

## CHAPTER 12

# Pan-genomics of fungi and its applications

Rodrigo Bentes Kato<sup>a,\*</sup>, Arun Kumar Jaiswal<sup>b,\*</sup>, Sandeep Tiwari<sup>b</sup>,  
Debmalya Barh<sup>c</sup>, Vasco Azevedo<sup>b</sup>, Aristóteles Góes-Neto<sup>a</sup>

<sup>a</sup>Molecular and Computational Biology of Fungi Laboratory, Department of Microbiology, Institute of Biological Sciences (ICB), Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

<sup>b</sup>PG Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

<sup>c</sup>Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Purba Medinipur, India

## 1 Introduction

Fungi are an evolutionary lineage within Opisthokonta, comprising one of the largest and most diverse groups of Eukarya on planet Earth [1]. Their multicellular non-motile bodies (mycelia) are constructed of apically growing, walled, and tubular cells (hyphae) or can be unicellular in which each adult individual is a single cell. These two distinct morphological groups are the so-called mycelial (or filamentous) fungi and yeasts, respectively [2]. The fungi are eukaryotic and chemoheterotrophic organisms, exhibiting osmotrophic nutrition with partial external digestion [3].

Fungi play a key role in the global carbon cycle especially in terrestrial biomes [4], and survive by using three basic trophic modes: (i) as saprotrophs, breaking down dead organic matter, (ii) as parasitic (and pathogens), or (iii) mutualistic symbionts with other living organisms [5]. Fungi and their by-products have a great economic importance. Fungi can be used to produce fermented food (e.g., beers, wines, breads, and cheeses) [6], primary and secondary metabolites (e.g., ethanol, hydrolases, and oxidoreductases enzymes, organic acids, and many vitamins, hypocholesterolemics, antineoplastics) [7], and inoculants and biocides (for mycorrhization and as biological control agents) [8]. Furthermore, fungi can be used for bioremediation of solid residues, effluents, and gaseous emissions [9], as well as to produce new biomaterials, such as mycocomposites [10] and nanomycomaterials [11].

Fungi constitute one of the major clades of organisms with approximately 145,000 species already described [12], however, there must be many more species since estimates suggest the existence of 5.1 million species, despite only about 10% were described [13]. Although there is still no consensus in the hierarchical taxonomical classification of the

\* These authors contributed equally to this work.

main groups inside of Kingdom Fungi since many recent classification schemes proposed different number of phyla [14–16], the great majority of the species are in the well-established and consensual phyla Ascomycota and Basidiomycota, which form the sub-kingdom Dikarya [17].

With the advancement and the low cost of high-throughput sequencing technology, these days leads in excess of 250,000 genome projects registered at the Genome Online Database (GOLD) (<https://gold.jgi.doe.gov/measurements>) [18]. These endeavors have made big change in the study of fungal genes and genomic association. The fungal genomic datasets can be exploited for adaptive and environmental behavior study. Substantial genomics and transcriptomics dataset of fungi have empowered the utilization of novel strategies and molecular evolution studies of fungi [19,20]. The combinations of experimental and computational methods have a great potential for point-by-point investigation of the fungal evolution and its biology [20].

The pan-genomics is a comparative genomics-based methodology that identifies the core and the dispensable genomes. The dispensable genome is composed of genes that present in some but not in all the strains studied, as well as the strain-specific genes. The dispensable genome helps to its fundamental way of life but rather present particular points of interest including antifungal resistance, niche adaptation, and the capacity to colonize new hosts [21]. As gene content and genome copy number can vary in distinct populations of a single species, the inventory of the variation at genomic level in different isolates is crucial to characterize the complete set of genes (core and accessory) that exists in a fungal species [20]. In this chapter, our work aimed to perform an extensive literature review and meta-analysis of this customized database in order to depict the state of the art of fungal pan-genomics.

## 2 Application of pan-genomics of fungi based on meta-analysis

The metadata related to genomics, comparative genomics and pan-genomics analyses on fungi, were mined through the literature. The NCBI (National Center for Biotechnology Information) genome and JGI (Joint Genome Institute) genome databases were also used for the search of data related to these areas. Thereafter, abstract and full-text level manual curations were performed. For the description of the data, histograms, and pie charts were constructed using the R version 3.5.0 software [22] and the ggplot2 version 2.2.1 package [23]. The localization on global map was done using the GPS Visualizer tool [24]. As a result, among the obtained 159 articles, only the most cited articles were considered for manual text curation. From the metadata from the databases, we found 97 species and 16.5% of these species has more than 16 distinct isolates fully sequenced. Then, we used this threshold (~16%) to our analyses. The obtained metadata related to fungal pan-genomics from published articles were divided in four groups.

- (1) *Technological*: This group contains fungal isolates that are used to produce several products related to pharmaceuticals, food, dairy products, and alcohol industries. In our search, we found many papers using fungi to produce foods as fermented products and alcohol beverages.
- (2) *Environmental*: It is a collection of published research results on the study of fungal diversity in the natural environment. We found some works that use fungi in biological process of biodegradation and biological treatment of lagoons or rivers. Industry and biodegradation areas have interesting in fungi that act in enzyme activities. These fungi are very important in agricultural and ecological contexts because they maintain the balance of the environment decomposing plant debris, degrade toxic substances, help plants to grow, and protect themselves against enemies.
- (3) *Host pathogen*: In this group, we discussed some studies on the fungal diversity in the host-pathogen interactions. Some of these fungi contribute to pharmacy industry to cure some disease.
- (4) *Laboratory*: some fungi were comparable with model, laboratory strains, and then we created this group.

Fig. 1 shows the frequency of number of works in each group. We found that most of the works were related to the technological importance fungal group (~61%). The reason behind the several studies in this area may be the presence of many industries, which invest large amount of money to its research and development. On the other hand, we found that host-pathogen related fungal research were the second bigger group with around 30%, mainly supported by agribusiness.

This metadata analysis was done to show the impact of pan-genomics in comparative fungal genomics. Furthermore, we found 1567 species related to 12 genera of fungi, and

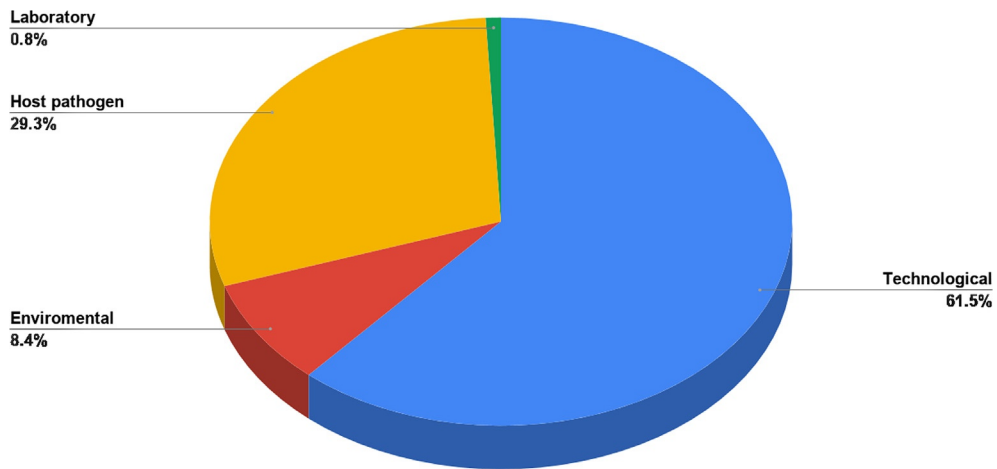
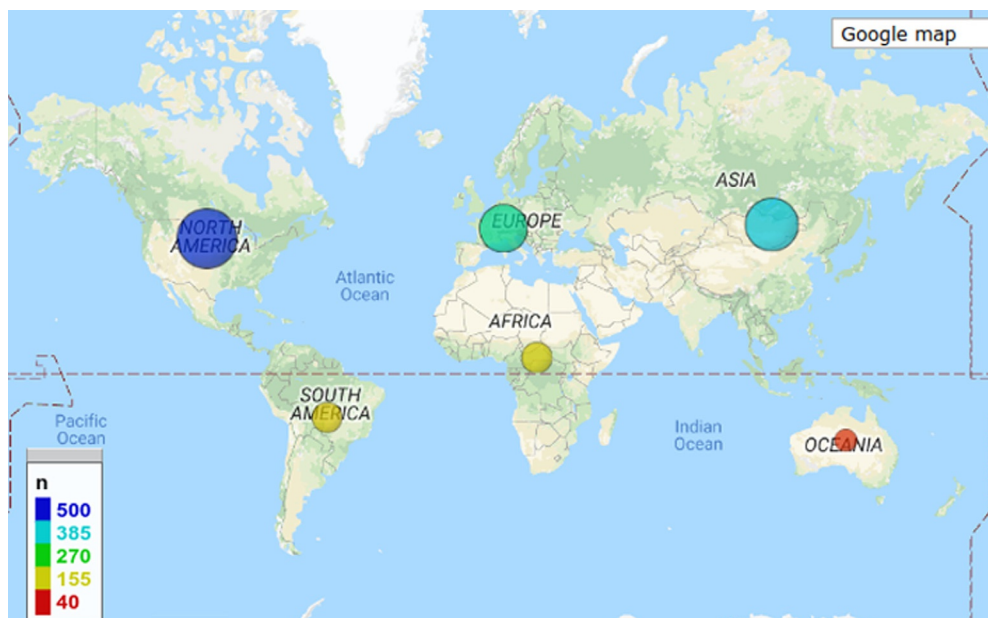


Fig. 1 The pie chart demonstrates the frequency of the studied fungi from each group.



**Fig. 2** The image shows the geographical a distribution of 1567 species of 12 different genera of fungal isolates.

they were distributed globally, where  $n$  means number of isolates for each country (Fig. 2).

Among these 12 genera we found that *Saccharomyces cerevisiae* is the most frequently (Fig. 3) studied genus. Around 83% of researches have been related to *S. cerevisiae* in the last 2 years of published research article [25].

## 2.1 Application of pan-genomics on advantageous fungus

The set of complete genes in all the strains of a specific species is known as pan-genome [20]. The genus *Saccharomyces* is among the most important and broadly studied model eukaryotic organisms. The fermented beverages production commonly used *S. cerevisiae* yeasts, dates at least as back 7000 BC, in china [26]. In order to comprehend the significance of selection during domestication and understand the levels of genetic diversity among wine yeasts, a number of pan-genome analyses have been done using commercial wine yeasts and industrial yeasts. A set of 83 strains of *S. cerevisiae* was used for the pan-genome analysis to identify the copy number variations in this yeast distributed in different industrial environments [26].

Another comparative work of 43 strains of *S. cerevisiae* isolated from fermenting grape was used to analyze genome renewal, and they propose that natural wine yeast strains can undergo such modifications and, thereby, change a multiple heterozygote into

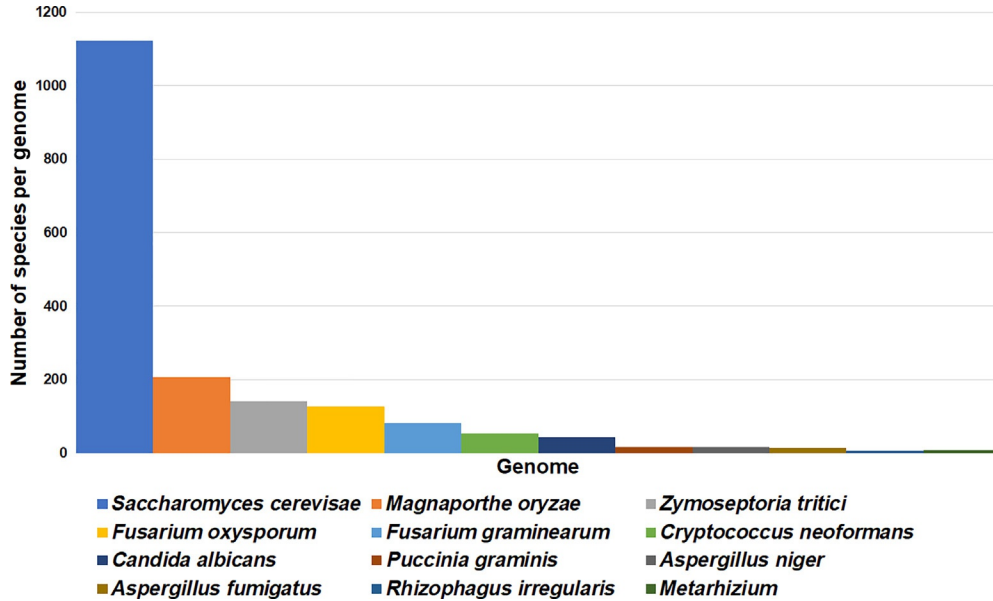


Fig. 3 The figure represents the most frequently studied fungi among the 12 genera.

completely homozygous diploids, some of which may replace the original heterozygous diploid [27]. The pan-genome studies of very polymorphic eukaryotic pathogens utilizing the accessory genome gives a better understanding for adaptive evolution. The genomics study of this yeast has enhanced our understanding of the evolutionary dynamics of natural populations when comparing with the domesticated strains, during infections, and during laboratory experiments [28]. Apart from *S. cerevisiae* [29], population genomic studies have also characterized the metabolic, genetic, and biogeographic diversity of *Saccharomyces paradoxus* [30], *Saccharomyces kudriavzevii* [31], and *Saccharomyces uvarum* [32]. As well as all the organisms, yeast genome sequences largely describe their genetic makeups; however, the comparative genomic studies have given better shape to the historical and genetic processes in their evolution [28].

## 2.2 Application of pan-genomics in disadvantageous fungus

Effect of chromosomal rearrangements on genes can lead to functional variation between individuals and influenced the expression of phenotypic attributes [33]. The inter and intraspecific structural variation among genomes of fungi has already been reported [34]. This structural variation among the pathogens can affect their host range. For instance, in the fungus *Melanopsichium pennsylvanicum*, gene loss are responsible for the hosts jump from dicotyledonous to monocotyledon plant hosts [33,35]. Unexpected

number of fungal and fungal-like diseases have been recently afflicted animals and plants, and some of them are the most severe die-offs and extinctions ever witnessed in wild species, and are a peril for food security [20]. Emerging infectious diseases (EIDs) brought about by fungi are progressively perceived as exhibiting a danger to food security around the world [36]. Until date, several fungal genomes are responsible to accurate and complete genome assemblies using long-read sequencing technologies [37,38]. Various symbiotic interactions have been recorded between insects and fungi. Although the genomics has already been elucidated in many fungi that expanded our knowledge on this group, there is still much to explore the genomic features of the insect-commensal relationships [39]. *Zymoseptoria tritici* is a pathogen of wheat causing Septoria tritici blotch, and a recently published work using pan-genome analysis of this pathogen identified that host specialization has evolved by gene deletions and chromosomal rearrangements. In this aforementioned study, the authors used five isolates for the pan-genome analysis and 15,749, 9,149, and 6600 nonredundant proteins were identified as pan-genome, core, and accessory genome, respectively [33] (Table 1).

**Table 1** Pan-genomics studies on different fungi

Fungi	Importance	Comparative Genomics	Strains/ isolates	Reference
<i>Saccharomyces cerevisiae</i>	Industrial important	1. Pangenome Analysis of <i>Saccharomyces cerevisiae</i>	25	[40]
		2. Report of the whole-genome sequencing and phenotyping of 1011 <i>Saccharomyces cerevisiae</i> isolates	1011	[25]
<i>Rhizophagus irregularis</i>	Plant pathogen	Genome assembly and gene annotation of the model strain <i>Rhizophagus irregularis</i> DAOM197198, and gene comparison with five different isolates of <i>Rhizophagus irregularis</i>	6	[41]
<i>Puccinia graminis</i> f. sp. <i>tritici</i>	Plant pathogen	Comparative genomics of Australian isolates of the wheat stem rust pathogen <i>Puccinia graminis</i> f. sp. <i>tritici</i> and draft genome was built for a founder Australian <i>Pgt</i> isolate	16	[42]

**Table 1** Pan-genomics studies on different fungi—cont'd

Fungi	Importance	Comparative Genomics	Strains/ isolates	Reference
<i>Zymoseptoria tritici</i>	Plant pathogen	Pangenome analysis of <i>Zymoseptoria tritici</i>	123	[33, 38]
<i>Metarhizium spp.</i>	Insect pathogen	Pangenome analysis for <i>Metarhizium spp.</i>	7	[43]
<i>Coccidioides posadasii</i> , <i>Coccidioides immitis</i> and other fungus of order Onygenales	Human fungal pathogen	Genome sequencing and comparison of the primary human pathogens <i>C. immitis</i> and <i>C. posadasii</i>	17	[44]
<i>Fusarium graminearum</i>	Cereal pathogen	Sequencing of genomes of 60 diverse <i>F. graminearum</i> isolates from North America, and also the assembly of the first pan-genome for <i>F. graminearum</i> to clarify population-level differences in gene content potentially contributing to pathogen diversity.	70	[45]
<i>Fusarium meridionale</i> / <i>Fusarium</i> <i>Asiaticum</i> / <i>Fusarium</i> <i>graminearum</i>	Plant/Cereal pathogen	Genomic comparison and gene content analysis of six newly isolates from the species complex, including the first available genomes of <i>F. asiaticum</i> and <i>F. meridionale</i> , with four other genomes	10	[46]

### 3 Conclusions and future prospective

Although people frequently demonstrate a nonmycophilic or even mycophobic relation with fungi, these group of organisms are vital on many aspects of human life, including medicine, food, and farming, and also play key roles in nature, such as in the carbon biogeochemical cycle. The comparative genomics approach based on sequence similarity with statistical analysis helps in identifying the essential genomic content common among all fungal isolates of a same species as well as the subset of genes encoding novel functions as variable genome. Biotechnology consists of the use of organisms for the development

of processes and products of economic or social interest. It is recognized as one of the technologies for the 21st century with higher potential impact on global problems (diseases, nutrition, and environmental pollution) and sustainable industrial development (use of renewable resources, “green technology,” and reduction of global warming). Based on the search and discovery of industrially exploitable biological resources, the scientific and technological advances achieved by fungal pan-genomics studies in recent years have revolutionized traditional approaches to the exploitation of biological resources for biotechnology.

## References

- [1] F. Badotti, F.S. de Oliveira, C.F. Garcia, A.B. Vaz, P.L. Fonseca, L.A. Nahum, et al., Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi), *BMC Microbiol.* 17 (1) (2017) 42.
- [2] D. Moore, 21st Century Guidebook to Fungi, *Q. Rev. Biol.* 87 (4) (2012) 396.
- [3] Sarah C. Watkinson NPM, Lynne Boddy. *The Fungi*. San Diego, Elsevier Science Publishing Co Inc.
- [4] G.M. Gadd, *Fungi in Biogeochemical Cycles*, CAB International, 2006.
- [5] N.H. Nguyen, Z. Song, S.T. Bates, S. Branco, L. Tedersoo, J. Menke, et al., FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild, *Fungal Ecol.* 20 (2016) 241–248.
- [6] M. Hofrichter, *The Mycota: A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research*, second ed., Springer, 2010.
- [7] *Fungal Biomolecules: Sources, Applications and Recent Developments*, Wiley-Blackwell, 2015.
- [8] T.M. Butt, C. Jackson, N. Magan (Eds.), *Fungi as Biocontrol Agents: Progress Problems and Potential*, CABI, 2001.
- [9] H. Singh, *Mycoremediation*, Wiley, 2006.
- [10] C. Girometta, A. Picco, R. Baiguera, D. Dondi, S. Babbini, M. Cartabia, et al., Physico-mechanical and thermodynamic properties of mycelium-based biocomposites: A review, *Sustainability* 11 (1) (2019).
- [11] R. Prasad, *Fungal Nanotechnology*, Springer, 2017.
- [12] R. F. Species 2000 & ITIS Catalogue of life, 20th February 2019. 2019.
- [13] M. Blackwell, The fungi: 1, 2, 3 ... 5.1 million species? *Am. J. Bot.* 98 (3) (2011) 426–438.
- [14] J.W. Spatafora, M.C. Aime, I.V. Grigoriev, F. Martin, J.E. Stajich, M. Blackwell, The fungal tree of life: from molecular systematics to genome-scale phylogenies, *Microbiol. Spectr.* 5 (5) (2017).
- [15] L. Tedersoo, S. Sánchez-Ramírez, U. Kõljalg, M. Bahram, M. Döring, D. Schigel, et al., High-level classification of the fungi and a tool for evolutionary ecological analyses, *Fungal Divers.* 90 (1) (2018) 135–159.
- [16] J. Choi, S.-H. Kim, A genome tree of life for the fungi kingdom, *Proc. Natl. Acad. Sci.* 114 (35) (2017) 9391–9396.
- [17] D.S. Hibbett, M. Blackwell, T.Y. James, J.W. Spatafora, J.W. Taylor, R. Vilgalys, Phylogenetic taxon definitions for Fungi, Dikarya, Ascomycota and Basidiomycota, *IMA Fungus* 9 (2018) 291–298.
- [18] Genome Online Database (GOLD) n.d. [Internet]. [cited March 2019]. Available from: <https://gold.jgi.doe.gov/measurements>.
- [19] D.S. Hibbett, J.E. Stajich, J.W. Spatafora, Toward genome-enabled mycology, *Mycologia* 105 (6) (2013) 1339–1349.
- [20] J.E. Stajich, Fungal genomes and insights into the evolution of the kingdom, *Microbiol. Spectr.* 5 (4) (2017).
- [21] H. Tettelin, D. Riley, C. Cattuto, D. Medini, Comparative genomics: the bacterial pan-genome, *Curr. Opin. Microbiol.* 11 (5) (2008) 472–477.



- [22] R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: <https://www.r-project.org/>.
- [23] Wickham H. ggplot2 2009.
- [24] Schneider, A GPS visualizer. Available from: [http://www.gpsvisualizer.com/map\\_input?form=data](http://www.gpsvisualizer.com/map_input?form=data).
- [25] J. Peter, M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, A. Bergström, et al., Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates, *Nature* 556 (7701) (2018) 339–344.
- [26] B. Dunn, C. Richter, D.J. Kvitek, T. Pugh, G. Sherlock, Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments, *Genome Res.* 22 (5) (2012) 908–924.
- [27] R.K. Mortimer, P. Romano, G. Suzzi, M. Polsinelli, Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts, *Yeast* 10 (12) (1994) 1543–1552.
- [28] C.T. Hittinger, A. Rokas, F.Y. Bai, T. Boekhout, P. Goncalves, T.W. Jeffries, et al., Genomics and the making of yeast biodiversity, *Curr. Opin. Genet. Dev.* 35 (2015) 100–109.
- [29] P.K. Strobe, D.A. Skelly, S.G. Kozmin, G. Mahadevan, E.A. Stone, P.M. Magwene, et al., The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen, *Genome Res.* 25 (5) (2015) 762–774.
- [30] G. Liti, D.M. Carter, A.M. Moses, J. Warringer, L. Parts, S.A. James, et al., Population genomics of domestic and wild yeasts, *Nature* 458 (7236) (2009) 337–341.
- [31] C.T. Hittinger, P. Gonçalves, J.P. Sampaio, J. Dover, M. Johnston, A. Rokas, Remarkably ancient balanced polymorphisms in a multi-locus gene network, *Nature* 464 (7285) (2010) 54–58.
- [32] P. Almeida, C. Gonçalves, S. Teixeira, D. Libkind, M. Bontrager, I. Masneuf-Pomarède, et al., A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*, *Nat. Commun.* 5 (1) (2014).
- [33] C. Plissonneau, F.E. Hartmann, D. Croll, Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome, *BMC Biol.* 16 (1) (2018).
- [34] M.E. Zolan, Chromosome-length polymorphism in fungi, *Microbiol. Rev.* 59 (4) (1995) 686–698.
- [35] R. Sharma, B. Mishra, F. Runge, M. Thines, Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the smut fungus *Melanopsichium pennsylvanicum*, *Genome Biol. Evol.* 6 (8) (2014) 2034–2049.
- [36] M.C. Fisher, D.A. Henk, C.J. Briggs, J.S. Brownstein, L.C. Madoff, S.L. McCraw, et al., Emerging fungal threats to animal, plant and ecosystem health, *Nature* 484 (7393) (2012) 186–194.
- [37] L. Faino, M.F. Seidl, E. Datema, G.C.M. van den Berg, A. Janssen, A.H.J. Wittenberg, et al., Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome, *MBio* 6 (4) (2015).
- [38] C. Plissonneau, A. Stürchler, D. Croll, The evolution of orphan regions in genomes of a fungal pathogen of wheat, *MBio* 7 (5) (2016).
- [39] Y. Wang, M. Stata, W. Wang, J.E. Stajich, M.M. White, J.-M. Moncalvo, et al., Comparative genomics reveals the core gene toolbox for the fungus–insect symbiosis, *MBio* 9 (3) (2018).
- [40] J. Schacherer, G. Song, B.J.A. Dickins, J. Demeter, S. Engel, B. Dunn, et al., AGAPE (automated genome analysis pipeline) for pan-genome analysis of *Saccharomyces cerevisiae*, *PLoS One* 10 (3) (2015).
- [41] E.C.H. Chen, E. Morin, D. Beaudet, J. Noel, G. Yildirim, S. Ndikumana, et al., High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*, *New Phytol.* 220 (4) (2018) 1161–1171.
- [42] N.M. Upadhyaya, D.P. Garnica, H. Karaoglu, J. Sperschneider, A. Nemri, B. Xu, et al., Comparative genomics of Australian isolates of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* reveals extensive polymorphism in candidate effector genes, *Front. Plant Sci.* 5 (2015).
- [43] X. Hu, G. Xiao, P. Zheng, Y. Shang, Y. Su, X. Zhang, et al., Trajectory and genomic determinants of fungal–pathogen speciation and host adaptation, *Proc. Natl. Acad. Sci.* 111 (47) (2014) 16796–16801.
- [44] T.J. Sharpton, J.E. Stajich, S.D. Rounsley, M.J. Gardner, J.R. Wortman, V.S. Jordar, et al., Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives, *Genome Res.* 19 (10) (2009) 1722–1731.

- [45] A.C. Kelly, T.J. Ward, Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen, PLoS One 13 (3) (2018).
- [46] S. Walkowiak, O. Rowland, N. Rodrigue, R. Subramaniam, Whole genome sequencing and comparative genomics of closely related Fusarium Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*, BMC Genomics 17 (1) (2016).