Cross kingdom analysis of putative quadruplex-forming sequences in fungal genomes: novel antifungal targets to ameliorate fungal pathogenicity?

Emily F. Warner,[1] Natália Bohálová,[2,3] Václav Brázda,[2] Zoë A. E. Waller,[4] and **Stefan Bidula**[5*]

[1] School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

[2] Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic

[3] Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czech Republic

[4] UCL School of Pharmacy, 29-39 Brunswick Square, London, United Kingdom

[5] School of Biological Sciences, University of East Anglia, Norwich, United Kingdom

***Correspondence to**: Stefan Bidula (s.bidula@uea.ac.uk)

Keywords: Fungi, G-quadruplexes, i-motifs, *in-silico*, virulence, drug resistance

Running title: Quadruplexes in fungi

Manuscript length: 5727 words

Number of Figures:10

Number of Tables:3

Abstract

Fungi contribute to upwards of 1.5 million human deaths annually, are involved in the spoilage of up to a third of food crops, and have a devastating effect on plant and animal biodiversity. Moreover, this already significant issue is exacerbated by a rise in antifungal resistance and a critical requirement for novel drug targets. Quadruplexes are four-stranded secondary structures in nucleic acids which can regulate processes such as transcription, translation, replication, and recombination. They are also found in genes linked to virulence in microbes, and quadruplex-binding ligands have been demonstrated to eliminate drug resistant pathogens. Using a computational approach, we identified putative quadruplex-forming sequences (PQS) in 1362 genomes across the fungal kingdom and explored their potential involvement in virulence, drug resistance, and pathogenicity. Here we present the largest analysis of PQS in fungi and identified significant heterogeneity of these sequences throughout phyla, genera, and species. Moreover, PQS were genetically conserved. Notably, loss of PQS in cryptococci and aspergilli was associated with pathogenicity. PQS in the clinically important pathogens *Aspergillus fumigatus, Cryptococcus neoformans,* and *Candida albicans* were located within genes (particularly coding regions), mRNA, repeat regions, mobile elements, tRNA, ncRNA, rRNA, and the centromere. Genes containing PQS in these organisms were found to be primarily associated with metabolism, nucleic acid binding, transporter activity, and protein modification. Finally, PQS were found in over 100 genes associated with virulence, drug resistance, or key biological processes in these pathogenic fungi and were found in genes which were highly upregulated during germination, hypoxia, oxidative stress, iron limitation, and in biofilms. Taken together, quadruplexes in fungi could present interesting novel targets to ameliorate fungal virulence and overcome drug resistance.

Introduction

Compared to viruses and bacteria, fungi are underappreciated [1]. They are key contributors to the food and drink, biotechnology, and textile industries, whilst also being an important source of novel antimicrobial compounds [2-5]. However, >1.5 million deaths per year are attributed to fungi in humans globally, more than malaria and on par with tuberculosis [6, 7]. Fungi can also cause blindness, serious skin conditions, promote allergic responses, and have been shown to cause secondary infections in cystic fibrosis, tuberculosis, Human Immunodeficiency Virus (HIV), and recently, SARS-CoV-2 patients [8, 9]. Fungal plant pathogens also destroy around a third of crops annually; enough to feed 600 million people [1]. Whilst fungal infections of amphibians, bats, bees, animals, and trees have a huge impact on biodiversity [7]. Thus, fungi are an important asset, but they also pose a devastating global burden, and a deeper understanding of their biology is therefore essential.

The negative effects of fungi are exacerbated by a lack of antifungals and the emergence of multidrug resistant pathogens with intrinsic resistance such as *Candida auris* and *Lomentospora prolificans* [7, 10]. Current classes of antifungals include azoles (e.g. fluconazole), echinocandins (e.g. caspofungin), and polyenes (e.g. amphotericin B), but resistance to these antifungals is ever more prevalent and there are no vaccines [11]. Indeed, a recent meeting of the World Health Organization (WHO) Expert Group on Identifying Priority Fungal Pathogens highlighted azole resistant *A. fumigatus* as one such priority pathogen [12]. Therefore, we have an urgent requirement to identify potential novel antifungal targets.

G-quadruplexes (G4s) and i-motifs (iMs) are intriguing four-stranded (quadruplex) secondary structures in nucleic acids that are enriched in regulatory regions, particularly the promoters and telomeric regions of prokaryotic and eukaryotic genomes [13, 14]. However, they can be found throughout the genome [15]. G4s can be found within guanine-rich regions of DNA and RNA. Here, four guanine bases associate through Hoogsteen hydrogen bonding to form the basic unit of the G4, the G-tetrad [16]. These can then stack on top of each other to form the G4 structure itself (Figure 1A-C). These stacks of G-tetrads are connected by loops of mixed-

sequence nucleotides and can form intramolecular or intermolecular associations [17, 18]. This structure is further stabilised by the presence of monovalent cations, especially potassium [19]. Moreover, the 5'- to 3'- directionality of the strands, glycosidic bonding in the G-tetrads, the cation present, and number of stacked G-tetrads contribute to the wide variation of observed G4 structures and topologies [13]. Conversely, iMs form within cytosine-rich regions of DNA and can typically be found on the complementary strand opposite a G4 [20]. Like G4s, they are four stranded structures but are composed of two intercalated hairpins which are stabilised by hemi protonated cytosine-cytosine^{+} (C·C+) base pairs (Figure 1D-F) [21]. Studies into iMs have been limited compared to G4s as it was thought they weren't physiologically relevant based on them being most stable in slightly acidic conditions. However, they have since been shown to form under physiological conditions, including neutral pH and molecular crowding, and have recently been identified within the nuclei of human cells [21-24]

Once deemed structural curiosities, G4s and iMs have been highlighted to participate within host-pathogen interactions where they have important roles within gene regulation and have been shown to trigger phase separation of RNA in cells [25-27]. However, the regulation of biological functions by quadruplexes is complex and is influenced by their position within DNA/RNA, the surrounding topology, and environmental factors within the cell. In recent years, there has been increased interest in the therapeutic potential of targeting quadruplexes following the implication of these secondary structures in disease, especially cancer, due to their prevalence in oncogene promoters [26]. However, there is also now a growing number of pathogens in which G4s have been shown to contribute to virulence phenotypes, including; viruses (Human Papilloma Virus; Epstein-Barr Virus, HIV, SARS-CoV-2), prokaryotic bacterial pathogens (*Staphylococcus aureus, Streptococcus pneumoniae, Enterococcus* spp., *Borrelia* spp*., Neisseria meningitidis* and *N. gonorrhoeae*), and eukaryotic pathogens (*Trypanosoma brucei*, *Plasmodium falciparum*) [25, 28]. Notably, G4 DNA binding agents have been shown to be active against methicillin-resistant *S. aureus* and vancomycin-resistant *Enterococcus* spp.; potentially providing a novel target to overcome drug resistance [29, 30]. On the contrary,

iMs have so far only been identified in the long terminal repeat promoter of the HIV-1 pro-viral genome where they modulate the transcription of viral genes [31]. Although, one could assume that this is only the tip of the iceberg, considering the ubiquitous nature of quadruplexes within the genomes of almost every organism.

Practically all organisms possess quadruplexes within their genomes but a thorough analysis of putative quadruplex-forming sequences (PQS) in fungi has not been conducted to date. Identifying PQS in key genes associated with virulence/drug resistance in fungal pathogens could highlight targets for G4/iM-binding ligands and a potential novel means to ameliorate fungal pathogenicity. In this study, we performed a computational investigation into the presence of PQS in 1362 genomes across the fungal kingdom and discussed their potential as novel antifungal targets within key pathogenic species.

Methods

Selection of genome sequences

The genomic sequences analysed in this study were obtained from publicly available repositories. The database (https://doi.org/10.5281/zenodo.3783970) was used for the Ascomycota [32]. Regarding the Ascomycota, publicly available genomes from a comprehensive study on Saccharomycotina were obtained, the NCBI's Genome Browser was used to obtain basic information on the strains, and draft genomes from the subphyla Pezizomycotina and Taphrinomycotina were compiled from GenBank via FTP access number (ftp://ftp.ncbi.nlm.nih.gov/genomes/). Published genomes from the Basidiomycota, Mucoromycota, Zoopagomycota, Chytridiomycota, Microsporidia, and Cryptomycota were obtained from the Joint Genome Institute MycoCosm portal (https://mycocosm.jgi.doe.gov/mycocosm/home; last accessed 16/7/2020) [33]. Only the genomes from isolates with the highest assembly level and latest release date were included.

The principle of the G4Hunter algorithm and process of analysis

The G4Hunter web application was used to identify PQS within fungal genomes [34]. The algorithm used in G4Hunter considers the G-richness and G-skewness of a genome. Each position within a sequence is given a score between -4 and 4, with G's giving a positive score and C's giving a negative score. A's and T's are neutral and have a score of 0. An increasing G4Hunter score (either positive or negative) correlates with increased propensity to form quadruplexes. A near-zero average score is indicative of a sequence that is most likely to form stable duplexes. The G4Hunter score is the arithmetic mean value of the sequence.

When analysing the genomes, the scored nucleic acid sequence was computed for a sliding window of 30 nucleotides and a threshold above 1.5 for stringent analyses, or a window of 25 nucleotides and thresholds between 1-2 for complete analyses. Regions in which the value of the mean score was above a threshold (either positive or negative) were extracted. Potential G4- or iM-forming sequences were all noted as PQS and were not treated differently.

To identify the location of PQS within annotated genomic features, the file containing the annotations for known genomic features within the genome of the selected fungi were downloaded from the NCBI database. The presence of PQS within a pre-defined genomic feature (e.g. gene, mRNA, mobile element) or within ±100 bp of these genomic features were analysed.

Analysis of the genomes for PQS was conducted using the G4Hunter DNA analyser web application (http://bioinformatics.ibp.cz:8888/#/) [34]. The location of PQS in known genomic features were identified using a publicly available script found at https://pypi.org/project/dna-analyser-ibp/. The gene ontologies and protein classes of genes containing PQS were determined via PANTHER™ GO slim v.15.0 and PANTHER™ Protein Class v.15.0, respectively [35, 36]. The sequences identified in G4Hunter were verified via QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/analyze.php) [37]. Sequences were analysed using the default settings of max length (30), min G-group (2) and loop size (0-36). The highest scoring sequences with the shortest loop length were selected in each case.

ChemDraw Ultra v12.0 was used to draw structures and The Protein Imager (https://3dproteinimaging.com/protein-imager/) [38] was used to generate 3D images from PDBs. These data were processed, and graphs were generated using GraphPad Prism software v6.01.

Transcriptome analysis

Upregulated genes in germinating conidia and hyphae [39], during hypoxia [40], and during iron limitation or oxidative stress [41] were identified using publicly available transcriptome datasets. These were analysed in FungiDB (https://fungidb.org/) [42]. Upregulated genes were identified by comparing a reference sample (control) with a comparison sample (test). In each case, the difference between the minimum expression value of each gene in the reference sample and maximum expression value in the comparison sample was quantified. All genes upregulated >2-fold were acquired. Upregulated transcripts in *A. fumigatus* biofilms [43] were

identified via the manuscript's Supplementary Data. PQS in the top 20 most upregulated genes were identified using G4Hunter and QGRS Mapper using the default search settings.

Statistics

Statistical analyses were conducted via parametric One-way ANOVA with Tukey's multiple comparisons, Student's t-test, or Pearson's correlation coefficient; or non-parametric Kruskal-Wallis tests with post-hoc Dunn's test and Bonferroni corrections using GraphPad Prism software v6.01. $p<0.05$ were considered statistically significant.

Results

There is large heterogeneity in the frequency and number of PQS across the fungal kingdom

A thorough analysis of putative quadruplex-forming sequences in fungi has not been conducted to date. Using G4Hunter, the number of PQS were quantified in 1362 fungal genomes of varying assembly, across 7 divisions, and 15 sub-divisions of fungi. Due to the high variability in genome size and chromosome number between fungal species, to normalise the data, the total number of PQS in addition to the frequency of PQS/kbp and number of PQS relative to the GC content (PQS/GC%) were noted.

Across the divisions, the Chytridiomycota had the largest average number of PQS (24,734 PQS) and highest PQS frequencies (0.555 PQS/kbp and 473 PQS/GC%; Figure 2). The Mucoromycota and Basidiomycota also had large numbers of PQS (19575 and 17108 PQS, respectively; Figure 2A and E). The Basidiomycota and Zoopagomycota had high PQS frequencies relative to genome size (0.445 and 0.373 PQS/kbp, respectively; Figure 2B and F). The Mucoromycota and Basidiomycota displayed high PQS frequencies relative to GC content (459 and 340 PQS/GC%, respectively; Figure 2C and G). Fungi within the Basidiomycota had the highest average GC content (53.3%; Figure 2D and H). The Microsporidia and Cryptomycota scored lowest for total number of PQS (300 and 372, respectively), PQS/kbp (0.091 and 0.029, respectively) and PQS/GC% (8 and 11, respectively; Figure 2). Moreover, they also had low GC content (39.6% and 35.0%, respectively).

Considering G4s and iMs form in guanine or cytosine rich regions, respectively, one would expect fungi with a higher genome GC content to have a higher PQS frequency by chance. To investigate this further, the frequency of PQS/kbp relative to the GC content in all fungi and their divisions were plotted. As expected, there was a positive correlation between GC content and PQS frequency amongst all the fungal species analysed ($r=0.5290$; $p<0.0001$; Figure 3A). Moreover, this positive correlation was observed amongst the Ascomycota, Basidiomycota,

and Mucoromycota (r=0.5619, r=0.3891, and r=0.5239 and r=0.2883, respectively; all p<0.05; Figure 3B-D). However, there was not a significant correlation observed within the Zoopagomycota, Chytridiomycota, Microsporidia, or Cryptomycota, possibly due to the limited sample size (Figure 3E-H).

There is large heterogeneity in the number and frequency of PQS within sub-divisions and in fungal genera containing important fungal pathogens

Within the Ascomycota, the average number of PQS in the sub-divisions Taphrinomycotina (n=14 species), Saccharomycotina (n=332 species), and Pezizomycotina (n=761 species) were 2799.7, 1079.0, and 17021.0, respectively (Figure 4A). The average PQS/kbp for each subphylum was 0.147, 0.075, and 0.436, respectively (Figure 4B). The average PQS/GC% for each subphylum was 42.4, 25.2, and 344.3, respectively (Figure 4C). Finally, the average GC% for each subphylum was 42.5%, 40.6%, and 49.6%, respectively (Figure 4D).

Within the Basidiomycota, the average number of PQS in the sub-divisions Agaricomycotina (n=134 species), Pucciniomycotina (n=24 species), and Ustilagomycotina (n=31 species) were 16538.7, 29434.8, and 5351.0, respectively (Figure 4A). The average PQS/kbp for each subphylum was 0.433, 0.610, and 0.293, respectively (Figure 4B). The average PQS/GC% for each subphylum was 319.2, 605.6, and 94.1, respectively (Figure 4C). Finally, the average GC% for each subphylum was 51.1%, 51.1%, and 57.6%, respectively (Figure 4D).

Within the Mucoromycota, the average number of PQS in the sub-divisions Glomeromycotina (n=9 species), Mortierellomycotina (n=2 species), and Mucoromycotina (n=20 species) were 2900.1, 16444.0, and 39382.7, respectively (Figure 4A). The average PQS/kbp for each subphylum was 0.023, 0.338, and 0.246, respectively (Figure 4B). The average PQS/GC% for each subphylum was 106.5, 349.8, and 919.7, respectively (Figure 4C). Finally, the average GC% for each subphylum was 27.2%, 44.9%, and 39.7%, respectively (Figure 4D).

Within the Zoopagomycota, the average number of PQS in the sub-divisions Zoopagomycotina (n=3 species), Entomophthoromycotina (n=2 species), and

Kickxellomycotina (n=7 species) were 8399.7, 6726.5, and 11315.1, respectively (Figure 4A). The average PQS/kbp for each subphylum was 0.746, 0.081, and 0.292, respectively (Figure 4B). The average PQS/GC% for each subphylum was 162.3, 162.7, and 305.4, respectively (Figure 4C). Finally, the average GC% for each subphylum was 54.4%, 35.3%, and 38.2%, respectively (Figure 4D).

Within the Chytridiomycota, the average number of PQS in the sub-divisions Chytridiomycetes (n=6 species), Monoblepharidomycetes (n=1 species), and Neocallimatigomycetes (n=5 species) were 23815.2, 50091, and 298, respectively (Figure 4A). The average PQS/kbp for each subphylum was 0.637, 1.027, and 0.003, respectively (Figure 4B). The average PQS/GC% for each subphylum was 446.7, 955.9, and 15.6, respectively (Figure 4C). Finally, the average GC% for each subphylum was 54.3%, 52.4%, and 18.9%, respectively (Figure 4D).

Within the Microsporidia (n=9 species) and Cryptomycota (n=2 species), the average number of PQS were 300.3 and 372, respectively (Figure 4A). The average PQS/kbp were 0.091 and 0.029, respectively (Figure 4B). The average PQS/GC% were 7.6 and 10.6, respectively (Figure 4C). Finally, the average GC% were 39.6% and 35.0%, respectively (Figure 4D).

Finally, we also highlighted the frequency of PQS in fungal genera which contained important human and plant pathogens. We found that there was also large heterogeneity in the frequency of PQS between species within genera containing human pathogens (e.g. *Aspergillus* spp., *Candida* spp., *Cryptococcus* spp., *Blastomyces* spp.) and plant pathogens (e.g. *Verticillium* spp., and *Fusarium* spp.; Figure 5). This variation was particularly wide within *Aspergillus* spp., and *Cryptococcus* spp.

PQS are evolutionarily conserved and are linked with genetic relatedness

Evolutionary conservation of genetic motifs within the genome are a hallmark of their fundamental importance to how that organism functions. Therefore, we endeavoured to explore whether there was evolutionary conservation of PQS within fungal genomes. We

chose to explore this relationship in *Aspergillus* spp*.,* due to the robustness and accuracy of the phylogenetic tree available [44].

Notably, we found that the frequency of PQS/kbp appeared to be intrinsically linked to how closely related species were, with species within the same section displaying similar PQS frequencies (Figure 6). Aspergilli in this tree were divided into 13 sections (range of PQS/kbp in brackets), the *Aspergillus* (0.364-0.461), the *Fumigati* (0.204-0.224), the *Candidi* (0.747-0.816), the *Circumdati* (0.395-0.467), the *Flavi* (0.225-0.289), the *Ochraceoros* (0.421-0.422), the *Usti* (0.385-0.396), the *Versicolores* (0.282-0.308), and the *Nigri* (0.495-0.782). The *Nigri* displayed the largest variation in PQS, however the lesser related species *A. carbonarius* and *A. aculeatus* skewed this range and if only the closest related species were noted, this range would be from 0.495-0.611 (Figure 6). As the *Nidulantes*, *Terrei, and Clavati* only contained one member each, correlations in these sections could not be made. Considering that we found a likely quadruplex-forming sequence within *cyp51A* in *A. fumigatus* and due to its importance in encoding a gene heavily involved in azole resistance, we chose to see if this sequence was conserved within the section *Fumigati*. Interestingly, *A. fischeri*, *A. novofumigatus*, and *A. lentulus* all retained the PQS observed in *A. fumigatus*, however this sequence could not be found in *A. udagawae* or *A. turcosus* (Figure 6).

Loss of PQS is associated with pathogenicity in *Aspergillus* and *Cryptococcus* species, but not *Candida*

The Ascomycota and Basidiomycota contain many of the most prevalent fungal pathogens of both plants and humans, including the genera *Aspergillus* spp*., Candida* spp*.,* and *Cryptococcus* spp*.,* which contain fungal species that account for most fungal-related deaths in humans. Although, not all species within these genera are potential pathogens and we found high variation in their PQS frequency. Therefore, we compared the PQS frequency between

pathogenic and non-pathogenic species to explore whether there was a link with pathogenicity.

Indeed, comparing 72 species of *Aspergillus* (22 pathogenic, 50 non-pathogenic) pathogenic species had a significantly lower frequency of PQS/kbp (0.364 vs. 0.523) and PQS/GC% (249.9 vs. 380.5) on average, compared to their non-pathogenic counterparts (Figure 7A and D). Further analysis highlighted that *A. unguis* had the lowest frequency of PQS/kbp (0.203 PQS/kbp), whilst *A. ellipticus* had the highest (0.997 PQS/kbp). Furthermore, the most pathogenic species, *A. fumigatus,* had a lower than average frequency of PQS (0.221 PQS/kbp; Supplementary Material). Notably, loss of PQS in the subphylum Pezizomycotina appeared to be associated with pathogenicity in humans overall (Supplementary Material).

Similarly, comparing 16 species of *Cryptococcus* (9 pathogenic, 7 non-pathogenic) we also found that pathogenic species had a significantly lower frequency of PQS/kbp (0.480 vs. 1.176) and PQS/GC% (187.1 vs. 465.9) on average, compared to their non-pathogenic counterparts (Figure 7B and E). However, the lowest frequency of PQS/kbp was found in the non-pathogenic species *C. depauperatus* (0.084 PQS/kbp). Regarding pathogenic species, the *C. neoformans* AD hybrid was found to have the lowest PQS frequency (0.381 PQS/kbp). The highest PQS frequency was observed in *C. floricola* (1.64 PQS/kbp; Supplementary Material).

Conversely, this phenomenon was not observed in *Candida* spp., or closely related *Clavispora* and *Debaryomyces* species. Comparing 47 species (14 pathogenic, 33 non-pathogenic) we found that there was no association between PQS frequency and pathogenicity (Figure 7C and F). Pathogenic and non-pathogenic species had equivalent PQS/kbp (0.074 vs. 0.079) and PQS/GC% (26.3 vs. 22.3; Figure 7C and F). Moreover, this trend was also observed for pathogenic and non-pathogenic species across the entire subphylum Saccharomycotina (Supplementary Material). Further analysis showed that *C. orba* had the lowest frequency of PQS (0.006 PQS/kbp), whereas *C. fructus* had the highest (0.261 PQS/kbp; Supplementary Material). The most prevalent pathogenic species, *C. albicans,* had a higher than average

337 frequency of PQS (0.101 PQS/kbp; Supplementary Material). Conversely, the multi-drug
338 resistant species *C. auris* had both a lower frequency of PQS/kbp (0.061) and PQS/GC%
339 (17.3), compared to the average.

340 **PQS are found in numerous genomic features in *A. fumigatus*, *C. neoformans*, and *C.***
341 ***albicans***

342 To evaluate the position of PQS within *A. fumigatus, C. neoformans,* and *C. albicans* (the most
343 prevalent pathogens of their genera) the annotation information was obtained from the NCBI
344 database and the presence of PQS within defined genomic features, or within 100 bp before
345 or after these features were analysed. Both the total number of PQS per feature and the
346 frequency of PQS/kbp relative to the combined genomic length of the described features were
347 noted.

348 When only total PQS were considered, the largest number of PQS in all three fungal species
349 could be found within the coding regions (CDS), genes, and mRNA, with few PQS found in
350 other genomic features (Figure 8A, B, and C). However, this was not the same when
351 considering the frequency of PQS/kbp of the genomic features. In *A. fumigatus,* the greatest
352 frequency of PQS could be found in the repeat regions (Figure 8D). The lowest frequency
353 could be found within the tRNA. In *C. neoformans,* the highest PQS frequencies were still in
354 the CDS, genes, and mRNA, with a very low frequency found within the tRNA (Figure 8E). In
355 *C. albicans,* the highest frequency of PQS could be found in the rRNA, followed by repeat
356 regions and ncRNA (Figure 8F). There were no PQS found in the tRNA and low frequencies
357 were again found in the mobile elements. The total number and frequency of PQS 100 bp
358 before and after the annotated genomic features appeared to be evenly distributed (Figure 8).

359 **PQS are found in genes encoding proteins involved in metabolism, nucleic acid**
360 **binding, cell transport, and protein modification**

361 As we knew the genomic location of the PQS, we could then identify the number and identity
362 of the genes which contained these sequences. This further enabled us to identify the classes

of proteins associated with PQS-containing genes. In *A. fumigatus*, 35.1% of genes contained at least one PQS. In *C. neoformans*, this number was almost double, with 59.9% of genes containing PQS. Conversely, PQS were only found in 5.6% of genes in *C. albicans*.

Despite the discrepancies in the number of genes where PQS can found between the organisms, in all cases, PQS were primarily located in genes which encoded proteins involved in metabolism, nucleic acid binding, cell transport, and protein modification (Figure 9). They were least likely to be found in genes encoding for calcium-binding proteins, extracellular matrix proteins, cell adhesion molecules, and defense/immunity proteins (Figure 9).

In all organisms, PQS could be found in the highest frequency in genes associated with metabolite interconversion enzymes. In *A. fumigatus*, the number of genes associated with metabolite interconversion enzymes was 3.7-fold higher than the next represented protein class (434 genes vs. 117 genes for nucleic acid binding proteins and transporters; Figure 9A). In *C. neoformans* the number of genes associated with these enzymes was 2.1-fold higher compared to nucleic acid binding proteins (491 genes vs. 231 genes, respectively; Figure 9B). In *C. albicans*, the difference in the number of PQS-containing genes associated with metabolite interconversion enzymes and nucleic acid binding proteins was much lower (26 genes vs. 21 genes, respectively; 1.2-fold; Figure 9C). Surprisingly, when categorising genes based on gene ontology terms, there was an almost identical distribution of genes involved in biological functions, molecular functions, and cellular components between species (Supplementary Material).

PQS are found in genes linked to virulence, drug resistance, or key biological processes in *A. fumigatus*, *C. neoformans,* and *C. albicans*

G4s have been discovered in virulence genes from several pathogens and have arisen as a promising target for antimicrobial therapy and overcoming drug resistance [25]. Moreover, an iM in the promoter of the HIV-1 pro-viral genome has also been recently been described [31]. Thus, whether PQS could be found in genes associated with virulence/drug resistance in *A.*

fumigatus, *C. neoformans*, and *C. albicans* was explored. Although the list is not exhaustive (there are many proteins still yet to be characterised), there were many interesting candidates that arose from the analysis. In total, PQS were found in over 100 genes associated with the virulence, drug resistance, or key biological processes of *A. fumigatus* (39 genes), *C. neoformans* (41 genes), and *C. albicans* (27; Tables 1-3).

In *A. fumigatus,* PQS could be found in notable genes, including the 14-α sterol demethylases (*cyp51A* and *cyp51B*), the 1,3-β-glucan synthase catalytic subunit *fks1,* and ABC drug exporter *atrF*, which are involved in drug resistance. In addition to genes involved in virulence, including transcription factors *stuA, hapX,* and *pacC*, genes involved in pigment biosynthesis (*pksP, arp2, abr1, abr2,* and *ayg1*), a master regulator of secondary metabolism *laeA,* and *gliN and gliP* which are involved in the synthesis of gliotoxin (Table 1).

As PQS could be found in almost two-thirds of *C. neoformans* genes, it was not surprising that PQS could be found in those associated with virulence. These included the ABC transporter *afr1* (which is associated with fluconazole resistance), the protein kinases *fsk* and *hog1*, the calcineurin-associated genes *crz1* and *cna1*, *pacC/rim101* like in *A. fumigatus*, and numerous capsule-associated genes (the main virulence factor of *Cryptococcus*) including *cap2, cap5, cap10, cap59, cap60, cap64, cas31, cas33,* and *cxt1* (Table 2).

There were very few genes in *C. albicans* that contained sequences likely to form quadruplexes, and thus, quadruplexes might be less important in this organism. Notable genes included the iron permeases *ftr1* and *ftr2*, and a gene associated with flucytosine resistance (*rrp9*; Table 3).

The highest scoring potential quadruplex-forming sequences for each of these genes were then re-analysed in an alternative PQS predictive algorithm called QGRS Mapper. In this instance, the scores of known quadruplex-forming sequences were compared to scores of the PQS in fungi. This was conducted to provide further insight into whether these sequences were likely to form quadruplex structures.

Known quadruplex-forming sequences (n=94) were shown to predominantly have QGRS scores of 21, 42, or 63 (depending on the number of G-tetrads – 2, 3, or 4, respectively; Supplementary Figures). The genes containing PQS from both *A. fumigatus* and *C. neoformans* had more sequences scoring between 21 and 42 compared to *C. albicans* and are thus more likely to form quadruplex structures (Supplementary Figures).

Genes in *A. fumigatus* that are upregulated during fungal germination, in response to stress, and in biofilms are enriched in PQS compared to the whole genome

Although PQS could be found in important fungal genes, we aimed to link the presence of PQS within these organisms with potential biological and pathophysiological functions. To investigate this, transcriptome datasets of *A. fumigatus* during germination [39], hypoxia [40], iron limitation and oxidative stress [41], or in biofilms [43] were analysed. In each instance, the top 20 upregulated genes (compared to dormant/unstressed *A. fumigatus* controls) were investigated for the presence and frequency of PQS. This was achieved via examining the sequences in both QGRS Mapper and G4Hunter using the default search settings. Gene sequences which were identified as containing PQS by both algorithms were used for analysis.

Notably, PQS could be found in 77.5% of the genes investigated (Figure 10A). This included genes which were upregulated in germinating conidia (16/20; 80.0%), hyphae (14/20; 70.0%), after 12 hours of hypoxia (17/20; 85.0%), following iron limitation (17/20; 85.0%), undergoing oxidative stress (14/20; 70.0%), and in biofilms (15/20; 75.0%). These genes were investigated further to assess whether they were enriched with PQS and contained a higher frequency of PQS comparative to the average frequency found in the *A. fumigatus* genome. Using the default G4Hunter settings (window size 25 nt and threshold of 1.2) the PQS frequency of the *A. fumigatus* genome was 1.55 PQS/kbp (red dashed line; Figure 10B). In all cases, the average PQS frequencies in the upregulated genes were higher than the average PQS observed throughout the entire genome (Figure. 10B). The average PQS frequencies in upregulated PQS-containing genes were 2.97 PQS/kbp (germinating conidia), 3.72 PQS/kbp (hyphae), 2.80 PQS/kbp (12 h hypoxia), 2.66 PQS/kbp (iron starvation), 2.32 PQS/kbp

(oxidative stress), and 2.26 PQS/kbp (biofilms; Figure 10B). Although, there were a range of PQS frequencies observed between the genes from 0.34 to 11.90 PQS/kbp. The genes containing the highest PQS frequencies for each condition were AFUA_8G01710 in germinating conidia and hyphae (11.90 PQS/kbp), AFUA_4G09580 in hypoxic fungi (5.59 PQS/kbp), AFUA_3G03650 during iron limitation (8.50 PQS/kbp), AFUA_5G10220 during oxidative stress (5.28 PQS/kbp), and AFUA_8G01980 in biofilms (5.90 PQS/kbp). Interestingly, each of these genes were upregulated in at least 3 out the 6 conditions investigated.

Discussion

In this study, the number of potential quadruplex-forming sequences within the genomes of fungi were computationally predicted and their potential involvement in pathogenicity was discussed. Several important observations were made. This was the first study to identify the heterogeneity of PQS amongst genetically distinct fungal species. Moreover, we highlighted that pathogenic *Aspergillus* and *Cryptococcus* species contained fewer PQS compared to their non-pathogenic counterparts and these could be found throughout known genomic features, including genes, mRNA, repeat regions, tRNA, ncRNA, and rRNA. Genes containing PQS were associated with metabolism, nucleic acid-binding proteins, protein modifying enzymes, and transporters. Notably, PQS likely to form quadruplexes were identified in genes linked with fungal virulence or drug resistance, such as *cyp51A,* and could be found in genes upregulated during fungal growth and in response to stress.

The frequency of PQS throughout genomes is highly variable; for example, human genomes were shown to contain around 0.228 PQS/kbp, whereas the genomes of *Escherichia coli* contain around 0.028 PQS/kbp [15]. In this study we also found significant differences in the number and frequency of PQS throughout fungal genomes. For example, fungi from the Saccharomycotina (primarily composed of yeasts) contained a low frequency of PQS. Conversely, filamentous fungi from the Pezizomycotina (containing many important pathogens of both plants and humans) had much higher PQS frequencies and >15-fold more PQS than

fungi from the Saccharomycotina on average. It has previously been shown that G4s contribute to genetic instability in yeasts (*Saccharomyces* spp.,) and as suggested for bacteria, it's possible that G4s may also have been deselected through evolution in ascomycete yeasts [45, 46]. However, things are more complicated concerning basidiomycetous yeasts such as *Cryptococcus* spp. and *Malassezia* spp. Cryptococcal species such as *C. wingfieldii, C. amylolentus,* and *C. floricola* had some of the highest PQS frequencies in the study. Alternatively, all the *Malassezia* species investigated had much lower PQS frequencies than expected, which was surprising given their genomes were GC rich compared to many other fungi (>55% GC content).

Notably, the genomes of pathogenic aspergilli and cryptococci contained significantly lower PQS frequencies relative to their non-pathogenic counterparts. This trend was observed throughout the Pezizomycotina overall. Conversely, there was no correlation between PQS and pathogenicity between *Candida* species or within the Saccharomycotina. The exact reasons underlying the loss of PQS are not currently known. Fungi may have evolved to retain the most essential quadruplex-forming sequences or those that provided them with a pathogenic advantage. In the case of *Aspergillus* and *Cryptococcus* species, this loss of PQS might be linked with evasion of the immune response or enhanced pathogenicity. Furthermore, the observed loss of PQS within *Candida* and *Malassezia* species might be associated with their ability to form commensal relationships and avoid stimulating a host response. Interestingly, loss of PQS has recently also been observed in pathogenic *Coronaviridae* [47]. It has also been reported that host nucleolin (an RNA-binding protein) can bind and stabilise quadruplexes in the LTR promoter of HIV-1, which can silence viral transcription [48]. Therefore, in this situation, loss of quadruplexes would be beneficial for immune evasion. Contrarily, it has been suggested that iMs within HIV-1 are triggered after they promote the acidification of intracellular compartments, modulating viral processes [31]. In this instance, the formation of quadruplexes could be beneficial and promote pathogenicity. The importance of quadruplex stabilisation may therefore be situational and could be detrimental or beneficial

496 depending on the environment. It is known that fungal pathogens such as *A. fumigatus* and *C.*
497 *neoformans* can survive the acidic conditions within intracellular organelles to propagate
498 infection [49, 50]. Therefore, it could be interesting to investigate the dynamics/flux of
499 quadruplex formation under conditions that fungi would experience within the host and
500 whether they contribute to fungal survival and propagation.

501 Here, we identified many PQS within the CDS, genes, and mRNA of *A. fumigatus, C.*
502 *neoformans,* and *C. albicans*; but a higher frequency of PQS within the repeat regions of *A.*
503 *fumigatus,* and the repeat regions and rRNA of *C. albicans.* It has been observed that *S.*
504 *cerevisiae* promoters and open reading frames are enriched with sequences capable of
505 potentially forming intramolecular G4s [51]. A recent study by Čutová *et al.,* also demonstrated
506 that inverted repeats were enriched in the centromeres and rDNA/rRNA regions, and G4s
507 were enriched in the telomeres and tRNAs of *S. cerevisiae* [52].

508 There were large differences between the numbers of genes containing PQS in these
509 organisms, but PQS were found distributed amongst genes encoding proteins from similar
510 classes. The highest represented groups being metabolite interconversion enzymes, nucleic
511 acid binding proteins, transporters, and protein modifying enzymes in all. Furthermore, genes
512 containing PQS within *S. cerevisiae* promoter regions were primarily involved in metabolism.
513 Which supports our observations and is equally concordant with observations made in both
514 humans and bacteria [53, 54]. It is widely acknowledged that metabolism in fungi is central to
515 its virulence, pathogenicity, and survival [55]. The ability of a fungus to rapidly adapt to the
516 host microenvironment can be achieved through such processes as metabolic remodelling,
517 stress resistance, and the utilisation of amino acids [56]. This in turn has a significant effect
518 on the triggering of important virulence traits, like the production of hyphae, biofilm growth,
519 capsule formation in *C. neoformans*, and melanisation [56]. Moreover, metabolic pathways
520 influence fungal vulnerability to innate immune defences and can regulate susceptibility to
521 antifungal drugs [57]. Not only primary metabolism, secondary metabolism results in the
522 release of various secondary metabolites and toxins, such as gliotoxin, aflatoxin, and

candidalysin [58, 59]. Interestingly, PQS could also be found in many of the most upregulated genes during germination, in response to environmental stress, or in biofilms. Additionally, these upregulated genes were enriched in PQS. Thus, the enrichment of PQS within these genes is suggestive of the potential importance of quadruplexes in the regulation of biological functions and pathogenicity in fungi, and it would be interesting to explore the effects quadruplex-targeting ligands could have on these pathways.

PQS could also be found within key genes linked to virulence and drug resistance in these fungi. In *A. fumigatus,* PQS could be found in *cyp51A, cyp51B, atrF,* and *fsk1* (involved in resistance to azoles and echinocandins) [60-62], *laeA, gliN and gliP,* (involved in toxin and metabolite biosynthesis) [63, 64], *pksP, arp2, abr1, arb2,* and *ayg1* (involved in the production of the virulence factor DHN-melanin) [65], in addition to numerous important transcription factors (*meaB, hapX, pacC*) [66]. Similarly, in *C. neoformans,* PQS could be found in genes linked to azole resistance (*afr1*) [67], melanin production (*lac2*) [68], and transcription factors (*pacC/rim101*). The most notable virulence factor of *C. neoformans* is its polysaccharide capsule and PQS could be found in numerous capsule-associated genes (*cas31, cap60, cap59, cap64, cap2, cap5, cas33, cap10,* and *cxt1*) [69]. In *C. albicans* PQS could be found in genes such as the iron permeases *ftr1* and *ftr2* [70]. Notably, many of these genes contained PQS which have previously been shown to be capable of forming *bona fide* quadruplexes, such as the sequence GGAGGAGGAGG [71]. It is also interesting to highlight that these organisms contained many more $G_{2+}L_{1-12}$ compared to $G_{3+}L_{1-12}$ PQS sequences, which is a characteristic shared with *S. cerevisiae* [15].

There are now an ever-increasing number of G4s identified within genes linked to microbial pathogenicity. G4-forming motifs located in the *hsdS, recD,* and *pmrA* genes of *S. pneumoniae,* and *var* genes of *P. falciparum* have been identified to modulate host-pathogen interactions [72, 73]. Moreover, targeting G4s in *espK, espB*, and *cyp51* from *Mycobacterium tuberculosis* with the G4/iM-binding ligand TMPyP4 could inhibit transcription of these essential virulence genes [74]. Binding of several G4-targeting ligands to G4s in viruses have

been shown to limit the virulence of the Herpes simplex virus-1, HIV-1, Ebola, and Hepatitis C [75]. Interestingly, dinuclear ruthenium (II) complexes (well-characterised G4 and iM DNA binding agents) have been shown to be active against methicillin-resistant *S. aureus* and vancomycin-resistant *Enterococcus* spp [29, 30]. Notably, another G4/iM-ligand, berberine, has also demonstrated activity towards fluconazole-resistant *C. albicans* and *C. neoformans* [76]. Thus, quadruplexes may provide a novel target to overcome the issues surrounding drug resistance.

Conclusion

Identifying PQS within key virulence/drug resistance genes of fungal pathogens has highlighted structures that may be targeted by G4/iM-binding drugs to ameliorate fungal pathogenicity. However, we first need to identify whether these structures are formed and if they can modulate biological functions. Understanding whether quadruplexes could form under pathophysiological conditions, whether fungi themselves could release quadruplex-binding ligands to modulate host responses, or identifying whether there are fungal specific quadruplex-forming sequences could also be important factors to consider. Thus, targeting quadruplexes in fungi could present a novel target for the amelioration of fungal infection and drug resistance, although a more thorough investigation is necessary.

Acknowledgements

V. B was supported by the Czech Science Foundation (18-15548S). I would like to acknowledge Phil Spence for contributing the G4 and iM structures and for critical reviewing of the manuscript. S.B conceived the study. E. F. W., N. B., V. B., and S.B., performed the data analysis. E. F. W. and S. B wrote the manuscript in consultation with Z. A. E. W. All authors contributed to interpretation of the results, provided critical feedback, and helped to shape the research and manuscript.

References

1. *Correction: Stop neglecting fungi.* Nat Microbiol, 2017. **2**: p. 17123.
2. Skellam, E., *Strategies for Engineering Natural Product Biosynthesis in Fungi.* Trends Biotechnol, 2019. **37**(4): p. 416-427.
3. Hernández, V.A., et al., *Talaromyces australis and Penicillium murcianum pigment production in optimized liquid cultures and evaluation of their cytotoxicity in textile applications.* World J Microbiol Biotechnol, 2019. **35**(10): p. 160.
4. Hooker, C.A., K.Z. Lee, and K.V. Solomon, *Leveraging anaerobic fungi for biotechnology.* Curr Opin Biotechnol, 2019. **59**: p. 103-110.
5. Gallone, B., et al., *Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts.* Cell, 2016. **166**(6): p. 1397-1410.e16.
6. Brown, G.D., et al., *Hidden killers: human fungal infections.* Sci Transl Med, 2012. **4**(165): p. 165rv13.
7. Fisher, M.C., et al., *Threats Posed by the Fungal Kingdom to Humans, Wildlife, and Agriculture.* mBio, 2020. **11**(3).
8. Köhler, J.R., A. Casadevall, and J. Perfect, *The spectrum of fungi that infects humans.* Cold Spring Harb Perspect Med, 2014. **5**(1): p. a019273.
9. van Arkel, A.L.E., et al., *COVID-19-associated Pulmonary Aspergillosis.* Am J Respir Crit Care Med, 2020. **202**(1): p. 132-135.
10. Ramirez-Garcia, A., et al., *Scedosporium and Lomentospora: an updated overview of underrated opportunists.* Med Mycol, 2018. **56**(suppl_1): p. 102-125.
11. Berman, J. and D.J. Krysan, *Drug resistance and tolerance in fungi.* Nat Rev Microbiol, 2020. **18**(6): p. 319-331.
12. World Health, O., *First meeting of the WHO antifungal expert group on identifying priority fungal pathogens: meeting report*. 2020, Geneva: World Health Organization.
13. Varshney, D., et al., *The regulation and functions of DNA and RNA G-quadruplexes.* Nat Rev Mol Cell Biol, 2020. **21**(8): p. 459-474.
14. Abou Assi, H., et al., *i-Motif DNA: structural features and significance to cell biology.* Nucleic Acids Res, 2018. **46**(16): p. 8038-8056.
15. Marsico, G., et al., *Whole genome experimental maps of DNA G-quadruplexes in multiple species.* Nucleic Acids Res, 2019. **47**(8): p. 3862-3874.
16. Burge, S., et al., *Quadruplex DNA: sequence, topology and structure.* Nucleic Acids Res, 2006. **34**(19): p. 5402-15.
17. Dai, J., et al., *An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution.* J Am Chem Soc, 2006. **128**(4): p. 1096-8.

18. Víglaský, V., L. Bauer, and K. Tlucková, *Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats.* Biochemistry, 2010. **49**(10): p. 2110-20.
19. Lane, A.N., et al., *Stability and kinetics of G-quadruplex structures.* Nucleic Acids Res, 2008. **36**(17): p. 5482-515.
20. Brazier, J.A., A. Shah, and G.D. Brown, *I-motif formation in gene promoters: unusually stable formation in sequences complementary to known G-quadruplexes.* Chem Commun (Camb), 2012. **48**(87): p. 10739-41.
21. Gehring, K., J.L. Leroy, and M. Guéron, *A tetrameric DNA structure with protonated cytosine.cytosine base pairs.* Nature, 1993. **363**(6429): p. 561-5.
22. Wright, E.P., J.L. Huppert, and Z.A.E. Waller, *Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH.* Nucleic Acids Res, 2017. **45**(6): p. 2951-2959.
23. Rajendran, A., S. Nakano, and N. Sugimoto, *Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH.* Chem Commun (Camb), 2010. **46**(8): p. 1299-301.
24. Zeraati, M., et al., *I-motif DNA structures are formed in the nuclei of human cells.* Nat Chem, 2018. **10**(6): p. 631-637.
25. Saranathan, N. and P. Vivekanandan, *G-Quadruplexes: More Than Just a Kink in Microbial Genomes.* Trends Microbiol, 2019. **27**(2): p. 148-163.
26. Balasubramanian, S., L.H. Hurley, and S. Neidle, *Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?* Nat Rev Drug Discov, 2011. **10**(4): p. 261-75.
27. Zhang, Y., et al., *G-quadruplex structures trigger RNA phase separation.* Nucleic Acids Res, 2019. **47**(22): p. 11746-11754.
28. Bartas, M., et al., *The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria.* Molecules, 2019. **24**(9).
29. Li, F., et al., *The antimicrobial activity of inert oligonuclear polypyridylruthenium(II) complexes against pathogenic bacteria, including MRSA.* Dalton Trans, 2011. **40**(18): p. 5032-8.
30. Li, F., J.G. Collins, and F.R. Keene, *Ruthenium complexes as antimicrobial agents.* Chem Soc Rev, 2015. **44**(8): p. 2529-42.
31. Ruggiero, E., et al., *A dynamic i-motif with a duplex stem-loop in the long terminal repeat promoter of the HIV-1 proviral genome modulates viral transcription.* Nucleic Acids Res, 2019. **47**(21): p. 11057-11068.
32. Shen, X.-X., et al., *Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota.* bioRxiv, 2020: p. 2020.05.11.088658.

33. Grigoriev, I.V., et al., *MycoCosm portal: gearing up for 1000 fungal genomes.* Nucleic Acids Res, 2014. **42**(Database issue): p. D699-704.
34. Brázda, V., et al., *G4Hunter web application: a web server for G-quadruplex prediction.* Bioinformatics, 2019. **35**(18): p. 3493-3495.
35. Mi, H., et al., *PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools.* Nucleic Acids Res, 2019. **47**(D1): p. D419-d426.
36. Thomas, P.D., et al., *Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W645-50.
37. Kikin, O., L. D'Antonio, and P.S. Bagga, *QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W676-82.
38. Tomasello, G., I. Armenia, and G. Molla, *The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities.* Bioinformatics, 2020. **36**(9): p. 2909-2911.
39. Hagiwara, D., et al., *Comparative transcriptome analysis revealing dormant conidia and germination associated genes in Aspergillus species: an essential role for AtfA in conidial dormancy.* BMC Genomics, 2016. **17**: p. 358.
40. Hillmann, F., et al., *The novel globin protein fungoglobin is involved in low oxygen adaptation of Aspergillus fumigatus.* Mol Microbiol, 2014. **93**(3): p. 539-53.
41. Kurucz, V., et al., *Additional oxidative stress reroutes the global response of Aspergillus fumigatus to iron depletion.* BMC Genomics, 2018. **19**(1): p. 357.
42. Stajich, J.E., et al., *FungiDB: an integrated functional genomics database for fungi.* Nucleic Acids Res, 2012. **40**(Database issue): p. D675-81.
43. Gibbons, J.G., et al., *Global transcriptome changes underlying colony growth in the opportunistic human pathogen Aspergillus fumigatus.* Eukaryot Cell, 2012. **11**(1): p. 68-78.
44. Steenwyk, J.L., et al., *A Robust Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the Genera Aspergillus and Penicillium.* mBio, 2019. **10**(4).
45. Lopes, J., et al., *G-quadruplex-induced instability during leading-strand replication.* Embo j, 2011. **30**(19): p. 4033-46.
46. Guo, J.U. and D.P. Bartel, *RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria.* Science, 2016. **353**(6306).
47. Bartas, M., et al., *In-Depth Bioinformatic Analyses of Nidovirales Including Human SARS-CoV-2, SARS-CoV, MERS-CoV Viruses Suggest Important Roles of Non-*

canonical Nucleic Acid Structures in Their Lifecycles. Front Microbiol, 2020. **11**: p. 1583.

48. Tosoni, E., et al., *Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription.* Nucleic Acids Res, 2015. **43**(18): p. 8884-97.
49. Wasylnka, J.A. and M.M. Moore, *Aspergillus fumigatus conidia survive and germinate in acidic organelles of A549 epithelial cells.* J Cell Sci, 2003. **116**(Pt 8): p. 1579-87.
50. Coelho, C., A.L. Bocca, and A. Casadevall, *The intracellular life of Cryptococcus neoformans.* Annu Rev Pathol, 2014. **9**: p. 219-38.
51. Hershman, S.G., et al., *Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae.* Nucleic Acids Res, 2008. **36**(1): p. 144-56.
52. Čutová, M., et al., *Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae.* Genomics, 2020. **112**(2): p. 1897-1901.
53. Eddy, J. and N. Maizels, *Gene function correlates with potential for G4 DNA formation in the human genome.* Nucleic Acids Res, 2006. **34**(14): p. 3887-96.
54. Rawal, P., et al., *Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation.* Genome Res, 2006. **16**(5): p. 644-55.
55. Brown, A.J., et al., *Metabolism impacts upon Candida immunogenicity and pathogenicity at multiple levels.* Trends Microbiol, 2014. **22**(11): p. 614-22.
56. Ene, I.V., et al., *Metabolism in fungal pathogenesis.* Cold Spring Harb Perspect Med, 2014. **4**(12): p. a019695.
57. Parente-Rocha, J.A., et al., *Antifungal Resistance, Metabolic Routes as Drug Targets, and New Antifungal Agents: An Overview about Endemic Dimorphic Fungi.* Mediators Inflamm, 2017. **2017**: p. 9870679.
58. Bignell, E., et al., *Secondary metabolite arsenal of an opportunistic pathogenic fungus.* Philos Trans R Soc Lond B Biol Sci, 2016. **371**(1709).
59. Moyes, D.L., et al., *Candidalysin is a fungal peptide toxin critical for mucosal infection.* Nature, 2016. **532**(7597): p. 64-8.
60. Zhang, J., et al., *The Fungal CYP51s: Their Functions, Structures, Related Drug Resistance, and Inhibitors.* Front Microbiol, 2019. **10**: p. 691.
61. Meneau, I., A.T. Coste, and D. Sanglard, *Identification of Aspergillus fumigatus multidrug transporter genes and their potential involvement in antifungal resistance.* Med Mycol, 2016. **54**(6): p. 616-27.
62. Jiménez-Ortigosa, C., et al., *Emergence of Echinocandin Resistance Due to a Point Mutation in the fks1 Gene of Aspergillus fumigatus in a Patient with Chronic Pulmonary Aspergillosis.* Antimicrob Agents Chemother, 2017. **61**(12).

722 63. Bok, J.W., et al., *LaeA, a regulator of morphogenetic fungal virulence factors.* Eukaryot
723 Cell, 2005. **4**(9): p. 1574-82.
724 64. Bok, J.W. and N.P. Keller, *LaeA, a regulator of secondary metabolism in Aspergillus*
725 *spp.* Eukaryot Cell, 2004. **3**(2): p. 527-35.
726 65. Stappers, M.H.T., et al., *Recognition of DHN-melanin by a C-type lectin receptor is*
727 *required for immunity to Aspergillus.* Nature, 2018. **555**(7696): p. 382-386.
728 66. Bultman, K.M., C.H. Kowalski, and R.A. Cramer, *Aspergillus fumigatus virulence*
729 *through the lens of transcription factors.* Med Mycol, 2017. **55**(1): p. 24-38.
730 67. Sanguinetti, M., et al., *Role of AFR1, an ABC transporter-encoding gene, in the in vivo*
731 *response to fluconazole and virulence of Cryptococcus neoformans.* Infect Immun,
732 2006. **74**(2): p. 1352-9.
733 68. Missall, T.A., et al., *Distinct stress responses of two functional laccases in*
734 *Cryptococcus neoformans are revealed in the absence of the thiol-specific antioxidant*
735 *Tsa1.* Eukaryot Cell, 2005. **4**(1): p. 202-8.
736 69. Zaragoza, O., et al., *The capsule of the fungal pathogen Cryptococcus neoformans.*
737 Adv Appl Microbiol, 2009. **68**: p. 133-216.
738 70. Ziegler, L., et al., *Functional characterization of the ferroxidase, permease high-affinity*
739 *iron transport complex from Candida albicans.* Mol Microbiol, 2011. **81**(2): p. 473-85.
740 71. Zahin, M., et al., *Identification of G-quadruplex forming sequences in three manatee*
741 *papillomaviruses.* PLoS One, 2018. **13**(4): p. e0195625.
742 72. Mishra, S.K., et al., *Characterization of highly conserved G-quadruplex motifs as*
743 *potential drug targets in Streptococcus pneumoniae.* Sci Rep, 2019. **9**(1): p. 1791.
744 73. Harris, L.M., et al., *G-Quadruplex DNA Motifs in the Malaria Parasite Plasmodium*
745 *falciparum and Their Potential as Novel Antimalarial Drug Targets.* Antimicrob Agents
746 Chemother, 2018. **62**(3).
747 74. Mishra, S.K., et al., *Characterization of G-Quadruplex Motifs in espB, espK, and cyp51*
748 *Genes of Mycobacterium tuberculosis as Potential Drug Targets.* Mol Ther Nucleic
749 Acids, 2019. **16**: p. 698-706.
750 75. Ruggiero, E. and S.N. Richter, *Viral G-quadruplexes: New frontiers in virus*
751 *pathogenesis and antiviral therapy.* Annu Rep Med Chem, 2020. **54**: p. 101-131.
752 76. da Silva, A.R., et al., *Berberine Antifungal Activity in Fluconazole-Resistant Pathogenic*
753 *Yeasts: Action Mechanism Evaluated by Flow Cytometry and Biofilm Growth Inhibition*
754 *in Candida spp.* Antimicrob Agents Chemother, 2016. **60**(6): p. 3551-7.

755

756

Table 1. PQS containing genes in *A. fumigatus* linked to virulence, drug resistance, or key biological processes.

Identifier	Gene	Description	Highest scoring sequence
AFUA_6G12400	*fks1*	1,3-beta-glucan synthase catalytic subunit	GGAAGGGGCTCGGGCATGGG*
AFUA_6G11390	*gel2*	1,3-beta-glucanosyltransferase	GGGCTTGGCGGAGGTCTGGGTTGTGTTGG*
AFUA_2G12850	*gel3*	Putative GPI anchored beta(1-3)glucanosyltransferase, belongs to the 7-member GEL family	GGAGGAGGAGG
AFUA_4G06890	*cyp51A*	14-alpha sterol demethylase	GGGTCCCGGGAACGGGCAAGGGG*
AFUA_7G03740	*cyp51B*	14-alpha sterol demethylase	GGGGTAGGGGGGCCCAGGG*
AFUA_1G15780	*leu2A*	3-isopropylmalate dehydrogenase with a predicted role in leucine biosynthesis	GGGGCGGAGG*
AFUA_4G06670	*aspf7*	Allergen Asp f 7	GGCCTGCGGAGTAGGGTAGATGG*
AFUA_5G11990	*alg3*	Ortholog(s) have dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase activity	GGTCTACGGGGCCCTTGG
AFUA_6G04360	*atrF*	Putative ABC transporter	GGGTATTTGTTTGGGGGGATGGGG
AFUA_3G03430	*sitT*	ABC multidrug transporter SitT, putative	GGCGAGGTCGGCCAGG*
AFUA_2G07900	*stuA*	Putative APSES domain transcription factor	GGGATCTGGCAAGCGGGGACGGG
AFUA_8G01670	*cat2*	Putative bifunctional catalase-peroxidase	GGTCGGGGTCTTGGTCCAGG*
AFUA_6G12150	*atfD*	Putative bZip-type transcription factor with similarity to atfA	GGGAGGTAGGCGTAGG*
AFUA_3G10930	*meaB*	Ortholog(s) have double-stranded DNA binding, sequence-specific DNA binding activity	GGGCACGCGGGAGACTCTGGGACCCGGG*
AFUA_5G03920	*hapX*	bZIP transcription factor HapX	GGAGCGGGAGGCGG*
AFUA_2G11780	*creA*	Carbon catabolite repressor A	GGCGGCGGCGG*
AFUA_4G09080	*sebA*	Putative transcription factor	GGGTGATGGAGAAAGGGATGGGCAGGG*
AFUA_3G11250	*ace2*	C2H2 transcription factor with a role in conidiophore development, pigment production, germination and virulence	GGTGGGGGTGG*
AFUA_3G11970	*pacC*	C2H2 finger domain transcription factor	GGTGGTGGCTACGGCCCTGG
AFUA_2G17560	*arp2*	1,3,6,8-tetrahydroxynaphthalene reductase arp2	GGGGAGGGGACCCTTGCATGGAGGGGGGG*
AFUA_2G17540	*abr1*	Multicopper oxidase abr1	GGCCGGTGGCTGG*
AFUA_2G17530	*arb2*	Laccase abr2	GGCGAGCGGTACTCGGTCATGG
AFUA_2G17550	*ayg1*	Heptaketide hydrolyase ayg1	GGACGGGGTTGGACTGG*
AFUA_2G17600	*pksP/alb1*	Conidial pigment polyketide synthase alb1	GGAATGCCGGGAGGACGGAAACACGG
AFUA_8G07080	*mep*	Putative secreted metalloprotease	GGCAAGGCGGCCTGG
AFUA_1G16190	*aspf9*	Cell wall glucanase	GGCTCTGGCTCCGGCTCCGG
AFUA_5G08030	*celA*	Ortholog(s) have spore wall localization	GGTTGTTGGGACGTGGGAGACGG*
AFUA_6G08640	*mip1*	Ortholog(s) have metalloendopeptidase activity, role in cellular iron ion homeostasis	GGGAGGTGGGTCGATGGGGATTGGG
AFUA_6G09720	*gliN*	N-methyltransferase gliN	GGGGGGCGCTTGGCTGGGAAGGCGGTTGGG*
AFUA_4G08580	*prx1*	Mitochondrial peroxiredoxin	GGTGTTACGGCCAGTGGAGGCGGG*
AFUA_6G09660	*gliP*	Nonribosomal peptide synthetase gliP	GGACCAGGCCGGGGCCCGGG*
AFUA_3G03420	*sidD*	Nonribosomal peptide synthetase 4	GGAAGGCATGGGAGG
AFUA_1G17200	*sidC*	Fusarinine C non-ribosomal peptide synthetase (NRPS), putative	GGGTATGGGATGCTGGATCCAGG*
AFUA_6G04820	*pabA*	Para-aminobenzoic acid synthetase, an enzyme catalyzing a late step in the biosynthesis of folate	GGTGGCGGGGG
AFUA_2G11970	*plaA*	Ortholog(s) have calcium-dependent phospholipase A2 activity	GGGCGGTGGGGTTGGGCAGTAGGG*
AFUA_2G16520	*pld2*	Putative phospholipase D	GGGCATTGTATTGTTGGGGGGAGGG*
AFUA_1G14660	*laeA*	Protein with similarity to protein methyltransferases, involved in regulation of secondary metabolism	GGGATGTGGGACAGGGATTTGGG
AFUA_2G09030	*dppV*	Secreted dipeptidyl peptidase	GGGAGGGAGAGGGAGGG*
AFUA_1G14550	*sod3*	Putative manganese superoxide dismutase	GGGTAGGGGAGGGGTGGG*

* Denotes that the sequence can be found on the complementary strand

761 **Table 2. PQS containing genes in *C. neoformans* linked to virulence, drug resistance,**
762 **or key biological processes.**

Identifier	Gene	Description	Highest scoring sequence
CNA00180	*vcx1*	Vacuolar calcium exchanger	GGCGGCGGAGG*
CNA00880	*itr1*	Myo-inositol transporter	GGCGATGGATTGTGGGACTTGG
CNA06800	*cas31*	Protein involved in gxm O-acetylation	GGAAACGGCGGATCCTTCGGGTGGGAAGG
CNA05840	*cap60/cel1*	Putative glycosyl hydrolase	GGACCCAGGAAACCTGGCCGTGGG*
CNA07000	*cap59*	Capsular associated protein	GGATAGGAGCTGGTACGG
CNA07090	*afr1*	ATP-binding cassette transporter	GGTAGGAAGGTGG*
CNA07240	*cap64/hrk1*	Serine/threonine-protein kinase	GGGGGAGGCGG
CNA07920	*sit1*	Siderophore iron transporter	GGAGGCGGTGG
CNB00720	*gsk3*	Similar to glycogen synthase kinase	GGGCGCAGGGCTGGGGGCTGGGGG*
CNB03210	*ktr3*	Alpha-1,2-mannosyltransferase	GGATGGAAGGGGGG
CNB03880	*cap2*	F-actin capping protein beta subunit	GGTGGGGGAGG
CNC02470	*crz1*	Phosphatase, putative	GGGAGGCTGGGAGGTGTTGGGAGAGGGG*
CNC04980	*cap64*	Capsular associated protein	GGGATGGGCCGGGCGGGG*
CNC06590	*hog1*	Mitogen-activated protein kinase, putative	GGATGGTGGATGGCGGATGGCGG
CND00550	*kre6*	Putative beta-glucan synthase	GGAGGTGGCGG
CND01080	*ctr4*	Copper uptake transporter	GGAAGGATGGTGGG
CND02320	*vph1*	Vacuolar (H+)-ATPase subunit	GGGTTGCAGGGACAGGATAATGGGGGG
CND02810	*cap2*	Capsular related protein	GGATGTTATGGTGCCAGATGGAGAGGGAGG
CND03510	*pmc1*	Calcium-transporting ATPase	GGGCCGAGCGGGGCGGAGGGGAGTCTGGG*
CND04030	*cap5*	Beta-1,2-xylosyltransferase	GGTTGGAGCGGGCGG
CNE00350	*cpk1*	Mitogen-activated protein kinase	GGGCGGGAGGGGGG*
CNF02940	*fsk*	Serine/threonine protein kinase	GGTAGGATGGTGGG*
CNG01250	*lac2*	Laccase 2	GGGGAGGAGAAGGGGAAGGGGAGGGG*
CNG01860	*zip2*	Zrt1 protein, putative	GGTCTTGGTCTAGGCGCTAGG
CNG02560	*uxs1*	UDP-glucuronic acid decarboxylase	GGGCTGGGCGGGAGCGGG*
CNG04090	*cmt1*	Alpha-1,3-mannosyltransferase	GGATCGGGGTGGGAGAGG*
CNG04130	*sec14-1*	Sec14 cytosolic factor	GGGGTTGGGAGGGTTGGAGGG
CNH00970	*rim101*	pH-response transcription factor pacC/RIM101	GGAGAGGGAGAGGGTTGAGG
CNH02270	*chs3*	Putative chitin synthase	GGGACAGGGCGGGATGGAGGG*
CNH00870	*usv101*	RNA polymerase II transcription factor	GGATGAGGGTGAGGATGAGG*
CNI02290	*cps1*	Polysaccharide synthase	GGGAAGAGAGGGCGGGGCGGG*
CNJ00330	*fzc27*	PRO1 protein	GGAGGAGGGGG*
CNJ01930	*pmt2*	Dolichyl-phosphate-mannose-protein mannosyltransferase	GGGCAAGTGGGAAAAGGGGTACGGCGCGGG
CNJ02230	*cna1*	Calcineurin A catalytic subunit	GGAGGCGGTGG
CNJ02920	*cir1*	Iron-responsive gata-type transcription factor	GGGAGAAACGTGAGAGGGAAAGGGAAAGGG
CNK00730	*cas33*	Capsular associated protein	GGGATGGAGAGGGCGGGGGGG*
CNK01140	*cap10*	Capsular associated protein	GGCGGGTAGGGAGTGG
CNL04070	*nrg1*	Transcriptional regulator	GGGGGGAAGAGACGCGGGGTGGTGAGGG
CNL04750	*cxt1*	Cryptococcal xylosyltransferase 1	GGGAGCTTGAGGGGGGAGGAGAAGGG
CNL04830	*ugt1*	Putative UDP-galactose transporter	GGATCTGGAGCTGGATTTGG
CNN01400	*ccc2*	Copper-exporting ATPase	GGCGGTACGGTAAACGGATTGGGG

763

764 * Denotes that the sequence can be found on the complementary strand

765

766

767

768 **Table 3.** PQS containing genes in *C. albicans* linked to virulence, drug resistance, or key
769 biological processes.

Identifier	Gene	Description	Highest scoring sequence
3643320	*set3*	NAD-dependent histone deacetylase complex component	GGTGGTGGCAGTGGTGGGGG
3639325	*crd1*	Cardiolipin synthase	GGGTTGGGGTTGGTTGTGG
3639369	*tup1*	Transcriptional repressor of filamentous growth	GGTGGGGCGGCAGG*
3634775	*arg81*	Zn(II)2Cys6 transcription factor	GGAATTGGAATTGGAACTGG
3643303	*ftr2*	High-affinity iron permease	GGCTAGGGGCGGGGCTAGG
3639306	*mvd1*	Mevalonate diphosphate decarboxylase	GGTTTGTGGCCTGGGAAATGGG
3643318	*ftr1*	High-affinity iron permease	GGCTAGGGGCGGGGCTAGGG
3640755	*ole1*	Fatty acid desaturase	GGGCTGGGGTGACTACTGGGGAGGG
3643806	*inp51*	Putative phosphatidylinositol-4,5-bisphosphate phosphatase	GGGGATTGGGGCAGTGATGGTGATGG
3637340	*rad32*	Protein similar to S. cerevisiae protein with role in nucleotide excision repair	GGTGGTGGTGG*
3646918	*ilv1*	Serine and threonine dehydratase	GGAGGAGGTGG
3643456	*lem3*	Putative membrane protein	GGTGGGTGGAGG
3638520	*rpd31*	SIN3 histone deacetylase complex component	GGTGGAGGTGG
3635225	*sfp1*	C2H2 transcription factor involved in regulation of biofilm formation	GGTTCGTGGCTAAAGGATGAGG
3637934	*mgm1*	Putative mitochondrial GTPase	GGGGGAGGGGG
3637862	*arg83*	GAL4-like Zn(II)2Cys6 transcription factor	GGTGGTGGTAGTGGTGGCGG
3643765	*pikA*	Phosphatidyl inositol kinase-like orf at the a mating type-like locus	GGGGATTTGGGGG
3636515	*psd2*	Phosphatidylserine decarboxylase	GGATTGGAGTGTGGTTGG
3646284	*atm1*	Member of MDR subfamily of ABC family	GGGGTGGAGCATTGGGG*
3639268	*sap5*	Biofilm-specific aspartyl protease	GGAAAAGGAATTCGGAAACAGG*
3647864	*yaf9*	Subunit of the NuA4 histone acetyltransferase complex	GGTGGTGAGGGTGG
3643166	*nrg1*	Represses yeast-hypha morphogenesis and hypha-specific gene expression	GGTACTTGGGAAGGATGGGGATAAGG*
3645897	*gcn2*	Kinase; has nonessential role in amino acid starvation response	GGTGGTGGTGG
3642576	*rrp9*	Ribosomal protein; mutation confers resistance to 5-fluorocytosine (5-FC), 5-fluorouracil (5-FU), and tubercidin (7-deazaadenosine)	GGAGGAGGAGG
3641877	*pex13*	Protein required for peroxisomal protein import mediated by PTS1 and PTS2 targeting sequences	GGTTATGGGAGTGGATATGG
3647296	*mss11*	Transcriptional regulator of filamentous growth MAPK cascade and cAMP-PKA pathways	GGAGGAGGAGG
3639727	*lrg1*	Similar to rho/rac GTPase-activating family of proteins	GGTGCTGGTGCTGGTGTAGG

770

771 * Denotes that the sequence can be found on the complementary strand

772

773

774

775

Figure Legends

Figure 1. G-quadruplex (G4) and i-motif (iM) structures. Representative examples of G4s and iMs. **(A)** Side **(B)** and top-down view of the human telomere DNA quadruplex in K^+ solution hybrid-1 form (PDB:2HY9). **(C)** The basic structure of the G-tetrad. **(D)** Side **(E)** and top-down view of an intramolecular i-motif DNA structure with $C{\cdot}C^+$ base pairing (PDB 1A83). **(F)** $C{\cdot}C^+$ base pairing found in i-motif structures. Images were generated using the Protein Imager software.

Figure 2. Heterogeneity of PQS across fungal divisions. The total number and frequency of PQS within 1362 fungal genomes were analysed using G4Hunter with a threshold of 1.5 and a window size of 30. The average number of PQS (**A** and **E**), PQS/kbp (**B** and **F**), PQS/GC% (**C** and **G**) and GC content (**D** and **H**) in fungi from the Ascomycota (n=1107), Basidiomycota (n=189), Mucoromycota (n=31), Zoopagomycota (n=12), Chytrdiomycota (n=12), Microsporidia (n=9), and Cryptomycota. (n=2). E to H contain boxplots with Tukey whiskers. The outliers are indicated by dots and the line within the boxplot is representative of the median value.

Figure 3. Higher genome GC-content is positively correlated with the frequency of PQS. The frequency of PQS relative to the GC content of fungi was plotted for **(A)** all fungal genomes in the study, **(B)** the Ascomycota, **(C)** the Basidiomycota, **(D)** the Mucoromycota, **(E)** the Zoopagomycota, **(F)** the Chytridiomycota, **(G)** the Microsporidia, and **(H)** the Cryptomycota. The Pearson correlation coefficient was used to determine the association between PQS and GC content. $P<0.05$ was considered statistically significant.

Figure 4. Heterogeneity of PQS across fungal sub-divisions. The total number and frequency of PQS within fungal sub-divisions were quantified using G4Hunter with a threshold of 1.5 and window size of 30. The average number of PQS **(A)**, PQS/kbp **(B)**, PQS/GC% **(C)** and GC content **(D)** in fungi from the Taphrinomycotina (n=14), Saccharomycotina (n=332), Pezizomycotina (n=761), Pucciniomycotina (n=24), Ustilagomycotina (n=31),

Agaricomycotina (n=134), Glomeromycotina (n=9), Mortierellomycotina (n=2), Mucoromycotina (n=20), Zoopagomycotina (n=3), Entomophthoromycotina (n=2), Kickxellomycotina (n=7), Chytridiomycetes (n=6), Monoblepharidomycetes (n=1), Neocallimastigomycetes (n=5), Microsporidia (n=9), and Cryptomycota (n=2). A to D contain boxplots with Tukey whiskers. The outliers are indicated by dots and the line within the boxplot is representative of the median value.

Figure 5. PQS distribution in fungal genera containing key species which cause disease in humans and plants. The frequency of PQS within genera containing key fungal pathogens were quantified using G4Hunter with a threshold of 1.5 and window size of 30. Fungal genera containing key plant pathogens are shown in blue, whilst genera containing key human pathogens are shown in pink. The boxplots display Tukey whiskers. The outliers are indicated by dots and the line within the boxplot is representative of the median value.

Figure 6. PQS in *Aspergillus* spp. are associated with genetic relatedness. *Aspergillus* spp., were categorised into sections based upon a phylogenetic tree generated by Steenwyk *et al*., [44]. The frequency of PQS was shown to be closely associated with the genetic relatedness of these fungi. Moreover, the PQS found in the *cyp51A* gene was conserved amongst species, but not found in *A. udagawae* or *A. turcosus*

Figure 7. Pathogenic *Aspergillus* and *Cryptococcus* species have lower PQS frequencies compared to their non-pathogenic counterparts. The frequency of PQS/kbp and PQS/GC% were quantified and compared between pathogenic and non-pathogenic species within *Aspergillus* spp., *Cryptococcus* spp., and *Candida* spp. The PQS/kbp in pathogenic and non-pathogenic species of **(A)** *A. fumigatus*, **(B)** *C. neoformans,* and **(C)** *C. albicans*. The PQS/GC% in pathogenic and non-pathogenic species of **(D)** *A. fumigatus*, **(E)** *C. neoformans,* and **(F)** *C. albicans.* Dots represent individual species within a genus. The error bars represent the SD. Asterisks indicate $p<0.05$.

827 **Figure 8. PQS in *A. fumigatus* Af293, *C. neoformans* JEC21, and *C. albicans* SC5314**
828 **can be found located throughout annotated genomic features.** The location of PQS found
829 100 bp before, within, and 100 bp after annotated genomic features with a G4Hunter score
830 >1.2. The total number of PQS in known genomic features in **(A)** *A. fumigatus*, **(B)** *C.*
831 *neoformans*, and **(C)** *C. albicans*. The frequency of PQS comparative to the genomic length
832 of the annotated features in **(D)** *A. fumigatus*, **(E)** *C. neoformans*, and **(F)** *C. albicans*.

833 **Figure 9. PQS can be found in genes encoding proteins involved in metabolite**
834 **interconversion, protein modification, nucleic acid binding, and transporter activity.**
835 The number of genes containing PQS were quantified and protein classes associated with the
836 genes were identified using PANTHER. The most highly represented groups are reported for
837 **(A)** *A. fumigatus* Af293, **(B)** *C. neoformans* JEC21, **(C)** and *C. albicans* SC5314.

838 **Figure 10. PQS can be found in genes which are highly upregulated during fungal**
839 **germination, in response to environmental stresses, and in biofilms in *A.***
840 ***fumigatus*.** Transcriptomes of *A. fumigatus* during germination, hypoxia, iron limiting
841 conditions, oxidative stress, and in biofilms, were analysed and the 20 most upregulated genes
842 were further investigated using QGRS Mapper and G4Hunter. (A) Genes that were highly
843 upregulated (compared to dormant or untreated *A. fumigatus* conidia) contained PQS. Data
844 points represent genes that contain PQS. (B) The PQS frequency in genes upregulated in
845 these conditions is higher when compared to the average PQS/kbp for the entire genome (red-
846 dashed line), analysed with the default G4Hunter settings (window size 25; threshold 1.2).
847 Data points represent PQS frequency of an individual gene. Boxplot represents the median,
848 maximum, and minimum values.

849

850

851

852

Supplementary Figure Legends

Supplementary Figure 1. (A-G) A breakdown of PQS frequencies per kilobase pair (kbp) for fungal genera from the Ascomycota. Each individual dot represents a species within the genera.

Supplementary Figure 2. (A-G) A breakdown of PQS frequencies per kilobase pair (kbp) for fungal genera from the Basidiomycota. Each individual dot represents a species within the genera.

Supplementary Figure 3. A breakdown of PQS frequencies per kilobase pair (kbp) for fungal genera from the **(A)** Mucoromycota, **(B)** Zoopagomycota, **(C)** Chytridiomycota, **(D)** Microsporidia, and **(E)** Cryptomycota. Each individual dot represents a species within the genera**.**

Supplementary Figure 4. The frequency of PQS per kilobase pair (kbp) **(A)** and relative to GC content **(B)** in *Aspergillus* species. Blue bars represent non-pathogenic species, whilst pink bars represent pathogenic species.

Supplementary Figure 5. The frequency of PQS per kilobase pair (kbp) **(A)** and relative to GC content **(B)** in fungi from the Pezizomycotina that are pathogenic in animals or plants, or non-pathogenic. Points represent individual species. Error bars represent the SD. * indicates a significant difference ($p<0.05$).

Supplementary Figure 6. The frequency of PQS per kilobase pair (kbp) **(A)** and relative to GC content **(B)** in *Cryptococcus* species. Blue bars represent non-pathogenic species, whilst pink bars represent pathogenic species.

Supplementary Figure 7. The frequency of PQS per kilobase pair (kbp) **(A)** and relative to GC content **(B)** in *Candida* species. Blue bars represent non-pathogenic species, whilst pink bars represent pathogenic species.

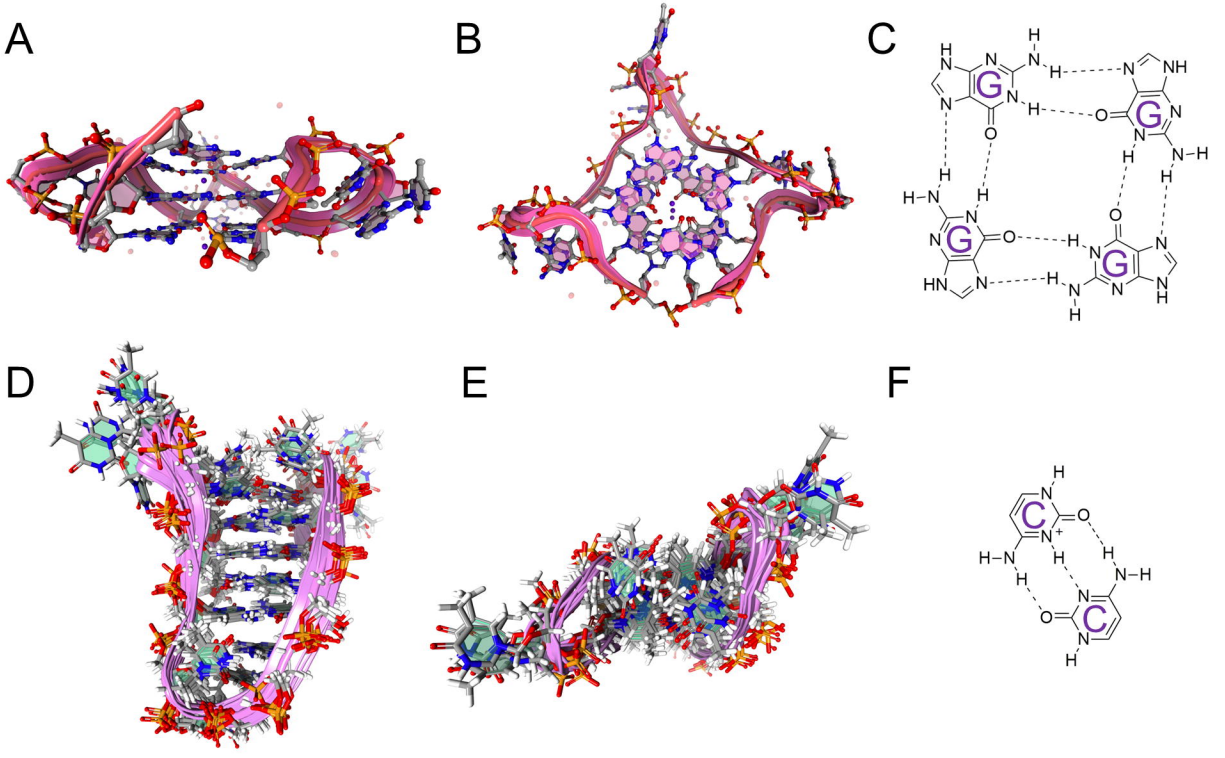
Supplementary Figure 8. The frequency of PQS per kilobase pair (kbp) **(A)** and relative to GC content **(B)** in fungi from the Saccharomycotina that are pathogenic in animals or plants, or non-pathogenic. Points represent individual species. Error bars represent the SD. ns indicates no significant difference ($p>0.05$).

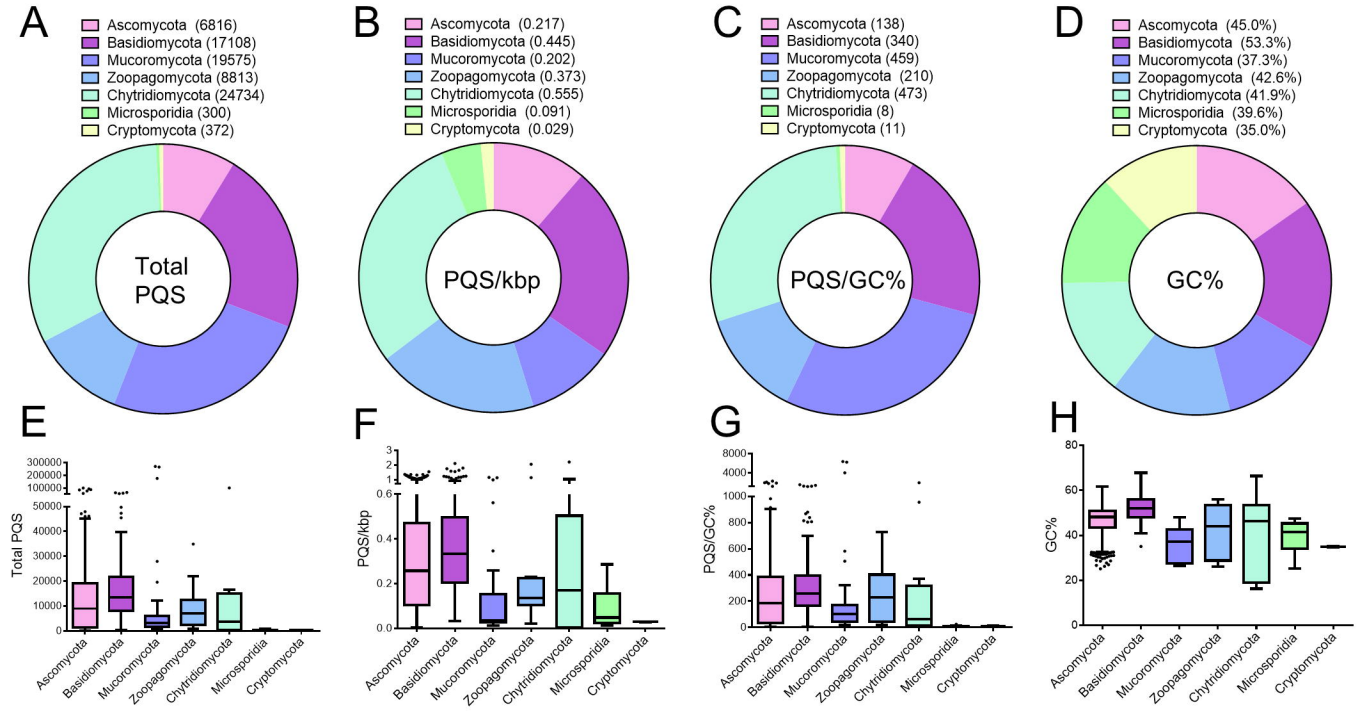
Supplementary Figure 9. The distribution and number of genes involved in biological processes in **(A and D)** *A. fumigatus,* **(B and E)** *C. neoformans,* and **(C and F)** *C. albicans.*

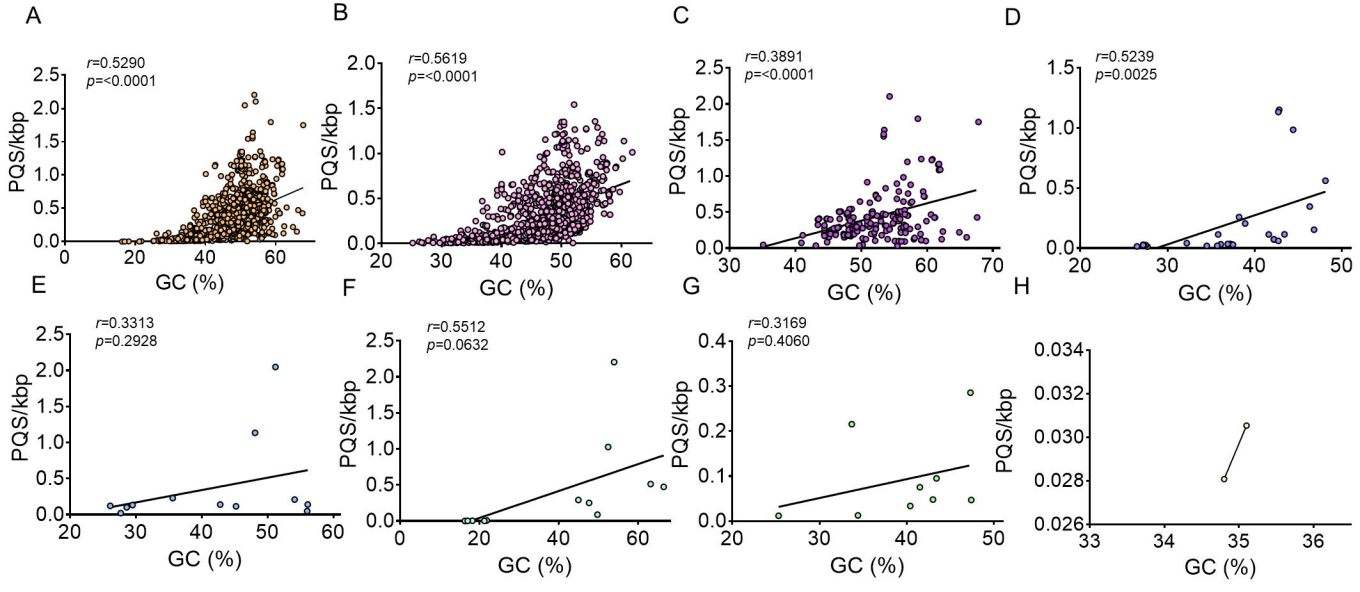
Supplementary Figure 10. The distribution and number of genes involved in molecular functions in **(A and D)** *A. fumigatus,* **(B and E)** *C. neoformans,* and **(C and F)** *C. albicans.*

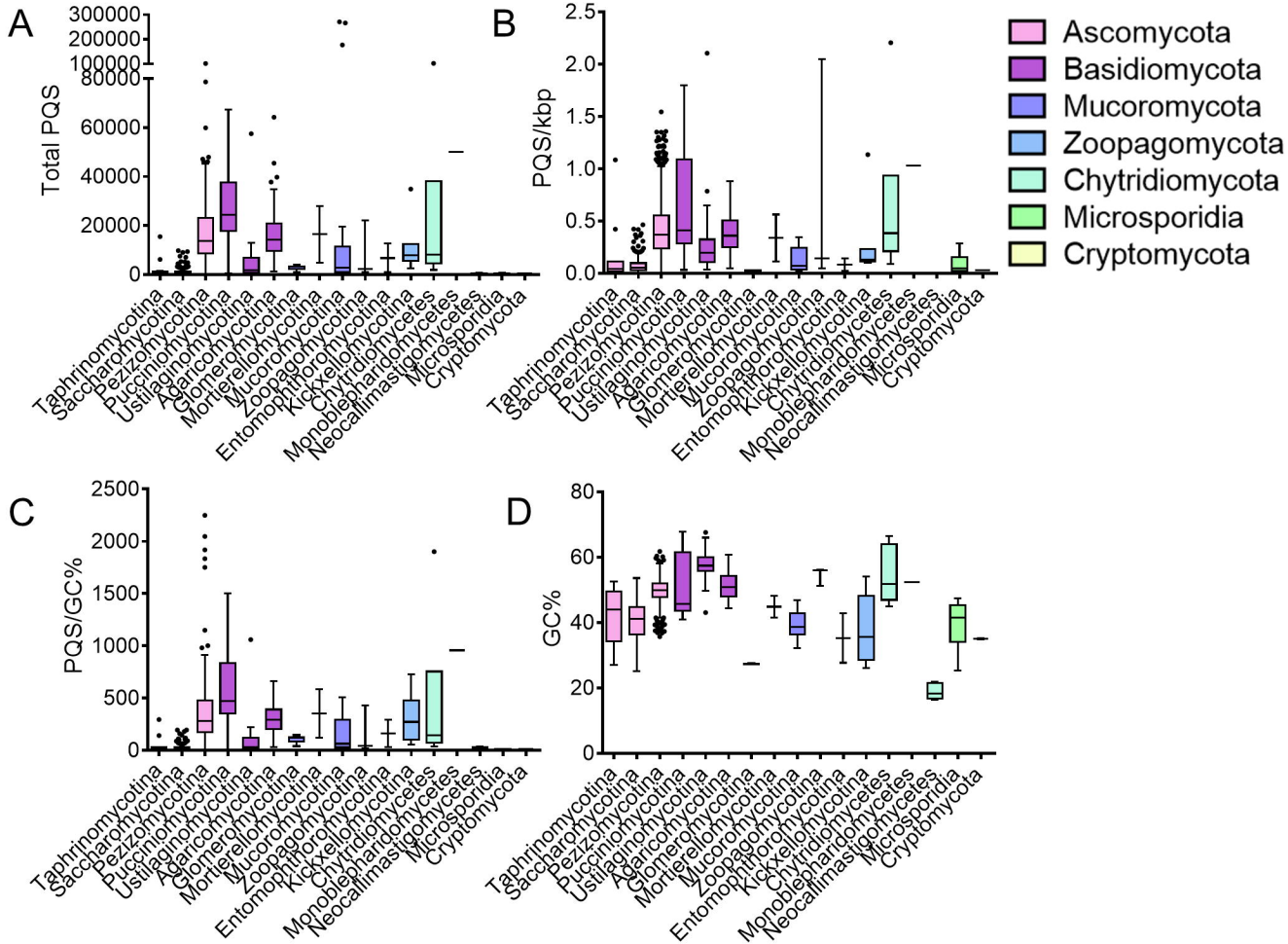
Supplementary Figure 11. The distribution and number of genes involved in cellular components in **(A and D)** *A. fumigatus,* **(B and E)** *C. neoformans,* and **(C and F)** *C. albicans.*

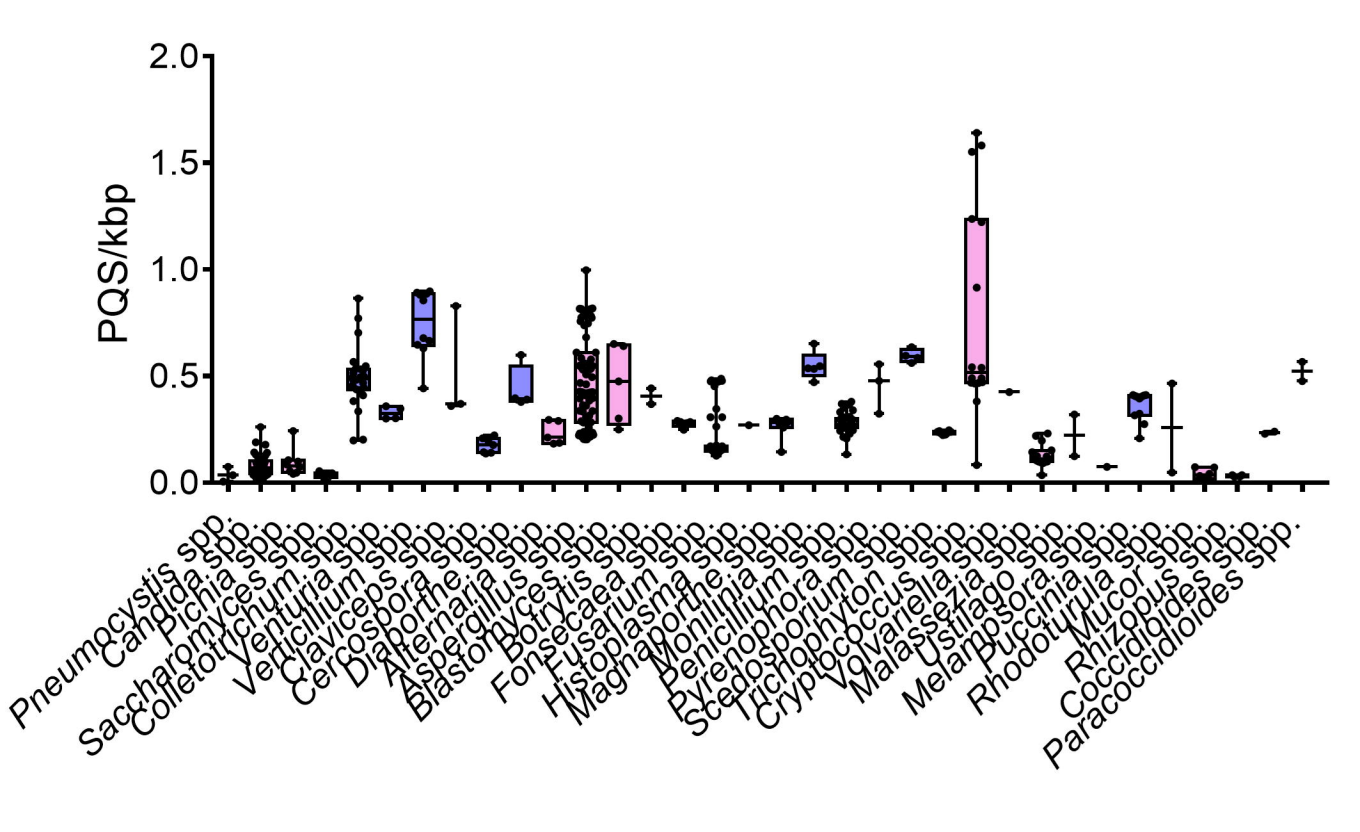
Supplementary Figure 12. The scores of PQS found in fungi compared to known quadruplex forming sequences. PQS in the identified genes were scored in QGRS Mapper and compared against the scores of known quadruplexes (n=94) to predict the propensity of PQS sequences to form quadruplex structures. Sequences with scores of 21, 42, and 63 in QGRS Mapper could form G4s containing 2, 3, and 4-tetrads, respectively. Sequences containing $G_{2+}L_{1-12}$ generally produced scores of 20 or 21, while those containing $G_{3+}L_{1-12}$ produced scores between 38-42. F and R refer to the coding and complementary strands, respectively.

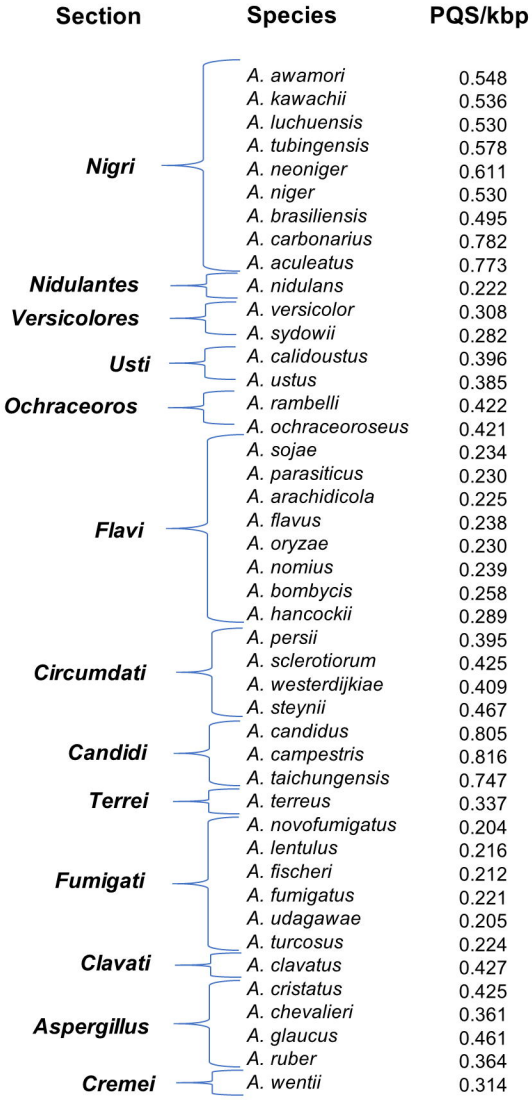

A
B
C
D
E
F


A
Ascomycota (6816)
Basidiomycota (17108)
Mucoromycota (19575)
Zoopagomycota (8813)
Chytridiomycota (24734)
Microsporidia (300)
Cryptomycota (372)
Total PQS
B
Ascomycota (0.217)
Basidiomycota (0.445)
Mucoromycota (0.202)
Zoopagomycota (0.373)
Chytridiomycota (0.555)
Microsporidia (0.091)
Cryptomycota (0.029)
PQS/kbp
C
Ascomycota (138)
Basidiomycota (340)
Mucoromycota (459)
Zoopagomycota (210)
Chytridiomycota (473)
Microsporidia (8)
Cryptomycota (11)
PQS/GC%
D
Ascomycota (45.0%)
Basidiomycota (53.3%)
Mucoromycota (37.3%)
Zoopagomycota (42.6%)
Chytridiomycota (41.9%)
Microsporidia (39.6%)
Cryptomycota (35.0%)
GC%
E
Total PQS
F
PQS/kbp
G
PQS/GC%
H
GC%
Ascomycota
Basidiomycota
Mucoromycota
Zoopagomycota
Chytridiomycota
Microsporidia
Cryptomycota


A
r=0.5290
p=<0.0001
PQS/kbp
GC (%)
B
r=0.5619
p=<0.0001
PQS/kbp
GC (%)
C
r=0.3891
p=<0.0001
PQS/kbp
GC (%)
D
r=0.5239
p=0.0025
PQS/kbp
GC (%)
E
r=0.3313
p=0.2928
PQS/kbp
GC (%)
F
r=0.5512
p=0.0632
PQS/kbp
GC (%)
G
r=0.3169
p=0.4060
PQS/kbp
GC (%)
H
PQS/kbp
GC (%)

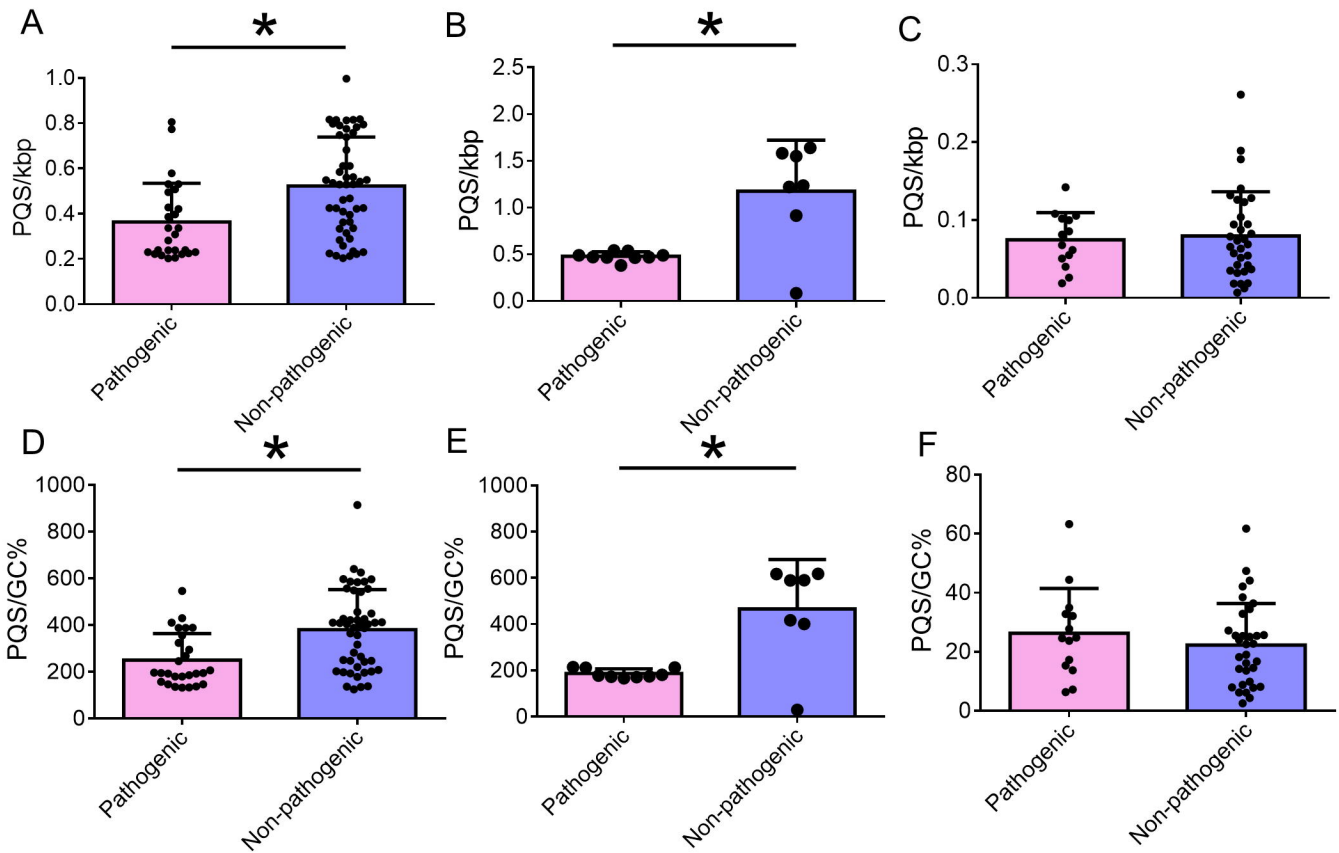

A
Total PQS
300000
200000
100000
80000
60000
40000
20000
0
B
PQS/kbp
2.5
2.0
1.5
1.0
0.5
0.0
C
PQS/GC%
2500
2000
1500
1000
500
0
D
GC%
80
60
40
20
0
Ascomycota
Basidiomycota
Mucoromycota
Zoopagomycota
Chytridiomycota
Microsporidia
Cryptomycota
Taphrinomycotina
Saccharomycotina
Pezizomycotina
Pucciniomycotina
Ustilaginomycotina
Agaricomycotina
Glomeromycotina
Mortierellomycotina
Mucoromycotina
Zoopagomycotina
Entomophthoromycotina
Kickxellomycotina
Chytridiomycetes
Monoblepharidomycetes
Neocallimastigomycetes
Microsporidia
Cryptomycota

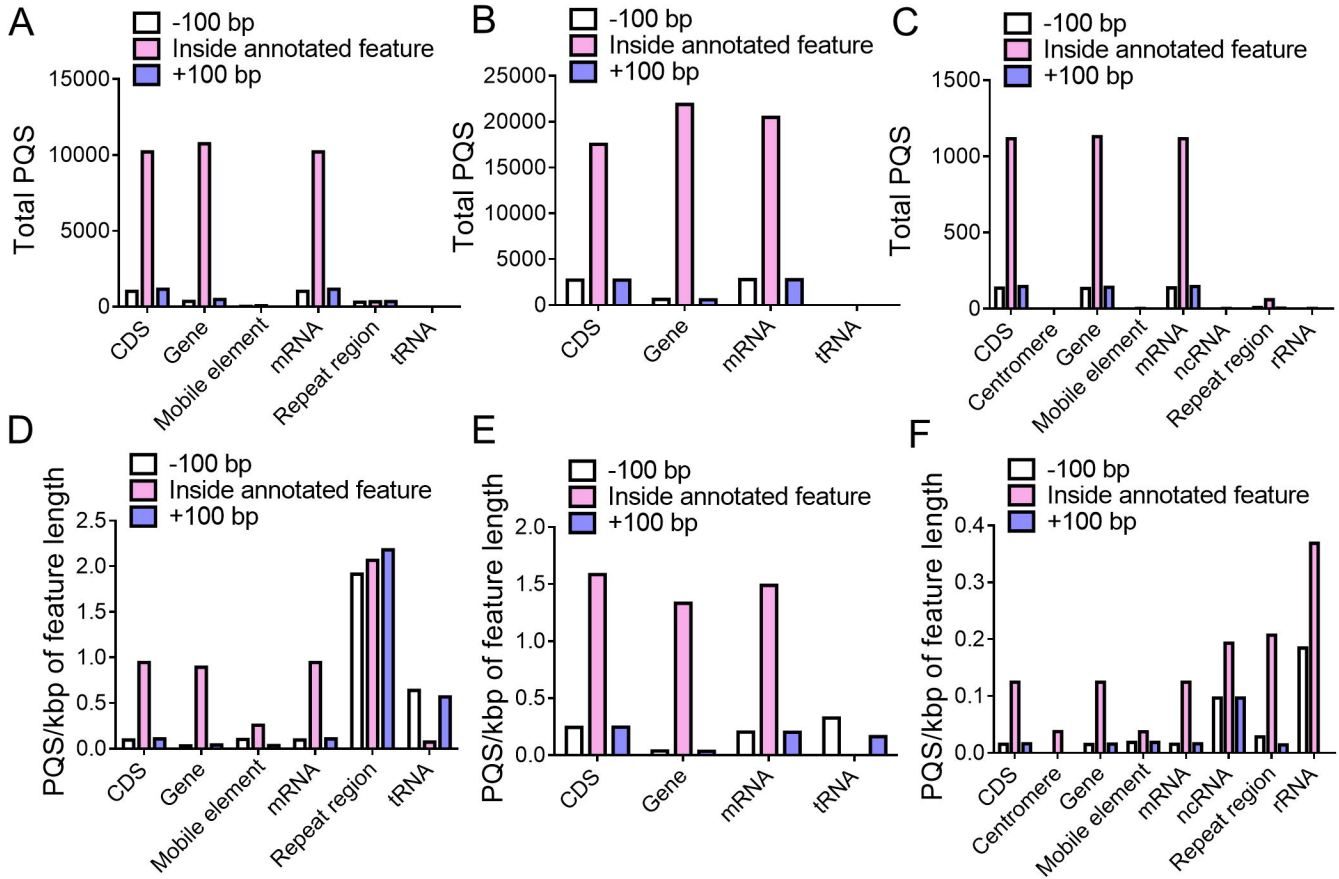

PQS/kbp
2.0
1.5
1.0
0.5
0.0
Pneumocystis spp.
Candida spp.
Pichia spp.
Saccharomyces spp.
Colletotrichum spp.
Venturia spp.
Verticillium spp.
Claviceps spp.
Cercospora spp.
Diaporthe spp.
Alternaria spp.
Aspergillus spp.
Blastomyces spp.
Botrytis spp.
Fonsecaea spp.
Fusarium spp.
Histoplasma spp.
Magnaporthe spp.
Monilinia spp.
Penicillium spp.
Pyrenophora spp.
Scedosporium spp.
Trichophyton spp.
Cryptococcus spp.
Volvariella spp.
Malassezia spp.
Ustilago spp.
Melampsora spp.
Puccinia spp.
Rhodotorula spp.
Mucor spp.
Rhizopus spp.
Coccidioides spp.
Paracoccidioides spp.

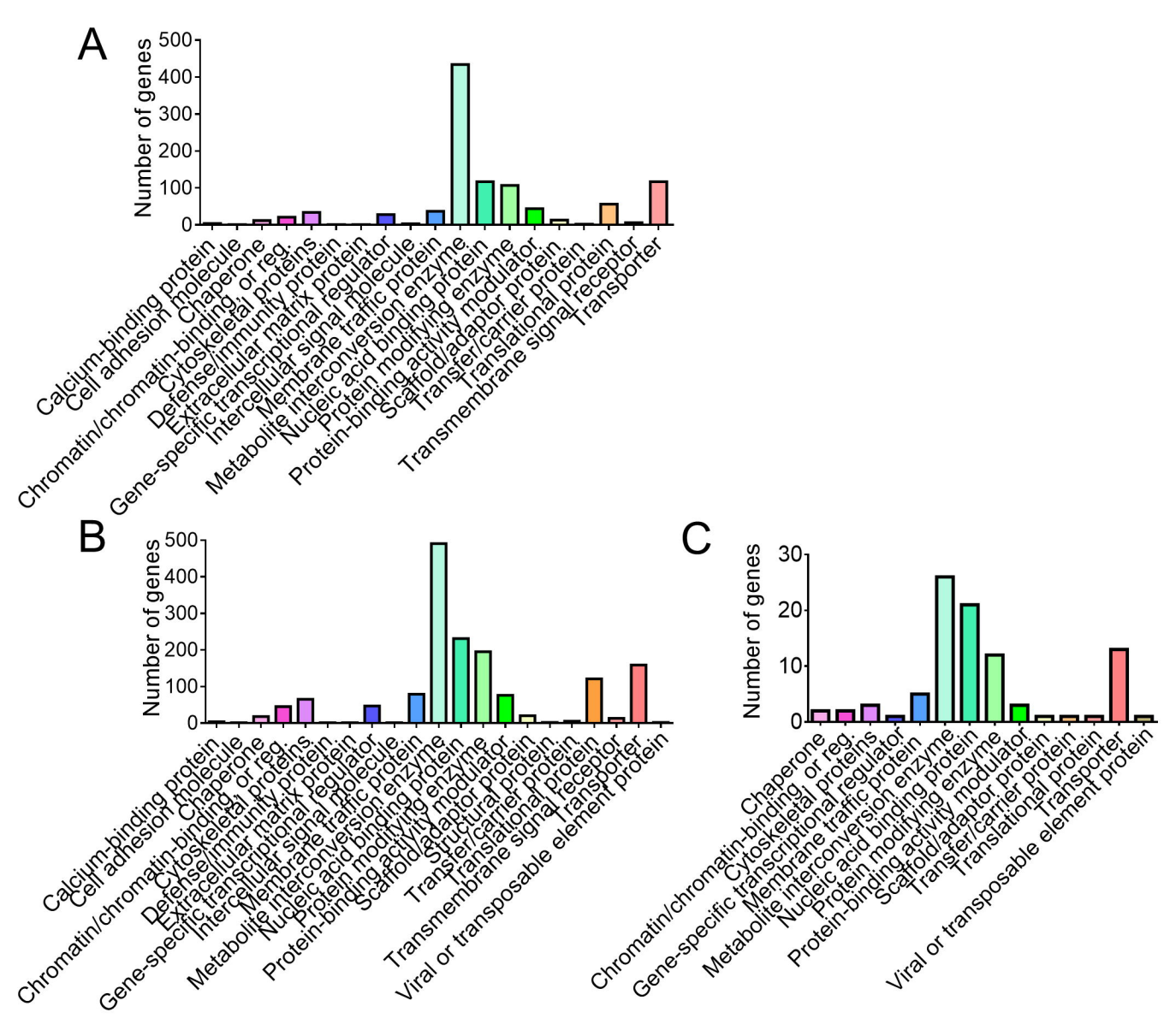


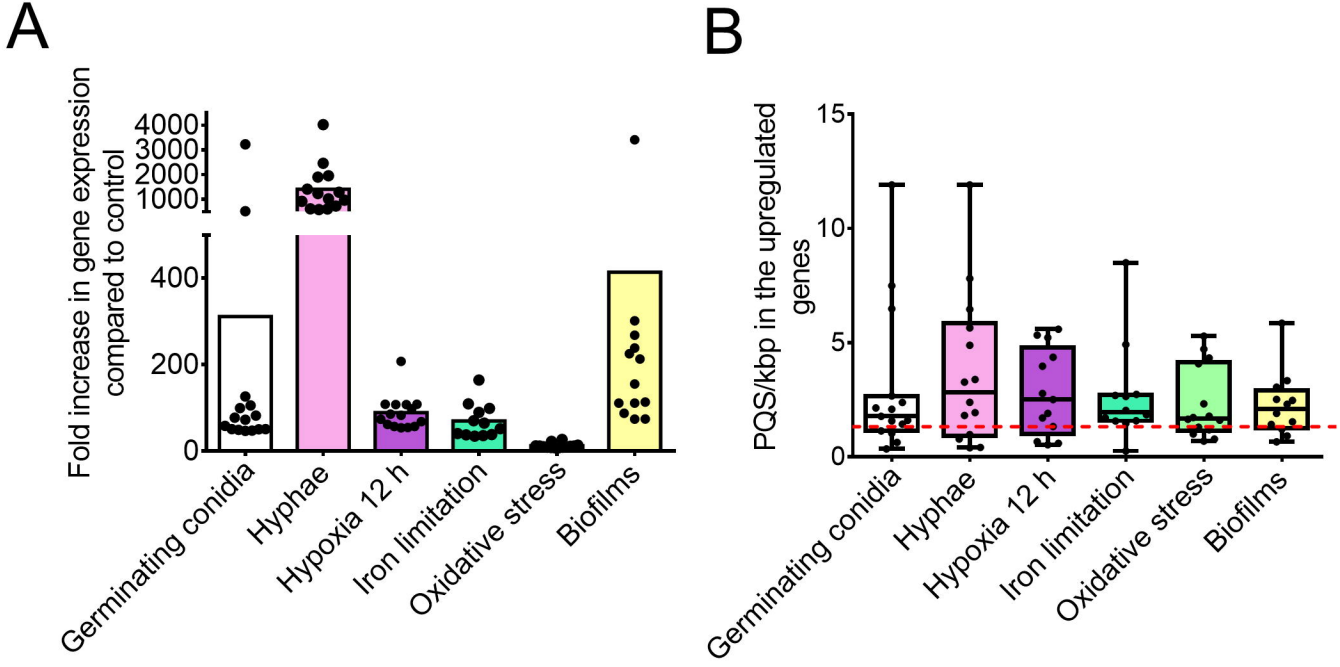
A. fumigatus – **GGG**TCCC**GGG**AA**GGG**CAA**GGGG** Position 408
A. novofumigatus – **GGG**CTTCC**GGG**AAC**GGG**CAA**GGG** Position 407
A. lentulus - **GGG**TTCC**GGG**AAC**GGG**CAA**GGGG** –Position 408
A. fischeri – **GGG**TTCC**GGG**AA**GGG**CAA**GGGG** – Position 408

A. udagawae and *A. turcosus* - this PQS cannot be found


A
B
C
D
E
F
*
PQS/kbp
PQS/GC%
Pathogenic
Non-pathogenic


A
-100 bp
Inside annotated feature
+100 bp
Total PQS
15000
10000
5000
0
CDS
Gene
Mobile element
mRNA
Repeat region
tRNA
B
-100 bp
Inside annotated feature
+100 bp
Total PQS
25000
20000
15000
10000
5000
0
CDS
Gene
mRNA
tRNA
C
-100 bp
Inside annotated feature
+100 bp
Total PQS
1500
1000
500
0
CDS
Centromere
Gene
Mobile element
mRNA
ncRNA
Repeat region
rRNA
D
-100 bp
Inside annotated feature
+100 bp
PQS/kbp of feature length
2.5
2.0
1.5
1.0
0.5
0.0
CDS
Gene
Mobile element
mRNA
Repeat region
tRNA
E
-100 bp
Inside annotated feature
+100 bp
PQS/kbp of feature length
2.0
1.5
1.0
0.5
0.0
CDS
Gene
mRNA
tRNA
F
-100 bp
Inside annotated feature
+100 bp
PQS/kbp of feature length
0.4
0.3
0.2
0.1
0.0
CDS
Centromere
Gene
Mobile element
mRNA
ncRNA
Repeat region
rRNA


A
Number of genes
500
400
300
200
100
0
Calcium-binding protein
Cell adhesion molecule
Chaperone
Chromatin/chromatin-binding, or reg.
Cytoskeletal proteins
Defense/immunity protein
Extracellular matrix protein
Gene-specific transcriptional regulator
Intercellular signal molecule
Membrane traffic protein
Metabolite interconversion enzyme
Nucleic acid binding protein
Protein modifying enzyme
Protein-binding activity modulator
Scaffold/adaptor protein
Transfer/carrier protein
Translational protein
Transmembrane signal receptor
Transporter
B
Number of genes
500
400
300
200
100
0
Calcium-binding protein
Cell adhesion molecule
Chaperone
Chromatin/chromatin-binding, or reg.
Cytoskeletal proteins
Defense/immunity protein
Extracellular matrix protein
Gene-specific transcriptional regulator
Intercellular signal molecule
Membrane traffic protein
Metabolite interconversion enzyme
Nucleic acid binding protein
Protein modifying enzyme
Protein-binding activity modulator
Scaffold/adaptor protein
Structural protein
Transfer/carrier protein
Translational protein
Transmembrane signal receptor
Transporter
Viral or transposable element protein
C
Number of genes
30
20
10
0
Chaperone
Chromatin/chromatin-binding, or reg.
Cytoskeletal proteins
Gene-specific transcriptional regulator
Membrane traffic protein
Metabolite interconversion enzyme
Nucleic acid binding protein
Protein modifying enzyme
Protein-binding activity modulator
Scaffold/adaptor protein
Transfer/carrier protein
Translational protein
Transporter
Viral or transposable element protein


A
Fold increase in gene expression compared to control
4000
3000
2000
1000
400
200
0
Germinating conidia
Hyphae
Hypoxia 12 h
Iron limitation
Oxidative stress
Biofilms
B
PQS/kbp in the upregulated genes
15
10
5
0
Germinating conidia
Hyphae
Hypoxia 12 h
Iron limitation
Oxidative stress
Biofilms