Genomics of parasitic and symbiotic fungi

Anders Tunlid* and Nicholas J Talbot†

Complete and partial genome sequence information is underway in several parasitic and symbiotic fungi that infect humans, other animals and plants. Comparative analyses of these sequences will provide new insights into the genomic plasticity and evolution of parasitism and mutualism in fungi.

Addresses
*Department of Microbial Ecology, Lund University, Ecology Building, SE-223 62 Lund, Sweden; e-mail: anders.tunlid@mbioekol.lu.se
†School of Biological Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter, EX4 4QG, UK;
e-mail: N.J.Talbot@exeter.ac.uk

Current Opinion in Microbiology 2002, **5**:513–519

1369-5274/02/$ – see front matter

Abbreviations

AM	arbuscular mycorrhiza
BAC	bacterial artificial chromosome
EST	expressed sequence tag
HGT	horizontal gene transfer
mC	methylated cytosine

Introduction

Approximately 100,000 species of fungi have been described so far, and approximately 10% of these obtain nutrients by living in close association with other organisms, such as plants and animals, including humans. Many fungal infections are parasitic and can lead to severe diseases. Other infections are mutualistic symbioses that are beneficial to the host organism. This group includes infections caused by the mycorrhizal fungi that infect the roots of many important crops and forest trees. These fungi improve the growth of the host plants by facilitating the uptake of nutrients such as nitrogen and phosphate from the soil.

Our understanding of how parasitic and symbiotic fungi infect their hosts, including the mechanisms of host recognition, development of infection structures, control of host defense reactions, and penetration and colonization of the host tissues, is limited. However, it can be expected that this situation will change rapidly in the coming years, because a large amount of information from the genome sequences of fungal pathogens and symbionts will shortly become available. Over the past five years, a corresponding flow of information about prokaryotes has had a major impact on the research of bacterial pathogenesis and symbiosis [1,2]. Comparative genomics of strains and species of bacteria has also provided new insights into the evolution of virulence and host adaptations. The concurrent development of post-genomic methods to determine gene function has transformed research into bacteria–host interactions from a piecemeal study of individual genes and proteins to a more systematic analysis of the entire gene and protein complements of microbial pathogens.

Since completion of the *Saccharomyces cerevisiae* genome in 1996 [3], progress on the sequencing of other fungal genomes has been limited. However, early this year, the annotated genome of the fission yeast *Schizosaccharomyces pombe* was published [4], and genome sequencing of several fungal species is nearing completion. These species include the filamentous fungus *Neurospora crassa* (http://www-genome.wi.mit.edu/annotation/fungi/neurospora/), the human pathogens *Candida albicans* and *Cryptococcus neoformans*, and the phytopathogen *Magnaporthe grisea* (the causal agent of rice blast). Genome sequence information and expressed sequence tag (EST) collections from several other parasitic and symbiotic fungi that infect humans, other animals and plants are also becoming more widespread (Table 1). In this review, we discuss the recent achievements in fungal genomic analyses and how such data can provide new insights into genomic plasticity and the evolution of parasitic and mutualistic life styles.

Genome diversity of parasitic and symbiotic fungi

Compared with the genome sizes of other eukaryotes such as animals and plants, the genome sizes of fungi are small. *S. cerevisiae* and *S. pombe* have genome sizes of 13.7 Mb and 13.8 Mb, respectively [3,4]. Except for the filamentous ascomycete *Ashbya gossypii*, which has a genome size of 8.9 Mb, other filamentous ascomycetes and basidiomycetes have genome sizes between 13–42 Mb [5,6]. Thus, the genome sizes of fungi are approximately one-third of those of *Caenorhabditis elegans* and *Arabidopsis thaliana*, and are an order of magnitude smaller than the genome of rice (*Oryza sativa*). Furthermore, fungal genomes have a high gene density, and a low proportion of repetitive sequences. For example, *S. cerevisiae* contains a gene approximately every 2 kb [3], whereas the larger genome of *N. crassa* contains a gene every 4 kb (http://www-genome.wi.edu/annotation/fungi/neurospora). The gene density in *M. grisiae* is estimated to be approximately one gene every 4.2 kb [7], and for the ectomycorrhizal fungus *Paxillus involutus*, one gene every 2.8 kb [8].

The genomes of the arbuscular mycorrhizal (AM) fungi have unusual sizes and structures. AM fungi are obligate symbionts and are all found in the order Glomales. The genome sizes of these fungi have been estimated to be 100–1000 Mb, which is significantly larger than those of ascomycetes and basidiomycetes [9]. The GC (guanine and cytosine) content of AM fungi is also significantly lower (30–35%) than the range of 40–56% reported for other fungi [6,10]. Compared with other fungi, the genomes of glomalean fungi have a higher proportion of methylated cytosine (mC), a fact that offers some explanation for the evolution of the large, adenine and thymine (AT)-rich genomes of AM fungi [9,10]. Mutation of mC

Table 1

Sequencing projects on fungal pathogens and symbionts.

Species	Source URL and/or Reference	Status (June 2002)
Human pathogens		
Candida albicans	http://www-sequence.stanford.edu/group/candida [22••]	10.4x genome coverage (shotgun)
(Opportunistic infections)	http://www.sanger.ac.uk/projects/C_albicans/	7 cosmids
Aspergillus fumigatus	http://www.sanger.ac.uk/projects/A_fumigatus/	7 BAC, 5548 BAC end sequences
(Pulmonary pathogen)	http://tigrblast.tigr.org/ufmg/	3 × genome coverage (shotgun)
Cryptococcus neoformans	http://www.genome.ou.edu/cneo.html	4000 ESTs
(Causing meningitis)	http://www-sequence.stanford.edu/group/C.neoformans/	156 genomic DNA contigs, 17.1 Mb
	http://tigrblast.tigr.org/ufmg/	Shotgun sequencing complete; closure of gaps under way
	http://rcweb.bcgsc.bc.ca/cgi-bin/cryptococcus/cn.pl	4867 BAC end sequences
Pneumocystis carinii	http://biology.uky.edu/Pc	3896 ESTs
(Opportunistic infections, pneumonia)	http://www.sanger.ac.uk/projects/P_carinii/	
Plant pathogens		
Blumeria graminis f.sp *hordei*	http://cogeme.ex.ac.uk/ [46]	2701 unisequences (ESTs)
(Barley powdery mildew)		
Botrytis cinerae	http://www.genoscope.cns.fr	2858 unisequences (ESTs)
(Grey mold)	http://cogeme.ex.ac.uk	
Cladosporium fulvum	http://www.ncbi.nlm.nih.gov/dbEST	595 ESTs
(Tomato leaf mould)		
Fusarium graminearium	http://cogeme.ex.ac.uk [47]	2831 unisequences (ESTs)
(Fusarium head blight)		
Fusarium sporotrichioides	http://www.genome.ou.edu/fsporo.html	7495 ESTs
(Fusarium head blight)		
Magnaporthe grisae	http://cogeme.ex.ac.uk	5142 unisequences (ESTs)
(Rice blast)	http://www.ncbi.nlm.nih.gov/dbEST	13 008 ESTs
	http://www.fungalgenomics.ncsu.edu [48,49]	619 ESTs
	[7]	1 BAC
Mycosphaerella graminicola	http://cogeme.ex.ac.uk	2926 ESTs
(Wheat leaf blotch)	http://www.ncbi.nlm.nih.gov/dbEST [50]	
Phytophthora infestans	http://www.ncbi.nlm.nih.gov/dbEST [51]	2139 ESTs
(Late blight on potato and tomato)		
Phytophthora sojae	http://www.ncbi.nlm.nih.gov/dbEST [52]	2004 ESTs
(Stem and root rot on soybean)		
Other pathogens		
Metarhizium anisopliae	http://www.ncbi.nlm.nih.gov/dbEST	1693 ESTs
(Entomopathogen)		
Dactylaria candida	Anders Tunlid [53]	2006 unisequences (ESTs)
(Nematode-trapping fungus)		
Symbiotic fungi		
Paxillus involutus	Tomas Johansson*	6000 ESTs
(ectomycorrhiza)	[8]	1 cosmid
Pisolithus	http://www.ncbi.nlm.nih.gov/dbEST	379 ESTs
(ectomycorrhiza)	http://mycor.nancy.inra.fr/pages/DNAdatabases/ectomycorrhizaDB.html/ [37•]	850 ESTs
Glomus intraradices	http://www.ncbi.nlm.nih.gov/dbEST [54]	2963 ESTs
(AM, endomycorrhiza)		

*e-mail:Tomas.Johansson@mvbioekol.lu.se. AM, arbuscular mycorrhiza; BAC, bacterial artificial chromosome; EST, expressed sequence tag.

to T requires only C-4 deamination. Such random mC→T transitions will induce rapid divergence in mC-rich repeated sequences and, thus, increase the AT content of the genome. Furthermore, nucleotide divergence may reduce recombination events between repeats and, thus, lead to the accumulation of non-coding DNA in the genomes of AM fungi.

Another unique feature of AM fungi is the presence of large, vegetative, unicellular spores containing several hundreds or thousands of nuclei [9]. AM fungi appear to have been asexual for at least 400 million years. In the absence of recombination, it has therefore been hypothesized that individual AM fungi evolved to harbor many genetically divergent genomes. By using fluorescent DNA–DNA *in situ* hybridization (FISH) on nuclei from spores of the AM fungus *Scutellospora castanea*, Kuhn *et al.* [11••] recently demonstrated that AM fungi are multigenomic — a population of genetically different nuclei co-exists within an individual fungus. Furthermore, by using phylogenetic analysis, it was shown that the genetic variation that occurs within an individual AM fungus has evolved through accumulation of mutations in an essentially clonal genome with some infrequent recombination events.

The list of currently ongoing genome sequencing projects in Table 1 includes human, animal and plant pathogens, as well as mutualistic species. Although the majority of these projects are publicly supported, there are numerous genome sequences present in databases belonging to companies, with severe limitations imposed on access. For example, the genome sequence of the corn smut fungus *Ustilago maydis* is currently held in two private databases belonging to Bayer Ag (Leverkusen, Germany) and Exelixis (San Francisco, CA, USA). In the next few years, numerous fungal genomes are scheduled to be sequenced, owing largely to the Fungal Genome Initiative at the Whitehead Institute (Cambridge, MA, USA), which promises to sequence up to 15 fungal genomes selected on the basis of economic, taxonomic and academic criteria. This initiative should provide a platform for elaborate comparative genomic analysis across the fungal taxa, including an effective comparison of saprotrophic, pathogenic and mutualistic species.

In the current absence of fully sequenced genomes, the genome diversity of several fungal pathogens and symbionts has been explored by generating collections of ESTs (see Table 1). ESTs are single-pass, partial sequences of either the 5′ or 3′ ends of cDNA clones. Because the sequences are derived from genes expressed at a specific stage or in a certain tissue of the fungi, EST sequencing also generates some information on expression levels of genes. However, a large dataset needs to be analysed before any statistically significant differences in expression level can be inferred from frequencies of EST sequences [12]. EST sequences are often clustered into a set of non-redundant sequences (unisequences or contigs), in which each assembled contig represents a putative unique gene in the genome of the organism. Methods for obtaining EST sequences, including clustering, annotations and data mining, related to studies of gene diversity in fungal plant pathogens have been recently reviewed by Skinner *et al.* [13].

Comparative and evolutionary genomics

A striking observation from comparisons of available genome sequences from fungi and other organisms is that a significant proportion of the sequences exhibits no similarity to protein or DNA sequences present in databases. For example, 40–60% of the unisequences of ESTs from fungal plant pathogens display no or little similarity to proteins of known function [13]. Such genes have been called orphans (open reading frames [ORFs] of no known function) and are also commonly found in the genomes of eukaryotic model organisms. For example, it has been estimated that about one-third of all predicted protein-coding regions in *S. cerevisiae* are orphans [14]. The function of these orphans needs to be determined by genetic or biochemical approaches and is one of the major challenges for functional genomics. Although orphans most probably represent the limits of current empirical investigations of cellular function, from an evolutionary standpoint, there are two alternative major and mutually non-exclusive explanations for the high proportion of orphans in the genome of an organism. First, they may represent genes whose phylogenetic distribution is restricted to certain evolutionary lineages. Second, orphan genes might represent genes that rapidly diverge between closely related species. Proteins encoded by such genes can be unconstrained in sequence evolution or subjected to directional selection, whereas their structure and function might be conserved even between distantly related organisms [15].

In a large-scale comparative study of fungal genome and EST sequences, Braun and colleagues [16•] presented evidence to show that there is a higher proportion of orphans in *N. crassa* than in *S. cerevisiae*. Some of these orphans represent genes that are not present in *S. cerevisiae*, which may reflect the acquisition or maintenance of novel genes and is consistent with the larger genome size and morphological complexity of *N. crassa*. Whether or not the orphans in *N. crassa* also include rapidly evolving genes — a second explanation — has to be investigated by comparing the genomes of more closely related species than *S. cerevisiae*. Comparisons of sequences from closely related species has, for example, revealed the presence of fast-evolving genes in budding yeast [17] and in other eukaryotes, including *Drosophila* [15].

The order and transcriptional orientation of genes along a chromosome can change during evolution by DNA inversions, transpositions and by chromosomal translocations. Comparison of genome sequences between *C. albicans* and *S. cerevisiae* demonstrated that inversions of small segments of DNA, less than 10 genes long, has been the major cause of gene rearrangement [18•]. It was estimated that 1100 single-gene inversions occurred since the divergence (140 million–330 million years ago) between the two species. Furthermore, it was calculated that gene order has been broken as frequently by local rearrangements as by chromosomal translocations or long-distance transpositions. The proportion of local gene inversions has also been estimated by comparing data from a low-coverage random sequencing project of 13 species of hemiascomycetes (http://cbi.labri.u-bordeaux.fr/genoleuvures). According to the analysis of these sequences, the impact of gene inversions is limited between closely related species (within the genus *Saccharomyces*), whereas small-size inversions are a major cause of genomic reorganisation between distantly related species (such as *S. cerevisiae* and *C. albicans*) [19]. Recently, it has been proposed that local gene inversions result from a mechanism of gene duplication, but in an inverted orientation followed by loss of the original copy [20•]. Gene loss was mainly caused by the accumulation of mutations in one of the two copies rather than by DNA deletion. Overall, the rate of genome shuffling appears to be higher in eukaryotes than in prokaryotes, but appears to take place on a more local scale that often involves only single genes. Gross chromosomal rearrangements like translocations may therefore be more rare in fungi than in bacteria [21]. There are examples of conservation of gene order between distantly related fungi. Synteny between a 53 kb region of a sequenced bacterial artificial chromosome

(BAC) clone of *M. grisea* and a portion of the *N. crassa* genome has, for example, recently been demonstrated [7], but this has not been observed with any gene regions of *S. cerevisiae* and *C. albicans*.

Comparative genomics has recently been used to examine the possibility of a sexual life cycle in *C. albicans* [22••]. *C. albicans* is a diploid fungal pathogen with no known sexual stage and studies have indicated that naturally occurring populations are predominantly clonal [23]. Comparative genomic analysis showed that *C. albicans* possesses many homologues of genes that are found in sexual pathways of other eukaryotes, including *S. cerevisiae*, which indicate that *C. albicans* may have the genetic capacity to carry out a sexual cycle in nature [22••]. However, some of the genes known to be involved in meiosis in *S. cerevisiae* were lacking in *C. albicans*, whereas other meiotic gene homologues present in filamentous fungi and other eukaryotes were found in the *C. albicans* genome.

Molecular phylogeny has shown that parasitic and symbiotic fungi are found in many taxonomical groups, which suggests that these life styles have evolved repeatedly within the fungal kingdom. On the genomic level, there are basically three compatible mechanisms that can account for the multiple emergence and adaptations to parasitic and symbiotic growth in fungi. They are as follows: first, parasitism and symbiosis are associated with the presence of novel genes. Such genes may have a specific role during host infection and could be acquired by gene duplication or horizontal gene transfer (HGT). Second, adaptations to the parasitic and symbiotic habits may result from differences in the regulation of gene expression. Third, parasitism and symbiosis are associated with gene loss and deletions.

There are several studies that have demonstrated that parasitic fungi have unique pathogenicity factors. Among the best examples are the so-called host-specific toxins produced by several species of plant pathogens. Each of these molecules is produced by only one genotype of one fungal species, and is required for pathogenesis by that genotype [24]. The alternative explanation, in which parasitism/symbiosis is due to differential gene expression, can be exemplified by the presence of conserved elements of signaling pathways for infection-related development in diverse pathogenic species. The MAP kinase PMK1, for example, is related to the pheromone response MAPK FUS3 from *S. cerevisiae*, whereas in the phytopathogen *M. grisea*, it is required for appressorium development and invasive growth [25]. Significantly, PMK1 has functional homologues in a variety of other pathogenic species, many of which exhibit diverse morphology and infection habits. In each case, so far, the MAP kinase homologue has been shown to be essential for virulence [25]. There is also increasing evidence that gene loss or genome degradation has been of importance during the evolution of parasitism and symbiosis in bacteria [26]. There is a consistent correlation between genome size and the obligate associations with the host cell. Genome reduction in these organisms does not appear to be an adaptation for living inside the host, but seems merely to be due to a lack of selection for maintaining genes in these specialised microorganisms. Whether or not similar mechanisms occur in obligate fungal pathogens and symbionts is not yet known.

Analyses of genome sequences in bacteria have demonstrated that many of the genes required for virulence are restricted to pathogenic organisms and that they have been introduced into the genomes by HGT [2]. There is evidence for HGT in fungi [27]. In bacteria, HGT often involves transfer of a whole cassette of genes, ranging in size from 5 to 100 kb. If these gene cassettes contribute to virulence, they have been called ‘pathogenicity islands’ [2]. Notably, a similar type of island, called a ‘symbiosis island’, that contains a type III secretion system has recently been identified in an insect endosymbiont [28]. Pathogenicity islands are further defined by their genetic instability, restricted phylogenetic distribution, proximity to mobile genetic elements, atypical GC content relative to the rest of the genome, and affiliated repeated sequences. Several pathogenicity genes are known to be clustered in plant pathogenic fungi, including genes encoding plant growth regulators, toxins and secondary metabolites [27]. Some of these clusters, such as the pea pathogenicity (*PEP*) genes of *Nectria haematococca* [29] and the Tox1 locus in *Cochliobolus heterostrophus* [30], have features shared by prokaryotic pathogenicity islands, including differences in codon usage and GC content from other positions of the genome and the presence of highly repetitive DNA. The potential mechanisms by which transfer of such islands may occur in fungi are not known. The *PEP* cluster is located on a supernumerary chromosome and such chromosomes have been shown to have the capacity for transfer between genetically isolated individuals of a plant pathogenic fungus [31].

Other suspected cases of HGT in fungi invoke non-infective selfish genetic elements such as plasmids, introns and transposons. In addition, gene transfer through mechanisms similar to DNA transformation have been reported to take place in culture and in sterile soil [27]. The underlying basis of most studies to date is to identify features indicating that the evolutionary history of a gene differs from that of ancestral (vertically transmitted) genes. However, in many cases alternative explanations may also be consistent with the observations made [27].

Functional genomics of parasitic and symbiotic fungi

The acquisition and analysis of complete genome sequences is, of course, merely a starting point for generating new hypotheses on the mechanisms of pathogenesis and symbiosis. Results inferred by DNA and/or protein similarities provide, in most cases, only a small clue to putative function, and the avalanche of genome sequence data has to be combined with genome-wide experimental approaches to determine gene function. Several methods have been developed, such as large-scale mutagenesis, nucleic acid hybridisation technologies and protein chemistry. Some of these methods have been adopted for the analyses of fungal pathogens and symbionts [32].

The large-scale gene disruption strategies developed in *S. cerevisiae* are not yet applicable to most filamentous fungi because of their larger genomes and significantly lower rates of targeted integration during transformation. An elegant method for genome-wide mutagenesis studies in filamentous fungi based on *in vitro* transposition into cosmid libraries has, however, recently been described [33••]. The transposon carries a selectable marker for expression in the fungus (hygromycin resistance) and the transposon insertion sites can be determined by DNA sequencing. Thus, a library of insertional mutagenesis vectors can be generated and used as a means of both rapidly sequencing a fungal genome and simultaneously providing vectors for gene disruption studies. In an industrial setting, this provides a rapid means of identifying genes that are essential for pathogenesis and defining effective fungicide targets. An alternative strategy for altering gene expression is to use artificial gene suppression. Post-transcriptional gene silencing has been observed in several fungi [32] but, recently, De Backer *et al.* [34] showed that such an approach could be used to alter gene expression on a genome-wide scale in *C. albicans*. Gene suppression was achieved by combining antisense RNA inhibition and promoter interference.

The availability of genome sequences has made it possible to construct DNA microarrays for the simultaneous analysis of expression levels of large sets of genes in several species of fungal pathogens and symbionts. An array based on the *C. albicans* genome sequence data [22••] has recently been used to identify new cellular targets of several transcriptional regulators that play a central role in the control of metabolism and yeast→hypha morphogenesis in *C. albicans* [35]. De Backer *et al.* [36] used a *C. albicans* Gene Expression Microarray from Incyte Genomics, Inc. (Palo Alto, CA, USA) to identify possible targets in the fungus to itraconazole treatment. In the study of mutualistic associations, microarray analysis identified several novel symbiosis-regulated genes in the *Eucalyptus globulus–Pisolithus tinctorius* ectomycorrhiza [37•]. Comparisons of signals from free-living partners and symbiotic tissues revealed that 17% of the analysed genes were differentially regulated in the mycorrhizal root tissue.

Proteome analysis offers the possibility to directly identify the abundance of proteins, their localisation, interactions and post-translational modifications. Bestel-Corre *et al.* [38] followed the expression of proteins during the interaction between the model plant *Medicago truncatula* and the AM fungus *Glomus mossae*, and found that approximately 500 proteins could be resolved on two-dimensional (2-D) gels with 24 proteins that showed abundance differences or modifications during the development of the mycorrhiza. Proteome analyses have also recently been used for analysing the expression of proteins in cell walls of parasitic fungi [39–41].

Conclusions

Genome sequence information is currently being generated from several parasitic and mutualistic fungi that infect humans, other animals and plants. Analysis of this wealth of information is certain to provide unique insight into infection biology, host adaptation and the evolution of fungal pathogens and symbionts. Genome-wide comparisons have, however, to be combined with experimental approaches to assess individual gene function in a detailed manner. Several recent papers have shown that such methods can be adopted for large-scale analyses of gene function in parasitic and symbiotic fungi, and that several of these methods can precede the generation of a complete genome sequence. For example, a DNA array of randomly picked cDNA clones can identify sets of genes that are uniquely expressed during infection or, alternatively, are unique to certain strains or species of fungi that represent avariants in terms of host specificity or development.

A current bottleneck that limits the progress in the analysis of fungal genomes is the insufficiency of bioinformatics tools to analyse genome sequence data and genome-wide expression data [42]. The utility of the yeast genome has been aided enormously by custom-designed databases, such as the yeast proteome database (http://www.incyte.com/sequence/proteome/index.shtml) and the Stanford genome database (http://genome-www.stanford.edu/Saccharomyces). Soanes *et al.* [42] have developed a relational database for comparative analyses of EST sequences of phytopathogenic fungi. The database has been constructed as a part of the Consortium for the Genomes of Microbial Eukaryotes (COGEME) project (http://cogeme.man.ac.uk) and will be incorporated with the Genome Information Management System (GIMS) developed at the University of Manchester (http://www.cs.man.ac.uk/~norm/gims) [43]. The COGEME database allows rapid gene identification from six different phytopathogenic species and advanced querying to explore the conservation of putative gene functions among these pathogens. More specific databases have also been constructed for genome analyses of the plant pathogens *M. grisea* and *Phytophthora* spp [44,45].

Finally, it should be emphasised that genome sequences are available for several of the hosts of important plant pathogens and symbionts, including human, rice, poplar and legumes. These data can enable analysis of the dynamic and complex interactions between fungi and their hosts, providing further information for the development of novel methods to control fungal infections on humans, animals and plants. It is clear that the key challenge ahead will be to harness the power of genomic approaches to address the fundamental question in host–microbe interactions that has remained elusive to experimental study, namely, what determinants are required to allow a fungus to grow and proliferate within another living organism.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wren BW: **Microbial genome analysis: insights into virulence, host adaptation and evolution.** *Nat Genet* 2000, **1**:30-39.

2. Ochman H, Moran NA: **Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis.** *Science* 2001, **292**:1096-1098.

3. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M *et al.*: **Life with 6000 genes.** *Science* 1996, **274**:546-567.

4. Wood V, Gwillan R, Rajandream M-A, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S *et al.*: **The genome sequence of *Schizosaccharomyces pombe*.** *Nature* 2002, **415**:871-880.

5. Kupfer DM, Reece A, Clifton SW, Roe BA, Prade RA: **Multicellular ascomycetous fungal genomes contain more than 8000 genes.** *Fungal Genet Biol* 1997, **21**:364-372.

6. Yoder OC, Turgeon BG: **Fungal genomics and pathogenicity.** *Curr Opin Plant Biol* 2001, **4**:315-321.

7. Hamer L, Pan H, Adachi K, Orbach MJ, Page A, Ramamurthy L, Woessner JP: **Regions of microsynteny in *Magnaporthe grisea* and *Neurospora crassa*.** *Fungal Genet Biol* 2001, **33**:137-143.

8. Le Quéré A, Johansson T, Tunlid A: **Genome size and complexity of the ectomycorrhizal fungus *Paxillus involutus*.** *Fungal Genet Biol* 2002, **36**:234-241.

9. Gianinazzi-Pearson V, van Tuinen D, Dumas-Gaudot E, Dulieu H: **Exploring the genome of Glomelean fungi.** In *The Mycota IX Fungal Associations*. Edited by Hock B. Berlin: Springer Verlag; 2001:3-17.

10. Hosny M, de Barros J-PP, Gianinazzi-Pearson V, Dulie H: **Base composition of DNA from glomalean fungi: high amounts of methylated cytosine.** *Fungal Genet Biol* 1997, **22**:103-111.

11. •• Kuhn G, Hiljri M, Sanders IR: **Evidence for the evolution of multiple genomes in arbuscular mycroorhizal fungi.** *Nature* 2001, **413**:745-748.

By using fluorescent DNA–DNA *in situ* hybridization (FISH) on nuclei from spores of the AM fungus *Scutellospora castanea*, the authors demonstrated that this AM fungus is multigenomic, that is, it contains a population of genetically different nuclei that co-exists in one individual. This finding should be considered in any future studies on the genomics of AM fungi.

12. Audic S, Calverie J-M: **The significance of digital gene expression profiles.** *Genome Res* 1997, **7**:986-995.

13. Skinner W, Keon J, Hargreaves J: **Gene information for fungal pathogens from expressed sequences.** *Curr Opin Microbiol* 2001, **4**:381-386.

14. Oliver SG: **From DNA sequence to biological function.** *Nature* 1996, **379**:597-600.

15. Schmid KJ, Aquadro CF: **The evolutionary analysis of 'orphans' from the *Drosophila* genome identifies rapidly diverging and incorrectly annotated genes.** *Genetics* 2001, **159**:589-598.

16. • Braun EL, Halpern AL, Nelson MA, Natvig DO: **Large-scale comparison of fungal sequence information: mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*.** *Genome Res* 2000, **10**:416-430.

The paper indicates the potential for using large-scale comparative genomics for identifying possible gene innovations and gene loss in fungi.

17. Ozier-Kalegogeropoulos O, Malpertuy A, Boyer J, Tekaia F, Dujon B: **Random exploration of the *Kluveromyces lactis* genome and comparison with that of *Saccharomyces cerevisiae*.** *Nucleic Acids Res* 1998, **26**:5511-5524.

18. • Seoighe C, Federspiel N, Jones T, Hansen N, Bivolarovic V, Surzycki R, Tamse R, Komp C, Huizar L, Davis RW *et al.*: **Prevalence of small inversions in yeast order evolution.** *Proc Natl Acad Sci USA* 2000, **97**:14433-14437.

Evolution of gene order was studied by comparing the genomes of *C. albicans* and *S. cerevisiae*, two species separated by 140 million–330 million years. Gene order was substantially different between the two species. This is the first large-scale study that documents the importance of local inversions in shuffling the eukaryotic genome.

19. Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S *et al.*: **Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*.** *FEBS Lett* 2000, **487**:122-133.

20. • Fischer G, Neuvéglise C, Durrens P, Gaillardan C, Dujon B: **Evolution of gene order in the genomes of two related yeast species.** *Genome Res* 2001, **11**:2009-2019.

On the basis of sequences in the *S. cerevisiae* genome and a low coverage sequencing of the closely related *Saccharomyces baynus* var. *uvarum*, possible mechanisms for local gene inversions were identified (compare with [18•]).

21. Huynen MA, Snel B, Bork P: **Inversion and the dynamics of eukaryotic gene order.** *Trends Genet* 2001, **17**:304-306.

22. •• Tzung KW, Williams RM, Scherer S, Federspiel N, Jones T, Hansen N, Bivolarevic V, Huizar L, Komp C, Surzycki R *et al.*: **Genomic evidence for a complete sexual cycle in *Candida albicans*.** *Proc Natl Acad Sci USA* 2001, **98**:3249-3253.

The authors have sequenced the genome of *C. albicans* to 10.4-fold coverage. Comparative analyses of these sequences showed that *C. albicans* has many homologues of genes that are found in sexual pathways of other eukaryotes, including *S. cerevisiae*. These findings are of major interests because a sexual phase of *C. albicans* has not yet been detected.

23. Gräser Y, Volovsek M, Arrington J, Schönian G, Presber W, Mitchell TG, Vilgalys R: **Molecular markers reveal that population structure of the human pathogen *Candida albicans* exhibits both clonality and recombination.** *Proc Natl Acad Sci USA* 1996, **93**:12473-12477.

24. Kohmoto K, Yoder OC: **Molecular genetics of host-specific toxins in plant disease.** Dordrecht: Kluwer Academic Publishers; 1998.

25. Xu JR: **Map kinases in fungal pathogens.** *Fungal Genet Biol* 2000, **31**:137-152.

26. Tamas I, Klasson LM, Sandström JP, Andersson SGE: **Mutualists and parasites: how to paint yourself into a (metabolic) corner.** *FEBS Lett* 2001, **498**:135-139.

27. Rosewich UL, Kistler HC: **Role of horizontal gene transfer in the evolution of fungi.** *Annu Rev Phytopathol* 2000, **38**:325-363.

28. Dale C, Young SA, Haydon DT, Welbourn SC: **The insect symbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion.** *Proc Natl Acad Sci USA* 2001, **98**:1883-1888.

29. Han Y, Liu X, Benny U, Kistler HC, Van Etten HD: **Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectria haematococca*.** *Plant J* 2001, **25**:305-314.

30. Yang G, Rose MS, Turgeon BG, Yoder OC: **A polyketide synthase is required for fungal virulence and production of the polyketide T-toxin.** *Plant Cell* 1996, **8**:2139-2150.

31. He C, Rusu AG, Poplawsi AM, Irwin JAG, Manners JM: **Transfer of supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*.** *Genetics* 1988, **150**:1459-1466.

32. Sweigard JA, Ebbole DJ: **Functional analysis of pathogenicity genes in a genomics world.** *Curr Opin Microbiol* 2001, **4**:387-392.

33. •• Hamer L, Adachi K, Montenegro-Chamorro MV, Tanzer MM, Mahanty SK, Lo C, Tarpey RW, Skalchunes AR, Heiniger RW, Frank SA *et al.*: **Gene discovery and gene function assigment in filamentous fungi.** *Proc Natl Acad Sci USA* 2001, **98**:5110-5115.

The authors of this paper describe a novel method of genome sequencing and gene disruption in a seamless high-throughput process. Using *in vitro* transposition to generate insertions into cosmid clones, a library of insertional mutations scattered across the genome can be generated for systematic transformation into the host.

34. De Backer MD, Nelissen B, Logghe M, Viaene J, Loonen I, Vandoninck S, de Hoogt R, Dewaele S, Simons FA, Verhasselt P *et al.*: **An antisense-based functional genomics approach for identification of genes critical for growth of *Candida albicans*.** *Nat Biotechnol* 2001, **19**:235-241.

35. Murad AMA, d´Enfert C, Gaillardin C, Tournu H, Tekaia F, Talibi D, Marechal D, Marchais V, Cottin J, Brown AJP: **Transcript profiling in *Candida albicans* reveals new cellular functions for the**

transcriptional repressors CaTup1, CaMig1 and CaNrg1. *Mol Microbiol* 2001, **42**:981-993.

36. De Backer MD, Ilyiana T, Ma X-J, Vandoninck S, Luyten WHML, Vanden Bossche H: **Genomic profiling of the response of *Candida albicans* to itraconazole treatment using a DNA microarray.** *Antimicrob Agents Chemother* 2001, **45**:1660-1670.

37. • Voiblet C, Duplessis S, Encelot N, Martin F: **Identification of symbiosis-regulated genes in *Eucalyptus globulus-Pisolithus tinctorius* ectomycorrhiza by differential hybridization of arrayed cDNAs.** *Plant J* 2001, **25**:181-191.

This is the first paper to demonstrate the usefulness of cDNA arrays in analysing patterns of gene expression at a global level in mutualistic fungus–plant interactions.

38. Bestel-Corre G, Dumas-Gaudot E, Poinsot V, Dieu M, Dierick J-F, van Tuinen D, Remacle J, Gianinazzi-Pearson V, Gianinazzi S: **Proteome analysis and identification of symbiosis-related proteins from *Medicago truncatula* Gaertn. by two-dimensional electrophoresis and mass spectrometry.** *Electrophoresis* 2002, **23**:122-137.

39. Apoga D, Ek B, Tunlid A: **Analysis of proteins in the extracellular matrix of the plant pathogenic fungus *Bipolaris sorokiniana* using 2-D gel electrophoresis and MS/MS.** *FEMS Microbiol Lett* **197**:145-150.

40. Lim D, Hains P, Walsh B, Bergquist P, Nevalainen H: **Proteins associated with the cell envelope of *Trichoderma reesei*: a proteomic approach.** *Proteomics* 2001, **1**:899-910.

41. Bruneau J-M, Magnin T, Tagat E, Legrand R, Bernard M, Diaquin M, Fudali C, Latgé J-P: **Proteome analysis of *Aspergillus fumigatus* identifies glycosylphosphatidylinistol-anchored proteins associated to the cell wall biosynthesis.** *Electrophoresis* 2001, **22**:2812-2823.

42. Soanes DM, Skinner W, Keon J, Hargreaves J, Talbot NJ: **Genomics of phytopathogenic fungi and the development of bioinformatic resources.** *Mol Plant-Microb Interact* 2002, **15**:421-427.

43. Paton NW, Khan SA, Hayes A, Moussouni F, Brass A, Eilbeck K, Goble CA, Hubbard S, Oliver SG: **Conceptual modeling of genomic information.** *Bioinformatics* 2000, **16**:548-558.

44. Martin SL, Blackmon BP, Rajagopalan R, Houfek TD, Sceeles RG, Denn SO, Mitchell TK, Brown DE, Wing RA, Dean RA: **MagnaportheDB: a federated solution for integrating physical and genetic map data with BAC end derived sequences for the rice blast fungus *Magnaporthe grisea*.** *Nucleic Acids Res* 2002, **30**:121-124.

45. Waugh M, Hraber P, Weller J, Wu Y, Chen G, Inman J, Kiphart D, Sobral B: **The *Phytophthora* genome initiative database: informatics and analysis of distributed pathogenomic research.** *Nucleic Acids Res* 2000, **28**:87-99.

46. Thomas SW, Rasmussen SW, Glaring MA, Rouster JA, Christiansen SK, Oliver RP: **Gene identification in the obligate fungal pathogen *Blumeria graminis* by expressed sequence tag analysis.** *Fungal Genet Biol* 2001, **33**:195-211.

47. Kruger WM, Pritsch C, Chao S, Muehlbauer GJ: **Functional and comparative bioinformatic analyses of expressed genes from wheat spikes infected with *Fusarium graminearum*.** *Mol Plant-Microb Interact* 2002, **15**:445-455.

48. Rauyaree P, Choi W, Fang E, Blackmon B, Dean RA: **Genes expressed during early stages of rice infection with the rice blast fungus *Magnaporthe grisea*.** *Mol Plant Pathol* 2001, **2**:347-354.

49. Kim S, Ahn Il-P, Lee Y-H: **Analysis of genes expressed during rice-*Magnaporthe grisea* interactions.** *Mol Plant-Microbe Interact* 2001, **14**:1340-1346.

50. Keon J, Bailey A, Hargreaves J: **A group of expressed cDNA sequences from the wheat fungal leaf blotch pathogen, *Mycosphaerella graminicola* (*Septoria tritici*).** *Fungal Genet Biol* 2000, **29**:118-133.

51. Kamoun S, Hraber P, Sobral B, Nuss D, Govers F: **Initial assessment of gene diversity for the oomycete pathogen *Phytophthora infestans* based on expressed sequences.** *Fungal Genet Biol* 1999, **28**:94-106.

52. Quotob D, Hraber PT, Sobral BWS, Gijzen M: **Comparative analysis of expressed sequences in *Phytophthora sojae*.** *Plant Physiol* 2000, **123**:243-253.

53. Tunlid A, Ahrén D: **Application of genomics to the improvement of nematode pathogenic fungi.** In *NATO Advanced Research Workshop Enhancing Biocontrol Agents and Handling Risks.* Edited by Vurro M *et al.* Amsterdam: IOS Press; 2001:193-200.

54. Sawaki H, Saito M: **Expressed sequences in the extraradical hyphae of an arbuscular mycorrhizal fungus, *Glomus intraradices*, in the symbiotic phase.** *FEMS Microbiol Lett* 2001, **195**:109-113.