9 Fungal Genomics

Robin A. Ohm[1]

CONTENTS

I. Introduction

In the past two decades, genomics has developed into a formidable tool to study various aspects of fungi. In the year 1996, *Saccharomyces cerevisiae* was the first fungal genome to be sequenced (Goffeau et al. 1996), and since then the number of publicly available genome sequences has increased to 2128 in GenBank and 1398 in MycoCosm (at the time of writing in September 2019) (Grigoriev et al. 2014; Clark et al. 2016). Fungal genomics as a research field was kick-started by the sequencing efforts of institutes and consortia, including the Fungal Genome Initiative of the BROAD Institute of MIT and Harvard (Cuomo and Birren 2010) and the Fungal Program of the US DOE Joint Genome Institute (Grigoriev et al. 2011). In recent years, sequencing costs have decreased considerably, placing genome sequencing and analysis well within the reach of smaller labs.

After the first fungal genome of *S. cerevisiae* (Goffeau et al. 1996) was published, genome sequencing efforts initially focused on other previously established model systems. Examples include *Neurospora crassa* (Galagan et al. 2003), various species of *Aspergillus* (Galagan et al. 2005), the human pathogen *Cryptococcus neoformans* (Loftus et al. 2005), the plant pathogen *Fusarium graminearum* (Cuomo et al. 2007), and *Trichoderma reesei* (Martinez et al. 2008). These genome sequences are still an indispensable tool for studying these important model systems. Among many other things, they facilitate high-throughput experiments such as RNA-Seq to study genome-wide gene expression or ChIP-Seq to study various aspects of epigenetic regulation. In combination, these approaches aim to assign functions to regions of the genome and are called Functional Genomics.

Moreover, the increasing number of available genome sequences (including those of non-model organisms) allowed for a comparative genomics approach. By comparing genomes of related species, new insights can be gained into genome evolution, gene evolution, gene association with a particular lifestyle, as well as phylogeny (examples of this are described below).

In general, a genome sequencing project starts with sequencing the genomic DNA using next-generation sequencing technologies. This is followed by genome assembly, which aims to computationally reconstruct the

[1] Microbiology, Department of Biology, Utrecht University, Utrecht, the Netherlands; e-mail: r.a.ohm@uu.nl

Genetics and Biotechnology, 3rd Edition
The Mycota II
J.P. Benz, K. Schipper (Eds.)
© Springer Nature Switzerland AG 2020

genome from the (relatively short) sequence reads. Next, the genome assembly is annotated. This entails the identification of repetitive sequences, genes, and other functional elements in the genome. The predicted genes are subsequently annotated by assigning a putative function, usually based on homology to known genes and domains.

This chapter will describe advances in technologies underlying fungal genome sequencing, annotation, and analysis. Furthermore, the impact of fungal genome sequencing is illustrated using examples from several fields of biotechnology.

II. Advances in Genome Sequencing Technologies

In the past decade, sequencing technologies have improved dramatically, radically changing the landscape of fungal genome sequencing. **Sanger** sequencing was the first sequencing technique that was used for genome sequencing (Sanger et al. 1977b). It was used to sequence landmark genomes such as the first bacteriophage φX174 (Sanger et al. 1977a), the first bacterium *Haemophilus influenza* (Fleischmann et al. 1995), the first eukaryote (and first fungus) *Saccharomyces cerevisiae* (Goffeau et al. 1996), the first plant *Arabidopsis thaliana* (Arabidopsis Genome Initiative 2000), the first animal *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium 1998), as well as the human genome (International Human Genome Sequencing Consortium 2001). The sequencing of these genomes was generally a multi-year undertaking and was performed by large consortia of collaborating labs. Sequencing reads that were obtained with Sanger technology were relatively long (up to approximately 1500 bp) and were relatively straightforward to assemble using assembly software such as Jazz (Aparicio et al. 2002) or Arachne (Batzoglou et al. 2002).

More recently, since the mid-2000s, several new sequencing platforms were developed that are collectively known as "**next-generation sequencing**" (NGS). Initially, these techniques included the now mostly defunct technologies Roche 454 (Margulies et al. 2005), IonTorrent (Life Technologies) and SOLiD (Applied Biosystems). Currently the most prominent short read sequencing technology, however, is **Illumina** (Bennett 2004). Although the sequences generated by Illumina technology were initially too short for efficient genome sequencing (up to 25 bp), this has increased to currently 2 × 300 bp on an Illumina MiSeq machine. New assembly approaches and software were developed for these short reads, such as Velvet (Zerbino and Birney 2008), ABySS (Simpson et al. 2009), SOAPdenovo (Luo et al. 2012), and SPAdes (Bankevich et al. 2012).

Rather paradoxically, the assemblies generated from early NGS techniques were not nearly as good as the ones generated from Sanger reads, with respect to assembly fragmentation. Especially repetitive genomic regions (e.g., originating from transposable elements) were challenging to assembly using short reads. However, crucial advantages of NGS technologies are that they are considerably faster and cheaper than Sanger sequencing (Ghurye and Pop 2019). This meant that genome sequencing became affordable to core facilities and even individual researchers, as opposed to the large sequencing consortia that were required for Sanger-based genome sequencing. This is illustrated by the following back-of-the-envelope calculation: sequencing a typical fungal genome of 30 Mbp with 100-fold coverage (each bp sequenced on average 100 times) currently costs less than 250 euro per genome on an Illumina NextSeq500 machine (if 35 genomes are pooled onto one lane). This is a stark difference with the multi-million euro Sanger sequencing efforts of the past (Goffeau et al. 1996).

Since the early 2010s, new technologies have become commercially available that produce considerably longer reads than Illumina. **Pacific Biosciences** (PacBio) is based on single-molecule sequencing and can produce reads of on average 5 kbp and a maximum of 20 kbp (Eid et al. 2009). **Oxford Nanopore** further revolutionized sequencing by vastly reducing the size of the machine to a mere USB flash drive (Jain et al. 2016). This MinION machine produces reads of over 100 kbp. However, both

PacBio and Oxford Nanopore reads have a considerably higher error rate (up to 15% errors) than Illumina technology (Mardis 2017). Although this error rate will likely improve as new protocols become available, it is problematic for accurate genome sequencing and assembly. Approaches for assembly using these long reads are either high coverage sequencing (Chin et al. 2013) or a hybrid approach that uses Illumina reads to correct sequencing and assembly errors (Walker et al. 2014). As these long-read NGS technologies mature further, it is likely that obtaining genome assemblies with telomere-to-telomere chromosomes will become trivial and affordable within a few years.

III. Genome Annotation

Sequencing a genome is only the first step, and the even more important next step is to annotate the genome. This process generally includes the identification of regions of repetitive DNA, the prediction of genes, and a function prediction for these genes and domains. These individual steps can be strung together into a **pipeline**. Several pipelines exist for eukaryotic genome annotation, and two examples of frequently used pipelines for fungal genome annotation are MAKER (Cantarel et al. 2008) and the pipeline used by the US DOE Joint Genome Institute (Haridas et al. 2018). This section describes the steps of genome annotation in more detail.

A. Repeats

The term "repeat" may refer to various types of sequences: "low-complexity regions" (sometimes called "simple repeats") such as a homopolymeric run of nucleotides, as well as transposable (mobile) elements (transposons) (Kapitonov and Jurka 2008). These transposable elements can essentially copy themselves and thus spread throughout the genome. They can be subdivided into two classes, depending on their mode of proliferation (Wicker et al. 2007). Class I elements use an RNA-intermediate (reminiscent of a retrovirus) and move via a "copy-paste" mechanism. They include long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and long terminal repeats (LTRs). Class II elements move via a DNA intermediate and include helitrons and terminal inverted repeats (Kapitonov and Jurka 2001, 2008).

Since repetitive regions are markedly different from gene-coding regions, it is common practice to "mask" the repetitive regions before commencing gene prediction. Masking ensures that any spurious open reading frames that may be present in the repeats will not confound (the training of) the gene predictor. Repeats can be identified in a newly sequenced genome using either homology-based or de novo tools. Homology-based tools rely on a database of known repetitive elements such as Repbase (Jurka et al. 2005) and a search algorithm such as RepeatMasker (Smit et al. 2015). Novel or genome-specific repeats can be identified using de novo tools such as Repeatscout (Price et al. 2005), which looks for sequences that occur repeatedly throughout the genome. Since transposable elements tend to be relatively AT-rich, their proliferation can result in large AT-rich regions. Those regions can be distinguished from gene-coding GC-rich regions by tools such as OcculterCut (Testa et al. 2016).

The repetitive content of the genome varies widely between fungi. For example, the very compact 13.6 Mbp genome of the fern pathogen *Mixia osmundae* has a repetitive content of <1% (Toome et al. 2014), whereas the 177.6 Mbp genome of the mycorrhizal ascomycete *Cenococcum geophilum* consist for 81% of repetitive sequences (Peter et al. 2016). Repetitive sequences are usually predominantly found in centromeric and sub-telomeric regions but may be spread throughout the assembly. Generally, self-replicating repeats are considered deleterious since their spread may interrupt genes. Fungi have evolved a defense mechanism that recognizes repeats and inactivates these by causing point mutations (repeat-induced point mutations, or RIP) (Clutterbuck 2011; Castanera et al. 2016). Intriguingly, genome sequencing of several plant pathogens has revealed that pathogenesis-related genes frequently co-

localize with repetitive sequences in these species. A potential evolutionary benefit of this co-localization is a higher rate of mutation due to RIP, which in turn may lead to a higher rate of evolution. This may allow these pathogens to adapt more quickly to the host plant's defenses (Rouxel et al. 2011; Ohm et al. 2012).

B. Gene Prediction

Genes are (arguably) the most important functional elements in a fungal genome. However, their accurate identification is non-trivial. The presence of introns in fungal genes precludes simple scanning for open reading frames (ORFs), which is a common initial approach in gene prediction in prokaryotes. The structure of protein-coding genes varies widely between eukaryotes (Yandell and Ence 2012) and even between fungi. Differences include GC content of the coding regions, splicing acceptor and donor sites, intron length, number of introns per gene, gene length, etc. For example, the ascomycete yeast *S. cerevisiae* has 6576 predicted genes with a median gene length of 1071 bp, of which 4.2% contain an intron (Goffeau et al. 1996). In contrast, the basidiomycete mushroom-forming fungus *Schizophyllum commune* has 16204 predicted genes with a median gene length of 1517 bp, of which 86.3% contain an intron (Ohm et al. 2010). Therefore, the gene-finding approach needs to be tailored to each organism individually.

Gene prediction algorithms can be divided into two categories: evidence-driven and ab initio approaches. Evidence-driven predictors take external evidence to identify the locations of protein-coding genes. This evidence usually takes the form of sequenced cDNA (Haas et al. 2003) or homology with known proteins of related species (Birney et al. 2004). Sequenced cDNA (in this context usually referred to as Expressed Sequence Tags or ESTs) are aligned to the assembly, and exons and intron splice sites are inferred. This approach has the advantage that it uses evidence specific to the organism but has the disadvantage that unexpressed genes are less likely to be identified correctly. Homology-based gene predictors rely on the alignment of known proteins from related organisms to identify exons. Advantages of this approach are that it is cheap (since no cDNA sequencing is required) but has the disadvantage that organism-specific genes are less likely to be identified correctly. An ab initio approach uses a mathematical model of the gene structure to predict genes. These algorithms require training, which means that they need to learn what a gene looks like (e.g., typical gene length, intron length, GC content of coding regions, etc.) from a subset of known genes. This poses a problem, since for most fungal genomes there is no prior knowledge available. Modern approaches use a hybrid strategy in which RNA-Seq data is used as evidence to train an ab initio gene predictor. The algorithm **BRAKER**, for example, only requires aligned RNA-Seq reads and a genome assembly and no other prior knowledge (Hoff et al. 2016). It uses these data to train the ab initio predictors GeneMark (Lomsadze et al. 2014) and Augustus (Stanke and Waack 2003) and subsequently generates a high-quality gene prediction.

Various gene prediction algorithms may predict different genes at the same locus. Although these sometimes represent alternative splicing variants (especially when the gene predictor uses expression data as evidence), it is more likely that only one variant is correct. Various methods have been published that aim to select the correct gene prediction at each locus; examples include MAKER (Cantarel et al. 2008), the US DOE Joint Genome Institute pipeline (Haridas et al. 2018), and FunGAP (Min et al. 2017).

The quality and completeness of the set of predicted genes can be assessed by determining the percentage of highly conserved eukaryotic genes that are found in the predicted gene set. Since these highly conserved genes (histones, DNA polymerase, etc.) are expected to be present among the genes of the newly sequenced fungus, their absence can be indicative of an incompleteness of the genome assembly or the gene prediction. CEGMA (Core Eukaryotic Genes Mapping Approach) was initially a popular tool to determine completeness (Parra et al. 2007). However, a key issue with CEGMA

was that the conserved genes were identified from only six eukaryotic species. BUSCO takes a clade-specific approach that is based on more eukaryotic genomes, and fungi-specific conserved gene sets are available (Simão et al. 2015). More recently, FGMP (Fungal Genome Mapping Project) was developed that provides a computational framework and sequence resource specifically designed to assess the completeness of fungal genomes (Cissé and Stajich 2019). It is based on 246 fungal genomes and can be used to assess assembly and annotation completeness as well as suggest assembly improvements.

C. Functional Annotation of the Predicted Genes

Once a reliable set of genes has been predicted, the next step is to determine the putative role of the encoded proteins. This is referred to as functional annotation of the predicted proteins. It is important to note, however, that automated function predictions should be interpreted with care. Lab experiments may be required to definitively confirm the function of individual genes (e.g., an enzyme activity assay to confirm the predicted activity of a putative enzyme).

Functional annotation usually starts with homology searches in a database of known proteins, for example, using Blast (Altschul et al. 1990) to search for homologs in GenBank (Clark et al. 2016) or UniProt/Swiss-Prot (Bateman et al. 2017). Moreover, conserved protein domains can be identified using InterPro (Hunter et al. 2009), which comprises a collection of domain databases that includes PFAM (Finn et al. 2016). Cellular localization of the proteins can be predicted using SignalP (Petersen et al. 2011), TMHMM (Krogh et al. 2001), and WoLF PSORT (Horton et al. 2007). Proteases/peptidases can be identified by homology to known enzymes in the MEROPS database (Rawlings et al. 2014). More generally, Gene Ontology (GO) aims to provide a hierarchical functional annotation of the predicted proteins, based on their molecular function, cellular localization, and the biological process they are involved in Ashburner et al. (2000). Similarly, KEGG (Kyoto Encyclopedia of Genes and Genomes) provides a classification system into metabolic pathways, including predicted enzyme activities based on the Enzyme Commission (EC) system (Kanehisa and Goto 2000).

Several functional annotation approaches have been developed that aim to identify genes that are involved in the lifestyle of fungi. The **CAZy** (carbohydrate-active enzymes) database focuses on enzymes that assemble, modify, or break down polysaccharides (Lombard et al. 2014). CAZymes are especially important in the context of plant biomass breakdown, for example, in lignocellulose degradation and plant disease (further discussed below). Fungi are known to produce a wide variety of secondary metabolites and other natural products (further discussed below). The genes involved in this process are frequently clustered in the genome, and these biosynthetic gene clusters can be identified by tools like AntiSMASH (Blin et al. 2017) or SMURF (Khaldi et al. 2010).

D. Data Visualization, Analysis, and Manual Curation

Large amounts of data are generated by genome sequencing and annotation. These can be challenging to interpret unless they are visualized. Genome sequencing consortia and/or institutes generally make the data accessible to the public by means of a centrally hosted web database, which allows users to analyze the genome sequence, gene predictions, and functional annotations. Examples include the genus-specific websites Saccharomyces Genome Database (SGD) and the Aspergillus Genome Database (AspGD) (Cherry et al. 2012; Cerqueira et al. 2014). MycoCosm hosts all fungal genome portals of the US DOE Joint Genome Institute (Grigoriev et al. 2014). FungiDB hosts numerous published fungi (Basenko et al. 2018). Upon publication of a genome, the data is generally submitted to NCBI GenBank, which has therefore amassed a large collection of fungal genome data (Clark et al. 2016).

Typically, a genome portal contains tools to visualize and analyze the genome data. These tools include Blast to search for homology (Altschul et al. 1990), a search function for functional annotations and a genome browser. Since the generated data will likely contain errors in gene prediction, it is important that these predictions can be fixed manually, based on external evidence. Genome browsers can facilitate this in an intuitive way. Data of various origins can be displayed, evaluated, and (if needed) manually corrected. This process is referred to as manual curation. Data that can be visualized include gene predictions, expression data, regions of homology (e.g., blast hits), genome synteny, etc.

An early example of a web-based genome browser was the UCSC (University of California, Santa Cruz) Genome Browser, which was originally developed to visualize the human genome (Kent et al. 2002) and is also used in MycoCosm. Later, GBrowse was developed (Stein 2013), which was designed to integrate well with the Generic Model Organism Database suite (www.gmod.org). Its successor JBrowse (Buels et al. 2016) offers an intuitive and flexible genome browser that can be easily installed and used in small-scale genome sequencing initiatives. Web Apollo (later renamed to Apollo) (Lee et al. 2013) is a plugin for JBrowse that facilitates the manual curation (correction) of gene predictions as well as other genomic features, making it a valuable tool for genome visualization, analysis, and curation. All corrections are stored in a centralized database, allowing collaborators from all over the world to simultaneously work on the same genome.

IV. Genomics and Biotechnology

Fungi play important roles in a wide range of fields that are interesting from a biotechnological perspective. Genome sequencing and annotation has greatly facilitated the development of these fields by revealing the genes involved in these processes. Examples of biotechnologically relevant topics include secondary metabolites, carbohydrate-active enzymes, mushroom development, and plant interactions. Obviously, this is by no means an exhaustive list of biotechnological topics. This section will discuss the roles genome sequencing and analysis have played in these important fields of study.

A. Secondary Metabolites or Natural Products

Fungi can produce a wide range of secondary metabolites, which are relatively small molecules that are not directly encoded by genes. In the context of biotechnology, they are frequently referred to as natural products. These metabolites can play an important role in processes such as pathogenesis, defense, interactions, pigmentation, etc. Often, they play an ecological role and help the fungi to colonize a niche. From a biotechnology perspective, they are interesting for their antibacterial, antifungal, and antitumor activities. Some well-known examples of natural products are the antibiotic penicillin, which is produced by species of *Penicillium* (Bennett and Chung 2001), and the cholesterol-lowering drug lovastatin (Downs et al. 1998).

Secondary metabolites are not directly encoded by genes, but instead they are generally produced by a set of enzymes that synthesize the metabolite in a conveyor belt-like fashion. These enzymes include polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), terpene cyclases (TC), dimethyl-allyltryptophan synthetases (DMATS), and a range of accessory enzymes including methyltransferases (Keller et al. 2005; Keller 2019). Intriguingly, the genes encoding these enzymes are frequently clustered in the genome, which makes them relatively easy to identify (Nützmann et al. 2018). These gene clusters are known as biosynthetic gene clusters. AntiSMASH is a commonly used tool to identify these clusters (Blin et al. 2017). It first identifies core genes (PKS, NRPS, TC, and DMATS) and then looks for putative accessory genes involved in the production of the secondary metabolite. Moreover, the identified putative clusters can be compared to known clusters in other organisms. This homology and the gene families in

the cluster are used to predict the type of secondary metabolite that may be produced, although this is currently still rather inaccurate.

The wide diversity among members of the fungal kingdom is also reflected in the wide range of natural products they produce, making fungi an interesting source for novel drugs. Genome sequencing has resulted in a large catalog of biosynthetic gene clusters (Keller 2019). Unfortunately, most natural products are not produced under lab conditions, complicating their identification in high-throughput screens (Keller et al. 2005). Several companies (e.g., Hexagon Bio, USA) are currently using high-throughput genome sequencing to identify novel natural products, purely based on their gene content. Interesting candidate gene clusters are then heterologously expressed in production species using a synthetic biology approach, thus circumventing the problem of low production of the natural products in their natural host. *S. cerevisiae* and *Aspergillus nidulans* are examples of production species (Billingsley et al. 2016; Clevenger et al. 2017; Harvey et al. 2018). This approach illustrates the power of large-scale genome sequencing and analysis.

B. Carbohydrate-Active Enzymes

Fungi are heterotrophs: they feed on organic matter. A large source of organic matter is plant biomass, or, more specifically, polysaccharides in lignocellulose (including cellulose, hemicellulose, and pectin). Fungi have evolved a wide range of extracellular enzymes to break down these recalcitrant polysaccharides into smaller compounds (monosaccharides and oligosaccharides) that can be transported over the cell membrane. Collectively, these enzymes are known as carbohydrate-active enzymes (CAZymes) and are organized in a special database, the CAZy database. CAZy describes the families of structurally related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds (Lombard et al. 2014). More generally, CAZymes are enzymes involved in the breakdown, biosynthesis, and modification of carbohydrates.

Based on their domain structure CAZymes are classified into glycoside hydrolases (GH), glycosyl transferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), and enzymes with auxiliary activities (AA). Each of these categories is subdivided into numerous families with predicted enzyme activities (Lombard et al. 2014). Although their identification is based on sequence homology (and therefore relatively straightforward), it is important to note that even within families there can be a range of predicted enzyme activities. It may therefore be necessary to confirm the enzyme activity of the predicted CAZyme with lab experiments.

From a biotechnology perspective, CAZymes are interesting due to their ability to break down (unfermentable) polysaccharides into oligosaccharides and monosaccharides that can be fermented into ethanol by *S. cerevisiae*. As such, CAZymes play an important role in converting plant biomass into biofuel. Moreover, fungal pathogens of plants use CAZymes as an important weapon in their arsenal to attack their host. In the case of pathogens of important agricultural crops, the CAZyme content of fungal genome can lead to important insights (discussed below).

Initial genome sequencing efforts focused on established model systems used to study CAZymes. Examples include *Aspergillus niger* (Pel et al. 2007) and *Neurospora crassa* (Galagan et al. 2003). This resulted in a wide range of well-characterized enzymes (Coutinho et al. 2009). Furthermore, several key regulators involved in the regulation of CAZyme gene expression were identified (Benocci et al. 2017). Later, large-scale sequencing efforts focused on fungi that break down plant polysaccharides. An important sequencing effort is the 1000 Fungal Genomes Project by the Joint Genome Institute (Grigoriev et al. 2014), resulting in a large number of genomes from across the fungal kingdom, including many plant biomass degrading fungi. More targeted sequencing efforts have focused on groups of fungi, such as the genus *Aspergillus* (Vesth et al. 2018) or the class *Agaricomycetes* (Floudas et al. 2012; Ohm et al. 2014), which includes potent degraders of lignocellulose.

The combined genome sequencing efforts have resulted in a large catalog of putative CAZymes, maintained in the CAZy database (Lombard et al. 2014). Similar to how it was described above for secondary metabolism, this catalog can be screened using a high-throughput synthetic biology approach. Putatively interesting CAZymes can be expressed in a production host, and the enzyme activity can be assayed. This approach precludes the need to grow the original host fungus.

C. Mushroom Development

Mushrooms are the sexual reproductive structures of fungi (predominantly) of the phylum *Basidiomycota* or, more specifically, the class *Agaricomycetes* (Kües and Liu 2000; Kües and Navarro-González 2015). Mushrooms are a nutritious and sustainable food source for a growing world population. They can be cultivated on low-quality agricultural waste streams (e.g., manure, saw dust or straw), which they convert into high quality food. As such, they contribute to a circular economy (Grimm and Wösten 2018) and are interesting from a biotechnology perspective. Examples of edible mushrooms include the white button mushroom (*Agaricus bisporus*), the oyster mushroom (*Pleurotus ostreatus*), and shiitake mushroom (*Lentinula edodes*).

Few mushroom-forming fungi are genetically accessible, but notable exceptions are *Schizophyllum commune* and *Coprinopsis cinerea*, both of which have been used as model systems for decades (Kües and Navarro-González 2015). This has resulted in the identification of structural proteins involved in mushroom development, such as hydrophobins (Wösten 2001), as well as multiple developmental regulators (Terashima et al. 2005; Ohm et al. 2011, 2013).

The number of available genomes of mushroom-forming fungi has dramatically increased in recent years, although it should be noted that most mushroom-forming fungi were sequenced due to their capacity to degrade lignocellulose (Ohm et al. 2014). Comparative genomics studies have given important new insights into the phylogeny of mushroom-forming fungi (Varga et al. 2019), showing that morphological diversification occurred especially in the Cretaceous and Paleocene. Moreover, numerous novel gene families have been identified that may be involved in mushroom development (Sipos et al. 2017; Krizsán et al. 2019; Almási et al. 2019), based on their conservation in mushroom-forming species as well as their gene expression profile during mushroom development. These genes are currently studied in more detail, which is facilitated by the recent development of CRISPR/Cas9 genome editing tools (Sugano et al. 2017; Vonk et al. 2019).

D. Plant Interactions

Many fungi interact with plants in one way or another. This can be beneficial for the host plant, for example, in the case of **mycorrhizal fungi** that form a symbiosis with plant roots. In contrast, **fungal pathogens** can be detrimental to plant health. Both these fungal lifestyles are important from a biotechnology perspective, since they can strongly impact the yield of agricultural crops.

Although plant pathogens are found across the fungal kingdom, many destructive pathogens belong to the phylum *Ascomycota*. Examples include various species of *Fusarium* and *Verticillium*, which were early targets of genome sequencing (Cuomo et al. 2007; Ma et al. 2010; Klosterman et al. 2011). Comparative genome analysis allowed the reconstruction of gene evolution of pathogenesis-related genes, which are generally called effector genes. More recently all *Verticillium* species were sequenced, and the subsequent analysis revealed frequent chromosomal rearrangements as well as gene family losses. Moreover, in these species only about 200–600 species-specific genes occurred, which are markedly different from the conserved genes and are likely candidates for host specificity (Shi-Kunne et al. 2018). The class *Dothideomycetes* harbors many pathogens, including the wheat pathogen *Zymoseptoria tritici* (formerly known as *Mycosphaerella graminicola*), tomato pathogen *Passalora fulva* (formerly known as *Clados-*

porium fulvum), pine pathogen *Dothistroma septosporum*, and maize pathogen *Bipolaris maydis* (formerly known as *Cochliobolus heterostrophus*) (Goodwin et al. 2011; de Wit et al. 2012; Condon et al. 2013). *P. fulva* and *D. septosporum* are closely related but have very different host plants (tomato and pine, respectively) and lifestyles (hemibiotroph and necrotroph, respectively). Genome sequencing revealed the evolution of a gene cluster involved in the production of dothistromin toxin by *D. septosporum*, as well as effector genes specific to *P. fulva*. Comparing the two genomes suggests that these pathogens had a common ancestral host but have since diverged into different hosts and lifestyles by a differentiation in gene content, pseudogenization, as well as gene regulation (de Wit et al. 2012). More generally, a comparative analysis of members of the class *Dothideomycetes* showed that genome evolution follows a pattern of frequent short intra-chromosomal inversions and few inter-chromosomal rearrangements (Hane et al. 2011; Ohm et al. 2012).

In contrast to plant pathogens, mycorrhizal fungi form a symbiosis that is beneficial to the plant host. Generally, during this symbiosis the fungus provides micronutrients to the plant, while the plant provides carbohydrates (sugars produced by photosynthesis) to the fungus. This mycorrhizal lifestyle evolved independently several times across the fungal kingdom, in species as diverse as the mushroom-forming *Basidiomycete Laccaria bicolor*, the *Dothideomycete Cenococcum geophilum*, and the Périgord black truffle *Tuber melanosporum* (Martin et al. 2008, 2010; Peter et al. 2016). Although there are many differences between these mycorrhizal fungi, a general pattern is that (compared to their non-mycorrhizal relatives) the number of plant cell wall degrading CAZymes decreased, while the number of lineage-specific genes increased (especially genes that were differentially expressed during symbiosis). Nevertheless, mycorrhizal fungi have retained a unique set of CAZymes, which suggests that they are still capable of degrading lignocellulose and therefore are not fully reliant on their plant host (Kohler et al. 2015; Martino et al. 2018).

The genus *Trichoderma* contains several mycoparasitic species that promote plant growth. To some extent this can be explained by the fact that they parasitise on deleterious plant pathogens. However, several strains also induce root branching and increase shoot biomass (Kubicek et al. 2011; Druzhinina et al. 2011; Contreras-Cornejo et al. 2016).

V. Conclusions

This chapter described recent improvements in sequencing technologies that are used to sequence fungal genomes. As these sequencing technologies mature further, it will soon be trivial and affordable to obtain a high-quality telomere-to-telomere assembly. Accurate gene prediction and data analysis is still a challenge, although algorithms and pipelines continue to improve. Currently fungal genome sequencing is already affordable to small labs and individual researchers. For those who are interested in starting with fungal genome sequencing, the following pipeline has proven to work very well in my lab (as an example): we routinely sequence fungal genomes using Illumina (occasionally supplemented with long read from Oxford Nanopore) and genome assembly is done with SPAdes (Bankevich et al. 2012). Gene prediction is preferably done with BRAKER in combination with RNA-Seq expression data (Hoff et al. 2016). Basic functional annotation is done with InterProScan (Hunter et al. 2009) and supplemented with other algorithms, depending on the scientific questions.

Important next steps include a functional genomics approach, which relies heavily on an accurate genome sequence. In functional genomics, high-throughput (sequencing-based) techniques are used in an effort to assign function to elements of the genome (usually genes). These techniques may include RNA-Seq (to study gene expression), ChIP-Seq (to study various aspects of epigenetics), as well as high-throughput gene inactivations. Gene inactivations and other genome editing approaches have been greatly facilitated by the develop-

ment of CRISPR/Cas9 across the fungal kingdom (Shi et al. 2017). I expect that in the coming years, large improvements will be made in techniques related to functional genomics, further accelerating discoveries across the fungal kingdom.

Fungal genome sequencing, comparative genomics and functional genomics are Big Data sciences and require a specific skill set: most bioinformatics tools run in a Linux environment and programming skills (e.g. in Python and R) are essential for advanced analyses. Most universities now include these aspects in their curriculum, ensuring that the next generation of researchers will be skilled in both experimental lab work and computational biology.

References

Almási É, Sahu N, Krizsán K, Bálint B, Kovács GM, Kiss B, Cseklye J, Drula E, Henrissat B, Nagy I, Chovatia M, Adam C, LaButti K, Lipzen A, Riley R, Grigoriev IV, Nagy LG (2019) Comparative genomics reveals unique wood-decay strategies and fruiting body development in the Schizophyllaceae. New Phytol. https://doi.org/10.1111/nph.16032

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aparicio S, Chapman J, Stupka E, Putnam N, Chia J-M, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MDS, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJK, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1301–1310. https://doi.org/10.1126/science.1072104

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815. https://doi.org/10.1038/35048692

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29. https://doi.org/10.1038/75556

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, Warrenfeltz S, Aurrecoechea C, Stoeckert CJ, Kissinger JC, Roos DS, Hertz-Fowler C (2018) FungiDB: an integrated bioinformatic resource for fungi and oomycetes. J Fungi 4:39. https://doi.org/10.3390/jof4010039

Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Bye-A-Jee H, Cowley A, Da Silva A, De Giorgi M, Dogan T, Fazzini F, Castro LG, Figueira L, Garmiri P, Georghiou G, Gonzalez D, Hatton-Ellis E, Li W, Liu W, Lopez R, Luo J, Lussi Y, MacDougall A, Nightingale A, Palka B, Pichler K, Poggioli D, Pundir S, Pureza L, Qi G, Renaux A, Rosanoff S, Saidi R, Sawford T, Shypitsyna A, Speretta E, Turner E, Tyagi N, Volynkin V, Wardell T, Warner K, Watkins X, Zaru R, Zellner H, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter M-C, Boeckmann B, Bolleman J, Boutet E, Breuza L, Casal-Casas C, de Castro E, Coudert E, Cuche B, Doche M, Dornevil D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G, Lara V, Lemercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nouspikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey A-L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh L-S, Zhang J (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169. https://doi.org/10.1093/nar/gkw1099

Batzoglou S, Jaffe DB, Stanley K, Butler J, Gnerre S, Mauceli E, Berger B, Mesirov JP, Lander ES (2002) ARACHNE: a whole-genome shotgun assembler. Genome Res 12:177–189. https://doi.org/10.1101/gr.208902

Bennett S (2004) Solexa Ltd. Pharmacogenomics 5:433–438. https://doi.org/10.1517/14622416.5.4.433

Bennett JW, Chung KT (2001) Alexander Fleming and the discovery of penicillin. Adv Appl Microbiol 49:163–184

Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, de Vries RP (2017) Regulators of plant biomass degradation in ascomycetous fungi. Biotechnol Bio-

fuels 10:152. https://doi.org/10.1186/s13068-017-0841-x

Billingsley JM, DeNicola AB, Tang Y (2016) Technology development for natural product biosynthesis in *Saccharomyces cerevisiae*. Curr Opin Biotechnol 42:74–83. https://doi.org/10.1016/j.copbio.2016.02.033

Birney E, Clamp M, Durbin R (2004) GeneWise and genomewise. Genome Res 14:988–995. https://doi.org/10.1101/gr.1865504

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH (2017) AntiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 45:W36–W41. https://doi.org/10.1093/nar/gkx319

Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH (2016) JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17:66. https://doi.org/10.1186/s13059-016-0924-1

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196. https://doi.org/10.1101/gr.6743907

Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J, Pérez G, Pisabarro AG, Grigoriev IV, Stajich JE, Ramírez L (2016) Transposable elements versus the fungal genome: impact on whole-genome architecture and transcriptional profiles. PLoS Genet 12: e1006108. https://doi.org/10.1371/journal.pgen.1006108

Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G, Simison M, Miyasato SR, Binkley J, Orvis J, Shah P, Wymore F, Sherlock G, Wortman JR (2014) The *Aspergillus* Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. Nucleic Acids Res 42:D705–D710. https://doi.org/10.1093/nar/gkt1029

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. Nucleic Acids Res 40:D700–D705. https://doi.org/10.1093/nar/gkr1029

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474

Cissé OH, Stajich JE (2019) FGMP: assessing fungal genome completeness. BMC Bioinformatics 20:184. https://doi.org/10.1186/s12859-019-2782-9

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. Nucleic Acids Res 44:D67–D72. https://doi.org/10.1093/nar/gkv1276

Clevenger KD, Bok JW, Ye R, Miley GP, Verdan MH, Velk T, Chen C, Yang K, Robey MT, Gao P, Lamprecht M, Thomas PM, Islam MN, Palmer JM, Wu CC, Keller NP, Kelleher NL (2017) A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat Chem Biol 13:895–901. https://doi.org/10.1038/nchembio.2408

Clutterbuck AJ (2011) Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. Fungal Genet Biol 48:306–326. https://doi.org/10.1016/J.FGB.2010.09.002

Condon BJ, Leng Y, Wu D, Bushley KE, Ohm RA, Otillar R, Martin J, Schackwitz W, Grimwood J, MohdZainudin N, Xue C, Wang R, Manning VA, Dhillon B, Tu ZJ, Steffenson BJ, Salamov A, Sun H, Lowry S, LaButti K, Han J, Copeland A, Lindquist E, Barry K, Schmutz J, Baker SE, Ciuffetti LM, Grigoriev IV, Zhong S, Turgeon BG (2013) Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. PLoS Genet 9:e1003233. https://doi.org/10.1371/journal.pgen.1003233

Contreras-Cornejo HA, Macías-Rodríguez L, del-Val E, Larsen J (2016) Ecological functions of *Trichoderma* spp. and their secondary metabolites in the rhizosphere: interactions with plants. FEMS Microbiol Ecol 92:fiw036. https://doi.org/10.1093/femsec/fiw036

Coutinho PM, Andersen MR, Kolenova K, vanKuyk PA, Benoit I, Gruben BS, Trejo-Aguilar B, Visser H, van Solingen P, Pakula T, Seiboth B, Battaglia E, Aguilar-Osorio G, de Jong JF, Ohm RA, Aguilar M, Henrissat B, Nielsen J, Stålbrand H, de Vries RP (2009) Post-genomic insights into the plant polysaccharide degradation potential of *Aspergillus nidulans* and comparison to *Aspergillus niger* and *Aspergillus oryzae*. Fungal Genet Biol 46(Suppl 1): S161–S169

Cuomo CA, Birren BW (2010) The fungal genome initiative and lessons learned from genome sequencing. Methods Enzymol 470:833–855. https://doi.org/10.1016/S0076-6879(10)70034-3

Cuomo CA, Guldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma L-J, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang Y-L, DeCaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes H-W, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MIG,

Seong K-Y, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317:1400–1402. https://doi.org/10.1126/science.1143708

de Wit PJGM, van der Burgt A, Ökmen B, Stergiopoulos I, Abd-Elsalam KA, Aerts AL, Bahkali AH, Beenen HG, Chettri P, Cox MP, Datema E, de Vries RP, Dhillon B, Ganley AR, Griffiths SA, Guo Y, Hamelin RC, Henrissat B, Kabir MS, Jashni MK, Kema G, Klaubauf S, Lapidus A, Levasseur A, Lindquist E, Mehrabi R, Ohm RA, Owen TJ, Salamov A, Schwelm A, Schijlen E, Sun H, van den Burg HA, van Ham RCHJ, Zhang S, Goodwin SB, Grigoriev IV, Collemare J, Bradshaw RE (2012) The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. PLoS Genet 8: e1003088. https://doi.org/10.1371/journal.pgen.1003088

Downs JR, Clearfield M, Weis S, Whitney E, Shapiro DR, Beere PA, Langendorfer A, Stein EA, Kruyer W, Gotto AM Jr, for the AFCAPS/TexCAPS Research Group for the AR (1998) Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels. JAMA 279:1615. https://doi.org/10.1001/jama.279.20.1615

Druzhinina IS, Seidl-Seiboth V, Herrera-Estrella A, Horwitz BA, Kenerley CM, Monte E, Mukherjee PK, Zeilinger S, Grigoriev IV, Kubicek CP (2011) Trichoderma: the genomics of opportunistic success. Nat Rev Microbiol 9:749–759. https://doi.org/10.1038/nrmicro2637

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344

Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, Bult C, Tomb J, Dougherty B, Merrick J, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512. https://doi.org/10.1126/science.7542800

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Górecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A, Kües U, Kumar TKA, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, Rokas A, Ruiz-Dueñas FJ, Sabat G, Salamov A, Samejima M, Schmutz J, Slot JC, St John F, Stenlid J, Sun H, Sun S, Syed K, Tsang A, Wiebenga A, Young D, Pisabarro A, Eastwood DC, Martin F, Cullen D, Grigoriev IV, Hibbett DS (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science 336:1715–1719. https://doi.org/10.1126/science.1221748

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D, Li W, Pratt RJ, Osmani SA, DeSouza CPC, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren B (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422:859–868. https://doi.org/10.1038/nature01554

Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman JR, Batzoglou S, Lee S-I, Baştürkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Peñalva MÁ, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW (2005) Sequencing of *Aspergillus nidulans* and compara-

tive analysis with *A. fumigatus* and *A. oryzae*. Nature 438:1105–1115. https://doi.org/10.1038/nature04341

Ghurye J, Pop M (2019) Modern technologies and algorithms for scaffolding assembled genomes. PLoS Comput Biol 15:e1006994. https://doi.org/10.1371/journal.pcbi.1006994

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274(546):563–567. https://doi.org/10.1126/SCIENCE.274.5287.546

Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der Lee TAJ, Grimwood J, Aerts A, Antoniw J, Bailey A, Bluhm B, Bowler J, Bristow J, van der Burgt A, Canto-Canché B, Churchill ACL, Conde-Ferràez L, Cools HJ, Coutinho PM, Csukai M, Dehal P, De Wit P, Donzelli B, van de Geest HC, van Ham RCHJ, Hammond-Kosack KE, Henrissat B, Kilian A, Kobayashi AK, Koopmann E, Kourmpetis Y, Kuzniar A, Lindquist E, Lombard V, Maliepaard C, Martins N, Mehrabi R, Nap JPH, Ponomarenko A, Rudd JJ, Salamov A, Schmutz J, Schouten HJ, Shapiro H, Stergiopoulos I, Torriani SFF, Tu H, de Vries RP, Waalwijk C, Ware SB, Wiebenga A, Zwiers L-H, Oliver RP, Grigoriev IV, Kema GHJ (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogeness. PLoS Genet 7:e1002070. https://doi.org/10.1371/journal.pgen.1002070

Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, Kuske C, Magnuson JK, Martin F, Spatafora JW, Tsang A, Baker SE (2011) Fueling the future with fungal genomics. Mycology 2:192–209. https://doi.org/10.1080/21501203.2011.584577

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I (2014) MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42:D699–D704. https://doi.org/10.1093/nar/gkt1183

Grimm D, Wösten HAB (2018) Mushroom cultivation in the circular economy. Appl Microbiol Biotechnol 102:7795–7803. https://doi.org/10.1007/s00253-018-9226-8

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31:5654–5666. https://doi.org/10.1093/nar/gkg770

Hane JK, Rouxel T, Howlett BJ, Kema GHJ, Goodwin SB, Oliver RP (2011) A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. Genome Biol 12:R45. https://doi.org/10.1186/gb-2011-12-5-r45

Haridas S, Salamov A, Grigoriev IV (2018) Fungal genome annotation. Methods Mol Biol 1775:171–184

Harvey CJB, Tang M, Schlecht U, Horecka J, Fischer CR, Lin H-C, Li J, Naughton B, Cherry J, Miranda M, Li YF, Chu AM, Hennessy JR, Vandova GA, Inglis D, Aiyar RS, Steinmetz LM, Davis RW, Medema MH, Sattely E, Khosla C, St Onge RP, Tang Y, Hillenmeyer ME (2018) HEx: a heterologous expression platform for the discovery of fungal natural products. Sci Adv 4:eaar5459. https://doi.org/10.1126/sciadv.aar5459

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2016) BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35:W585–W587. https://doi.org/10.1093/nar/gkm259

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2009) InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211–D215. https://doi.org/10.1093/nar/gkn785

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. https://doi.org/10.1038/35057062

Jain M, Olsen HE, Paten B, Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239. https://doi.org/10.1186/s13059-016-1103-0

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467. https://doi.org/10.1159/000084979

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27

Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci USA 98:8714–8719. https://doi.org/10.1073/pnas.151269298

Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented

in Repbase. Nat Rev Genet 9:411–412. https://doi.org/10.1038/nrg2165-c1

Keller NP (2019) Fungal secondary metabolism: regulation, function and drug discovery. Nat Rev Microbiol 17:167–180. https://doi.org/10.1038/s41579-018-0121-1

Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism—from biochemistry to genomics. Nat Rev Microbiol 3:937–947. https://doi.org/10.1038/nrmicro1286

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res 12:996–1006. https://doi.org/10.1101/gr.229102

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol 47:736–741. https://doi.org/10.1016/j.fgb.2010.06.003

Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z, Henrissat B, Lee Y-H, Park J, Garcia-Pedrajas MD, Barbara DJ, Anchieta A, de Jonge R, Santhanam P, Maruthachalam K, Atallah Z, Amyotte SG, Paz Z, Inderbitzin P, Hayes RJ, Heiman DI, Young S, Zeng Q, Engels R, Galagan J, Cuomo CA, Dobinson KF, Ma L-J (2011) Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog 7:e1002137. https://doi.org/10.1371/journal.ppat.1002137

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, Colpaert J, Copeland A, Costa MD, Doré J, Floudas D, Gay G, Girlanda M, Henrissat B, Herrmann S, Hess J, Högberg N, Johansson T, Khouja H-R, LaButti K, Lahrmann U, Levasseur A, Lindquist EA, Lipzen A, Marmeisse R, Martino E, Murat C, Ngan CY, Nehls U, Plett JM, Pringle A, Ohm RA, Perotto S, Peter M, Riley R, Rineau F, Ruytinx J, Salamov A, Shah F, Sun H, Tarkka M, Tritt A, Veneault-Fourrey C, Zuccaro A, Tunlid A, Grigoriev IV, Hibbett DS, Martin F, Martin F (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet 47:410–415. https://doi.org/10.1038/ng.3223

Krizsán K, Almási É, Merényi Z, Sahu N, Virágh M, Kószó T, Mondo S, Kiss B, Bálint B, Kües U, Barry K, Cseklye J, Hegedüs B, Henrissat B, Johnson J, Lipzen A, Ohm RA, Nagy I, Pangilinan J, Yan J, Xiong Y, Grigoriev IV, Hibbett DS, Nagy LG (2019) Transcriptomic atlas of mushroom development reveals conserved genes behind complex multicellularity in fungi. Proc Natl Acad Sci USA 116:7409–7418. https://doi.org/10.1073/PNAS.1817822116

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/JMBI.2000.4315

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Coulpier F, Deshpande N, von Döhren H, Ebbole DJ, Esquivel-Naranjo EU, Fekete E, Flipphi M, Glaser F, Gómez-Rodríguez EY, Gruber S, Han C, Henrissat B, Hermosa R, Hernández-Oñate M, Karaffa L, Kosti I, Le Crom S, Lindquist E, Lucas S, Lübeck M, Lübeck PS, Margeot A, Metz B, Misra M, Nevalainen H, Omann M, Packer N, Perrone G, Uresti-Rivera EE, Salamov A, Schmoll M, Seiboth B, Shapiro H, Sukno S, Tamayo-Ramos JA, Tisch D, Wiest A, Wilkinson HH, Zhang M, Coutinho PM, Kenerley CM, Monte E, Baker SE, Grigoriev IV (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. Genome Biol 12:R40. https://doi.org/10.1186/gb-2011-12-4-r40

Kües U, Liu Y (2000) Fruiting body production in basidiomycetes. Appl Microbiol Biotechnol 54:141–152. https://doi.org/10.1007/s002530000396

Kües U, Navarro-González M (2015) How do Agaricomycetes shape their fruiting bodies? 1. Morphological aspects of development. Fungal Biol Rev 29:63–97. https://doi.org/10.1016/J.FBR.2015.05.001

Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE (2013) Web Apollo: a web-based genomic annotation editing platform. Genome Biol 14:R93. https://doi.org/10.1186/gb-2013-14-8-r93

Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, Vamathevan J, Miranda M, Anderson IJ, Fraser JA, Allen JE, Bosdet IE, Brent MR, Chiu R, Doering TL, Donlin MJ, D'Souza CA, Fox DS, Grinberg V, Fu J, Fukushima M, Haas BJ, Huang JC, Janbon G, Jones SJM, Koo HL, Krzywinski MI, Kwon-Chung JK, Lengeler KB, Maiti R, Marra MA, Marra RE, Mathewson CA, Mitchell TG, Pertea M, Riggs FR, Salzberg SL, Schein JE, Shvartsbeyn A, Shin H, Shumway M, Specht CA, Suh BB, Tenney A, Utterback TR, Wickes BL, Wortman JR, Wye NH, Kronstad JW, Lodge JK, Heitman J, Davis RW, Fraser CM, Hyman RW (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science 307:1321–1324. https://doi.org/10.1126/science.1103773

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178

Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm.

Nucleic Acids Res 42:e119–e119. https://doi.org/10.1093/nar/gku557

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim W-B, Woloshuk C, Xie X, Xu J-R, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RAE, Chapman S, Coulson R, Coutinho PM, Danchin EGJ, Diener A, Gale LR, Gardiner DM, Goff S, Hammond-Kosack KE, Hilburn K, Hua-Van A, Jonkers W, Kazan K, Kodira CD, Koehrsen M, Kumar L, Lee Y-H, Li L, Manners JM, Miranda-Saavedra D, Mukherjee M, Park G, Park J, Park S-Y, Proctor RH, Regev A, Ruiz-Roldan MC, Sain D, Sakthikumar S, Sykes S, Schwartz DC, Turgeon BG, Wapinski I, Yoder O, Young S, Zeng Q, Zhou S, Galagan J, Cuomo CA, Kistler HC, Rep M (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464:367–373. https://doi.org/10.1038/nature08850

Mardis ER (2017) DNA sequencing technologies: 2006–2016. Nat Protoc 12:213–218. https://doi.org/10.1038/nprot.2016.182

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. https://doi.org/10.1038/nature03959

Martin F, Aerts A, Ahrén D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buée M, Brokstein P, Canbäck B, Cohen D, Courty PE, Coutinho PM, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbé J, Lin YC, Legué V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculita-Hirzel H, Oudot-Le Secq MP, Peter M, Quesneville H, Rajashekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kües U, Lucas S, Van de Peer Y, Podila GK, Polle A, Pukkila PJ, Richardson PM, Rouzé P, Sanders IR, Stajich JE, Tunlid A, Tuskan G, Grigoriev IV (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature 452:88–92. https://doi.org/10.1038/nature06556

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, Porcel B, Rubini A, Amicucci A, Amselem J, Anthouard V, Arcioni S, Artiguenave F, Aury J-M, Ballario P, Bolchi A, Brenna A, Brun A, Buée M, Cantarel B, Chevalier G, Couloux A, Da Silva C, Denoeud F, Duplessis S, Ghignone S, Hilselberger B, Iotti M, Marçais B, Mello A, Miranda M, Pacioni G, Quesneville H, Riccioni C, Ruotolo R, Splivallo R, Stocchi V, Tisserant E, Viscomi AR, Zambonelli A, Zampieri E, Henrissat B, Lebrun M-H, Paolocci F, Bonfante P, Ottonello S, Wincker P (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature 464:1033–1038. https://doi.org/10.1038/nature08867

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barabote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS, Brettin TS (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. Hypocrea jecorina). Nat Biotechnol 26:553–560. https://doi.org/10.1038/nbt1403

Martino E, Morin E, Grelet G-A, Kuo A, Kohler A, Daghino S, Barry KW, Cichocki N, Clum A, Dockter RB, Hainaut M, Kuo RC, LaButti K, Lindahl BD, Lindquist EA, Lipzen A, Khouja H-R, Magnuson J, Murat C, Ohm RA, Singer SW, Spatafora JW, Wang M, Veneault-Fourrey C, Henrissat B, Grigoriev IV, Martin FM, Perotto S (2018) Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. New Phytol 217:1213–1229. https://doi.org/10.1111/nph.14974

Min B, Grigoriev IV, Choi I-G (2017) FunGAP: fungal genome annotation pipeline using evidence-based gene model evaluation. Bioinformatics 33:2936–2937. https://doi.org/10.1093/bioinformatics/btx353

Nützmann H-W, Scazzocchio C, Osbourn A (2018) Metabolic gene clusters in eukaryotes. Annu Rev

Genet 52:159–183. https://doi.org/10.1146/annurev-genet-120417-031237

Ohm RA, De Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, De Vries RP, Record E, Levasseur A, Baker SE, Bartholomew KA, Coutinho PM, Erdmann S, Fowler TJ, Gathman AC, Lombard V, Henrissat B, Knabe N, Kües U, Lilly WW, Lindquist E, Lucas S, Magnuson JK, Piumi F, Raudaskoski M, Salamov A, Schmutz J, Schwarze FWMR, vanKuyk PA, Horton JS, Grigoriev IV, Wösten HAB (2010) Genome sequence of the model mushroom *Schizophyllum commune*. Nat Biotechnol 28:957–963

Ohm RA, de Jong JF, de Bekker C, Wösten HAB, Lugones LG (2011) Transcription factor genes of *Schizophyllum commune* involved in regulation of mushroom formation. Mol Microbiol 81:1433–1445. https://doi.org/10.1111/j.1365-2958.2011.07776.x

Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, Hesse CN, Kosti I, LaButti K, Lindquist EA, Lucas S, Salamov AA, Bradshaw RE, Ciuffetti L, Hamelin RC, Kema GHJ, Lawrence C, Scott JA, Spatafora JW, Turgeon BG, de Wit PJGM, Zhong S, Goodwin SB, Grigoriev IV (2012) Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog 8:e1003037. https://doi.org/10.1371/journal.ppat.1003037

Ohm RA, Aerts D, Wösten HAB, Lugones LG (2013) The blue light receptor complex WC-1/2 of *Schizophyllum commune* is involved in mushroom formation and protection against phototoxicity. Environ Microbiol 15:943–955. https://doi.org/10.1111/j.1462-2920.2012.02878.x

Ohm RA, Riley R, Salamov A, Min B, Choi I-G, Grigoriev IV (2014) Genomics of wood-degrading fungi. Fungal Genet Biol 72:82–90. https://doi.org/10.1016/J.FGB.2014.05.001

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071

Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albang R, Albermann K, Andersen MR, Bendtsen JD, Benen JAE, van den Berg M, Breestraat S, Caddick MX, Contreras R, Cornell M, Coutinho PM, Danchin EGJ, Debets AJM, Dekker P, van Dijck PWM, van Dijk A, Dijkhuizen L, Driessen AJM, d'Enfert C, Geysens S, Goosen C, Groot GSP, de Groot PWJ, Guillemette T, Henrissat B, Herweijer M, van den Hombergh JPTW, van den Hondel CAMJJ, van der Heijden RTJM, van der Kaaij RM, Klis FM, Kools HJ, Kubicek CP, van Kuyk PA, Lauber J, Lu X, van der Maarel MJEC, Meulenberg R, Menke H, Mortimer MA, Nielsen J, Oliver SG, Olsthoorn M, Pal K, van Peij NNME, Ram AFJ, Rinas U, Roubos JA, Sagt CMJ, Schmoll M, Sun J, Ussery D, Varga J, Vervecken W, van de Vondervoort PJJ, Wedler H, Wösten HAB, Zeng A-P, van Ooyen AJJ, Visser J, Stam H (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. Nat Biotechnol 25:221–231. https://doi.org/10.1038/nbt1282

Peter M, Kohler A, Ohm RA, Kuo A, Krützmann J, Morin E, Arend M, Barry KW, Binder M, Choi C, Clum A, Copeland A, Grisel N, Haridas S, Kipfer T, LaButti K, Lindquist E, Lipzen A, Maire R, Meier B, Mihaltcheva S, Molinier V, Murat C, Pöggeler S, Quandt CA, Sperisen C, Tritt A, Tisserant E, Crous PW, Henrissat B, Nehls U, Egli S, Spatafora JW, Grigoriev IV, Martin FM (2016) Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. Nat Commun 7:12662. https://doi.org/10.1038/ncomms12662

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701

Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. Bioinformatics 21:i351–i358. https://doi.org/10.1093/bioinformatics/bti1018

Rawlings ND, Waller M, Barrett AJ, Bateman A (2014) *MEROPS*: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 42:D503–D509. https://doi.org/10.1093/nar/gkt953

Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S, Cozijnsen AJ, Ciuffetti LM, Degrave A, Dilmaghani A, Duret L, Fudal I, Goodwin SB, Gout L, Glaser N, Linglin J, Kema GHJ, Lapalu N, Lawrence CB, May K, Meyer M, Ollivier B, Poulain J, Schoch CL, Simon A, Spatafora JW, Stachowiak A, Turgeon BG, Tyler BM, Vincent D, Weissenbach J, Amselem J, Quesneville H, Oliver RP, Wincker P, Balesdent M-H, Howlett BJ (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. Nat Commun 2:202. https://doi.org/10.1038/ncomms1189

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison CA, Slocombe PM, Smith M (1977a) Nucleotide sequence of bacteriophage φX174 DNA. Nature 265:687–695. https://doi.org/10.1038/265687a0

Sanger F, Nicklen S, Coulson AR (1977b) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467. https://doi.org/10.1073/PNAS.74.12.5463

Shi T-Q, Liu G-N, Ji R-Y, Shi K, Song P, Ren L-J, Huang H, Ji X-J (2017) CRISPR/Cas9-based genome editing of the filamentous fungi: the state of the art.

Appl Microbiol Biotechnol 101:7435–7443. https://doi.org/10.1007/s00253-017-8497-9

Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF (2018) Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene loss. Environ Microbiol 20:1362–1373. https://doi.org/10.1111/1462-2920.14037

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123. https://doi.org/10.1101/gr.089532.108

Sipos G, Prasanna AN, Walter MC, O'Connor E, Bálint B, Krizsán K, Kiss B, Hess J, Varga T, Slot J, Riley R, Bóka B, Rigling D, Barry K, Lee J, Mihaltcheva S, LaButti K, Lipzen A, Waldron R, Moloney NM, Sperisen C, Kredics L, Vágvölgyi C, Patrignani A, Fitzpatrick D, Nagy I, Doyle S, Anderson JB, Grigoriev IV, Güldener U, Münsterkötter M, Nagy LG (2017) Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. Nat Ecol Evol 1:1931–1941. https://doi.org/10.1038/s41559-017-0347-8

Smit AFA, Hubley R, Green P (2015) RepeatMasker Open-4.0 2013-2015. http://www.repeatmasker.org

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19:ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080

Stein LD (2013) Using GBrowse 2.0 to visualize and share next-generation sequence data. Brief Bioinform 14:162–171. https://doi.org/10.1093/bib/bbt001

Sugano SS, Suzuki H, Shimokita E, Chiba H, Noji S, Osakabe Y, Osakabe K (2017) Genome editing in the mushroom-forming basidiomycete *Coprinopsis cinerea*, optimized by a high-throughput transformation system. Sci Rep 7:1260. https://doi.org/10.1038/s41598-017-00883-5

Terashima K, Yuki K, Muraguchi H, Akiyama M, Kamada T (2005) The dst1 gene involved in mushroom photomorphogenesis of *Coprinus cinereus* encodes a putative photoreceptor for blue light. Genetics 171:101–108. https://doi.org/10.1534/genetics.104.040048

Testa AC, Oliver RP, Hane JK (2016) OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. Genome Biol Evol 8:2044–2064. https://doi.org/10.1093/gbe/evw121

The C elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012–2018. https://doi.org/10.1126/SCIENCE.282.5396.2012

Toome M, Ohm RA, Riley RW, James TY, Lazarus KL, Henrissat B, Albu S, Boyd A, Chow J, Clum A, Heller G, Lipzen A, Nolan M, Sandor L, Zvenigorodsky N, Grigoriev IV, Spatafora JW, Aime MC (2014) Genome sequencing provides insight into the reproductive biology, nutritional mode and ploidy of the fern pathogen *Mixia osmundae*. New Phytol 202:554–564. https://doi.org/10.1111/nph.12653

Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L, Andreopoulos W, Angelini C, Antonín V, Barry KW, Bougher NL, Buchanan P, Buyck B, Bense V, Catcheside P, Chovatia M, Cooper J, Dämon W, Desjardin D, Finy P, Geml J, Haridas S, Hughes K, Justo A, Karasiński D, Kautmanova I, Kiss B, Kocsubé S, Kotiranta H, LaButti KM, Lechner BE, Liimatainen K, Lipzen A, Lukács Z, Mihaltcheva S, Morgado LN, Niskanen T, Noordeloos ME, Ohm RA, Ortiz-Santana B, Ovrebo C, Rácz N, Riley R, Savchenko A, Shiryaev A, Soop K, Spirin V, Szebenyi C, Tomšovský M, Tulloss RE, Uehling J, Grigoriev IV, Vágvölgyi C, Papp T, Martin FM, Miettinen O, Hibbett DS, Nagy LG (2019) Megaphylogeny resolves global patterns of mushroom evolution. Nat Ecol Evol 3:668–678. https://doi.org/10.1038/s41559-019-0834-1

Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, Hoof JB, Brandl J, Salamov A, Riley R, Gladden JM, Phatale P, Nielsen MT, Lyhne EK, Kogle ME, Strasser K, McDonnell E, Barry K, Clum A, Chen C, LaButti K, Haridas S, Nolan M, Sandor L, Kuo A, Lipzen A, Hainaut M, Drula E, Tsang A, Magnuson JK, Henrissat B, Wiebenga A, Simmons BA, Mäkelä MR, de Vries RP, Grigoriev IV, Mortensen UH, Baker SE, Andersen MR (2018) Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section Nigri. Nat Genet 50:1688–1695. https://doi.org/10.1038/s41588-018-0246-1

Vonk PJ, Escobar N, Wösten HAB, Lugones LG, Ohm RA (2019) High-throughput targeted gene deletion in the model mushroom *Schizophyllum commune* using pre-assembled Cas9 ribonucleoproteins. Sci Rep 9(1):7632. https://doi.org/10.1038/s41598-019-44133-2

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M,

Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. https://doi.org/10.1038/nrg2165

Wösten HAB (2001) Hydrophobins: multipurpose proteins. Annu Rev Microbiol 55:625–646. https://doi.org/10.1146/annurev.micro.55.1.625

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13:329–342. https://doi.org/10.1038/nrg3174

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107