Fungal genome sequencing: basic biology to biotechnology

Krishna Kant Sharma

Laboratory of Enzymology and Recombinant DNA Technology, Department of Microbiology, Maharshi Dayanand University, Rohtak, Haryana, India

Abstract
The genome sequences provide a first glimpse into the genomic basis of the biological diversity of filamentous fungi and yeast. The genome sequence of the budding yeast, *Saccharomyces cerevisiae*, with a small genome size, unicellular growth, and rich history of genetic and molecular analyses was a milestone of early genomics in the 1990s. The subsequent completion of fission yeast, *Schizosaccharomyces pombe* and genetic model, *Neurospora crassa* initiated a revolution in the genomics of the fungal kingdom. In due course of time, a substantial number of fungal genomes have been sequenced and publicly released, representing the widest sampling of genomes from any eukaryotic kingdom. An ambitious genome-sequencing program provides a wealth of data on metabolic diversity within the fungal kingdom, thereby enhancing research into medical science, agriculture science, ecology, bioremediation, bioenergy, and the biotechnology industry. Fungal genomics have higher potential to positively affect human health, environmental health, and the planet’s stored energy. With a significant increase in sequenced fungal genomes, the known diversity of genes encoding organic acids, antibiotics, enzymes, and their pathways has increased exponentially. Currently, over a hundred fungal genome sequences are publicly available; however, no inclusive review has been published. This review is an initiative to address the significance of the fungal genome-sequencing program and provides the road map for basic and applied research.

Introduction
The first completed whole genome sequence of *Saccharomyces cerevisiae* strain S288C was a landmark in basic biology and fungal genomics (Goffeau et al., 1996). Thereafter, the genome of fission yeast, *Schizosaccharomyces pombe* (Wood et al., 2002), was sequenced, following *S. cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, and *Homo sapiens*. A comparative genomic approach has helped to identify and characterize the function of novel ORFs in *S. cerevisiae*, and has shown its probable function in the evolution of yeast species (Cai et al., 2008). To better understand its evolution and natural history, Rhind et al. (2011) have compared the genomes and transcriptomes of all known fission yeasts, i.e. *S. pombe*, *Schizosaccharomyces japonicus*, *Schizosaccharomyces octosporus*, and *Schizosaccharomyces cryophilus*. Interestingly, the comparative genome analysis of strains of *S. cerevisiae* and isolates of *Cryptococcus neoformans* showed variations in the structure and the function of several proteins (Engel & Cherry, 2013; Ormerod et al., 2013).

Despite the necessity, progress in sequencing fungal genomes had been slow at initial stages. However, during the last decade, the sequencing studies took pace due to the involvement of a consortium of mycologists in collaboration with scientists from the Whitehead Institute, MIT Center for Genome Research (now the Broad Institute). They launched the Fungal Genome Initiative (FGI-http://www.broad.mit.edu/annotation/fgi/), with the goal to sequence the genomes of fungi throughout the kingdom. Thereafter, a number of specific databases have been regularly updated, which provide great assistance to the different areas of fungal biology. The Genome Online Database (GOLD) is a site totally dedicated to whole genome-sequencing projects (http://www.genomeline.org) and has a mirror site at the Institute of Molecular Biology and Biotechnology, Crete, Greece (http://gold.imbb.forth.gr/) (Liolios et al., 2007). The fungal genome provides a deeper sequence understanding of the fundamental cellular processes common to all eukaryotes. It has solved some of our most urgent and major challenges in health, agriculture, enzyme biotechnology, bioenergy, and ecological diversity (Figure 1).

The fungal phylum Basidiomycota encompasses diverse organisms, from the single-celled *C. neoformans* to the large size *Armillaria* spp. They can be plant pathogens, or animal pathogens; some of them produce wonderful metabolites (laccases and peroxidases), including pigments and toxins (Sanodiya et al., 2009; Sharma & Kuhad, 2008). Despite their
importance, basidiomycete biology is poorly understood in comparison with ascomycete biology.

Further, the genomic information has provided the basic mechanism in understanding the cellular commands for the manufacture and maintenance of eukaryotic systems needed to control and operate, as well as a tool to make designer organism, which has potential to bring revolution in the pharmaceutical and fermentation industries (Annaluru et al., 2014; Frazier et al., 2003). This review is an attempt to highlight the importance of fungal genome-sequencing programs, in basic and applied research, which was disparaged for a long period of time.

**Fungal genome databases**

During the last decade, realizing the importance of fungal genomics, a consortium of mycologists in the year 2000, in collaboration with scientists from the Broad Institute, launched the Fungal Genome Initiative (FGI-http://www.broad.mit.edu/annotation/fgi/) (Table 1). The *Saccharomyces* Genome Database (SGD) is the community resource for whole genome of *S. cerevisiae* strain S288C, which provides information on functional annotations, mapping and sequence information, protein domains and structure, expression data, mutant phenotypes, physical and genetic interactions, and the primary literature from which these data are derived (http://www.yeastgenome.org/) (Goffeau et al., 1996). The Textpresso (http://textpresso.yeastgenome.org) search tool is available to search approximately 50 000 papers collected by the SGD project (Muller et al., 2004). The Pathway Tools interface provides a complete description of each pathway, with molecular structures, ‘enzyme commission numbers’ and full reference listing (http://pathway.yeastgenome.org) (Cherry et al., 2012). SGD uses the GBrowse, a genome browser to display diverse types of genomic information (http://browse.yeastgenome.org) (Stein et al., 2002). All the curated interaction data in SGD are loaded in bulk from the BioGRID database of physical and genetic interactions (Biological General Repository for Interaction Datasets; http://www.thebiogrid.org/) (Breitkreutz et al., 2008). Access to the fungal genomic data is also available through multiple online resources (Table 1).

The US Department of Energy-JGI, in 2012, completed 2635 projects, a three-fold increase over previous year, and generated >56 trillion nucleotides of genome-sequence data from microbial communities, fungi, algae, and plants (Nordberg et al., 2014). In the past year alone, JGI has added 650 genomes to the public databases out of which 63 belong to fungi (Figure 2). The 1000 fungal genome project is one of the latest JGI large-scale genomic initiatives focused at divergent fungal species to provide compressive data for studying complex metagenome.

The majority of fungal genomic resources generated in the past were for the dikarya (http://www.ncbi.nlm.nih.gov/ genomes/leuks.cgi) and mostly focused on fungi that are pathogenic, *Phytophthora infestans*, an Oomycetes, comprises a unique branch of destructive eukaryotic plant pathogens that are responsible for causing a number of the world’s most devastating diseases of dicot plants (Kamoun, 2003). These diseases are not only difficult to manage but they also cause enormous economic damage to important crop species such as

---

**Figure 1.** Economically important fungi with their genomes sequenced.
| S. no. | Genome            | Database                                                                 | Functions                                                                                                                                  | References                                                                 |
|-------|-------------------|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| 1.    | Candida albicans  | http://www.candidagenome.org/ (http://genolist.pasteur.fr/CandidaDB/)   | Functional information about genes and proteins; biochemical pathways and the textpresso full-text literature search tool                  | (Skrzypek et al., 2010)                                                    |
| 2.    | Aspergillus spp.  | http://www.aspergillus-genomes.org.uk, http://www.cadre-genomes.org.uk | Manage genomic data and to offer web-based tools for analysis and visualization of genomic features; provides access to annotated genomes, which are of particular importance to the Aspergillus genetics community (A. fumigatus Af293 and A1163, A. nidulans, A. oryzae, and A. niger); of these, two are clinical isolates (A. fumigatus Af293 and A1163) | (Gilsenan et al., 2012)                                                    |
| 3.    | Phytophthora spp. | http://www.pfgd.org                                                      | Allows intuitive navigation of transcript and genomic data; automated analyses are performed on all sequence data                          | (Gajendran et al., 2006)                                                    |
| 4.    | Saccharomyces cerevisiae | http://genome-www.stanford.edu/Saccharomyces/, http://www.yeastgenome.org, http://pathway.yeastgenome.org | SGD collects and organizes biological information about genes and proteins of the yeast S. cerevisiae. It provides genome-wide computational analyses for S. cerevisiae, or for comparative studies; genetics and cellular biology of other fungal genera that are important to basic research, human health and industry | (Cherry et al., 1998, 2012; Christie et al., 2004; Engel & Cherry 2013; Engel et al., 2010) |
| 5.    | Fungi             | http://FungiDB.org                                                      | Functional genomic resource for pan-fungal genomes; FungiDB provides a data-mining interface to the comparative and functional genomic data of multiple species of fungi that differs from the species-focused resources of SGD, CGD, and AspGD | (Stajich et al., 2012)                                                    |
| 6.    | TIGR fungal database | www.tigr.org.tbd/fungal                                               | Collection of species-specific databases that use a highly refined protocol to analyze EST sequences                                         | (Galagan et al., 2005)                                                    |
| 7.    | Fungi             | http://www.genomesonline.org                                           | Centralized monitoring of genome and metagenome projects worldwide                                                                       | (Pagani et al., 2012)                                                    |
| 8.    | Fungi             | http://www.broad.mit.edu/FGI/                                          | Genome sequence of fungi from throughout the kingdom                                                                                      | (Rhind et al., 2011)                                                    |
| 9.    | Saccharomyces cerevisiae | http://www.thebiogrid.org/                                              | BioGRID contains high-throughput interaction data for S. cerevisiae and other model organisms                                             | (Breitkreutz et al., 2008)                                                |
| 10.   | Fungi             | http://mips.gsf.de/projects/fungi/                                     | Munich Information Center for Protein Sequences                                                                                           | (Mewes et al., 2010)                                                    |
| 11.   | Ustilago maydis   | http://mips.gsf.de/genre/proj/ustilago/                               | MUMDB: Ustilago maydis genome database                                                                                                   | (Mewes et al., 2010)                                                    |
| 12.   | Schizosaccharomyces pombe | http://www.pombase.org/                                                  | PomBase is a comprehensive database for the fission yeast Schizosaccharomyces pombe, providing structural and functional annotation, literature curation and access to large-scale data sets                   | (Rhind et al., 2011)                                                    |
| 13.   | Fungi             | http://fgi.doe.gov/fungi                                               | MycoCosm is a fungal genomics portal, developed by the US Department of Energy Joint Genome Institute to support integration, analysis and dissemination of fungal genome sequences and other “omics” data by providing interactive web-based tools. JGI Genome Portal automatically generates and monitors BioSample and BioProject submissions to NCBI | (Grigoriev et al., 2014; Nordberg et al., 2014) |
potato, tomato, and soybean. The *Phytophthora* Functional Genomics Database (PFGD; http://www.pfgd.org) has been integrated with the Solanaceae Genomics Database (SolGD; http://www.solgd.org) to provide insight into the mechanisms of infection and resistance, specifically as they relate to the genus *Phytophthora* pathogens and their plant hosts (Gajendran et al., 2006). The *Candida* Genomes Database (CGD, http://www.candidagenome.org/) is another online resource of gene and protein information for the opportunistic fungal pathogen, *Candida albicans* (Jones et al., 2004). The comparative genomic analysis of eight closely related *Candida* species has revealed new features to existing CGD, i.e. Biochemical Pathways and the Textpresso (Butler et al., 2009; Skrzypek et al., 2010). Merging of *Aspergillus* website (http://www.aspergillus.org.uk) and the Central *Aspergillus* Data Repository, CADRE (http://www.cadre-genomes.org.uk) has benefited from extensive cross-linking with medical information to create a unique resources, spanning genomics, industrial, and clinical aspects of the genera (Gilsenan et al., 2012).

### Need to sequence fungal genome

Completion of the *S. cerevisiae* genome sequencing by Goffeau et al. (1996) and several prokaryotic genomes has led to a great incentives for the cause to sequence other fungal genomes. In one-and half-decades, there has been an impressive rise in fungal genomics that has greatly expanded our understanding of the genetic, physiological, and ecological diversity of these organisms. Many of these sequenced species...
form clusters of related organisms designed to enable comparative studies. The genome data explore an unparalleled opportunity to study the genetic composition and genomic comparison of the medically, industrially, and environmentally important fungi to unravel its evolutionary relatedness. Comparative studies of filamentous fungal genomes with essential and highly conserved components of normal development in animals have suggested that fungal and animal lineages may have diverged from their common originator, well before the emergence of any multicellular arrangement (Moore & Meškauskas, 2006; Wainright et al., 1993).

The filamentous fungi and yeasts have enormous potential for useful metabolites such as organic acids, enzymes, and a variety of pharmaceuticals. Citric acid fermentation by A. niger has been studied for nearly 100 years, but with the availability of genome data, the central carbon metabolism has been modeled (Pel et al., 2007). Genomic data produced evidence of genetic redundancy in economically important fungi, which can eventually decrypt the secret of differential enzyme productions, secretion pathway, and other key factors for pathogenicity. The genome sequence of the model fungus, Neurospora crassa, was the starting point in the study of animal and plant pathogenesis, biotechnology, molecular genetics, biochemistry, physiology, molecular cell biology, photobiology, circadian rhythms, gene silencing, ecology, and evolution (Galagan et al., 2003). A high-quality draft sequence of the N. crassa genome has approximately 10,000 protein-coding genes, which is more than twice as many as in the fission yeast, S. pombe and only about 25% fewer than in the fruit fly, Drosophila melanogaster (Galagan et al., 2003). Many predicted Neurospora proteins have no homologues in the yeasts, S. cerevisiae and S. pombe, but are similar to proteins in animals, plants, and other filamentous fungi. These genomic features approve Neurospora to be an excellent eukaryotic model system for the studies of numerous aspects of biology.

Fungal genome and human health

Fungi have been associated with a number of human diseases, some acute and others chronic. Therefore, the initial priority of genome sequencing was for fungi that present a serious threat to human health or serve as important models for biomedical research (Table 2). Candida albicans, one of the first eukaryotic pathogens selected for genome sequencing, is the most commonly encountered human fungal pathogen, causing skin and mucosal infections in generally healthy individuals and life-threatening infections in persons with severely compromised immune function. The significance of the extensive allelic differences in C. albicans is unknown, but may function to increase genetic diversity and contribute to the evolution of drug resistance (Cowan et al., 2002). Candida genome sequence (14 Mb) reveals several genes for environmental sensing (calmodulin signaling pathway, including a protein kinase), adaptation (pH regulatory genes), and encodes a small family of chloride channels with members resembling to a variety of mammalian tissues (Jones et al., 2004). Comparative genomic analyses could provide important clues about the evolution of the pathogens and its mechanisms of pathogenesis, which is needed for disease management (Jones et al., 2004).

Coccidioides immitis and Coccidioides posadasi are primary pathogens of immunocompetent mammals, including humans, which infect at least 150,000 people annually; 40% of whom develop a pulmonary infection (Hector & Laniado-Laborin, 2005; Sharpton et al., 2009). There genome sequences suggesting that pathogenic Coccidioides species are not only soil saprophytes but the adaptive changes to existing genes and the acquisition of a small number of new genes that has transformed them from plant based to animal based for nutritional support (Sharpton et al., 2009). Another Ascomycetes fungus, Aspergillus fumigatus, is a prototypical airborne opportunistic pathogen, affecting a wide range of susceptible patient groups, particularly those with neutropenia, receiving corticosteroids, with T cell defects (Ronning et al., 2005). Much of the basic biology of this organism was poorly understood, but the recent completion of its genome sequence and comparative genomic studies of two clinical isolates of A. fumigatus revealed species-specific chromosomal islands, which contribute to their rapid adaptation to heterogeneous environments (Fedorova et al., 2008; Ronning et al., 2005).

The Basidiomycetous yeast, Cryptococcus neoformans, causes life-threatening infections in the lungs and central nervous system, and is regarded as one of the most important pathogens in human fungal infections. Loftus et al. have sequenced ~20 Mb genome of C. neoformans, basidiomycetous model yeast for fungal pathogenesis, which contains ~6500 intron-rich gene structures and encodes a transcriptome abundant in alternatively spliced and antisense messages (Loftus et al., 2005). Interestingly, transposon richness in the genome may be responsible for karyotype instability and phenotypic variation, which eventually exhibits changes in key virulence factors, nutrient acquisition and metabolic profiles (Loftus et al., 2005; Ormerod et al., 2013). Genome of 118 isolates of C. gattii have recently been sequenced, including the North American Pacific Northwest (PNW) subtypes and the global diversity of molecular type VGII, to better ascertain the natural source and genomic adaptations leading to the emergence of infection in the PNW (Engelthaler et al., 2014). It also illustrates the importance of population diversification as a result of micro-evolution in pathogenic organisms and the clonal expansion under differential selection pressure (Ormerod et al., 2013).

The genome of two basidiomycete fungi Malassezia globosa and Malassezia restricta, responsible for dandruff and seborrheic dermatitis, and systemic infections in patients with repressed immune system have recently been sequenced (Martinez et al., 2012). The Malassezia genome lacks fatty acid synthase and contains several secreted lipases and hydrolases, which are helpful in adapting to human skin. Phylogenetic analysis using available whole genome sequences revealed that M. globosa, a human fungal pathogen closely related to maize pathogen Ustilago maydis, implying an ancestral shift from plant to animal host preference (Xu et al., 2007). It is remarkable that clusters of genes for secreted proteins from M. globosa are also found in the U. maydis genome, although these proteins appear to function in plant-specific interactions. The genome
of *Trichophyton rubrum* has also been sequenced and compared with the related species to find out the candidate genes involved in human skin and nail infections (Martinez et al., 2012). Further, the genome sequence of the dermatophytes and related species, i.e. *Trichophyton tonsurans*, *Trichophyton equinum*, *Microsporum canis*, and *Microsporum gypseum*, has revealed that the dermatophytes entire genome has duplicated in the course of evolution and were enriched with genes for proteases, kinases, secondary metabolites, and lysine motif (LysM), that may contribute to
the ability of these fungi to cause disease (Martínez et al., 2012).

**Fungal genome and plant disease**

Overall, a crop loss due to fungal infection exceeds $200 billion annually. Consequently, the control of plant diseases is crucial for sustainable production of food and reductions in agricultural use of land, water, fertilizers, and fuel. However, the disease control has been hampered by a limited understanding of the genetic and biochemical basis of pathogenicity, including mechanisms of infection and of resistance in the host. The top 10 most important plant-pathogenic fungi on the basis of economic importance are: (i) *Magnaporthe oryzae*, (ii) *Botrytis cinerea*, (iii) *Puccinia* spp., (iv) *Fusarium graminearum*, (v) *Fusarium oxysporum*, (vi) *Blumeria graminis*, (vii) *Mycosphaerella graminicola*, (viii) *Colletotrichum* spp., (ix) *Ustilago maydis*, and (x) *Melampsora lini* (Dean et al., 2012) (Table 3).

It is estimated that each year enough rice is destroyed by rice blast disease caused by fungus *M. oryzae*, to deprive feed for 60 million people (Dean et al., 2012). Analysis of the genes set and active transposable elements from the genome sequence of *M. oryzae* (38.8 Mb) provides an insight into the invasion and coexistence with widespread paddy cultivation by a fungus to cause disease (Dean et al., 2005). Recent reports of a robust gene expression dataset of enzymes that break down plant cell walls; this may represent

The rust fungi are highly destructive plant pathogens and are obligated on their plant host. The genome sequences of four rust fungi (two *Melampsoraceae* and two *Pucciniaceae*) have been analyzed so far. Genome-wide analyses of these species, as well as transcriptomics performed on a broader range of rust fungi, revealed hundreds of small secreted proteins considered as rust candidate secreted effector proteins (CSEPs). *Puccinia* is the largest genus of rust fungi and currently contains approximately 4000 species. *Puccinia graminis, P. triticina,* and *P. striiformis* represent distinct lineages within the cereal and grass rusts. The genome of multiple isolates of *P. striiformis f. sp. tritici* has been sequenced to identify CSEPs and relate them to their distinct virulence profiles (Cantu et al., 2013). Integration of genomics, transcriptomics, and effector-directed annotation of *P. striiformis f. sp. tritici* isolates has enabled the development of a framework for mining effector proteins in closely related isolates and relate these to their virulence profiles. Recent analysis of the genome size of 30 rust species (representing eight families) revealed the occurrence of very large genome sizes, including the two largest fungal genomes ever reported, *Gymnosporangium fusiform* (893.2 Mb) and *Puccinia chrysanthemi* (806.5 Mb) (Tavares et al., 2014). Another plant fungus, *F. graminearum*, has become adapted to different kinds of hosts and environments, which has resulted in major economic losses in the grain production worldwide. Analysis of the 36 Mb genomes provides deep insight into how this fungus adapts and interacts with its host. Comparison and analysis of the genomes of three pheno-typically diverse species, i.e. *F. graminearum, Fusarium verticilloides,* and *F. oxysporum* f. sp. *lycopersici* revealed lineage-specific and nutrition-specific genomic regions, which may be responsible for the conversion of non-pathogenic strains into a potential pathogen (Ma et al., 2010; Pain & Hertz-Fowler, 2008).

The plant-pathogenic fungus *Mycosphaerella graminicola* (asexual stage: *Septoria tritici* or *S. nodorum*) causes *septoria tritici* blotch, a disease that globally reduces the yield and quality of wheat (Goodwin et al., 2011). The 21-chromosome comprising 39.7 Mb genome of *M. graminicola* has revealed an apparently novel origin for dispensable chromosomes by horizontal transfer followed by extensive recombination, a possible mechanism of undetected pathogenicity and exciting new aspects of genome structure. It was found that structural rearrangements have strongly affected eight small dispensable chromosomes and positive selection in a small number of genes, while the essential chromosomes were syntenic (Stukenbrock et al., 2010). A surprising feature of the *M. graminicola* genome is a low number of genes for enzymes that break down plant cell walls; this may represent an evolutionary response to evade detection by plant defense mechanisms (Goodwin et al., 2011).

**Fungal genome in bioactive compound production**

The fungal kingdom includes many species with unique and unusual biochemical pathways for economically important, low molecular-weight bioactive compounds classified together as secondary metabolites (Keller et al., 2005). The main groups of fungal secondary metabolites are (1) peptides, (2) alkaloids, (3) terpenes, and (4) polyketides (Keller et al., 2005). Large-scale genome-sequencing projects have helped to anticipate the capacity of basidiomycetes to synthesize polyketides (Lackner et al., 2012). The polyketide pathway is known for the production of biotechnologically important antibiotics (e.g. tetracycline and erythromycin), immunosuppressant, antitumor agents (e.g. epothilone), and lipid-lowering drugs (e.g. lovastatin). In addition, various pigments, polyphenols, as well as a plethora of mycotoxins, such as the aflatoxins and fumonisins, are also produced (Crawford & Townsend, 2010; Lackner et al., 2012).

Genome-mining efforts indicate that the capability of fungi to produce secondary metabolites has been substantially underestimated because their gene clusters are silent under standard cultivation conditions (Brakhage, 2013). Bioinformatic algorithms such as SMURF (Khalid, 2010), antiSMASH (Medema et al., 2011), and FungiFun (Priebe et al., 2011) thus allow identification of secondary metabolism gene clusters. Recently, Chen et al. (2012) have reported the complete genome sequence of medicinal mushroom, *Ganoderma lucidum* strain 260125-1 (43.3 Mb), and identified a large set of genes and potential gene clusters involved in putative secondary metabolism and wood degradation (Sanodiya et al., 2009). The organism is well known for lignin
Table 3. Whole genome sequencing of fungi causing plant disease.

| S. No. | Organism | Genome size (Mb) | GenBank accession/release date | Funder/consortium | Significance | References |
|--------|----------|-----------------|-------------------------------|-------------------|-------------|------------|
| 1.     | *Pyrenophora semeniperda* (anamorph *Drechslera campulata*) | 32.5 Mb | ATLS00000000 | National Research Initiative, United States Department of Agriculture | Necrotrophic, fungal seed pathogen; has a wide host range within the Poaceae family | (Soliai et al., 2014) |
| 2.     | *Phytophthora infestans* | 240 | AATU01000000/9/2009 | Broad Institute | Most destructive disease of potato; annual world-wide potato crop losses due to late blight are conservatively estimated at $6.7 billion | (Haas et al., 2009) |
| 3.     | *Aspergillus flavus* | – | – | – | A. flavus infection causes ear rot in corn and yellow mold in peanuts | Unpublished |
| 4.     | *Magnaporthe grisea* 70-15 | 40 | 10/31/2003 | International Rice Blast Genome Consortium | Cause rice blast fungus, rice rotten neck, rice seedling blight, blast of rice, oval leaf spot of Gramineae, pitting disease, ryegrass blast | (Dean et al., 2005) |
| 5.     | *Fusarium graminearum* | | | Broad Institute | Cause head blight, crown rot, and root rot of cereals | (Cuomo et al., 2007) |
| 6.     | *Fusarium oxysporum f. sp. melonis* NRRL 26406 | 54.03 | AGNE00000000.1 | Broad Institute | Important group of fungal plant pathogens, cause various diseases on nearly every economically important plant species and pose health hazard to humans and livestock. Produce mycotoxins. Besides their economic importance, species of *Fusarium* also serve as key model organisms for biological and evolutionary research | (Ma et al., 2014) |
| 7.     | *F. pseudograminearum* CS3220 | 37.16 | CBMC010000000/2013 | Bioplatforms, Australia | Cause crown rot disease; in Australia, crown rot is responsible for annual losses of approx $100 million in the wheat and barley industries | (Moolhuijzen et al., 2013) |
| 8.     | *Ustilago maydis*521 | 20.5 | 07/29/2003 | Broad Institute | Causes smut on maize and teosinte | (Kämper et al., 2006) |
| 9.     | *Puccinia graminis* | – | – | – | Disease of bread wheat, durum wheat, barley, and triticale | (Tavares et al., 2014) |
| 10.    | *Mycosphaerella graminicola* (anamorph *Septoria tritici*) | 39.7 | 2011 | DOE Joint Genome Institute | Wheat plant pathogen causing septoria leaf blotch | (Goodwin et al., 2011) |
| 11.    | *Alternaria brassicicola* ATCC 96836 | 29.6 | 03/10/2009 | Washington University Genome Center (WUGC) | Causes black spot disease on virtually every important cultivated Brassica species | Unpublished |
| 12.    | *Alternaria arborescens* CBS 102605 | 34.0 | AIC01000000 | Ohio Agricultural Research and Development Center’s Research Enhancement Competitive Grants Program (SEED) | Fungal plant pathogens cause serious agricultural losses worldwide, major pathogen of tomato | (Hu et al., 2012) |
degrading enzymes, i.e. laccases, lignin-oxidase, and manganese-oxidase (Sharma et al., 2013). Another study reports 28.15 Mb genome sequences of Omphalotus olearius revealing a diverse network of sesquiterpene synthases and two metabolic gene clusters associated with anticancer allidin sesquiterpenoids biosynthesis (Wawrzyn et al., 2012). By combining genomic and biochemical information, Wawrzyn et al. (1992b) had developed a predictive framework that will allow researchers to tap into the vast terpenomes from Basidiomycota, and target-specific biosynthetic genes for heterologous pathway engineering in order to produce novel bioactivities (Ward et al., 1992b).

**Fungal genome in food and feed**

The two industrial Aspergilli species (Aspergillus niger and Aspergillus oryzae) contain the highest percentage of extracellular enzymes used in the food and feed industry, i.e. cellulases, hemicellulases, pectinases, amylases, inulinases, lipases, and proteases (Pel et al., 2007). The ability to secrete large amounts of proteins, development of a transformation system and inability of producing aflatoxin, has qualified A. oryzae in modern biotechnology for the production of traditional fermented foods and beverages (Ward et al., 1992b). Aspergillus oryzae, unlike A. flavus, does not produce aflatoxin, and its long history of use in the food industry has proved its importance for mankind (Ward et al., 1992b). The genome (37 Mb) of A. oryzae contains 12074 genes and has additional 7–9 Mb sequence in comparison with the genomes of A. nidulans and A. fumigatus (Machida et al., 2005). The genome sequences are enriched with genes involved in the synthesis of secondary metabolites, amino acid, and sugar uptake transporters to further sustain A. oryzae, as an ideal microorganism for fermentation (http://www.aspergillus-genomes.org.uk) (Machida et al., 2005). The genomes of A. oryzae, A. fumigatus, and A. nidulans revealed 135, 99, and 90 secreted proteinase genes, respectively, which constitute roughly 1% of the total genes in each genome. Similarly, A. oryzae possesses more secretory proteinase genes that function in acidic pH, including aspartic proteinase, peptatin-insensitive proteinase, serine-type carboxypeptidase, and aorin. This may be responsible for the increase in the adaptation of A. oryzae to an acidic pH during the course of its domestication (Machida et al., 2005). Considering the industrial importance, 33.9 Mb long genome of A. niger was sequenced and, thereafter, the annotated genome data were transformed into a model of central carbon metabolism, i.e. citrate synthesis (Pel et al., 2007). Different glycosyl hydrolase, lyase, and esterase families involved in the polysaccharide degradation in the Aspergilli sequenced were identified using the carbohydrate-active enzymes (CAZy) classification (http://www.cazy.org/). The A. niger genome revealed abundance of proteins involved in proteolytic degradation, including a variety of secreted aspartyl endopeptidases, serine carboxypeptidases, di- and tripeptidylaminopeptidases, and several secondary metabolite clusters. Another filamentous fungus, Ashbya gossypii, is currently used as an attractive model to study filamentous growth and in the food industry for the production of vitamin B2. Dietrich et al. (2004) had sequenced and annotated the genome of A. gossypii. With a size of only 9.2 Mb, encoding 4718 protein-coding genes, it is the smallest genome of a free-living eukaryote, characterized to date. More than 90% of A. gossypii gene shows both homology and a particular pattern of synteny with S. cerevisiae, qualifying the additional role in basic research (Dietrich et al., 2004).

The hypogeous fruiting body and an ectomycorrhizal symbiont, i.e. Tuber melanosporum Vittad (truffle), and true morel Morchella conica are in worldwide demand for its delicacy. Identification of processes and conditions that trigger fruit body formation can be facilitated by a thorough analysis of truffle genomic sequences. To obtain a better understanding of the biology and evolution of the ectomycorrhizal symbiosis, 125 Mb long sequence of the haploid genome of T. melanosporum has been completed (Martin et al., 2010). The genome contains ~7500 protein-coding genes with very rare multi-gene families. The upregulation of genes encoding for lipases and multi-copper oxidases suggest that T. melanosporum degrades its host cell walls during colonization (Martin et al., 2010). Further, many Morchella species are highly valued as edible species, but their cultivation is still a challenge. Recent fungal complete genome sequencing of M. conica CCBAS932 (http://1000. fungalgenomes.org/home/) may provide a potential clue in solving cultivation problems.

**Fungal genome and lignin degradation**

Lignin is a heterogeneous phenolic polymer that provides strength and rigidity to wood, and protects cellulose and hemicellulose from microbial attack. The wood rotting basidiomycete, P. chrysosporium, has a plethora of genes that enable wood degradation. First to be sequenced was the P. chrysosporium genome and has revealed several new isozymes for many previously identified lignocellulolytic enzymes, such as manganese peroxidase (MnP) and copper radical oxidases (Kersten & Cullen, 2007; Martinez et al., 2004). It has also revealed new putative flavine adenine dehydrogenase (FAD)-dependent oxidases, which have been predicted to take part in lignocellulose degradation (Martinez et al., 2004). The majority of genome sequencing projects after P. chrysosporium (Martinez et al., 2004) is focused on lignocellulose degraders, such as G. lucidum (Chen et al., 2012) and brown-rot, such as Postia placenta (Martinez et al., 2009), and Serpula lacrymans (Eastwood et al., 2011). Another basidiomycete, i.e. Schizophyllum commune, has been reported to be a pathogen of humans and trees, but it mainly adopts a saprobic lifestyle by causing white-rot. Sequencing of the 38.5 Mb genome assembly of S. commune strain H4-8 revealed 11.2% repeat content. Compared with the genomes of other basidiomycetes, S. commune has the highest number of FOLyme genes, glucose oxidase, lignocellulosic genes, glycoside hydrolyases, and polysaccharide lyases, which explain its abundance (Ohm et al., 2010).

Till date more than 1000 complete fungal genomes have been publicly released, out of which approximately 40 genomes belong to the Basidiomycota. The phylum Basidiomycota contains roughly 30,000 described species, accounting for 37% of the true fungi (Kirk et al., 2001).
As part of the JGI Fungal Genomics Program, 12 species of wood-decaying Agaricomycotina are now being analyzed, including seven white-rot species and five brown-rot species (http://genome.jgi.doe.gov/programs/fungi/). Currently, complete genome sequences of the lignin degrading basidiomycetes species are from the order, Agaricales, Boletales or Polyporales, e.g. *Pleurotus ostreatus*, *Laccaria bicolor*, *Schizophyllum commune*, *Postia placenta*, *Coprinopsis cinerea*, *Serpula lacrymans*, *Ceriporiopsis subvermispora*, *G. lucidum*, *O. olearius*, and *Agaricus bisporus* (Chen et al., 2012; Eastwood et al., 2011; Fernández-Fueyo et al., 2012; Kersten & Cullen, 2007; Lorenz et al., 2014; Martin et al., 2008; Martinez et al., 2009; Morin et al., 2012; Stajich et al., 2010; Wawrzyn et al., 2012). Comparative analyses of lignolytic fungal genomes suggest that lignin-degrading peroxidases expanded in the lineage leading to the ancestor lignolytic fungal genomes suggest that lignin-degrading peroxidases expanded in the lineage leading to the ancestor *S. commune*, which is reconstructed as a white-rot species (Floudas et al., 2012). Moreover, comparative analyses of lignin degrading enzymes in *A. bisporus* with other white-rot fungus, *P. chrysosporium*, the brown-rot fungus, *Postia placenta*, the coprophilic litter fungus, *C. cinerea*, and the ectomychorzial fungus, *L. bicolor*, revealed enzyme diversity consistent with adaptation to substrates rich in humic substances and lignocellulose (Doddapaneni et al., 2013). Annotation of another ligninolytic fungus, *G. lucidum* genome revealed a set of 36 ligninolytic oxidoreductases. Interestingly, *G. lucidum* possesses a large set of ligninolytic peroxidases, along with laccases and a cellobiose dehydrogenase compared with well-known lignin degraders, *P. chrysosporium* and *S. commune*. The presence of these enzymes suggests that *G. lucidum* may exploit different strategies for the breakdown of lignin, including oxidation by hydrogen peroxide in a reaction catalyzed by class-II peroxidases (Chen et al., 2012). Contrary to this, *S. commune*, a white-rot species, lacks class-II peroxidases and, in this regard, is more similar to the brown-rot *P. placenta* as well as the ectomycorrhizal *L. bicolor* (Grigoriev et al., 2011; Martin et al., 2008). These findings from just 10 species under three orders suggest that there has been extensive diversification in the decay mechanisms in Basidiomycota, thereby proposing a broad sampling of the phylum for whole genome sequencing.

**Role of fungal genome in ecology and environment**

Several fungal genome-sequencing projects have provided the evidence and reasoning for essential ecosystem functions, such as decomposing organic matter, nutrient cycling, and in the case of mycorrhizal species, also nutrient transfer and land colonization by plants. In forest ecosystems, they are largely responsible for the breakdown of large biopolymers, i.e. cellulose, hemicellulose, lignin, and chitin (Figure 1) (Dix & Webster, 1995; Hättenschwiler et al., 2005; Kellner & Vandenbol, 2010; Martin et al., 2008, Steffen et al., 2002).

Ectomycorrhizal soil fungus, *L. bicolor*, forms a truly symbiotic association with trees and thus has a beneficial impact on plant growth in natural and agroforestry ecosystems (Martin et al., 2008). *Laccaria bicolor* with 65 Mb genome sequences is equipped to process the diverse nitrogen both from plant and animal sources, which are found in decaying organic matter. Another model fungus, i.e. *A. bisporus*, which is rich source of protein in human diet has successfully adapted to humic-rich environment. Comparative transcriptomics of mycelium grown on defined medium, casing-soil, and compost revealed genes encoding enzymes involved in xylan, cellulose, pectin, and protein degradation are more highly expressed in compost. The gene repertoire and the expression of hydrolytic enzymes in humic-rich substrates reveal that *A. bisporus* is substantially diverse from the taxonomically related ectomycorrhizal symbiont *L. bicolor* (Morin et al., 2012).

Cell survival depends on an organism’s ability to sense and respond to environmental stresses. In saline environments, organisms respond to osmolarity changes through multiple signaling pathways (Hohmann, 2009). The relatively recent discovery of fungi in hypersaline environments has enabled the study of salt tolerance (Gunde-Cimerman et al., 2009; Hohmann, 2009). The Dead Sea is one of the most hypersaline habitats for *Eurotium rubrum* with a genome size of 26.2 Mb. The transcriptome analyses under different salt growth conditions revealed differentially expressed genes encoding ion and metabolite transporters (Kis-Papo et al., 2014). The genome composition is typical characteristic of the halophilic prokaryotes, supporting the theory of convergent evolution under extreme hypersaline stress.

Basidiomycetous fungi *Wallemia sebi* is a common food-borne contaminant that has been isolated from environments with different levels of water activity (*a*<sub>ω</sub>) (Padamsee et al., 2012; Pitt & Hocking, 2009). The *W. sebi* CBS 633.66 compact genome (9.8 Mb) has revealed the largest fraction of genes with functional domains compared with other Basidiomycota (Padamsee et al., 2012). Despite the seemingly reduced genome, several osmotic stress proteins and a high number of transporters were found that also provide clues to the ability of *W. sebi* to colonize harsh environments (Padamsee et al., 2012).

**Fungal genome in enzyme biotechnology and bioenergy**

The world market for enzymes has reach from $5.1 billion to $7 billion (6.3% annually) in 2013. Continued strong demand for specialty enzymes, as well as above average growth in the animal feed, paper, and pulp industry and ethanol production markets, will drive the enzyme market further (http://www.feedindustrynetwork.com/enzymes.aspx). Fungal enzymes are versatile and already being used in textile, detergent, food, and feed industry (Supplementary Table 1). The application of enzymes permits the industrial processes to be performed under milder conditions, using less energy and producing fewer toxic byproducts.

Regarding relevance to medicine and industry, and the desire to better understand this genus, the genomes of 10 Aspergilli have recently been sequenced, seven of which have been annotated by the collaborative effort (Gilsenan et al., 2012). The filamentous fungus *Aspergillus niger* has the ability to produce tremendous amounts of useful chemicals and enzymes. This fungus is the major source of citric acid for food, beverages, and pharmaceuticals and of several important commercial enzymes, including glucoamylase, which is widely used for the conversion of starch to food syrups and to
Fungal genome sequencing: basic biology to biotechnology

fermentative feedstocks for ethanol production (Cullen, 2007). The availability of whole genomes allows the search of enzymes with enhanced properties and provides invaluable aid toward improving the production of chemicals and enzymes in this organism (Table 4).

Fungi can produce all kinds of carbohydrate-active enzymes (CAZymes) (Cantarel et al., 2009). Among them, plant cell wall-degrading enzymes received special attention because of their importance in fungal pathogens for penetration and its application in the hydrolysis of woody material for bioethanol production. A number of studies have revealed that the activity of hydrolytic enzymes from different fungi showed preferences for different types of plant biomass and adaption to their lifestyles (Battaglia et al., 2011; King et al., 2011; Zhao et al., 2013). Recently, a complete and systematic comparative analysis of CAZymes across the fungal kingdom has been reported (Zhao et al., 2013). More than 300 CAZymes have been reported from the proteomes predicted from 187 CAZyme families. The distribution of some CAZyme families was found to be phylum specific. For example, 28 families were found in the Ascomycetes, whereas, 15 families appeared to be Basidiomycota specific (Zhao et al., 2013). The fungi of division Ascomycota Trichoderma reesei (teleomorph Hypocrea jecorina) are widely used in industry as a source of cellulases and hemicellulases for the hydrolysis of plant cell wall polysaccharides. Whereas, Podospora anserina has the highest number of carbohydrate-binding proteins, and lignin and cellulose-degrading enzymes that has been found in any fungal genome sequenced till date (Espagne et al., 2008).

Several bioprocesses, which currently produce biofuels and renewable products, are fungal based. The US Department of Energy (DOE) Joint Genome Institute (JGI) has launched a Fungal Genomics Program (FGP) aiming to scale up sequencing and analysis of fungal genomes to explore their diversity and applications for bioenergy (Table 4) (Grigoriev et al., 2011).

Fungal enzymes play an important role in second- and third-generation biofuel production processes in deconstructing lignocellulosic plant stocks. They are involved in maintaining feedstock status, plant biomass saccharification, enzyme production to bioprocesses development for producing ethanol, or higher alcohols. Genomic analyses have shown that white-rot species possess multiple lignin-degrading peroxidases (PODs) and expanded suites of enzymes attacking crystalline cellulose. To test the adequacy of the white and brown-rot categories, Grigoriev et al. analyzed 33 fungal genomes (Riley et al., 2014).

Bioethanol is produced through microbial fermentation of carbohydrates derived from agricultural feedstocks, mainly starch and sucrose. Yeasts conventionally used for industrial ethanol production such as Saccharomyces bayanus, and various hybrids, such as Saccharomyces carlsbergensis, produce ethanol rapidly from glucose, maltose, mannose, or sucrose, but they are not capable of fermenting xylose, arabinose, and cellobiose. Because of these deficiencies, a great deal of research has been devoted to metabolic engineering of Saccharomyces for improved xylose and cellobiose metabolism (Galazka et al., 2010; Hahn-Hägerdal et al., 2006; Jeffries, 2006). In approaching this problem from a fungal genomics perspective, researchers have examined numerous unconventional xylose fermenting species (Stephanopoulos, 2007), Pachysolen tannophilus (Schneider et al., 1981), Candida shehatae (Dupree & Vander Walt, 1983), and Scheffersomyces (Pichia) stipitis (Jeffries, 2006; Kurtzman & Robnett, 2010; Smith et al., 2008). One interesting discovery to emerge by sequencing the complete genome of P. stipitis is that this yeast possesses numerous genes for the rapid metabolism of cellobiose. To better understand xylose utilization for subsequent microbial engineering, DOE-JGI conducted the genome sequence of two xylose-fermenting, beetle-associated fungi, Spathaspora passalidarum, and Candida tenuis. A comparative genomic approach was applied to identify genes involved in xylose metabolism. Genomic expression profiling across five Hemiascomycete species with different xylose-consumption phenotypes implicated many genes and processes involved in xylose assimilation (Wohlhab et al., 2011). Comparison of the large number of S. cerevisiae strains also enabled the characterization of a cluster of five ORFs that have integrated into the genomes of the wine and bioethanol strains on diverse genomic locations (Borneman et al., 2011).

Pseudoczyma aphidis DSM 70725, a Basidiomycetous Fungus, is an efficient and a novel producer of mannosylerythritol lipids (MELs) and biosurfactant cellobiose lipid is also secreted during nitrogen limitation. Considering the biotechnological importance, P. aphidis DSM 70725 genome (17.92 Mb) has been recently sequenced and compared with the nucleotide contigs against the genome of the closely related species P. antarctica T34 (Lorenz et al., 2014). Strikingly, the complete MEL biosynthesis genes were found to be significantly conserved between P. antarctica T34 and P. aphidis (Lorenz et al., 2014). Potential biofuel fungi, Umbelopsis isabellina NBRC 7884 (genotype size, 37 Mb) and Mortierella alpina isolate CDC-B6842 (genotype size, 39.53 Mb), belong to the subdivision Mucoromycotina, many members of which have been shown to produce high levels of intracellular triacylglyceride accumulation, versatility in nutrient utilization, and high growth rate (Etienne et al., 2014; Takeda et al., 2014). Further, comparative genome study can be used to determine the genetic factors related to oleaginous characteristics.

Many fungi have the ability to increase the productivity of bioenergy crops, such as Miscanthus, switchgrass, and Populus, through mutually beneficial relationships called mycorrhizae (van der Heijden et al., 1998). Research advances on mychorrhizal and biocontrol fungi accelerated by genome-sequencing projects could make biomass energy crops more economically viable by lowering our dependence on pesticides and chemical fertilizers. Further, epidemics of poplar leaf rust, caused by Melampsora spp., is a major constraint on the development of bioenergy programs based on domesticated poplars as a result of the lack of durable host resistance. Rust fungi are obligate biotrophic parasites with a complex life cycle that often includes two phylogenetically unrelated hosts. Recently, DOE-JGI has sequenced and compared the 101 Mb genome of Melampsora larici-populina, the causal agent of poplar leaf rust, and the 89 Mb genome of Puccinia graminis f. sp. tritici, the causal agent of wheat and barley stem rust (Duplessisa et al., 2011).
| S. no. | Organism | Genome size (Mb) | GenBank accession/release date | Center/consortium | Economic importance | References |
|-------|----------|-----------------|-------------------------------|------------------|-------------------|------------|
| 1.    | *Saccharomyces cerevisiae* | 12 | 1996 | International effort | Most useful yeast, having been instrumental to wine-making, baking and brewing since ancient times; most intensively studied eukaryotic model organisms in molecular and cell biology | (Goffeau et al., 1996) |
| 2.    | *Trichoderma reesei* QM6a | 33 | 08/04/2005 | DOE- Joint Genome Institute | Secrete large amounts of cellulolytic enzymes | (Martinez et al., 2008) |
| 3.    | *Aspergillus niger* CBS 513.88 | 37 | 03/28/2006 | DSM, The Netherlands | Soil saprobe with a wide array of hydrolytic and oxidative enzymes involved in the breakdown of plant lignocellulose; known for citric-acid and gluconic acid production; causes black mold of onions | (Pel et al., 2007) |
| 4.    | *A. oryzae* RIB40 | 37 | 12/20/2005 | NITE (National Institute of Advanced Industrial Sciences and Technology) | Ferment soybeans, also used to saccharify rice, other grains, and potatoes in the making of alcoholic beverages such as huangjiu, sake, maktegolli, and rice vinegars | (Machida et al., 2005) |
| 5.    | *Penicillium chrysogenum* Wisconsin 54-1255 | 34.1 | – | DSM Anti-infectives, The Netherlands | Source of several β-lactam antibiotics, e.g. penicillin; industrially used to produce enzymes, i.e. polyamine oxidase, phospho-glucanate dehydrogenase, and glucose oxidase | (Jones et al., 2004) |
| 6.    | *Podospora anserina* | 35.5 | – | – | Biomass degradation | (Espagne et al., 2008) |
| 7.    | *Ashbya gossypii* | – | – | – | Riboflavin-overproduction | (Gattiker et al., 2007) |
| 8.    | *Phanerochaete chrysosporium* RP-78 | 30 | 2004 | DOE – Joint Genome Institute | Degrade lignin to carbon dioxide; used in bioremediation | (Martinez et al., 2004) |
| 9.    | *Wickerhamomyces anomalus* (formerly *Pichia anomala*) | 25.47 | 2012 | – | Bio-control agent; anti-fungal agent; wine fermentation; phytate digestion; source of protein and vitamins | (Passoth et al., 2006; Schneider et al., 2012) |
| 10.   | *Agaricus bisporus* | 36 | AEOK00000000 and AEOL00000000 | DOE – Joint Genome Institute | Xylan, cellulase, pectin decomposer; secondary decomposer fungus, plays an ecologically significant role in the biomass degradation; also used for bioremediation of substrates contaminated with heavy metals | (Morin et al., 2012) |
| 11.   | *Pichia angusta* CBS 4732 | 9.0 | BRG Grant (resources génétiques des micro-organismes No. 11-0926-99) Genolevures | Methylotrophic yeast, well studied for structure, function, and biogenesis of the peroxisomes | (Blandin et al., 2000) |
| 12.   | *P. farinose* CBS 7064 | 13.9 | Genolevures | Produce glycerol from glucose | Unpublished |
| 13.   | *P. guilliermondii* ATCC 6260 | 12 | 03/17/2005 | Broad Institute | Convert xylose to xylitol; riboflavin synthesis | Unpublished |

(continued)
Comparative genomic studies of *M. larici-populina* and *P. graminis* f. sp. *tritici* to other saprotrophic, pathogenic, and symbiotic basidiomycetes revealed that rust fungi lineages did not involve major changes in the ancestral repertoire of known conserved proteins (Duplessisa et al., 2011).

*Penicillium* species are ubiquitous filamentous ascomycetes important to the biotechnology, biomedical and food industries. They commonly occur as food spoilage agents and opportunistic pathogens and are widely used as versatile cell factories. Berg et al. (2008) had sequenced the 32.19 Mb genome of *P. chrysogenum* Wisconsin 54-1255 and identified numerous genes responsible for key steps in penicillin production. The recent sequencing of the two leading filamentous fungi used in cheese manufacture, *Penicillium roqueforti* and *Penicillium camembertii*, and comparison with the penicillin producer *Penicillium rubens* (previously known as *Penicillium chrysogenum*) reveals a 575 kb long genomic island in *P. roqueforti* (Cheeseman et al., 2014). This genomic island accommodates about 250 genes, some of which are probably involved in competition with other micro-organisms.

**Future prospects of fungal genome sequencing**

The science of genomics has largely been driven by the desire to understand the organization and function of fungal genome. Later, the development of novel high-throughput DNA sequencing methods has provided a new method for both mapping and quantifying transcriptomes in yeast (Nagalakshmi, 2008). Over the past 10 years, large-scale sequencing has been revolutionized by the development of several next-generation sequencing (NGS) technologies. NGS can provide unprecedented data on the composition of mixed microbial communities from important clinical infections with low absolute abundance in the samples and high levels of fungal DNA from contaminating sources. It can be used not only for determining the DNA sequence of a microbial genome, ribosomal rRNA gene segments from fungi and bacteria in DNA extracted from bronchiolar lavage samples and oropharyngeal wash but also for resolving higher-order structures within the eukaryotic nucleus (Bittinger et al., 2014). The use of NGS to obtain transcriptomics data is collectively known as RNA deep sequencing or RNA-seq. The first novel transcribed regions were reported in the yeasts *S. cerevisiae* and *Schizosaccharomyces pombe* in the year 2008 and since then RNA-seq has been extended to a number of other organisms (Wilhelm et al., 2008). Recently, comparative global transcriptional studies of entomopathogenic fungi *Metarhizium anisopliae* and *Metarhizium acridum* provided a broad-based analysis of gene expression during early colonization processes, particularly in terms of the genes involved in host recognition, metabolic pathways and pathogen differentiation (Gao et al., 2011). A fungal transcriptome study has started in 2007 as express sequence tag (EST), followed by gene expression profiling. Until today, 198 fungal transcriptomes have been sequenced by JGI alone in 3 years, with marked exponential growth (Figure 3). Although RNA-Seq is still a technology under active development, it offers several key advantages over existing technologies. First, unlike hybridization-based approaches, RNA-Seq is not limited to detecting transcripts that correspond to an existing genomic sequence. RNA-Seq can reveal the precise location of transcription boundaries and also give information about how two exons are connected (Wang et al., 2009). This makes RNA-Seq particularly attractive for non-model organisms with genomic sequences that are yet to be determined.
The legacy of more than 150 years of fungal research, coupled with the availability of molecular and genetic tools and the advancement in sequencing technologies, offers enormous potential for continued discovery. Fungal genome sequences from several ongoing (http://www.tigr.org/tdb/mdb/mdbinprogress.html) and planned (http://www.genome.wi.mit.edu/seq/fgi) projects will provide extraordinary opportunities for comparative analyses. This new era in fungal biology promises to yield insights into this important group of organisms, as well as to provide a deeper understanding of the fundamental cellular processes common to all eukaryotes. Construction of a synthetic genome (Gibson et al., 2008) and a designer chromosome III, of *S. cerevisiae* (Annaluru et al., 2014), demonstrates the possibility to develop a custom made genome for all economically important fungi in the coming days with the help of publicly available genome sequence database. Moreover, the ability to implement many simultaneous and directed changes to natural DNA sequences and to build and test synthetic systems will present researchers with a powerful new tool for de novo life form (Endy, 2008).

**Conclusion**

Studies on fungal biology have been supported by rapid enrichment of bio-informatics tools and genome sequence data. Novel approaches for extracting information on the functioning of metabolic pathways, gene expression, protein levels, sub-cellular localization, and functionality and disease forecasting provide a “genomic insight” of how an organism grows, reproduces, and responds to its surroundings. The genome sequence along with transcriptome sequence analysis of infected humans and plants lead to a more comprehensive understanding of the pathogenesis system, which eventually helps in developing more effective surveillance and disease management strategies for the most devastating pathogens. Results of comparative genomic data increase our understanding of eukaryotic genome evolution processes, using fungi as models. However, more data will probably be needed to understand the complexity of the genetic factors for various molecular and ecological adaptations of fungi. Next-generation sequencing technology may help to initiate a comparative study of multiple and related genomes to understand the evolution of gene for metabolic pathways, economically important enzymes, and the evolutionary relationships related to protein function. Considering their basic, economic, and biotechnological importance, fungal genome studies need to be more extensive and highly focused.

**Acknowledgements**

Author sincerely acknowledges Council of Scientific and Industrial Research (CSIR), Department of Biotechnology (DBT), and Department of Science and Technology (DST), New Delhi, India, for financial help and Maharshi Dayanand University, Rohtak, Haryana, India, for providing infrastructure to do the review work.

**Declaration of interest**

The author reports that he has no competing financial interests.

**References**

Almeida AJ, Matute DR, Carmona JA, et al. (2007). Genome size and ploidy of *Paracoccidioides brasiliensis* reveals a haploid DNA content: flow cytometry and GP43 sequence analysis. Fungal Genet Biol, 44, 25–31.

Annaluru N, Muller H, Mitchell LA, et al. (2014). Total synthesis of a functional designer eukaryotic chromosome. Science, 344, 55–8.

Battaglia E, Benoit I, van den Brink J, et al. (2011). Carbohydrate-active enzymes from the zygomycete fungus *Rhizopus oryzae*: a highly specialized approach to carbohydrate degradation depicted at genome level. BMC Genomics, 12, 38.

Berg MA, Albang R, Albermann K, et al. (2008). Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. Nat Biotechnol, 26, 1161–8.

Bittinger K, Charison ES, Loy E, et al. (2014). Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome Biol, 15, 487.
Blandin G, Llorente B, Malpertuy A, et al. (2000). Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*. FEBS Lett, 487, 76–81.

Borneman AR, Desany BA, Riches D, et al. (2011). Whole genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet, 7, e1001287.

Brakhage AA. (2013). Regulation of fungal secondary metabolism. Nat Rev Microbiol, 11, 21–32.

Burd W, Skarzynski C, Eppert F, et al. (2013). Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biol, 12, R7.

Butler G, Rasmussen MD, Lin MF, et al. (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature, 459, 657–62.

Cai J, Zhao R, Jiang H, et al. (2008). *De Novo* origination of a new protein coding gene in *Saccharomyces cerevisiae*. Genetics, 179, 487–96.

Cantarel BL, Coutinho PM, Rancurel C, et al. (2009). The Carbohydrate-Active EnZymes database (CAZY): an expert resource for glyco-enzymologists. Nucleic Acids Res, 37, D233–8.

Cantu D, Segovia V, MacLean D, et al. (2013). Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *Tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. BMC Genomics, 14, 270.

Cheeseman K, Kopars J, Renault P, et al. (2014). Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat Commun, 5, 2878.

Chen S, Xu J, Liu C, et al. (2012). Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. Nature Commun, 3, 913.

Cherry JM, Adler C, Ball C, et al. (1998). SGD: *Saccharomyces* Genome Database. Nucleic Acids Res, 26, 73–9.

Cherry JM, Hong EL, Amundsen C, et al. (2012). *Saccharomyces* Genome Database: the genomes resource of budding yeast. Nucleic Acids Res, 40, D700–5.

Christie KR, Weng S, Balakrishnan R, et al. (2004). *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. Nucleic Acids Res, 32, D311–4.

Cowen LE, Anderson JB, Kohn LM. (2002). Evolution of drug resistance in *Candida albicans*. Annu Rev Microbiol, 56, 139–65.

Crawford JM, Townsend CA. (2010). New insights into the formation of the hemiascomycetous yeasts: 13. *Phytophthora* genome. Science, 304, 304–7.

Dean RA, Talbot NJ, Ebbole DJ, et al. (2005). The genome sequence of *Magnaporthe grisea*. Science, 309, 290–3.

Dean R, Van Kan JAL, Pretorius ZA, et al. (2012). The top 10 features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci USA, 109, 20404–9.

Dietrich FS, Voegeli S, Brachat S, et al. (2004). The *Saccharomyces cerevisiae* genome annotation. Mycology, 2, 118–41.

Dix NJ, Webster J. (1995). Fungal ecology. London: Chapman and Hall.

Duplessisa S, Cuomob CA, Linc Y-C, et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci USA, 108, 9166–71.

Dupree JC, Vanderwart JP. (1983). Fermentation of xylose to ethanol by a strain of *Candida shehatae*. BioTechnol Lett, 5, 357–62.

Eastwood DC, Floudas D, Binder M, et al. (2011). The plant cell wall degrading machinery underlies the functional diversity of forest fungi. Science, 333, 762.

Endy D. (2008). Reconstruction of the genomes. Science, 319, 1196–7.

Engel SR, Balakrishnan R, Binkley G, et al. (2010). *Saccharomyces* Genome Database provides mutant phenotype data. Nucleic Acids Res, 38, D433–6.

Engel SR, Cherry JM. (2013). The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the *Saccharomyces* genome database 2013. doi:10.1093/database/bat012.

Engelthaler DM, Hicks ND, Gillee JD, et al. (2014). *Cryptococcus gatti* in North American Pacific Northwest: whole population genome analysis provides insights into species evolution and dispersal. mBio, 5, e01464–14.

Espagne E, Lespinet O, Malagnac F, et al. (2008). The genome sequence of the model ascomycete fungus *Podospora anserina*. Genome Biol, 9, R77.

Etienne KA, Chibucos MC, Su Q, et al. (2014). Draft genome sequence of *Mortierella alpina* isolate CDC-B6842. Genome Announc, 2, e01180–13.

Fedorova ND, Khalidi N, Joardar VN, et al. (2008). Genomic Islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet, 4, e100046.

Fernández-Fueyo E, Ruiz-Dueñas FI, Miki Y, et al. (2012). Lignin-degrading peroxidases from genome of selective ligninolytic fungus *Ceriporiopsis subvermispora*. J Biol Chem, 287, 16903–16.

Floudas D, Binder M, Riley R, et al. (2012). The paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science, 336, 1715.

Frazier ME, Johnson GM, Thomassen DG, et al. (2003). Realizing the potential of the genome revolution: the genomes to life program. Science, 300, 290–3.

Gajendran K, Gonzales MD, Farmer A, et al. (2006). *Phytophthora* functional genomics database (PFGD): functional genomics of *phytophthora*-plant interactions. Nucleic Acids Res, 34, D465–70.

Galagan JE, Calvo SE, Borkovich KA, et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. Nature, 422, 859–68.

Galagan JE, Henn MR, Ma L-J, et al. (2005). Genomics of the fungal kingdom: insights into eukaryotic biology. Genome Res, 15, 1620–31.

Galazka JM, Tian C, Beeson WT, et al. (2010). Celldextrin transport in yeast for improved biofuel production. Science, 330, 84–6.

Gao Q, Jin K, Ying S-H, et al. (2011). Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. PLoS Genet, 7, e1001264.

Gattiker A, Rischatsch R, Demougin P, et al. (2007). *Asphyx* Genome Database 3.0: a cross-species genome and transcriptome browser for yeast biologists. BMC Genomics, 8, 9.

Gibson DG, Benders GA, Axelrod KC, et al. (2008). One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Natl Acad Sci USA, 105, 20404–9.

Gilsenan JM, Cooley J, Bowyer P. (2012). CADRE: the Central *Aspergillus* Data Repository. Nucleic Acids Res, 40, D660–6.

Goffeau A, Barrell BG, Bussey H, et al. (1996). Life with 6000 genes. Science, 274, 563–67.

Goodwin SB, Ben M’Barek S, Dhillon B, et al. (2011). Finished genome of the fungal wheat pathogen *Magnaportha grisea* reveals dispersive structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet, 7, e1002070.

Grigoriev IV, Cullen D, Goodwin SB, et al. (2011). Fueling the future with fungal genomics. Mycology, 2, 192–209.

Grigoriev IV, Nikiit R, Haridas S, et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res, 42, D699–704.

Gunde-Cimerman N, Ramos J, Plemetinaus A. (2009). Halotolerant and halophilic fungi. Mycol Res, 113, 1231–41.

Haas BJ, Kamoun S, Zody MC, et al. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature, 461, 393–8.

Haas BJ, Zeng Q, Pearson MD, et al. (2011). Approaches to fungal genome annotation. Mycology, 2, 118–41.

Hahn-Hägerdal B, Galbe M, Gorrwa-Grauslund M-F, et al. (2006). Bioethanol: the fuel of tomorrow from the residues of today. Trends Biotechnol, 24, 549–56.
Priebe S, Linde J, Albrecht D, et al. (2011). FungiFun: a web-based application for functional categorization of fungal genes and proteins. Fungal Genet Biol, 48, 353–8.

Rhind N, Chen Z, Yassour M, et al. (2011). Comparative functional genomics of the fission yeasts. Science, 332, 930–6.

Riley R, Salamova AA, Brown DW, et al. (2014). Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. Proc Natl Acad Sci USA, 111, 9923–8.

Ronning CM, Fedorova ND, Bowyer P, et al. (2005). Genomics of Rileya R, Salamova AA, Brown DW, et al. (2014). Extensive sampling of Aspergillus fumigatus. Rev Iberoam Micol, 22, 223–8.

Sanodiya BS, Thakur GS, Baghel RK, et al. (2009). Ganoderma lucidum: a potent pharmacological macrofungus. Curr Pharm Biotechnol, 10, 717–42.

Schneider H, Wang PY, Maleszka CR. (1981). Conversion of D-xylose into ethanol by yeast Pachysolen tannophilus. Biotech Lett, 3, 89–92.

Schneider J, Rupp O, Trost E, et al. (2012). Genome sequence of Wickerhamomyces anomalus DSM 6766 reveals genetic basis of biotechnologically important antimicrobial activities. FEMS Yeast Res, 12, 382–6.

Schantz KD, Lin Y-C, Tiels P, et al. (2009). Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol, 27, 561–9.

Schaumhager CR, Kuhad RC. (2008). Laccase: enzyme revisited function redefined. Indian J Microbiol, 48, 309–16.

Sharpton TJ, Stajich JE, Rounsley SD, et al. (2009). Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. Genome Res, 19, 1722–31.

Skrzyypek MS, Arnaud MB, Costanzo MC, et al. (2010). New tools at the Candida Genome Database: biochemical pathways and full-text literature search. Nucleic Acids Res, 38, D428–32.

Smith DR, Quinlan AR, Peckham HE, et al. (2008). Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Res, 18, 1638–42.

Solaimani MM, Meyer SE, Udall JA, et al. (2014). De novo genome assembly of the fungal plant pathogen Pyrenophora semeniperda. PLoS One, 9, e87045.

Stajich JE, Harris T, Brunk BP, et al. (2012). FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Res, 40, D675–81.

Stajich JE, Wilke SK, Ahrén D, et al. (2010). Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom Coprinopsis cinerea (Coprinus cinereus). Proc Natl Acad Sci USA, 107, 11889–94.

Steffen KT, Hatakka A, Hofrichter M. (2002). Removal and mineralization of polycyclic aromatic hydrocarbons by litter-decomposing basidiomycetous fungi. Appl Microbiol Biotechnol, 60, 212–17.

Stein LD, Mungall C, Shu SQ, et al. (2002). The generic genome browser: a building block for a model organism system database. Genome Res, 12, 1599–610.

Stephanopoulos G. (2007). Challenges in engineering microbes for biofuels production. Science, 315, 801–4.

Stukenbrock EH, Jorgensen FG, Zala M, et al. (2010). Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen Mycosphaerella graminicola. PLoS Genet, 6, e1001189.

Takeda I, Tamano K, Yamane N, et al. (2014). Genome sequence of the Mucozymoryctina Fungus Umbelopsis isabellina, an effective producer of lipids. Genome Announc, 2, e0071–14.

Tavares S, Ramos AP, Pires AS, et al. (2014). Genome size analyses of Pucciniales reveal the largest fungal genomes. Front Plant Sci, 5, 422.

van der Heijden MGA, Klironomos JN, Ursic M, et al. (1998). Mycorrizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. Nature, 396, 69–72.

Wainwright PO, Hinkle G, Sugin ML, Stickel SK. (1993). Monophyletic origins of the metazoa: an evolutionary link with fungi. Science, 260, 340–2.

Wang F-Q, Zhong J, Zhao Y, et al. (2014). Genome sequencing of high-penicillin producing industrial strain of Penicillium chrysogenum. BMC Genomics, 15, S11.

Wang Z, Gerstein M, Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 10, 57–63.

Ward PP, Lo JY, Duke M, et al. (1992b). Production of biologically active recombinant human lactoferrin in Aspergillus oryzae. Biotechnology, 10, 784–9.

Wawrzyn GT, Quin MB, Choudhary S, et al. (2012). Draft genome of Omphalotus olearius provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota. Chem Biol, 19, 772–83.

Wilhelm BT, Marguerat S, Watt S, et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, 453, 1239–43.

Wohlboch D, Kuo A, Satob TK, et al. (2011). Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. Proc Natl Acad Sci USA, 108, 13212–17.

Wong P, Walter M, Lee W, et al. (2011). FGDB: revisiting the genome annotation of the plant pathogen Fusarium graminearum. Nucleic Acids Res, 39, D637–9.

Wood V, William R, Rajandream M-A, et al. (2002). The genome sequence of Schizosaccharomyces pombe. Nature, 415, 871–80.

Xu J, Saunders CW, Hu P, et al. (2007). Dandelion-associated Malassezia genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. Proc Natl Acad Sci USA, 104, 18730–5.

Zhao Z, Liu H, Wang C, Xu J-R. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics, 14, 274.

Supplementary material available online
Supplementary Table 1.