The interkingdom horizontal gene transfer in 44 early diverging fungi boosted their metabolic, adaptive and immune capabilities

Michał Ciach*[1,3] ,Julia Pawłowska[2], Paweł Górecki[3] and Anna Muszewska*[1]

[1] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5A, 02-106 Warsaw, Poland

[2] Department of Molecular Phylogenetics and Evolution, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, ul. Żwirki i Wigury 101, 02-089 Warsaw, Poland;

[3] Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Stefana Banacha 2, 02-097 Warsaw, Poland

Running title: HGT in 44 early diverging fungi

Keywords: Early Diverging Fungi, Horizontal Gene Transfer, Fungal immunity, Ecological adaptation, Unrooted Gene Trees, HGT patterns, associated bacteria

Abstract

Numerous studies have been devoted to individual cases of horizontally acquired genes in fungi. It has been shown that such genes expand their metabolic capabilities and contribute to their adaptations as parasites or symbionts. Some studies have provided a more extensive characterization of the horizontal gene transfer (HGT) in Dikarya. However, in the early diverging fungi (EDF), the overall influence of HGT on the ecological adaptation and evolution is largely unknown. In order to fill this gap, we have designed a computational pipeline to obtain a sample of over 600 phylogenetic trees with evidence for recent to moderately old HGT across multiple EDF genomes ranging from *Chytridiomycota* and *Blastocladiomycota* to *Mucoromycota*. Our pipeline is designed to obtain a small sample of reliable HGT events with a possibly minimal number of false detections that distort the overall statistical patterns. We show that ancestrally aquatic fungi are generally more likely to acquire foreign genetic material than terrestrial ones. A close ecological relationship with another organism is a predisposing condition for HGT, as exemplified by a transfer of NOD-like receptors between Mortierellomycotina and their symbiotic bacteria, but does not always result in an extensive gene exchange.

Importance

Although it is now recognized that horizontal gene exchange is a factor influencing the adaptation and evolution of eukaryotic organisms, the so far described cases in early diverging fungi (EDF) are fragmentary, and a large-scale comprehensive study is lacking. We have designed a methodology to obtain a reliable, statistical sample of inter-kingdom xenologs across the tree of life of EDF to give a preliminary characterization of their general

properties and patterns. We study how different fungal lineages vary in terms of the number of xenologs, what are their ecological associations, and the molecular properties of proteins encoded by the acquired genes. Our results help to better understand to what extent and in what way the incorporation of foreign genetic material shaped the present metabolic, adaptive and immune capabilities of fungi.

Introduction

Horizontal gene transfer (HGT) is a process of exchange of genetic material between distantly related species. It has been extensively studied in prokaryotes and proved to be an important mechanism that increases the genomic diversity, contributes to niche adaptation, antibiotic resistance, new metabolic pathways and numerous other traits. In eukaryotes, the significance and frequency of HGT remain mostly unclear, and the mechanisms of gene transfer beyond *Agrobacterium*-mediated gene transformation are still a matter of debate[1]. It has been shown, however, that HGT does occur both within eukaryotes (including animals) and between eukaryotes and prokaryotes [2].

Fungi represent very diverse life strategies, making them a unique model to study the connections between HGT and ecological adaptations. Genomes of fungi are protected against the incorporation of foreign DNA at multiple levels: starting from the use of alternative genetic codes (e.g. in CTG *Candida albicans* and *Pachysolen tannophilus* [3]), through foreign DNA elimination mechanisms RNAi and RIP [4], which are ubiquitous in fungi [5], to the physical separation of chromatin from the rest of the cell by the nuclear envelope. On the other hand, although the genomes of fungi are protected against the incorporation of foreign DNA at multiple levels [3–5], the osmotrophic lifestyle [6] and hyphal fusion [7,8] could make the intake of foreign DNA more common. The complex interplay of factors facilitating and preventing the incorporation of foreign DNA places fungi somewhere between bacteria and animals in terms of the rate of HGT and contributes to large differences in the numbers of HGT between different fungal lineages.

Several horizontally transferred genes have documented adaptive roles for fungal fitness. A fungal pathogen of insects, *Metarhizium robertsii*, has acquired sterol carrier and peptidase genes that facilitate the penetration of the victim's body [9,10]. Multiple transfers of toxins in plant-infecting fungi have improved their ability to infect new hosts [11,12]. Bacteria have been described as a source of extensive HGT in plant-associated *Colletotrichum* sp. [13], *Verticillium dahliae* [14] and animal skin pathogens *Malassezia* [15]. Spectacular examples of transfers of clusters of functionally-linked genes [16,17], genes facilitating metabolism expansions [18] or even conditionally dispensable chromosomes in pathogenic *Fusarium* species [19] have influenced the perception of genomic evolution in fungi. They showed that HGT can provide a faster way to gain the ability to colonize new hosts and degrade new compounds than vertical inheritance.

Existing studies of fungal HGT have mostly focused on the Dikarya fungi [20,21], and the abundance of HGT in early diverging fungi (EDF, non-Dikarya fungi) is relatively understudied. However, it is already known that it does occur and may have a profound impact on the recipient organism. Recent reports show that EDF experienced several

documented waves of transfers, like the acquisition of secondary metabolite clusters in *Basidiobolus* [22], *Mortierella* [23] and *Rhizophagus* [24], or a number of enzymes transferred to *Neocallimastigomycota* involved in the metabolism of nucleotides [25] and carbohydrates [26]. A more complex scenario was reported for *Rhizophagus irregularis* that acquired genes both from the host plant and its endosymbiotic bacterium [27]. Catalase genes of bacterial origin, which presumably increase ROS resistance, were found in the amphibian pathogen *Batrachochytrium dendrobatidis* [28]. However, despite numerous individual examples, no comprehensive characterization of HGT in EDF has been presented so far.

In this work, we study the patterns of gene exchange between fungi and non-fungal organisms on a representative set of 44 EDFs from a taxonomic, ecological, and molecular perspective. We have designed a custom phylogenetic-based pipeline to obtain a reliable sample of xenologs, focusing on limiting the number of false positive results that distort the observed signal while retaining a sufficient sample size to infer statistical trends. The pipeline is designed after a comparison of multiple existing approaches and a theoretical and experimental identification of potential pitfalls that can result in false detections. This allows us to make a comparative study of HGT in diverse EDF lineages.

Methods

Dataset

Proteomes for 44 early diverging fungi (Supplementary Table S1) were downloaded from NCBI in April 2018 [29]. Fungal ecological tagging was obtained from Fungal Traits [30]. A taxonomical annotation was obtained from the NCBI taxonomy database, with manual fine-tuning of taxonomic names and ranks when needed according to [31]. Genomic features were obtained from GFF files of the genome assemblies. A species tree was inferred from whole proteomes with OMA2 based on OMA groups of orthologs [32].

HGT identification

Low-complexity regions in all proteins were masked using ncbi-seg [33] (window length 12, trigger complexity 1.8, extension complexity 2.0). Proteins with less than 30 non-masked amino acids, more than 2500 characters (including masked amino-acids), or more than 50% masking were discarded as prone to produce false-positive hits. Homologous families within the 44 proteomes were detected using blastp suite [29] with an e-value threshold 1e-05. Proteins with homologs in a single fungal family were retained for further analysis.

Homologous sequences for the retained proteins were found in the Non-Redundant NCBI database (last updated December 01, 2020) using blastp suite with an e-value threshold 1e-04 and maximum 500 target sequences. Detected homologs were masked using ncbi-seg (default settings). Excessively short, long, or masked (70% coverage threshold) proteins were discarded as before. Only homologs from bacteria, archaea, or eukarya were retained, in order to discard viral or artificial sequences. Taxonomic lineage of each homolog was inspected using the NCBI Taxonomy database (last updated May 17, 2021). Homologs with ’unclassified’, ‘uncultured’ or ‘environmental’ keywords anywhere in their lineage, as

well as non-fungal homologs with the '*incertae sedis'* keyword, were discarded as prone to produce false-positive hits due to uncertain taxonomy (note that some of the fungi analyzed in this paper belong to *incertae sedis* families; their homologs were retained despite the uncertain taxonomy).

Remaining fungal sequences and their homologs were clustered using psi-cd-hit in a stepwise clustering procedure. First all homologs were clustered with cd-hit 90% identity threshold (-c 0.9 -n 5 -aL 0.4), then cd-hit 60% identity threshold (-c 0.6 -n 4 -aL 0.4) and finally psi-cd-hit with 30% sequence identity threshold (-c 0.3 -aL 0.5 -ce 1e-5) [34]. Proteins located on contigs in which no other protein had any homolog in the 44 target fungi were discarded as potential genome contaminants (first stage contaminant filter). Clusters with more than 20 fungal sequences were discarded. Fungal sequences without at least 20% sequence identity on a region of at least 20 amino acids to any outgroup protein were discarded. Outgroup sequences without at least 20% identity to any fungal protein were discarded as well. In each cluster, up to 60 outgroup sequences most similar to fungal proteins were taken, and the remaining outgroup sequences were discarded. For each remaining fungal protein, up to 5 other proteins located on the same contig were taken and their homologs were searched for in the Non-Redundant NCBI database. Contigs which did not contain any protein with a fungal top blast hit were discarded as potential contaminants (second stage contaminant filter).
Remaining clusters were aligned using mafft 3.7 [35] with local iterative mode (localpair, maxiterate 1000). Alignments were trimmed using GBlocks allowing for half of gap characters in each column (-b5=h) in order to remove poorly aligned regions which may disturb the phylogenetic tree topology [36]. In each alignment, sequences without at least 10 remaining non-gap characters were discarded. . Constant columns were discarded as lacking phylogenetic information in order to speed up tree computations. Alignments with less than 20 columns, less than 4 sequences, without sequences from the 44 target proteomes, or without sequences from at least three non-fungal species were discarded. Maximum likelihood (ML) trees were inferred using IqTree [37] with automated model selection.

Trees in which fungi did not form a single clade were discarded. Corresponding species trees were downloaded from a local copy of the NCBI Taxonomy database (last updated August 8, 2021). In order to identify trees with evidence for a horizontal gene transfer, in each gene tree-species tree pair, the clades adjacent to the fungal clade were compared with respect to their species composition (see Supplementary Methods for a detailed description of the rationale and algorithms used). The fungal clade was defined to be *displaced* if the clades had differing sets of species in both trees (indicative of a horizontally acquired gene) and defined to be *repositionable* if the species in one of the neighboring clades in the gene tree formed a clade in the species tree as well (indicative of a properly resolved gene tree topology at the root node of the fungal clade). An example of a displaced and repositionable fungal clade is shown in Fig. **1**.

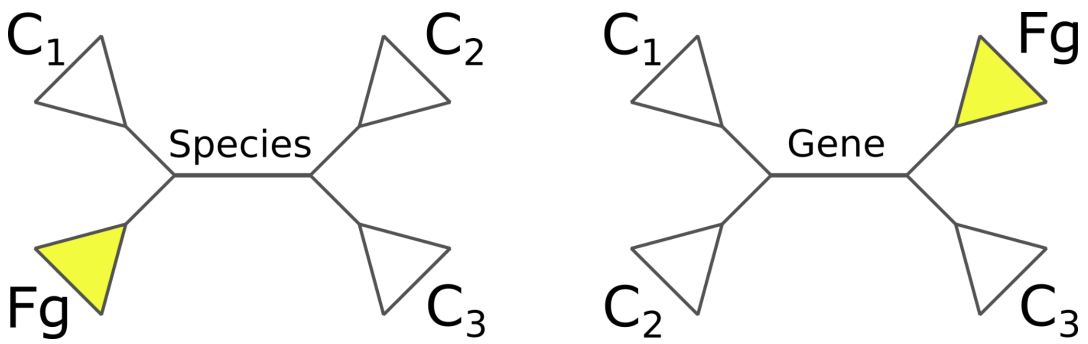

Fig. 1. An example of a displaced, repositionable fungal clade. Such clades are indicative of a horizontal gene transfer into the fungi (see the Supplementary Methods). Fg, fungal clade, C_1-C_3, neighboring clades. In this work, we define clades only by their species compositions, as clades may have different topologies in both trees due to gene tree inference errors.

Phylogenetic trees with a single fungal clade and at least two non-fungal species were classified into three separate groups of putative HGT events. First, if the branch joining the fungal clade with the rest of the tree was shorter than 0.001 substitution per site, the homologous family was considered as a recent transfer. Such families were not further analyzed in this study due to difficulties in determining the donor group of the transfer and the possibility of genome contamination, and were treated as partially supported cases of HGT. Second, if the fungal clade had the same location in the gene and the species tree, it was considered as another type of a possible, but only partially supported, case of HGT. Finally, in order to be considered a strongly supported case of HGT, the tree had to contain a single, displaced, respositionable fungal clade with SH-aLRT support of at least 85%, with the branch connecting the fungal clade to the outgroup shorter than two substitutions per site. A tree which did not satisfy any of those conditions was considered a third type of partially supported case of HGT.

Xenolog protein products were mapped against PFAM [38] using pfamscan.pl. Additionally, proteins were scanned with SignalP for signal peptides [39] and cellular localization with WolfPsort II [40]. Statistical and phylogenetic analyses were performed with the R statistical package (https://www.R-project.org/) and the Python 3 programming language using the Jupyter Notebook [41] and numpy [42], scipy [43] and ete3 [44] libraries. Plots were created using the matplotlib [45] library.

Results

We have detected 547 gene trees containing a strong indication of a horizontal gene transfer (**Fig.2**., see the Methods section and the **Supplementary Methods** for details). The protein clusters corresponding to those trees jointly contained 1203 fungal proteins, out of which 993 belonged to the 44 EDF analyzed in this study. We have additionally identified 1002 clusters with partial evidence for HGT, including 14 clusters with short branches connecting fungi to the outgroup (indicative of either a very recent transfer or an undetected contamination), and 572 clusters in which the fungal clade was not displaced with respect to the species tree (indicative of a vertical inheritance with extensive gene losses). Due to an incomplete evidence of HGT, these clusters are not analyzed in this paper, but their list is available in the **Supplementary Table S1**.

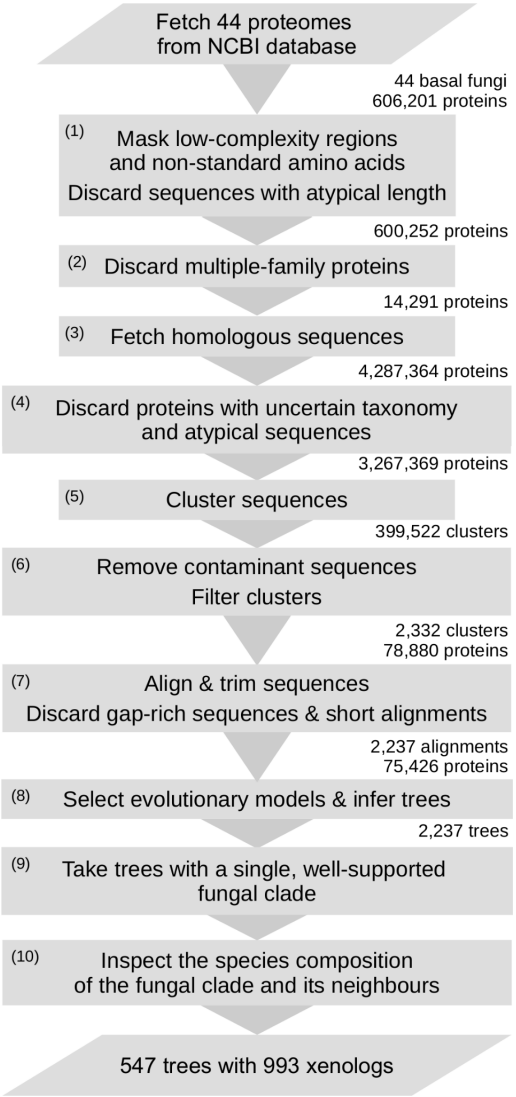

Fig. 2. A brief summary workflow of all of the steps in the HGT identification pipeline.

Over half of the trees contained a single fungal protein (**Fig. 3**), and the vast majority of xenologous parts of protein clusters have up to five proteins. However, the distribution of the numbers of xenologs in protein clusters closely follows a power-law distribution, characterized by infrequent occurrences of extremely large values. This suggests that protein clusters with much larger numbers of xenologs are likely to exist, but are rare and not present in our dataset due to our restrictive criteria. We did not observe a statistically significant correlation between the proteome size and the number of xenologs (ρ=0.21, p=0.17).

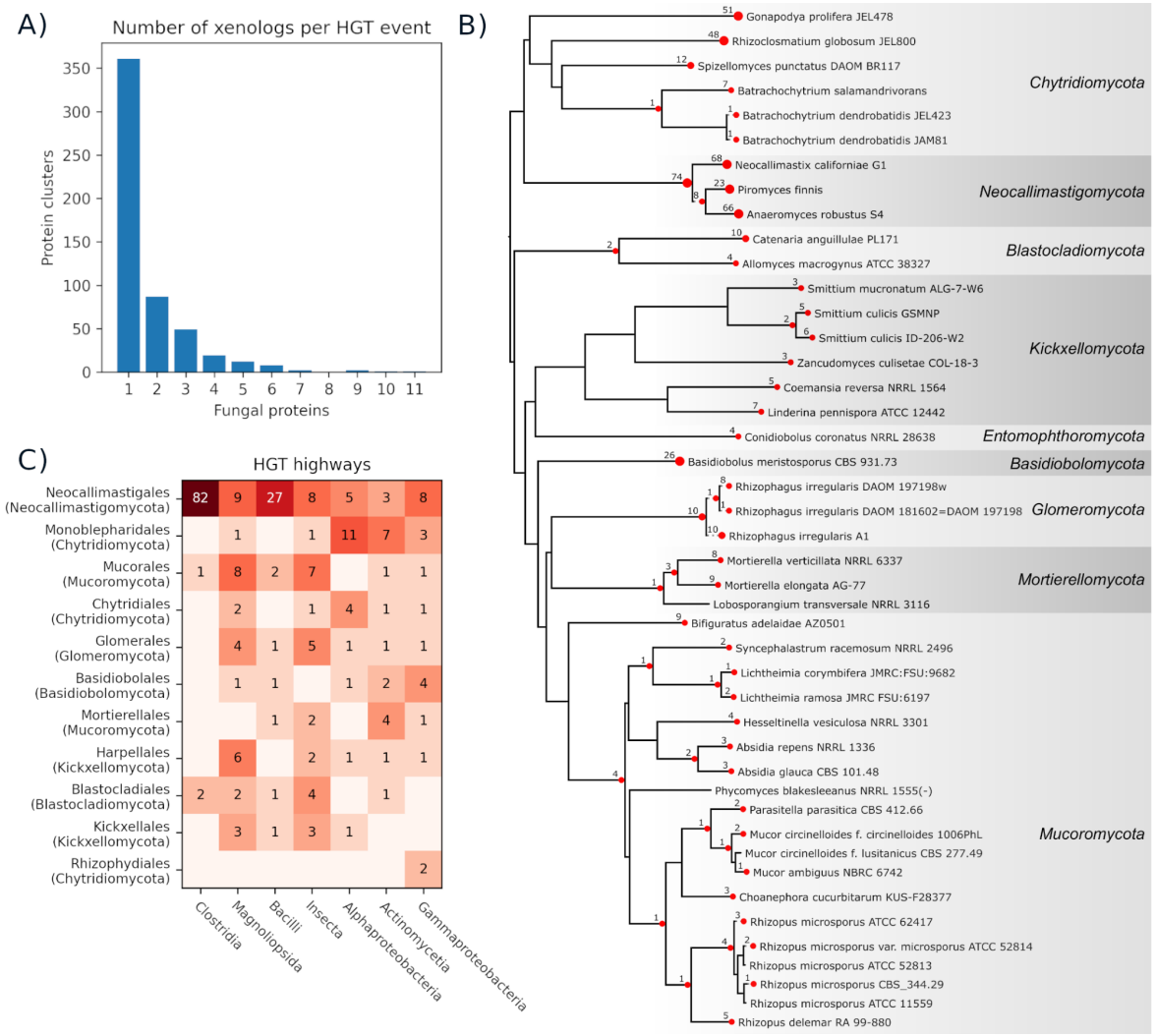

Fig. 3. The HGT patterns in 44 early diverging fungi. A) The number of fungal xenologs in protein clusters, restricted to 44 fungi analyzed in this study. B) The OMA2 phylogenetic tree of fungi annotated with the numbers of detected HGTs. C) The numbers of HGT into fungi per acceptor-donor pair. Taxonomic names based on NCBI Taxonomy with manual fine-tuning according to Voigt et al. [31].

The number of xenologs detected in each fungus is depicted in **Supplementary Fig. S1**. Symbiotic species of the gut-dwelling *Neocallimastigomycota*, the parasitic *Chytridiomycota* species and the *Basidiobolus* representative contained the most xenologs in their proteomes, followed by the plant-interacting *Glomeromycota* and the soil saprotrophic *Mortierellomycota.* . The number of xenologs was highly divergent across fungal species, with 6 fungi from those three groups accounting for almost 70% of all detected xenologs. As shown in the next sections, those xenologs are of both direct origin (when the given strain was the recipient of a transfer) and the ancestral one (when the gene was transferred to an ancestor of a strain), albeit with different proportions-mostly ancestral for *Neocallimastix californiae* G1, and mostly direct for chytrid species. Most of the transfers are placed shallowly in the tree, suggesting a recent occurrence of most HGT events and a later deletion of non-beneficial genes.

Transfer rates are highly divergent across the fungal tree of life. We have generated an OMA2 phylogenomic tree of the 44 EDFs and annotated its nodes with the numbers of HGTs they received, shown in **Fig. 3**. Note that the results are only an approximation of the true transfer acceptors, as the transfers can be more ancestral than the data suggests.

We have discovered a highly uneven distribution of HGTs across the fungi. There is a clear pattern dividing ancestrally aquatic taxa grouped in *Blastocladiomycota*, *Neocallimastigomycota* (living in the digestive tract of terrestrial herbivores) and *Chytridiomycota* with higher HGT rates compared to terrestrial fungi. The mainly saprotrophic *Mucoromycotina*, despite highest sampling in our dataset, shows the lowest number of HGT events per genome, just as the animal-related *Entomophthoromycota* and *Kickxellomycota* fungi. *Basidiobolus meristosporus* is a notable exception, with a higher share of xenologs compared to its closest relative *Conidiobolus coronatus*. Its massive HGT was recently described in the context of secondary metabolite clusters [22].

Transfer rates reflect physical proximity. To identify major HGT pathways, for each cluster we have matched the taxonomic order of its acceptor with the taxonomic class of the donor (**Fig. 3**, **Supplementary Fig. S2**). In the case of 59 clusters, the taxonomic rank of the LCA of the predicted donor group was higher than class. Those clusters were not included in the analysis of taxonomic HGT pathways.

Among the identified pathways, *Clostridia*, flowering plants, diverse *Proteobacteria*, *Actinobacteria* and insects are the main sources of xenologs to diverse fungal recipients. Our results confirm a major HGT event in the ancestor of *Neocallimastigomycota* that might have driven its evolution towards a gut-dwelling fungus [26]. This group of fungi is the only one that has obtained numerous transfers from the anaerobic *Clostridia*, which purportedly enabled them to adapt to the gut environment [26]. Moreover, *Neocallimastigomycota* obtained genes from a myriad of diverse bacteria groups, including *Bacteroidetes, Bacilli (Firmicutes), Slackia (Actinomycetota)* and *Aeromonas (Proteobacteria)* which may have contributed to their adaptive toolkit, a result that agrees with previous reports [48].
On the other hand, the representatives of *Bacilli* and *Burkholderiaceae* stand out as the main contaminants in our data set, (Supplementary Table S1). In *Mortierellomycota* species, where *Burkholderia*-related endobacteria (BRE) are common, HGT from these endosymbionts (i.e. *Mycoavidus cysteinexigens*) dominates. On the other hand, in the closely related *Glomeromycota* and *Mucoromycota* fungi, which do not have associated BRE, most xenologs come from loosely associated soil-inhabiting bacteria, such as *Actinomycetia, Bacilli,* and *Gammaproteobacteria*. In addition, we have detected xenologs from plant hosts in the endomycorrhizal *Glomeromycota*, but we did not detect any HGT from their mycoplasma endosymbionts (MRE), possibly due to the highly reduced genome of these bacteria which offers less possibilities for transfer. In the water-living *Chytridiomycota*, most xenologs come from loosely associated bacteria such as *Alphaproteobacteria, Gammaproteobacteria* and *Actinomycetia*.

Transfers from Eukarya are at least as common as from Bacteria. Bacteria, with 293 protein clusters, were the most represented donors, followed by 250 clusters from Eukarya and four clusters from Archaea (including a *Gonapodya prolifera* molybdopterin binding oxidoreductase KXS19169.1). We note, however, that our method may be biased towards transfers from distant donors, as discussed at the end of this article. In particular, our methodology does not detect gene exchange between the fungi. We have identified 46 Viridiplantae and 136 Metazoan clusters. Those numbers confirm a cross-kingdom transfer of genes between eukaryotes [49–51]. Interestingly, insects contributed uncharacterized genes to diverse fungal lineages, including both the parasitic *Basidiobolus* and mycorrhizal *Glomeromycota*.

Xenolog sequences evolve according to ordinary substitution models. For over 60% of protein clusters, iqTree selected the LG model as the best-fitting one, followed by WAG, JTT and VT models (**Supplementary Fig. S3**), which jointly accounted for nearly 90% of clusters. All of the four most common models are general models of amino acid substitutions suited for nuclear sequences. The less frequent models were nearly equally divided into chloroplast (3%), mitochondrial (3%), viral (2%) and other nuclear models (5%). This result suggests that xenolog evolution follows the same substitution patterns as most protein coding genes.

Low-complexity sequences are rarely transferred. The distribution of the proportions of protein sequences covered by low-complexity regions has a lighter right tail for xenologs than for the proteome background (**Fig. 4**). The results suggest that proteins with no low-complexity regions are transferred preferably, while extensive low-complexity regions tend to inhibit transfers. On the other hand, we found no difference in the distribution of sequence length of xenologs and background proteins.

Comparison of the average sequence lengths and low-complexity proportions of fungal xenologs and the donor group showed no visible trend (**Supplementary Fig. S4**). Therefore, the differences in low complexity seem to influence the likelihood of an HGT instead of being caused by it, while there seems to be no evolutionary pressure on sequence length either before or post transfer.

Xenologs are secreted more often than typical fungal proteins, but determining their precise subcellular location poses difficulties. TargetP 2.0 detected a signal peptide in approximately 18% of identified xenologs, compared to 7.7% for the background. This stands in agreement with the osmotrophic lifestyle of fungi and the massive fungal secretome. It is also consistent with the accessory function of many xenologs. We have detected only 19 xenologs with mitochondrial transfer peptides, accounting for 1.9% of xenologs, compared to 3.1% for the whole proteomes, consistent with the bacterial origin of many xenologs. We did not observe significant differences in numbers of transmembrane helices, predicted by TMHMM, between xenologs and background proteins.

The subcellular localization predicted by WolfPsort was consistent for 85% of clusters, compared to 95% for TargetP (**Fig. 4**, **Supplementary Fig. S5**). Compared to the proteome

background, we have noted a lower proportion of xenologs localized in the nucleus and the mitochondria, while a higher proportion localizes in the cytoplasm, the endoplasmic reticulum, and the extracellular space. In contrast to TargetP, WolfPSort predicts 137 proteins to localize in the mitochondria.

Interestingly, WolfPSort predicts the nucleus as the most common location of xenologs. This unexpected phenomenon could be caused by a bias of the method. To verify this hypothesis, we have generated a decoy dataset by randomly permuting each sequence from the 44 proteomes. We have found that the distribution of locations is very similar for the proteome background and the decoy dataset, with the nucleus as the most common location, and the decoy proteins attain similar prediction scores (**Supplementary Fig. 4**). This is likely caused by the fact that WolfPSort is based on the *k*-Nearest Neighbours classification algorithm, which, when faced with a new, randomly selected data point, tends to reflect the distribution of the underlying training dataset.
However, the location of transposon-derived proteins predicted by WolfPSort is, as expected, consistently nuclear. This suggests that the software classifies nuclear proteins correctly, but is biased towards nuclear/mitochondrial localization for proteins that are difficult to classify. Therefore, the actual proportions of cytosol and extracellular xenologs may be even higher.

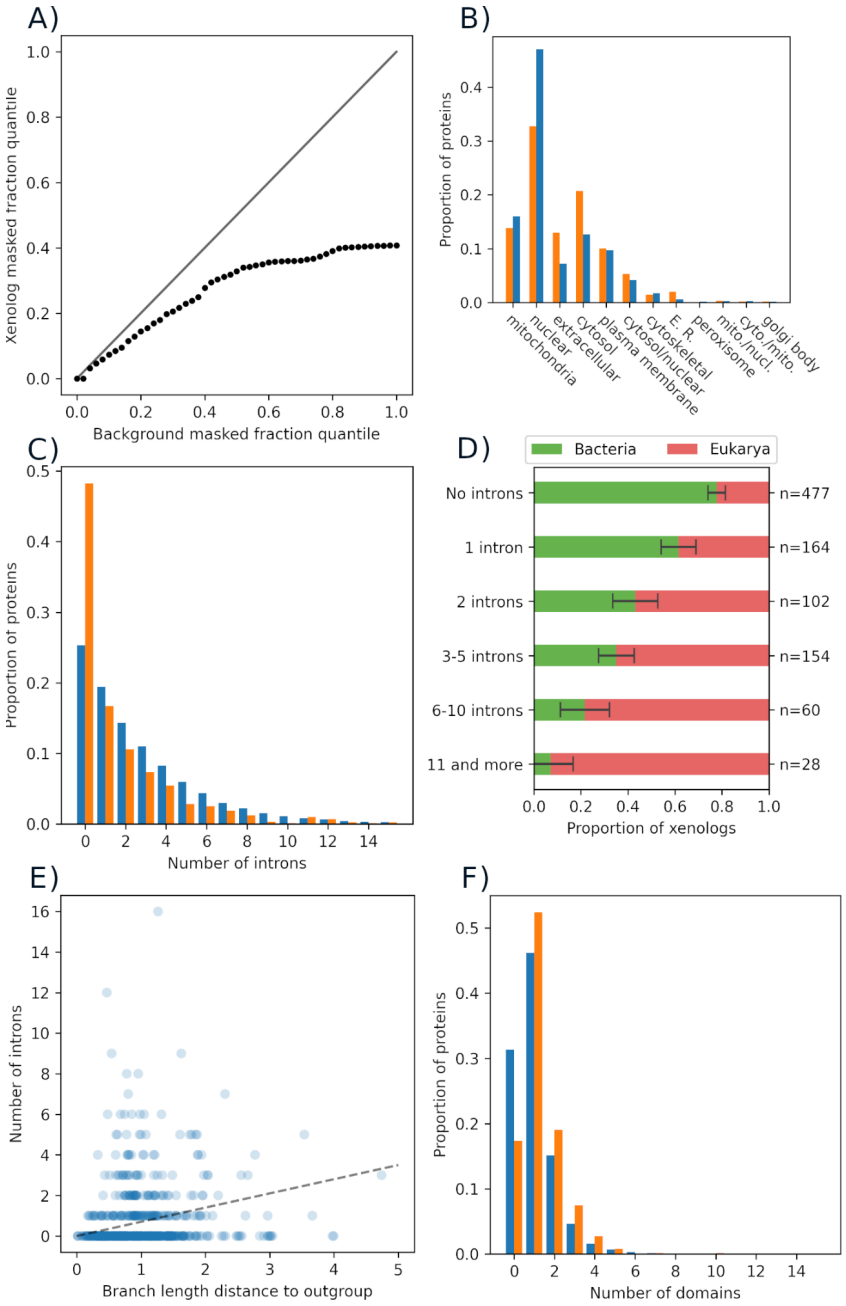

Fig. 4: Molecular characteristics of xenologs (orange) and background proteins (blue). A) A quantile plot comparing the sequence low-complexity fraction between the xenologs and the proteome background. B) Subcellular localizations. C) The number of introns. D) The dependence of the number of introns of a xenolog on the kingdom of the donor. E) The number of introns versus the average number of substitutions after transfer for xenologs with bacterial origin. The dashed line corresponds to a fitted linear model without intercept. F) The number of PFAM domains.

Xenologs typically contain few introns. Approximately 50% of xenologs did not contain any intron, and 75% contained at most two, compared to 25% and 59% for the background, respectively (**Fig. 4**). This result is in line with recommended HGT verification measures [13]. However, we have observed notable exceptions, especially in the case of *Gonapodya prolifera* and *Rhizoclosmatium globosum* (*Chytridiomycota*) which are particularly prone to

have 4 and more introns in their xenologs. These proteins come from both unicellular eukaryotic and prokaryotic donors. We have detected 28 xenologs, mostly of eukaryotic origin, with 11 or more introns (**Fig. 4**), including two *G. prolifera* xenologs: a phosphorylase with twelve introns transferred from green algae (KXS22044.1) and a copper dependent monooxygenase with eleven introns transferred from molluscs (KXS15796.1). In both cases, the donor homologs also contained multiple introns. Although there are chytrid species referred to as *intron-rich chytrids* [52], neither *G. prolifera* nor *R. globosum* have been described as such in their genomic papers [53,54].

The distribution of the number of introns in a xenolog seems to follow a zero-inflated geometric distribution, which is characteristic for a birth-and-death process with time-variable birth and death rates [55]. This suggests that xenologs acquire introns after transfer. Cases of such acquisition have been reported for genes transferred from bacteria [56]. In our data set, this also seems to be the case in two *R. globosum* xenologs, a *Parvularcula mediterranea* glutaminase with seven introns (ORY47769.1) and a deltaproteobacterium *Labilithrix luteola* magnesium transporter with twelve introns (ORY15655.1). In agreement with those observations, for xenologs with a bacterial origin, we have observed a positive correlation between the number introns and the average number of substitutions per site after transfer (measured as the total length of branches connecting a given xenolog to the outgroup), which is a measure of the evolutionary age of the transfer event. A linear model fitted without an intercept estimates 0.7 introns per unit branch length ($p < 0.001$). However, the overall picture is more complicated, as some xenologs quickly acquire multiple introns, while many others have none even after a long time. Together with the zero-inflated distribution of the numbers of introns, this suggests that some bacterial xenologs are particularly resistant to intronization.

Xenologs acquire domains post-transfer. Compared to the proteome background, fungal sequences obtained *via* HGT are more likely to map to one or more PFAM domains (**Fig. 4**). We have detected 520 xenologous sequences, accounting for over 50% of all detected xenologs, that map to a single domain, and 301 that map to at least two. However, we have found that not all domains of fungal xenologs originate from the gene transfer event. In order to inspect whether xenologs obtain new domains after transfer, in each cluster we have looked for domains present in at least one fungal sequence and which had no clan-level siblings in any non-fungal sequence in the cluster, referred to as *novel* domains. We have detected 80 different novel domains in 105 fungal proteins spread among 72 clusters, accounting for approximately 10% of xenologs and clusters respectively, indicating that addition of a domain after HGT is widespread. The most commonly added domain was the carbohydrate binding module CBM_10 (PF02013), a part of the dockerin cellulosome typical of anaerobic fungi, acquired post-transfer in 38 fungal proteins, followed by a typically eukaryotic Presenilin domain (PF01080), involved in calcium homeostasis in humans and extensively studied in the context of Alzheimer's disease (Pagnon de la Vega et al. 2022), acquired after transfer in 4 proteins. The latter domain was lost in Dikarya but occurs in animals, plants and eukaryotic microorganisms, including the EDF.

On the other hand, we have identified *forgotten domains*, defined as domains absent in fungi, but present (or having a clan-level sibling) in at least 20% of sequences in both clades

neighboring with the fungal one in the gene tree. Such domains can be assumed to be ancestral for the whole protein cluster, but not inherited in a transfer. We have identified 97 such domains in 66 clusters which jointly contained 112 xenologs.
Both results indicate that caution should be taken when analyzing domains of transferred proteins to elucidate how HGT has shaped the evolution of organisms. A domain found in a xenolog can be acquired after a HGT, rather than being obtained in the process, and, since HGT can lead to acquisition of gene fragments, an acquired gene can inherit only selected functions of the original gene. In total, we have found 203 proteins that had a novel or a forgotten domain (or, in the case of 14 xenologs, both). This observation indicates that approximately 20% of homologs can perform different functions in the recipient of the gene than in the donor. Notably, in our data set, cellulosome components CBM_10 were found in more xenologs than any other domain; in all cases, however, this domain was acquired post-transfer.

Metabolic, DNA binding, and transposon domains constitute the majority of xenologs. To analyze HGT-derived protein domains of fungal xenologs, for each cluster we have selected domains which were present in at least a half of fungal proteins and in at least 20% of proteins in one of its neighboring clades.

We have found numerous transfers of genes coding abhydrolases, acetyltransferases and esterases. Notably, many carbohydrate active enzymes, such as glycohydrolases and glycosyltransferases, were transferred to *Neocallimastigomycota*. On the other hand, peptidases were underrepresented in the dataset (peptidase xenologs n=36) compared to the carbohydrate active enzymes and carbohydrate binding modules (n=70), but were transferred to diverse recipients, including metallopeptidases transferred from *Thiofilum flexile* to *Basidiobolus meristosporus* (ORY01793.1) and from *Stegodyphus* to *Mortierella verticillata* (KFH70486.1)

Xenologs with DNA binding modules enriched the genomes of *Neocallimastix californiae, Rhizoclosmatium globosum*, *Rhizophagus* spp. and *Basidiobolus meristosporus.* We have also identified several phage and bacterial proteins without clear function in fungi (eg. Phage tail lysozyme, PF18013, in *Neocallimastix californiae* protein ORY21496.1). Domains of unknown function (DUFs/UPFs) constitute a numerous functional class with 35 diverse Pfam families found in 56 xenologs.

In line with the reports on ubiquitous horizontal transposon transfers (HTTs), we have detected several transfers (n=39) of different classes of transposases to *Smittium culicis* (OMJ28469.1, OMJ25253.1, Tigger) from Ecdysozoa, from *Sorghum bicolor* to *Lichtheimia ramosa* (CDS10934.1, MuLE) and from *Folsomia candida* to *Mucorales* (eg. *Parasitella parasitica* CEP16237.1, helitron). We have also detected a transfer of a Ty3/Gypsy LTR retrotransposon from plants to *L. corymbifera* (CDH59124.1). We have also observed transfers of homing endonucleases with a GIY-YIG domain from protists and bacteria to *R. irregularis* (PKC55335.1). The inteins found in *R. irregularis* seem to be distantly related to remaining fungal inteins, pointing to the mobile character of inteins in general.

We have detected several toxins, genes involved in resistance and secondary metabolism

classes of genes regarded as prone to HGT. Putative betalactamase resistance genes were transferred from *Clostidia* to different *Neocallimastigomycota* (eg. *Anaeromyces robustus* protein ORX75694.1), MFS_1 and ABC transporters to *Neocallimastigomycota* (eg. ORY49205.1). *Basidiobolus meristosporus* gained a potential resistance kinase from Dictyostelids (ORY05370.1). Several *Neocallimastigomycota* representatives gained ricin type sugar binding modules (eg. ORX83988.1). Similar proteins were recently documented in *Mucor* acting as toxins (mucoricin) [57]. Condensation domain (PF00668) containing candidate NRPS secondary metabolite clusters were transferred from *Glomeribacter* sp. to *Basidiobolus meristosporus* (ORY00193.1) and from *Listeria seeligeri* to several *Neocallimastigomycota* (e.g. *Anaeromyces robustus*, ORX81431.1). Adenylate forming (AMP-binding, PF00501) modules were also transferred to *Neocallimastigomycota*, potentially contributing to secondary metabolite clusters (Anaeromyces robustus, ORX46255.1). *Basidibiobolus* was recently reported as an acceptor of several transfers of metabolite clusters [22]. Four tocopherol cyclase genes with a PF14249 domain have been transferred from a photosynthetic bacterium *Rhodopila globiformis* to *Gonapodya prolifera* (KXS11127.1). Vitamin E related compounds are predominantly produced by plants and bacteria.

Mortierellomycota acquired the NOD-like receptors with NACHT-PPR-WD40 architecture from *Mycoavidus cysteinexigens* endosymbionts). These proteins were shown to play a role in fungal immunity [58,59]. Interestingly, all of the *Mortierellomycota* immunity proteins seem to be of a bacterial origin or at least shared with the endosymbiont. *Mycoavidus* homologs have ~50-45% of sequence identity and are the best reciprocal blast hits. It is an open question whether these proteins have a role in the symbiotic interaction.

Moreover, *Mortierellomycota* acquired fatty acid desaturases similar to delta-12 fatty acid dehydrogenases of bacterial origin (e.g. KFH67875.1) probably from *Deltaproteobacteria.* The representatives of this fungal group are known to produce long polyunsaturated fatty acids.

Discussion

Our aim in this work was to provide a general, statistical characterization of the properties of interkingdom HGT across diverse EDF. To design a pipeline appropriate for the context of this work, we have compared multiple existing approaches with respect to their applicability to a high-throughput screening for xenologs. We have then analyzed their key steps both theoretically and experimentally, by analyzing the results of each step and checking how often it produces false discoveries. We have concluded that multiple factors may inflate the false positive rate when HGT detection is done in a high-throughput setting, including:

- Relying on sequence similarity instead of phylogenetic trees,
- Excessively long or short sequences (motivating step 1 in Fig. 2),
- Viral and environmental homologs (motivating step 4),
- Choice of the clustering algorithm,
- Low-quality alignments (motivating step 7) and a wrong choice of the trimming software,
- Methods relying on rooted gene trees,
- Using reconciliation algorithms with improper event weights,
- Complex gene trees with multiple fungal clades (motivating steps 2 and 9 in Fig. 2).

How and why each of the above steps influences the false positive rate is described and discussed in detail in the **Supplementary Methods**. In order to avoid them, we have developed a custom-made algorithm for detection of gene trees with an unexpected location of the fungal clade, which allows us to detect HGT events and identify gene donors in unrooted trees. Due to the application of multiple filtering criteria and avoidance of risky steps in the pipeline, our data should provide a reliable source of fungal xenologs which can further be used to benchmark new HGT identification methods (**Supplementary Table S1**).

Our method has provided us with a dataset of xenologs that allows us to draw general conclusions about the nature of the moderately recent HGT into EDF. However, the requirement of a single fungal clade in the gene tree means that our methodology mostly detects transfers from distantly related donors. In particular, it does not detect transfers between fungal species, such as the massive transfer from a fungal parasite *Parasitella parasitica* to its host *Absidia glauca*, and may also have a limited sensitivity towards transfers of animal genes, many of which have a fungal homolog. Studies of gene exchange between closely related organisms most likely require a different, specialized approach.

We have observed some clear patterns in the properties of protein products of transferred genes. Xenologs are characterized by relatively short low-complexity regions. Regions with skewed amino acid abundance constitute 20% of eukaryotic proteins [66] and are involved in diverse molecular functions usually linked to regulation [67] and protein-protein interactions [68]. The depletion of low-complexity regions can be related to the so-called *complexity hypothesis*, which states that xenologs tend to be far from central hubs of interaction networks [69], and instead participate in secondary metabolism, drug resistance etc.

In agreement with the current notions, the xenologs in our dataset tend to contain few introns. We have, however, detected notable exceptions to this rule that may result from the acquisition of intron-rich genes or intronization of xenologs in chytrids.

We have found no correlation between the number of xenologs and proteome size. This implies that HGT seems to have no influence on the proteome sizes of EDF. Moreover, large proteomes do not predispose to HGT and the overall amount of genetic material does not seem to be a limiting factor for incorporation of transferred genes. The number of xenologs in a proteome does not seem to be subject to evolutionary selection even in the case of parasites with a reduced genome.

The receptiveness for foreign genes varies greatly between fungal lineages, with bacteria being the most common source of xenologs. Some lineages of EDF have associated bacteria, either endohyphal or acquired from the environment, which have been pointed as likely donors of xenologs [61]. There are several documented cases where the prokaryotic symbionts are co-evolving with the fungal host and transmitted vertically [62]. Such close endosymbiotic relations pose a natural opportunity for a large-scale gene exchange, with the prominent example of a massive gene transfer from the mitochondrion to the nucleus of eukaryotic organisms [63]. This pattern was confirmed in this study, with a notable example of *Mortierella verticillata* acquiring fatty acid desaturases probably from *Deltaproteobacteria* and NOD-like receptors from endohyphal *Mycoavidus cysteinexigens.* The latter suggest a parallel evolution of NOD-like receptors and perhaps a NOD-like receptor based immunity in

Dikarya and *Mortierellomycotina*. Moreover, NOD-like receptors may be involved in the host-bacteria interaction. *T*he transfers of an ABC transporter, a betalactamase, and a bacteriocin resistance cysteine peptidase (C39) from species of *Bacilli* to *Batrachochytrum* also suggest the role of HGT in fungal resistance.

Our results also confirm the reports on a massive enzyme encoding gene acquisition by *Neocallimastigomycota* [26] which has likely played a role in their transition to gut-dwelling mode of life.

In agreement with previous reports, we have found that fungi belonging to ancestrally aquatic groups experienced more HGT events than terrestrial lineages. This can be explained by the easier exchange of genetic material in the aquatic ecosystem, the ubiquity of unicellular aquatic life forms, and the possibility of forming biofilms with bacteria [64–66]. Another possible explanation is that terrestrial fungi may have lost some of the HGT predisposing traits or gained defensive mechanisms preventing foreign gene incorporation.

Despite the impact of the aquatic environment, the relationship with the host and the necessity to adapt to a new environment seem to have a higher influence on the rate of HGT. Within fungi with at least 10 xenologs, most have a documented long-term association with another organism. On the other hand, many mycorrhizal fungi did not experience extensive transfers These observations suggest that, while symbiosis or parasitism does not always result in gene exchange, it is a predisposing condition.

Acknowledgements

This work was supported by National Science Centre grants (#2017/25/B/NZ2/01880 to AM, #2019/33/B/ST6/00737 to Paweł Górecki supporting M.C., #2017/25/B/NZ8/00473 to J.P.)

Author Contributions

M.C and A.M. designed the study, M.C. implemented software, M.C. and A.M. prepared the dataset and performed genome analyses, M.C. performed statistical analyses, and M.C., J.P. and A.M. interpreted the data and wrote the manuscript.

Data Availability

Phylogenetic trees of supported and unsupported HGT are deposited under 10.5281/zenodo.6365634, all accessions and assemblies are listed in Supplementary Table S1.

References

1. Naranjo-Ortiz MA, Gabaldón T. Fungal evolution: cellular, genomic and metabolic complexity. Biol Rev Camb Philos Soc. 2020;95: 1198–1232.

2. Gabaldón T. Patterns and impacts of nonvertical evolution in eukaryotes: a paradigm

shift. Ann N Y Acad Sci. 2020;1476: 78–92.

3. Krassowski T, Coughlan AY, Shen X-X, Zhou X, Kominek J, Opulente DA, et al. Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. Nat Commun. 2018;9: 1887.

4. Clutterbuck AJ, John Clutterbuck A. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. Fungal Genetics and Biology. 2011. pp. 306–326. doi:10.1016/j.fgb.2010.09.002

5. Torres-Martínez S, Ruiz-Vázquez RM. The RNAi Universe in Fungi: A Varied Landscape of Small RNAs and Biological Functions. Annu Rev Microbiol. 2017;71: 371–391.

6. Richards TA, Talbot NJ. Horizontal gene transfer in osmotrophs: playing with public goods. Nature Reviews Microbiology. 2013. pp. 720–727. doi:10.1038/nrmicro3108

7. Priest SJ, Yadav V, Heitman J. Advances in understanding the evolution of fungal genome architecture. F1000Res. 2020;9. doi:10.12688/f1000research.25424.1

8. Feurtey A, Stukenbrock EH. Interspecific Gene Exchange as a Driver of Adaptive Evolution in Fungi. Annu Rev Microbiol. 2018;72: 377–398.

9. Zhao H, Xu C, Lu H-L, Chen X, St. Leger RJ, Fang W. Host-to-Pathogen Gene Transfer Facilitated Infection of Insects by a Pathogenic Fungus. PLoS Pathogens. 2014. p. e1004009. doi:10.1371/journal.ppat.1004009

10. Zhang Q, Chen X, Xu C, Zhao H, Zhang X, Zeng G, et al. Horizontal gene transfer allowed the emergence of broad host range entomopathogens. Proc Natl Acad Sci U S A. 2019;116: 7982–7989.

11. Mehrabi R, Bahkali AH, Abd-Elsalam KA, Moslem M, Ben M'Barek S, Gohari AM, et al. Horizontal gene and chromosome transfer in plant pathogenic fungi affecting host range. FEMS Microbiology Reviews. 2011. pp. 542–554. doi:10.1111/j.1574-6976.2010.00263.x

12. Ambrose KV, Koppenhöfer AM, Belanger FC. Horizontal gene transfer of a bacterial insect toxin gene into the Epichloë fungal symbionts of grasses. Sci Rep. 2014;4: 5562.

13. Jaramillo VDA, Armijos Jaramillo VD, Sukno SA, Thon MR. Identification of horizontally transferred genes in the genus Colletotrichum reveals a steady tempo of bacterial to fungal gene transfer. BMC Genomics. 2015. doi:10.1186/1471-2164-16-2

14. Shi-Kunne X, van Kooten M, Depotter JRL, Thomma BPHJ, Seidl MF. The Genome of the Fungal Pathogen Verticillium dahliae Reveals Extensive Bacterial to Fungal Gene Transfer. Genome Biol Evol. 2019;11: 855–868.

15. Ianiri G, Coelho MA, Ruchti F, Sparber F, McMahon TJ, Fu C, et al. HGT in the human and skin commensalMalassezia: A bacterially derived flavohemoglobin is required for NO resistance and host interaction. Proceedings of the National Academy of Sciences. 2020. pp. 15884–15894. doi:10.1073/pnas.2003473117

16. Wisecaver JH, Rokas A. Fungal metabolic gene clustersâ€"caravans traveling across genomes and environments. Frontiers in Microbiology. 2015. doi:10.3389/fmicb.2015.00161

17. Wisecaver JH, Slot JC, Rokas A. The evolution of fungal metabolic pathways. PLoS Genet. 2014;10: e1004816.

18. Gao S, Gold SE, Wisecaver JH, Zhang Y, Guo L, Ma L-J, et al. Genome-wide analysis of Fusarium verticillioides reveals inter-kingdom contribution of horizontal gene transfer to the expansion of metabolism. Fungal Genet Biol. 2019;128: 60–73.

19. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010. pp. 367–373. doi:10.1038/nature08850

20. Richards TA, Leonard G, Soanes DM, Talbot NJ. Gene transfer into the fungi. Fungal Biology Reviews. 2011. pp. 98–110. doi:10.1016/j.fbr.2011.04.003

21. Marcet-Houben M, Gabaldón T. Acquisition of prokaryotic genes by fungal genomes. Trends Genet. 2010;26: 5–8.

22. Tabima JF, Trautman IA, Chang Y, Wang Y, Mondo S, Kuo A, et al. Phylogenomic Analyses of Non-Dikarya Fungi Supports Horizontal Gene Transfer Driving Diversification of Secondary Metabolism in the Amphibian Gastrointestinal Symbiont,. G3 . 2020;10: 3417–3433.

23. Wurlitzer JM, Stanišić A, Wasmuth I, Jungmann S, Fischer D, Kries H, et al. Bacterial-Like Nonribosomal Peptide Synthetases Produce Cyclopeptides in the Zygomycetous Fungus Mortierella alpina. Applied and Environmental Microbiology. 2021. doi:10.1128/aem.02051-20

24. Venice F, Desirò A, Silva G, Salvioli A, Bonfante P. The Mosaic Architecture of NRPS-PKS in the Arbuscular Mycorrhizal Fungus Gigaspora margarita Shows a Domain With Bacterial Signature. Frontiers in Microbiology. 2020. doi:10.3389/fmicb.2020.581313

25. Alexander WG, Wisecaver JH, Rokas A, Hittinger CT. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. Proc Natl Acad Sci U S A. 2016;113: 4116–4121.

26. Murphy CL, Youssef NH, Hanafy RA, Couger MB, Stajich JE, Wang Y, et al. Horizontal Gene Transfer as an Indispensable Driver for Evolution of Neocallimastigomycota into a Distinct Gut-Dwelling Fungal Lineage. Appl Environ Microbiol. 2019;85. doi:10.1128/AEM.00988-19

27. Li M, Zhao J, Tang N, Sun H, Huang J. Horizontal Gene Transfer From Bacteria and Plants to the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis. Frontiers in Plant Science. 2018. doi:10.3389/fpls.2018.00701

28. Hansberg W, Salas-Lizana R, Domínguez L. Fungal catalases: function, phylogenetic origin and structure. Arch Biochem Biophys. 2012;525: 170–180.

29. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2017;45: D12–D17.

30. Põlme S, Abarenkov K, Henrik Nilsson R, Lindahl BD, Clemmensen KE, Kauserud H, et al. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles.

Fungal Diversity. 2020. pp. 1–16. doi:10.1007/s13225-020-00466-2

31. Voigt K, James TY, Kirk PM, Santiago ALCM de A, Waldman B, Griffith GW, et al. Early-diverging fungal phyla: taxonomy, species concept, ecology, distribution, anthropogenic impact, and novel phylogenetic proposals. Fungal Divers. 2021; 1–40.

32. Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Warwick Vesztrocy A, Dalquen DA, et al. OMA standalone: orthology inference among public and custom genomes and transcriptomes. Genome Res. 2019;29: 1152–1163.

33. Wootton JC, Federhen S. Statistics of local complexity in amino acid sequences and sequence databases. Computers & Chemistry. 1993. pp. 149–163. doi:10.1016/0097-8485(93)85006-x

34. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28: 3150–3152.

35. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics. 2018. pp. 2490–2492. doi:10.1093/bioinformatics/bty121

36. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17: 540–552.

37. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37: 1530–1534.

38. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Research. 2019. pp. D427–D432. doi:10.1093/nar/gky995

39. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology. 2019. pp. 420–423. doi:10.1038/s41587-019-0036-z

40. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35: W585–7.

41. Beg M, Taka J, Kluyver T, Konovalov A, Ragan-Kelley M, Thiery NM, et al. Using Jupyter for Reproducible Scientific Workflows. Computing in Science & Engineering. 2021. pp. 36–46. doi:10.1109/mcse.2021.3052101

42. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585: 357–362.

43. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020. pp. 261–272. doi:10.1038/s41592-019-0686-2

44. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol. 2016;33: 1635–1638.

45. Bisong E. Matplotlib and Seaborn. Building Machine Learning and Deep Learning Models on Google Cloud Platform. 2019. pp. 151–165.

doi:10.1007/978-1-4842-4470-8_12

46. Taktek S, Trépanier M, Servin PM, St-Arnaud M, Piché Y, Fortin J-A, et al. Trapping of phosphate solubilizing bacteria on hyphae of the arbuscular mycorrhizal fungus Rhizophagus irregularis DAOM 197198. Soil Biology and Biochemistry. 2015. pp. 1–9. doi:10.1016/j.soilbio.2015.07.016

47. Chen ECH, Morin E, Beaudet D, Noel J, Yildirir G, Ndikumana S, et al. High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytol. 2018;220: 1161–1171.

48. Alexandre FS, Della Flora LV, Henrique IG, da Silva DC, Mercedes AP, Silva AC, et al. Arbuscular Mycorrhizal Fungi (Rhizophagus clarus) and Rhizobacteria (Bacillus subtilis) Can Improve the Clonal Propagation and Development of Teak for Commercial Plantings. Frontiers in Plant Science. 2021. doi:10.3389/fpls.2021.628769

49. Schönknecht G, Weber APM, Lercher MJ. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. Bioessays. 2014;36: 9–20.

50. Van Etten J, Bhattacharya D. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? Trends in Genetics. 2020. pp. 915–925. doi:10.1016/j.tig.2020.08.006

51. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell. 2009;21: 1897–1911.

52. van de Vossenberg BTLH, Warris S, Nguyen HDT, van Gent-Pelzer MPE, Joly DL, van de Geest HC, et al. Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of Synchytrium endobioticum. Sci Rep. 2019;9: 8672.

53. Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, et al. Phylogenomic Analyses Indicate that Early Fungi Evolved Digesting Cell Walls of Algal Ancestors of Land Plants. Genome Biol Evol. 2015;7: 1590–1601.

54. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, et al. Widespread adenine N6-methylation of active genes in fungi. Nat Genet. 2017;49: 964–968.

55. Kendall DG. On the Generalized “Birth-and-Death” Process. Ann Math Stat. 1948;19: 1–15.

56. Lage J-LD, Da Lage J-L, Binder M, Hua-Van A, Janeček Š, Casane D. Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial α-amylase gene to Basidiomycetes. BMC Evolutionary Biology. 2013. p. 40. doi:10.1186/1471-2148-13-40

57. Soliman SSM, Baldin C, Gu Y, Singh S, Gebremariam T, Swidergall M, et al. Mucoricin is a ricin-like toxin that is critical for the pathogenesis of mucormycosis. Nature Microbiology. 2021. pp. 313–326. doi:10.1038/s41564-020-00837-0

58. Uehling J, Deveau A, Paoletti M. Do fungi have an innate immune response? An NLR-based comparison to plant and animal immune systems. PLoS Pathog. 2017;13: e1006578.

59. Dyrka W, Lamacchia M, Durrens P, Kobe B, Daskalov A, Paoletti M, et al. Diversity and

variability of NOD-like receptors in fungi. Genome Biol Evol. 2014;6: 3137–3158.

60. Carter ME, Carpenter SCD, Dubrow ZE, Sabol MR, Rinaldi FC, Lastovetsky OA, et al. A TAL effector-like protein of an endofungal bacterium increases the stress tolerance and alters the transcriptome of the host. Proceedings of the National Academy of Sciences. 2020. pp. 17122–17129. doi:10.1073/pnas.2003857117

61. Naito M, Morton JB, Pawlowska TE. Minimal genomes of mycoplasma-related endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota. Proc Natl Acad Sci U S A. 2015;112: 7791–7796.

62. Pawlowska TE, Gaspar ML, Lastovetsky OA, Mondo SJ, Real-Ramirez I, Shakya E, et al. Biology of Fungi and Their Bacterial Endosymbionts. Annu Rev Phytopathol. 2018;56: 289–309.

63. Anselmetti Y, El-Mabrouk N, Lafond M, Ouangraoua A. Gene tree and species tree reconciliation with endosymbiotic gene transfer. Bioinformatics. 2021. pp. i120–i132. doi:10.1093/bioinformatics/btab328

64. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. High frequency of horizontal gene transfer in the oceans. Science. 2010;330: 50.

65. Grüll MP, Mulligan ME, Lang AS. Small extracellular particles with big potential for horizontal gene transfer: membrane vesicles and gene transfer agents. FEMS Microbiology Letters. 2018. doi:10.1093/femsle/fny192

66. Abe K, Nomura N, Suzuki S. Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. FEMS Microbiol Ecol. 2020;96. doi:10.1093/femsec/fiaa031

67. Mier P, Paladin L, Tamana S, Petrosian S, Hajdu-Soltész B, Urbanek A, et al. Disentangling the complexity of low complexity proteins. Brief Bioinform. 2020;21: 458–472.

68. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16: 18–29.

69. Coletta A, Pinney JW, Weiss Solís DY, Marsh J, Pettifer SR, Attwood TK. Low-complexity regions within protein sequences have position-dependent roles. BMC Systems Biology. 2010. doi:10.1186/1752-0509-4-43

70. Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol Biol Evol. 2011;28: 1481–1489.