PLOS PATHOGENS

PEARLS

Unraveling the genomic diversity and virulence of human fungal pathogens through pangenomics

Marion Perrier[1,2], **Amelia E. Barber**[1,2]*

1 Junior Research Group Fungal Informatics, Institute of Microbiology, Friedrich Schiller University, Jena, Germany, **2** Cluster of Excellence Balance of the Microverse, Friedrich Schiller University, Jena, Germany

* amelia.barber@uni-jena.de

OPEN ACCESS

Citation: Perrier M, Barber AE (2024) Unraveling the genomic diversity and virulence of human fungal pathogens through pangenomics. PLoS Pathog 20(7): e1012313. https://doi.org/10.1371/journal.ppat.1012313

Editor: Anuradha Chowdhary, Vallabhbhai Patel Chest Institute, INDIA

Published: July 11, 2024

Funding: MP and AEB are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 20151 – Project-ID 390813860. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Fungi are ubiquitous in all major biomes where they play beneficial roles, but they can also infect plants, insects, and animals. Notably, only a small fraction of the fungal kingdom causes life-threatening infections in humans. Human pathogenic fungi are a growing threat to humanity as advances in modern medicine increase the number of patients who are at risk for fungal infections. The mortality rate associated with these infections is generally high, mainly due to limited therapies and increasing antifungal resistance. Fungal pathogens, like most microbes, exhibit a wide genetic diversity and a correspondingly broad range of phenotypic differences among strains of the same species, including varying levels of virulence. The cumulative impact of these differences on the virulence of fungal pathogens remains poorly understood. However, a better understanding of the genomic differences that underlie virulence differences may ultimately lead to improved management of human fungal infections.

What is a pangenome and why are they relevant to the study of human pathogenic fungi?

Pangenomes were first described in 2005 and have since been widely used to study prokaryotic microbes [1]. However, they have only recently been extended to eukaryotic organisms due to their larger genome size and complexity [2]. A pangenome is defined as the total collection of genes within a given phylogenetic group. It consists of a core genome, which are genes common to all strains, and an accessory genome, which are genes found only in a subset of strains (Fig 1). The core genome is mainly composed of essential genes and those involved in vital cell functions [3]. The function of accessory genes in eukaryotic organisms is not yet fully understood, but is hypothesized to be involved in fungal adaptation to the host or environment [4]. Accessory genes may also play roles in communication, pathogenicity, and antifungal resistance. The knowledge gap regarding the accessory genome in human pathogenic fungi prevents us from answering why some strains are more virulent than others and, more broadly, why strains of the same species exhibit different phenotypes (Fig 1). One way to answer these questions is to use pangenomes as a comparative method to analyze genomic differences between strains of the same species. Pangenomes can also serve as a reference for experimental studies that take diversity into account when investigating the roles of both conserved and variably present genes.

Pangenomes highlight interstrain genomic diversity

Fungi show considerable genomic variation, even within the same species [5]. Since pangenomes have only recently been extended to eukaryotes, only a few human fungal pathogens have been studied, but they have revealed highly variable ratios of core and accessory genes. The commensal fungi *Nakaseomyces glabratus* (formerly *Candida glabrata*) and *Candida albicans* have lower fractions of accessory genes with 6% and 9% of their total pangenomes, respectively (Fig 2). For environmental human fungal pathogens, the observed fraction of accessory genes is higher. In *Cryptococcus neoformans var. grubii*, accessory genes make up 19% of the pangenome and they account for 28% to 31% of the *Aspergillus fumigatus* pangenome [3,6,7]. A similarly wide range in the fraction of accessory genes is observed in the pangenomes of plant fungal pathogens, from 13% for *Pyrenophora teres f. teres* to 41% in *Zymoseptoria tritici*. The multi-kingdom plant and human pathogen *Aspergillus flavus* has the highest proportion of accessory genes described for a fungal pangenome, making up 59% of the total gene content for the species [8]. However, variable taxonomic definitions affect pangenome size and the degree of gene conservation reported. For example, 54% of genes were described as core in the *Fusarium oxysporum* pangenome [9], but other researchers would have classified the genomes analyzed in this study as belonging to many, distinct species as part of the *F. oxysporum* species complex. Interestingly, *Z. tritici*, *Fusarium* spp., and other plant-pathogenic fungi encode part of their accessory genome on discrete chromosomes that show presence–absence variation [9–11]. In human fungal pathogens, the accessory genome has so far only been described throughout the genome, with an enrichment in subtelomeric regions [3,6].

Pangenomes reduce bias introduced by single, linear reference genomes

Omics analysis of eukaryotic organisms typically relies on a linear reference genome, which is the genome from a single strain that has been previously assembled, annotated, and made publicly available. This approach reduces computational time and complexity by simply mapping the genomic or transcriptomic sequence to the reference. However, using a single genome as a reference has a major drawback: it introduces bias when analyzing sequencing data from genetically divergent, non-reference strains. Reads containing more sequence polymorphisms are mapped to the reference at a lower rate, leading to an underestimation of genetic divergence [17]. This bias is even more pronounced in species with high genomic diversity, such as *A. fumigatus*, where a large number of accessory genes are simply ignored because they are not in the reference genome. In light of this, recent work has focused on the development of reference pangenomes. These can be implemented simply as a collection of CDSs (i.e., coding sequences) or as graph-based pangenome structures, both of which can be used for the alignment of sequencing reads [18]. Regardless of the implementation, pangenomes encourage a move away from single reference strains and toward a better understanding of how genetic variation affects a strain's phenotype, including differences in virulence or fitness under stress.

Pangenomes can elucidate phenotypic variation within a species

For many human fungal pathogens, there is marked phenotypic heterogeneity among commonly used strains, including variation in virulence traits [5–7,19]. There is a lack of knowledge about the accessory genes of human fungal pathogens because most research is done with reference strains. In *A. fumigatus* and *A. flavus*, there is remarkable presence–absence variation in many functional categories, including secondary metabolism genes. In *A. fumigatus*, this variation includes the presence of the virulence-associated factor gliotoxin [6,20] and in *A.*

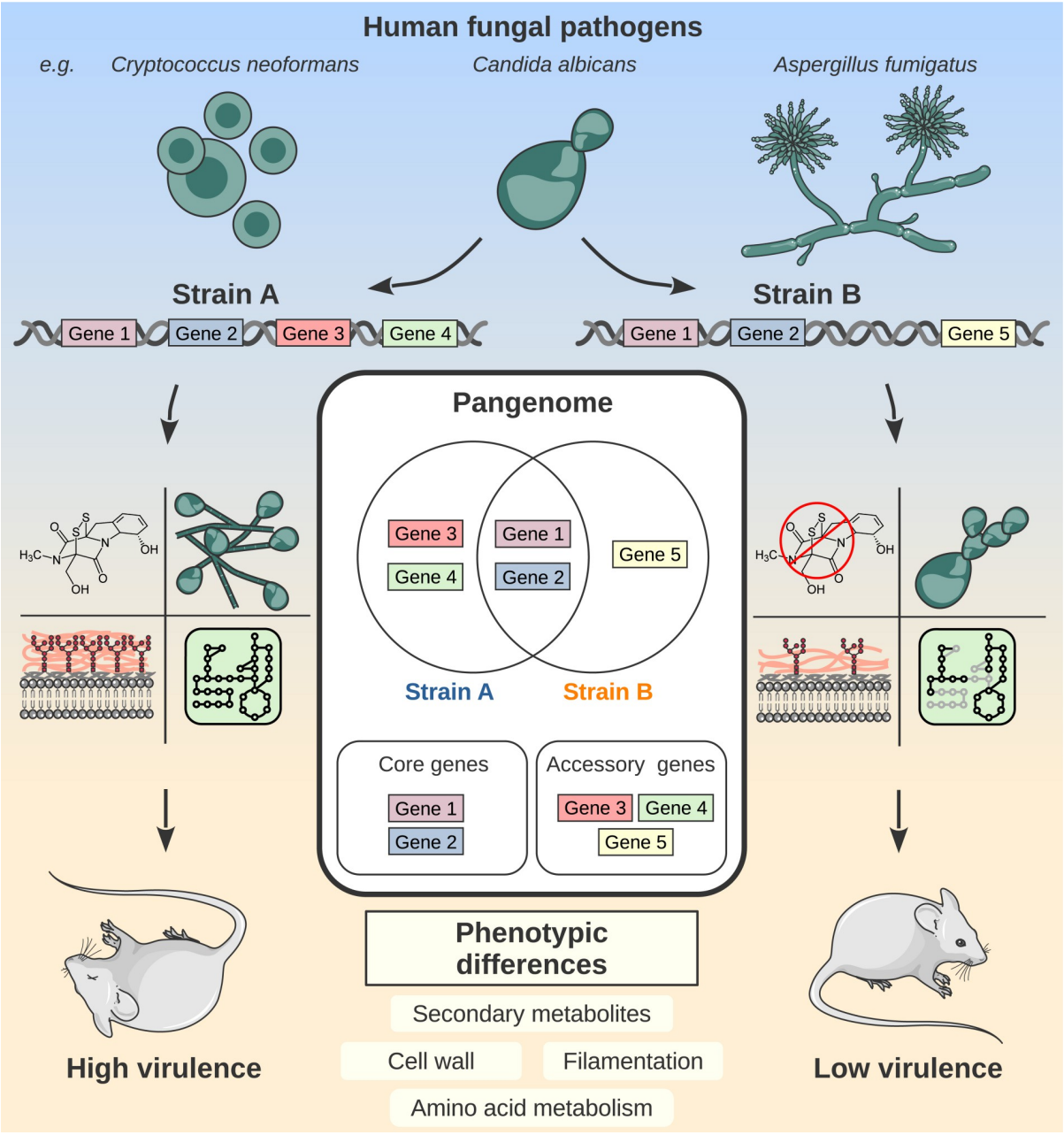

Fig 1. Pangenomic differences impact fungal virulence phenotypes. Illustrative example of the genomic differences between 2 strains of the same species and how they can lead to phenotypic differences in virulence. A schematic representation of a pangenome is shown in the middle. This figure uses original and modified drawings from Servier Medical Art, freely available under CC BY 4.0.

https://doi.org/10.1371/journal.ppat.1012313.g001

flavus the carcinogenic aflatoxin [21]. The accessory genome of *N. glabratus* contains many adhesion proteins, which are important virulence factors, and includes 4 novel adhesion groups [12,22]. In *C. albicans*, the natural loss of *ERG1*, a key regulator of filamentous growth and virulence, transforms a pathogen into an avirulent commensal [19]. A final example of the contribution of accessory genes to fungal virulence is demonstrated by the plant pathogen *F. oxysporum*. The virulence factors responsible for the ability to infect tomato plants are conserved only in strains that cause tomato wilt and are located on a small, accessory

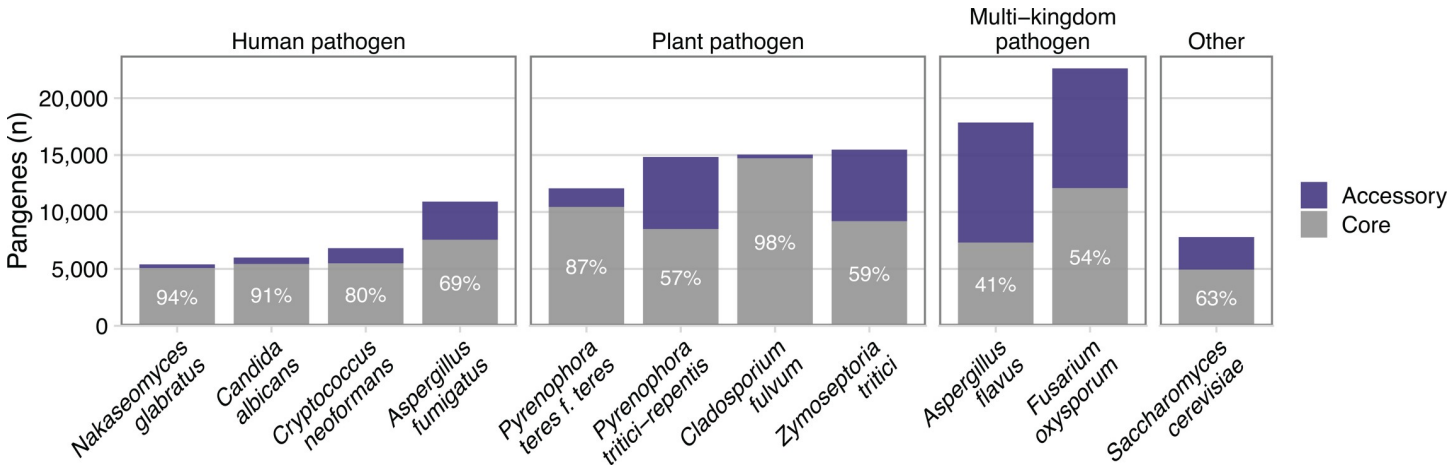

Fig 2. Variation in the pangenomes of human fungal pathogens. Bars show the total number of pangenes and their fill indicates the proportion of core and accessory genes. The core genome's relative proportion is also indicated as a numerical value. Pangenome data for *N. glabratus* from [12] *C. albicans* and *C. neoformans var grubbii* from [3], *A. fumigatus* from [6], *P. teres f. teres* from [13], *P. tritici-repentis* from [14], *C. fulvum* from [11], *Z. tritici* from [15], *A. flavus* from [8], *F. oxysporum* from [10], and *S cerevisiae* from [16].

https://doi.org/10.1371/journal.ppat.1012313.g002

chromosome. The accessory genes can even be transferred during co-incubation between a strain lacking them and one possessing them, transforming a non-pathogen into a pathogen of tomato [23].

Pangenomics helps to understand the evolutionary history of human fungal pathogens

The origin and maintenance of fungal accessory genes remains an open question. In fungal pathogens, as in eukaryotes in general, the accessory genome is primarily driven by both clonal and sexual recombination, gene duplication, and transposons and the differential contributions of these processes likely influences the proportion of accessory genes in a species. In contrast to prokaryotes, horizontal gene transfer (HGT) plays a lesser role in eukaryotic pangenome evolution [3,4,24]. The accessory genome likely contributes to the ongoing evolutionary arms race between host and pathogen, as demonstrated by the pangenomic studies of plant pathogenic fungi [10,13]. Hosts, including humans, are constantly evolving new strategies to recognize and eliminate pathogenic microbes, while at the same time pathogenic microbes are evolving new mechanisms to cause disease. For human-associated pathogenic fungi such as *N. glabratus*, new accessory adhesion genes may emerge due to host selective pressures if they confer a fitness advantage. In contrast, human fungal pathogens that primarily live in the environment are more likely to have pangenomes driven by their environmental fitness, rather than human habitats. The environment of these organisms is highly variable and constantly changing, potentially promoting the evolution of larger pangenomes to cope. Studies of the insect-pathogenic *Metarhizium* genus support this, as generalist species, which can survive in many habitats, have larger accessory genomes than their specialist counterparts [25,26]. This higher proportion of accessory genes in environmental organisms is also supported by a comparative study of 126 bacterial species, in which lifestyle was the largest determinant of pangenome evolution, and free-living species had larger and more fluid pangenomes than host-associated species [27]. How adaptive and/or neutral processes interact to collectively exert environmental impacts on pangenomes is an open question and may vary among species.

The path ahead: Achieving widespread implementation of pangenomics

The use of pangenomes opens exciting new perspectives for gaining important functional insights into human fungal pathogens. However, there are still challenges that need to be overcome to realize their full potential:

Lack of functional annotation and experimental characterization of accessory genes

The study of accessory genes is crucial to understanding phenotypic heterogeneity in fungal pathogens. However, researchers have historically used a limited number of reference strains. Pangenomes will improve our understanding of accessory genes and facilitate experimental work using non-reference strains. However, testing multiple strain backgrounds still requires considerable labor. Hopefully, high-throughput experimental techniques like robotics can offset this in the future.

Genomic data availability and quality vary

The number of genomes, their representativity within the species, and their quality all impact the accuracy of the pangenome. Demonstrating this, the percentage of accessory genes for *A. fumigatus* grew from 17% for 12 genomes [3] to 31% for 300 genomes [6]. The complex nature of fungal genomes is another challenge. The annotation of eukaryotic genomes is an intricate and multi-step process due to eukaryotic genome features such as repetitive elements, complex regulatory elements, and intron-exon gene structures. Poorly annotated or incomplete genomes may overestimate the fraction of accessory genes when core genes are missing or misannotated.

Lack of methodological "gold standard" and curation of pangenomic data

The field of pangenomics is relatively new for eukaryotic organisms. Bioinformatic tools in the field are constantly evolving and expanding and there is no "gold standard." The most common approach is to search for orthologous coding sequences, independent of their genomic location. Other tools use a syntenic approach, which reflects the origins of pangenomics in prokaryotic organisms where genes are organized into operons. Graph-based pangenomes have recently been introduced to encode the genomic variation into a single reference structure. Comparing findings across studies is difficult due to the lack of consensus in methodology. Furthermore, there is no database that curates pangenomic data for fungal pathogens. However, the *Saccharomyces* Genome Database recently implemented a method to incorporate accessory genes that are absent from the reference, providing a promising solution [28].

In conclusion, fungal pangenomes are a recent but rapidly expanding field with the potential to reveal novel insights into their evolution, pathogenesis, and phenotypic heterogeneity.

Acknowledgments

We apologize to the authors whose contributions could not be referenced due to space constraints. We would like to thank the members of the Barber Lab for their helpful discussions of this manuscript.

Author Contributions

Conceptualization: Marion Perrier, Amelia E. Barber.

Funding acquisition: Amelia E. Barber.

Supervision: Amelia E. Barber.

Visualization: Marion Perrier, Amelia E. Barber.

Writing – original draft: Marion Perrier, Amelia E. Barber.

Writing – review & editing: Marion Perrier, Amelia E. Barber.

References

1. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." Proc Natl Acad Sci U S A. 2005 Sep 27; 102(39):13950–5. https://doi.org/10.1073/pnas.0506758102 PMID: 16172379
2. Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. Trends Genet. 2020 Feb 1; 36(2):132–45. https://doi.org/10.1016/j.tig.2019.11.006 PMID: 31882191
3. McCarthy CGP, Fitzpatrick DA. Pan-genome analyses of model fungal species. Microb Genomics. 2019; 5(2):e000243. https://doi.org/10.1099/mgen.0.000243 PMID: 30714895
4. Badet T, Croll D. The rise and fall of genes: origins and functions of plant pathogen pangenomes. Curr Opin Plant Biol. 2020 Aug 1; 56:65–73. https://doi.org/10.1016/j.pbi.2020.04.009 PMID: 32480355
5. Freese J, Beyhan S. Genetic Diversity of Human Fungal Pathogens. Curr Clin Microbiol Rep. 2023 Jun 1; 10(2):17–28. https://doi.org/10.1007/s40588-023-00188-4 PMID: 37388463
6. Barber AE, Sae-Ong T, Kang K, Seelbinder B, Li J, Walther G, et al. Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection. Nat Microbiol. 2021 Dec; 6(12):1526–36. https://doi.org/10.1038/s41564-021-00993-x PMID: 34819642
7. Lofgren LA, Ross BS, Cramer RA, Stajich JE. The pan-genome of Aspergillus fumigatus provides a high-resolution view of its population structure revealing high levels of lineage-specific diversity driven by recombination. PLoS Biol. 2022 Nov 17; 20(11):e3001890. https://doi.org/10.1371/journal.pbio.3001890 PMID: 36395320
8. Gangurde SS, Korani W, Bajaj P, Wang H, Fountain JC, Agarwal G, et al. Aspergillus flavus pangenome (AflaPan) uncovers novel aflatoxin and secondary metabolite associated gene clusters. BMC Plant Biol. 2024 May 1; 24(1):354. https://doi.org/10.1186/s12870-024-04950-8 PMID: 38693487
9. van Westerhoven AC, Aguilera-Galvez C, Nakasato-Tagami G, Shi-Kunne X, Martinez de la Parte E, Chavarro-Carrero E, et al. Segmental duplications drive the evolution of accessory regions in a major crop pathogen. New Phytol. 2024; 242(2):610–625. https://doi.org/10.1111/nph.19604 PMID: 38402521
10. Plissonneau C, Hartmann FE, Croll D. Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 2018 Jan 11; 16(1):5. https://doi.org/10.1186/s12915-017-0457-4 PMID: 29325559
11. Zaccaron AZ, Stergiopoulos I. Analysis of five near-complete genome assemblies of the tomato pathogen Cladosporium fulvum uncovers additional accessory chromosomes and structural variations induced by transposable elements effecting the loss of avirulence genes. BMC Biol. 2024 Jan 29; 22(1):25. https://doi.org/10.1186/s12915-024-01818-z PMID: 38281938
12. Marcet-Houben M, Alvarado M, Ksiezopolska E, Saus E, de Groot PWJ, Gabaldón T. Chromosome-level assemblies from diverse clades reveal limited structural and gene content variation in the genome of Candida glabrata. BMC Biol. 2022 Oct 8; 20(1):226. https://doi.org/10.1186/s12915-022-01412-1 PMID: 36209154
13. Wyatt NA, Richards JK, Brueggeman RS, Friesen TL. A Comparative Genomic Analysis of the Barley Pathogen Pyrenophora teres f. teres Identifies Subtelomeric Regions as Drivers of Virulence. Mol Plant-Microbe Interactions. 2020 Feb; 33(2):173–88. https://doi.org/10.1094/MPMI-05-19-0128-R PMID: 31502507
14. Moolhuijzen PM, See PT, Shi G, Powell HR, Cockram J, Jørgensen LN, et al. A global pangenome for the wheat fungal pathogen Pyrenophora tritici-repentis and prediction of effector protein structural homology. Microb Genomics. 2022; 8(10):000872. https://doi.org/10.1099/mgen.0.000872 PMID: 36214662
15. Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. BMC Biol. 2020 Feb 11; 18(1):12. https://doi.org/10.1186/s12915-020-0744-3 PMID: 32046716

16. Peter J, De Chiara M, Friedrich A, Yue JX, Pflieger D, Bergström A, et al. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature. 2018 Apr; 556(7701):339–44. https://doi.org/10.1038/s41586-018-0030-5 PMID: 29643504

17. Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, et al. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics. 2009 Dec 15; 25 (24):3207–12. https://doi.org/10.1093/bioinformatics/btp579 PMID: 19808877

18. Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, et al. A draft human pangenome reference. Nature. 2023 May; 617(7960):312–24. https://doi.org/10.1038/s41586-023-05896-x PMID: 37165242

19. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, et al. Genetic and phenotypic intra-species variation in Candida albicans. Genome Res. 2015 Mar 1; 25(3):413–25. https://doi.org/10.1101/gr.174623.114 PMID: 25504520

20. Lind AL, Wisecaver JH, Lameiras C, Wiemann P, Palmer JM, Keller NP, et al. Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. PLoS Biol. 2017 Nov 17; 15(11): e2003583. https://doi.org/10.1371/journal.pbio.2003583 PMID: 29149178

21. Drott MT, Rush TA, Satterlee TR, Giannone RJ, Abraham PE, Greco C, et al. Microevolution in the pan-secondary metabolome of Aspergillus flavus and its potential macroevolutionary implications for filamentous fungi. Proc Natl Acad Sci U S A. 2021 May 25; 118(21):e2021683118. https://doi.org/10.1073/pnas.2021683118 PMID: 34016748

22. de Groot PWJ, Bader O, de Boer AD, Weig M, Chauhan N. Adhesins in Human Fungal Pathogens: Glue with Plenty of Stick. Eukaryot Cell. 2013 Mar 28; 12(4):470–81. https://doi.org/10.1128/EC.00364-12 PMID: 23397570

23. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010 Mar; 464(7287):367–73. https://doi.org/10.1038/nature08850 PMID: 20237561

24. Beavan A, Domingo-Sananes MR, McInerney JO. Contingency, repeatability, and predictability in the evolution of a prokaryotic pangenome. Proc Natl Acad Sci U S A. 2024 Jan 2; 121(1):e2304934120. https://doi.org/10.1073/pnas.2304934120 PMID: 38147560

25. Hu X, Xiao G, Zheng P, Shang Y, Su Y, Zhang X, et al. Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. Proc Natl Acad Sci U S A. 2014 Nov 25; 111(47):16796–801. https://doi.org/10.1073/pnas.1412662111 PMID: 25368161

26. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. The Ecology and Evolution of Pangenomes. Curr Biol. 2019 Oct 21; 29(20):R1094–103. https://doi.org/10.1016/j.cub.2019.08.012 PMID: 31639358

27. Dewar AE, Hao C, Belcher LJ, Ghoul M, West SA. Bacterial lifestyle shapes pangenomes. Proc Natl Acad Sci U S A. 2024 May 21; 121(21):e2320170121. https://doi.org/10.1073/pnas.2320170121 PMID: 38743630

28. Wong ED, Miyasato SR, Aleksander S, Karra K, Nash RS, Skrzypek MS, et al. Saccharomyces genome database update: server architecture, pan-genome nomenclature, and external resources. Genetics. 2023 May 2; 224(1):iyac191. https://doi.org/10.1093/genetics/iyac191 PMID: 36607068