genomes. The hemiascomycete yeasts, with their compact genomes but distinct physiologies, provide a unique opportunity for unravelling the traces of these events and understanding the history of individual lineages. After our first survey of Hemiascomycete genomic diversity using low coverage sequencing (Souciet *et al.*, 2000), we have now completed the sequences of four yeast species, selected to represent a broad evolutionary range that *a posteriori* proved to be comparable to that of the entire phylum of Chordates.

*Candida glabrata*, the second causative agent of human candidiasis, is a pathogenic yeast phylogenetically related to *S. cerevisiae*. *Kluyveromyces lactis* is an industrial yeast commonly used for genetic studies. *Debaryomyces hansenii* is a halotolerant yeast phylogenetically related to *Candida albicans* and other pathogenic yeasts. *Yarrowia lipolytica* is a very distantly-related yeast that shares a number of properties with filamentous fungi. The genomes of the four species contain a total of *ca.* 24,500 genes whose translation products have been annotated and classified into families together with *S. cerevisiae* proteins. A significant proportion of these families appear specific to Hemiascomycetes, and a number of physiologically significant family-size expansions and specific gene losses were found.

Comparisons of chromosome maps and genome redundancies reveal that the different yeast lineages have evolved through a remarkable interplay between several distinct mechanisms, including tandem gene repeat formation, massive genome duplication, segmental duplications and gene loss. Results of this large scale comparative analysis will be discussed in the light of recent experiments performed in *S. cerevisiae* (Kozul *et al.*, 2004).

## Reference:

- Souciet, J. L. *et al.* Genomic exploration of the hemiascomycetous yeasts. *FEBS Lett.* **487**, 3-147 (2000).  
Kozul, R., Caburet, S., Dujon, B. and Fischer, G. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23**, 234-243 (2004).

The Génolevures Consortium (GDR2354 from CNRS) is coordinated by J.L. Souciet (Université Louis Pasteur/CNRS, 28 rue Goethe, 67000 Strasbourg, France); and includes Génoscope, 2 rue Gaston Crémieux, 91057 Evry Cedex, France, Pasteur Génopole Ile-de-France, Institut Pasteur, 25 rue du Dr. Roux, 75015 Paris, France; Laboratoire Bordelais de Recherche en Informatique (LaBRI), UMR 5800 CNRS, 351 cours de la Libération, 33405 Talence cedex, France; Centre de Bioinformatique de Bordeaux (CBiB), 146 rue Léo Saignat, 33076 Bordeaux cedex, France; Unité de Génétique Moléculaire des Levures (URA 2171 CNRS and UFR 927 Univ. Pierre et Marie Curie), Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France; Génétique Moléculaire et Cellulaire, INRA UMR 216, CNRS URA 1925, INA-PG, P.O. Box 01, 78850 Thiverval-Grignon, France); Institut de Génétique Moléculaire, IGM-bat 400, Université d'Orsay, 91405 Orsay Cedex, France; Institut de Biochimie et Génétique Cellulaires (IBGC), UMR 5095 CNRS, 1 rue Camille Saint-Saëns, 33077 Bordeaux cedex, France; Service de Biochimie et de Génétique Moléculaire, CEA/Saclay, 91191 Gif-sur Yvette, France; Microbiologie et Génétique ERS 2009, CNRS/UCB/INSA, bat. 405 R2, Université Lyon I, 69622 Villeurbanne Cedex, France; Unité de Génétique des interactions macromoléculaires (URA2171, CNRS) Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France; Modélisations et Simulations des Acides Nucléiques, UPR CNRS 9002, IBMC, 15 rue René Descartes, 67000 Strasbourg, France; Groupe Logiciels et Banques de données, Institut Pasteur, 25 rue du Dr. Roux, 75015 Paris, France.

# **Genome-scale approaches to addressing important phylogenetic questions: incongruence and the effect of taxon sampling**

*Antonis Rokas*

Howard Hughes Medical Institute and Laboratory of Molecular Biology, R. M. Bock Laboratories, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA, arokas@facstaff.wisc.edu.

The presence of incongruence between phylogenies obtained using different datasets and the relative contribution of taxon sampling and sequence dataset size to confidence in phylogenetic inference represent two major issues in phylogenetic research. Recent advances in genome sequencing of multiple fungal species offers a unique opportunity to address these questions on a genomic scale. The presence of incongruence generates uncertainty as to what are the true phylogenetic associations of the organisms in question. We systematically investigated the extent of incongruence, and potential ways of overcoming it, by analyzing the recently published genome data from eight yeast species.

Our results show that there is a widespread incongruence between phylogenies obtained from individual genes. Perhaps surprisingly, none of the factors known to mislead phylogenetic reconstruction (such as base composition bias, gene length, etc.) could systematically account for the observed incongruence. In sharp contrast, analyses of the entire data set (a concatenation of 106 genes) yielded a single, fully resolved species tree with absolute support. Similar results were obtained with concatenation of a minimum of twenty genes. Therefore, reliance on single or a small number of genes has a significant probability of supporting incorrect relationships. The relative contribution of taxon sampling and sequence data set size to confidence in phylogenetic inference also has important implications for phylogenetic inference. The strategy favored by most phylogeneticists is increasing the number of taxa. This approach has gained support by simulation studies which have demonstrated that this is crucial in phylogenies containing long branches.

To examine the trade-offs involved in increasing the number of taxa while keeping the amount of sequence data stable and vice versa, we have taken advantage of published genomic data available from 17 closely related yeast species. We have devised multiple tests to evaluate the relative merits of each approach under a variety of conditions using biological sequence data on a genomic scale. Our data argue for a very significant effect of data set size on the confidence of the inferred phylogeny. In summary, our results suggest that resolution of contentious branches of the Tree of Life may be accomplished by the tremendous power offered by genome-wide datasets.

## Quantitative and qualitative patterns in the evolution of gene content

*M.A. Huynen*

Nijmegen Center for Molecular Life Sciences, Center for Molecular and Biomolecular Informatics, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands, M.Huynen@cmbi.kun.nl

We study the evolution of gene content on a quantitative and qualitative level. Quantitatively we observe that the number of genes that two genomes share is mainly determined by their size and by their phylogenetic relatedness. By correcting for genome size the phylogenetic signal in a genome's gene content becomes so strong that a (distance-based) phylogeny can be constructed that is solely based on the fraction of genes two genomes share, and that is remarkably similar to "classical" sequence-based phylogenies. Nevertheless some discrepancies appear. These can partly be resolved by not including phylogenetic "discordant" orthologous groups of genes, whose phylogenetic distribution is dissimilar from majority of the orthologous groups.

On a qualitative level we use the similar phylogenetic distribution of genes to predict functional interactions between the proteins they encode: genes that have a similar phylogenetic distribution are expected to encode proteins that functionally interact, e.g. by being part of the same pathway or protein complex. Some successful examples of this method have been reported, like the prediction of the involvement of the disease gene frataxin in the assembly of iron-sulfur clusters in mitochondria. Nevertheless, a large-scale analysis of the phylogenetic distribution of genes that are known to be involved in a pathway in one species, shows little evolutionary modularity of such biomolecular systems, implying a high flexibility in the organization of proteins into pathways in the cell.

## **Phylogenomics: from prokaryotes to eukaryotes**

*B. Snel*

Nijmegen Center for Molecular Life Sciences, Center for Molecular and Biomolecular Informatics, Toernooiveld 1, 6525 ED Nijmegen, the Netherlands, b.snel@cmbi.kum.nl

Phylogenetic trees based on gene content are remarkably similar to the current consensus of life history. To reliably construct phylogenetic trees in this manner we need a measure for the presence and absence of genes, i.e. we need orthologous (not just homologous) genes. Another important issue in constructing genome trees is the correlation between genome size and the number of shared genes. The controversial topic of horizontal gene transfer can be examined by iteratively filtering genome phylogenies for inconsistent orthologous gene families. Such filtering reveals that the majority of genes contain some phylogenetic signal.

The presence and absence of genes does not only allow the reconstruction of phylogenies. It also allows the study of the evolution of cellular pathways and protein complexes. We have explicitly probed this issue by quantifying modularity in the evolution of such biomolecular systems: we study to what extent functional modules behave cohesively across genomes thereby answering the question “are functional modules also evolutionary modules?” We thereby find that there is only limited modularity in the evolution of molecular systems and that a substantial proportion of the apparent modularity is in fact phylogenetic, rather than functional in nature.

## **Gene dynamics: speciation and regulation**

*Paulien Hogeweg*

Theoretical Biology and Bioinformatics Group, University Utrecht, Padualaan 8, Utrecht, The Netherlands, p.hogeweg@bio.uu.nl

We will discuss mutual interactions between evolutionary adaptation, gene regulation and speciation. In the discussion we will combine insights obtained from genome analysis with insights obtained from bioinformatic modeling.

## Comparative genomic hybridisation provides new insights into the molecular taxonomy of yeasts

*Stephen G Oliver*

School of Biological Sciences, University of Manchester, The Michael Smith Building, Oxford Road, Manchester, M13 9PT, UK, [steve.oliver@man.ac.uk](mailto:steve.oliver@man.ac.uk)

The science of taxonomy is constantly improving as new techniques are developed. Current practice is to construct phylogenetic trees based on the analysis of the DNA sequence of single genes, or parts of single genes. However, this approach has recently been brought into question as several tree topologies may be produced for the same clade when the sequences for various different genes are employed.

The availability of complete genome sequences for several organisms has seen the adoption of microarray technology to construct molecular phylogenies of bacteria, based on all of the genes. Similar techniques have been used to reveal the relationships between different strains of the yeast *Saccharomyces cerevisiae*. We have exploited microarray technology to construct a molecular phylogeny for the *Saccharomyces* 'sensu stricto' complex of yeast species, which is based on all of the protein-encoding genes revealed by the complete genome sequence of the paradigmatic species, *S. cerevisiae*. We have also analysed different strains of *S. cerevisiae*, itself, as well as the putative species *S. boulardii*. We show that in addition to the phylogeny produced, we can identify and analyse individual ORF traits and interpret the results to give a detailed explanation of evolutionary events underlying the phylogeny.

## Genomic approaches to fitness and speciation in yeast

*Daniela Delneri*

School of Biological Sciences, University of Manchester, 2205 Stopford Building, Oxford Road, Manchester M13 9PT, UK, d.delneri@bso.man.ac.uk

The *Saccharomyces* ‘sensu stricto’ group includes yeast species that will mate with one another, but the resulting hybrids are sterile. Several chromosomal rearrangements has been found among this group and retrospective analysis of their genomes revealed that translocations have occurred between closely related species, while more distant ones have collinear genomes. Although these findings are suggesting that the chromosomal rearrangements are not a prerequisite for speciation in yeast, the specific contributions that such translocations make to reproductive isolation and to adaptive processes between *Saccharomyces* ‘sensu stricto’ species is still to be determined.

The principal problem in testing the impact of translocations on speciation and fitness was the impossibility at separating the effects of these karyotypic rearrangements from genome-wide genetic differences. To overcome this technical difficulty, a modification of the *cre/loxP* system was used to construct in *S. cerevisiae* some of the translocations detected in the sibling species *S. mikatae*.

This imposed chromosome collinearity allowed the generation of interspecific hybrids that produce a large proportion of spores that are viable, although extensively aneuploid, reinforcing the concept that reciprocal translocations could intensify the post-zygotic barriers once a species has arisen by another route. These data also suggest a mechanism for the generation of redundancy in the *S. cerevisiae* genome, which involves the possibility of genome duplication by allopolyploidisation or widespread aneuploidy.

Engineered *S. cerevisiae* strains, mimicking the genome structure present in *S. mikatae*, were also used in competition experiments to determine the contribution of reciprocal translocations to the whole organism’s fitness. The vast majority of competition experiments undertaken provided evidence for positive fitness effects of reciprocal translocations, with the rearranged strains of *S. cerevisiae* out-competing the reference *S. cerevisiae* strain with no translocation, especially under glucose limitation. Although it is always difficult to extrapolate laboratory results to natural systems, these data indicate that chromosomal rearrangements in *Saccharomyces* may provide selective advantages in some conditions, independently of any other genetic differences. This is compatible with a model of adaptive fixation of translocations in the *Saccharomyces* ‘sensu stricto’ group of species.

## Functional genomics of plant infection by the rice blast fungus *Magnaporthe grisea*

Nicholas J. Talbot, Darren M. Soanes, Joanna M. Jenkinson, Zheng Yi Wang, Lucy J. Holcombe, Madhumita Barooah, Martin J. Gilbert, and Gurpreet Bhambra

School of Biological Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter, EX4 4QG, UK, N.J.Talbot@exeter.ac.uk

The rice blast fungus, *Magnaporthe grisea* causes one of the most serious diseases of cultivated rice, and understanding the early events of the infection is of paramount importance if durable control measures are to be developed. *Magnaporthe grisea* develops a specialised infection structure called an appressorium which is used to penetrate the tough outer cuticle of rice leaves allowing the fungus entry to the underlying tissue.

Work by our research team has been aimed at defining the molecular determinants of infection-related development by *M. grisea* and determining the regulatory networks that control plant infection. The draft genome sequence of *M. grisea*, which was generated at the Whitehead Institute and released in 2002, has allowed an unparalleled insight into the gene inventory required by a pathogen such as *M. grisea* to enable it to cause disease. The fungus shows specific gene family expansion, for example, of genes encoding enzymes associated with colonisation of plant tissue. We have utilized the *M. grisea* genome information in comparative genomic analyses with a range of other phytopathogenic, human pathogenic, and saprotrophic fungal species. This has revealed conservation of a number of pathogen-specific genes among a broad phylogenetic range of fungi, and also highlighted considerable diversity in functionally-related groups of gene products. As part of the Consortium for Functional Genomics of Eukaryotic Microorganisms (COGEME), we have attempted to integrate genomic information from phytopathogenic fungi into an accessible format for utilisation by the wider research community. We are also collating fungal genomic information into a dynamic genome information management system to allow more elaborate interrogation of a range of functional genomic data to be carried out.

Genomic information has had considerable impact on our physiological and molecular genetic studies of *M. grisea* and the strategies we are adopting to study appressorium formation and function will also be discussed.

### References:

- Tunlid, A., Talbot, N.J. (2002) Genomics of parasitic and symbiotic fungi. *Current Opinion in Microbiology* 5: 513-519
- Holcombe, M., Holcombe L, Gheorghe, M and Talbot, N.J. (2003) A hybrid machine model for rice blast fungus *Magnaporthe grisea*. *Biosystems* 68: 223-228
- Giles, P.M., Soanes, D.M., and Talbot, N.J. (2003) A relational database for the discovery of genes encoding amino acid biosynthetic enzymes in pathogenic fungi. *Comparative and Functional Genomics* 4:4-15
- Soanes, D.M., Skinner, W., Keon, J., Hargreaves, J., Talbot, N.J. (2002) Genomics of phytopathogenic fungi and the development of bioinformatic resources. *Molec. Plant-Microbe Interact.* 15: 421-427
- Talbot, N.J. (2003) On the trail of a cereal killer: investigating the biology of *Magnaporthe grisea*. *Annual Review of Microbiology* 57: 177-202

## Comparative Fungal Genomics

*J. Galagan*

MIT Center for Genome Research, 320 Charles street, Cambridge, MA 02139, USA,  
jgalag@mit.edu

With the ever-increasing number of eukaryotic genome sequences being produced, the prospects for comparative genomics are unprecedented. There is, however, a crucial need to develop new tools and methods that will allow us to take full advantage of this opportunity.

Fungal genomes are ideal organisms for this purpose. The biological diversity within this group of organisms offers numerous informative comparisons spanning very recent divergence to over 900 million years of evolution. Their small genome size means that multiple whole genome comparisons are now possible. And because fungi share many important genomic characteristics with other eukaryotes, these comparisons inform us about the biology of all eukaryotes. Furthermore, due to their relatively compact genome structures, techniques can be more easily developed and tested in fungi as compared to more complex organisms.

To take advantage of the potential of fungal genomic research, the Broad Institute, in collaboration with the fungal scientific community, launched the Fungal Genome Initiative (FGI). As part of the FGI, we have generated and released 10 fungal genomes, with 10 additional genomes planned for sequencing in 2004. My group has also been working to develop the tools and techniques for analyzing these genome sequences to generate biological insights. I will provide an update on the status of the FGI and describe the results of our ongoing comparative analyses focusing on (1) the investigation of the patterns of genome evolution, (2) the evolution of gene exon structure, (3) the identification of conserved sequence elements.