Patterns of conservation and diversification in the fungal polarization network

Eveline T. Diepeveen[1], Valérie Pourquié[1,2], Thies Gehrmann[2,3], Thomas Abeel[2,4] & Liedewij Laan[1*]

[1] Department of Bionanoscience, Kavli Institute of NanoScience, Faculty of Applied Sciences, Delft University of Technology, P.O. Box 5046, 2600 GA, Delft, the Netherlands

[2] Delft Bioinformatics Lab, Faculty of Electrical Engineering, Mathematics and Computer Science, Intelligent Systems, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, the Netherlands

[3] Westerdijk Fungal Biodiversity Institute (CBS-KNAW), P.O. Box 85167, 3508 AD, Utrecht, the Netherlands

[4] Broad Institute of MIT and Harvard, Genome Sequencing and Analysis Program, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA

[*] Author for Correspondence: Liedewij Laan, Department of Bionanoscience, Kavli Institute of NanoScience, Faculty of Applied Sciences, Delft University of Technology, Delft, the Netherlands, +31 (0)15-2782856, l.laan@tudelft.nl

Data Deposition

Sequence alignments used are available upon request.

Abstract

The combined actions of proteins in networks underlie all fundamental cellular functions. Deeper insights into the dynamics of network composition across species and their functional consequences are crucial to fully understand protein network evolution. Large-scale comparative studies with high phylogenetic resolution are now feasible through the recent rise in available genomic datasets of both model *and* non-model species. Here we focus on the polarity network, which is universally essential for cell proliferation and studied in great detail in the model organism, *Saccharomyces cerevisiae*. We examine 34 proteins, directly related to cell polarization, across 200 fungal strains/species to determine evolution of the composition of the network and patterns of conservation and diversification.

We observe strong protein conservation for a group of 16 core proteins: 95% of all strains/species possess at least 75% of these proteins, albeit in varying compositions. We find high levels of variation in prevalence and sequence identity in the remaining 18 proteins, resulting in distinct lineage-specific compositions of the network in the majority of strains/species. We test if the observed diversification in network composition correlates with potential underlying factors and find that lineage, lifestyle, and genetic distance co-vary with network size. Yeast, filamentous and basal unicellular fungi form distinctive groups based on these analyses, with substantial differences to their polarization network. Our study shows that evolution of the fungal polarization network is highly dynamic, even between closely related species, and that functional conservation is achieved by varying structural conservation of the fungal polarization proteins.

Keywords

Protein network evolution; Cell polarity; Protein network; Evolution, Fungi; Adaptation.

Introduction

All cells are maintained through fundamental cellular functions, such as respiration, biosynthesis, homeostasis and reproduction, that are crucial to the cell's overall existence. These complex functions are carried out by the combined action of proteins in protein networks with distinct cellular tasks (Papin et al. 2005; Pawson & Nash 2003). Through evolution of protein networks, by means of e.g., amino acid mutations, network expansion/reduction, and interaction effects, diversity is generated that can ultimately lead to the evolution of new functions (see Gladieux et al. 2014 for a list of reviews). Protein networks can evolve by changes in network composition, changes in levels of protein conservation and divergence and changes in expression levels, each with potential functional consequences (Schüler & Bornberg-Bauer 2010; Voordeckers et al. 2015). Comparative genomics and/or interaction studies of protein networks, such as the citric-acid cycle (Huynen et al. 1999), mitotic checkpoint (Vleugel et al. 2012), and the mitogen-activated protein kinase pathway (Mody et al. 2009) illustrate such patterns. Most of these studies have examined overall protein network evolution by means of cross kingdom comparisons, covering ~20 divergent species. Although such approaches are insightful for testing if proteins are commonly found in distant clades of the tree of life, patterns such as parallel evolution are difficult to disentangle because of the lack of phylogenetic resolution. Examining (the factors promoting) protein network evolution in more species, especially at phylogenetic dense levels, is essential to gain deeper insights into the dynamics of protein network evolution.

What factors promote protein (network) evolution? Numerous factors have been presented that influence the evolution of individual proteins (see Pál et al. 2006; Zhang & Yang 2015). To simplify, these factors can be divided into two broad categories: sources of genetic variation, those relating to regional genomic properties, such as variation in mutation

or recombination rate; and selection on genetic variation, factors dependent on specific protein properties, such as the proportion and distribution of sites that are involved in a specific functions, protein structure, expression level, and competition or adaptation (Pál et al. 2006). Although there are clear examples to illustrate these individual factors, the factors often do not act independently, making it hard to identify the relative importance of each factor. In yeast, for instance, the functional importance of a protein influences the rate of protein evolution (Wall et al. 2005; Drummond 2005; Hirsh & Fraser 2001), non-essential genes evolve on average faster than essential genes (Wall et al. 2005), and loci with more protein-protein interactions evolve on average slower (Jordan et al. 2003). Various studies have shown that expression rates have the most prominent effect on the rate of protein evolution (Wall et al. 2005; Drummond 2005).

Although selection acts on the outcome of fundamental cellular processes, specifically the phenotype, there are substantial differences in characteristics of proteins within a single network. As a consequence, selection can act differently on different proteins in the same network. For instance, proteins within a network can vary in the number and type of interactions, the position within the network (e.g., central versus peripheral), and the overall number of incorporated proteins in the network. Assessing the potential role of these factors, especially with multi-species comparisons, is crucial for understanding protein network evolution. Early comparative genomics studies, such as by Huynen *et al*. (Huynen et al. 1999), indicate that protein networks, even those involved in fundamental biological processes such as energy release by means of the citric-acid cycle, are characterized by variation in the composition and presence of specific proteins. Protein networks can change compositions by losing proteins or including novel proteins brought forward, for example through duplication events through neo- or subfunctionalization (Evlampiev & Isambert 2008; 2007). They can compensate for loss of proteins and new functions can evolve (Schüler

& Bornberg-Bauer 2010). By inferring the evolutionary dynamics of protein networks, various patterns have emerged. Proteins with many interactions in the network ('hub' proteins) evolve slower than average (Fraser et al. 2002; Kim et al. 2006), especially when they use multiple binding interfaces (Kim et al. 2006). Interacting proteins evolve at similar rates (Fraser et al. 2002). Overall, protein networks are characterized by conservation in topology and function, but also by substantial levels of divergence in network constitution among species (Liang et al. 2006; Vleugel et al. 2012). Protein networks are thus affected by a combination of factors, including positive selection, selection on protein function or structure and drift (Pál et al. 2006).

In this paper, we examine a fundamental protein network with the aim to elucidate its evolution and to identify the factors that contribute to it. We focus on polarity establishment, a process essential for proliferation in basically all unicellular and multicellular organisms. Polarity establishment, or the asymmetrical distribution of cellular components, has been described in great detail in the budding yeast *Saccharomyces cerevisiae* (Chang & Peter 2003; Martin & Arkowitz 2014; Madhani 2007). To polarize, cells need to break up the symmetrical distribution of cellular content and self-organize in a polarized way. The small GTPase, Cdc42, is a central key protein in this process (Etienne-Manneville 2004; Park & Bi 2007; Johnson 1999). The asymmetrical distribution of so-called polarization proteins, recruited by Cdc42, determines the site of local growth, or budding in the case of the budding yeast *S. cerevisiae*, which is essential for proper cell division and mating.

Cdc42 is a highly conserved protein throughout eukaryotes at both the sequence and functional level (Johnson 1999; Martin 2015) and its activity is regulated through well-documented feedback mechanisms (Martin 2015; Goryachev & Pokhilko 2008; Irazoqui et al. 2003; Wedlich-Soldner et al. 2003). The proteins that directly interact with Cdc42 can be divided into five groups: the GTPase activating proteins (GAPs), that hydrolyze GTP to GDP

and change Cdc42 to its inactive state; the guanine nucleotide exchange factors (GEFs), that catalyze the exchange of GDP for a new GTP molecule which activates Cdc42; the GDP dissociation inhibitors that extract Cdc42 from the membrane (Rdi1 is the only GDP dissociation inhibitor in budding yeast (Richman et al. 2004)); proteins involved in regulatory mechanisms, such as positive feedback (e.g., the scaffold protein Bem1 (Butty et al. 2002)); and a wide range of Cdc42 effector proteins which are activated by the active GTP bound state of Cdc42 (Figure 1A). Examples of Cdc42 effector proteins are the p21-associated kinases (PAK) Ste20, Cla4 and Skm1 (Johnson 1999), and the GTPase Interactive Components Gic1 and Gic2 (Brown et al. 1997). These proteins co-localize with Cdc42 during polarity establishment and form a protein complex by recruiting other proteins that are needed for proper actin and microtubule polarization (Johnson 1999; Brown et al. 1997; Drees et al. 2001). Various factors that could influence protein network evolution, such as the number of genetic and/or physical interactions and expression levels, have been determined for the proteins in this network in budding yeast (see Figure 1B). We therefore investigate the protein network evolution of polarization establishment among the ecologically and genetically highly diverse clade: the Fungi (Galagan et al. 2005; Ebersberger et al. 2012; Mueller & Schmit 2007).

The eukaryote kingdom of fungi is estimated >760 million years old (Lucking et al. 2009) and consists of up to 5.1 million estimated extant species (O'Brien et al. 2005). It includes an abundance of species with ecological, agricultural, medical and scientific relevance. Lifestyles can be restricted to a unicellular lifestyle, either yeast-like or non-yeast as observed in the basal clade of Microsporidia, or multicellular (i.e., filamentous species), or can consist of different stages, switching between two or more lifestyles (i.e., di-, trimorphic species). The wealth of different ecologies together with the available genomic and phenotypic resources and tools, such as the Saccharomyces Genome Database

148 (www.yeastgenome.org), make fungi an excellent tool for comparative studies between
149 ecologically highly diverse, but also relatively closely related species. A vast increase of
150 available fungal genomic datasets, especially fueled by initiatives such as the Fungal Genome
151 Initiative (Rhind et al. 2011), the FungiDB (Stajich et al. 2011) and the 1K fungal genomes
152 project (http://1000.fungalgenomes.org/home/; see also Sharma 2016), took place in the last
153 five years prior to this study and provides the desirable scale of data to reach high
154 phylogenetic resolution.

155 Although the processes of cell polarity and morphogenesis have been studied
156 extensively in *S. cerevisiae* (Bi & Park 2012; Chant 1999; Pruyne & Bretscher 2000; Pruyne
157 et al. 2004; Park & Bi 2007; Drees et al. 2001; Chang & Peter 2003; Martin & Arkowitz
158 2014; Madhani 2007), it is unknown to what extent the network's topology is conserved
159 across the fungal phylogeny, mainly because only a small number of divergent species has
160 been examined, which are characterized by variation in both network composition and
161 phenotypes (Diepeveen et al.). Due to its fundamental function in cell proliferation, the
162 polarization protein network is hypothesized to be a conserved system (Chang & Peter 2003;
163 Pruyne et al. 2004). In fact, several members of the polarization protein network, such as
164 Cdc42, Cdc24 and Sec15, are found to be essential in *S. cerevisiae* (Liu et al. 2015).
165 Interestingly, previously we showed that, under laboratory settings, the polarization network
166 in *S. cerevisiae* is able to adapt to genetic perturbations to one of the core proteins: Bem1,
167 which regulates Cdc42 (Laan et al. 2015). It is unknown to what extent this represents
168 adaption under natural conditions. Thus, there is some information available on the
169 conservation and evolvability (i.e., the ability of a species to evolve adaptive diversity) of a
170 small number of individual polarization proteins, but a larger screen to quantify the
171 evolutionary conservation across large phylogenetic distances is currently lacking.

172 Here we take advantage of the (newly) available tools and data in order to untangle
173 patterns of protein network evolution with high phylogenetic resolution within a single, but
174 phenotypically diverse kingdom. We aim to elucidate lineage-specific, independently
175 recurrent and/or conserved patterns of protein network composition, and levels of protein
176 conservation (i.e., both the presence of the protein in a species and sequence conservation)
177 and divergence of 34 polarization loci among 200 closely related fungal species. We aim to
178 elucidate factors involved in evolution of this particular and fundamental protein network.

Material and Methods

Focal (non-)polarization protein list and selected strains and species

We selected 34 polarization proteins (see Supplementary File 1, Figure 1) based on described physical or genetic interaction with the small GTPase, Cdc42, a key regulator of polarization (Etienne-Manneville 2004; Park & Bi 2007) and/or described functions of the protein in the polarization network on the Saccharomyces Genome Database (SGD; www.yeastgenome.org; June 2015; Cherry et al. 2011). For each of these proteins we downloaded the amino acid sequences from the SGD for *Saccharomyces cerevisiae* and used these sequences as reference or input query in the subsequent analyses (described below).

We obtained the full proteomes of 407 available fungal strains and species from Ensembl Fungi release 27 (June 2015; Supplementary File 1; Kersey et al. 2016) and created an initial proteome database (PDB) with the BLAST command line applications (version 2.2.31; Neumann et al. 2014). The proteome of *S. cerevisiae* (strain ATCC 204508/S288c) was downloaded from UniProt (www.uniprot.org). To perform the phylogenetic and reciprocal BLAST search analyses in an effective manner, we discarded strains and species with low quality proteomic data. We ran in-house scripts to determine the 'optimum' number of species that shared an 'optimum' number of communal proteins. In short, we calculated the number of shared protiens across the range of the 407 initial strains and species. We selected an 'optimum' of 200 strains and species (Supplementary File 1) for which we constructed the final PDB based on their proteomes as described above. The 200 strains and species had 86 proteins in common (Supplementary File 1). Note that these 86 proteins do not include any of the 34 polarization proteins. These protein sequences were used for the phylogenetic tree construction. We also constructed a *S. cerevisiae* (S288c) specific PDB for the reciprocal BLAST analysis.

204 *Phylogenetic tree construction*

205 We retrieved the amino acid sequence of the 86 shared proteins (combined length of
206 49484 aa in *S. cerevisiae*) for the 200 strains and species. Alignments were made for each of
207 the 86 proteins individually with Clustal Omega 1.2 (Sievers et al. 2011). The total length of
208 the full alignment was 158856 aa. We performed phylogenetic inference on a single
209 concatenated data file by means of the approximately maximum likelihood method as
210 implemented in FastTree (version 2.1.8; Price et al. 2009) with the JTT model of amino acid
211 evolution (Jones et al. 1992). The tree is mid-point rooted. The tree was visualized and edited
212 in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). We checked the obtained support
213 values and only reported them in the phylogeny, when they are < 0.9.
214
215 *Polarization protein conservation matrix*
216 We screened the final PDB (see above) for the presence of the 34 focal polarization
217 proteins through Reciprocal BLAST searches. We used the *S. cerevisiae* (strain ATCC
218 204508/S288c) amino acid sequence as input query. This was done to overcome
219 computational limitations with orthology prediction between the 200 species. Other tools,
220 such as DIAMOND (Buchfink et al. 2014) may ameliorate computational problems, but will
221 result in a loss of sensitivity which is required when examining such specific protein
222 networks. We performed local BLASTP searches without e-value constraints to be able to
223 find hits, even in highly diverged species. We took the associated protein sequence of all
224 obtained hits and performed local BLASTP searches again now against the *S. cerevisiae*-
225 specific PDB, again without e-value constraints. We selected the hit with the best e-value in
226 the initial BLAST search and called the hit a 'true match' only if the same protein ID was
227 retrieved as the original *S. cerevisiae* query. If there was no match in protein ID between hit
228 and query, then we called it a 'no match'. For the 'true matches' we corrected the similarity

scores (i.e., the number of positive-scoring matches) of the best BLAST hit to the query protein length of *S. cerevisiae*, thereby obtaining the fraction of similarity per hit. For the 'no-matches' we assigned a similarity score of 0. We generated a matrix of similarity scores by combining all the obtained scores and organizing them according to the species order as observed in our constructed phylogeny for each protein. These steps were performed with in-house python scripts. The generated data matrix was displayed as gray-scale matrix in R version 3.1.2 (R Core Team 2014).

We constructed a separate protein matrix for the group of 36 strains/species with high genome quality (i.e., for which the number of chromosomes is known and the number of scaffolds (roughly) equals the number of chromosomes; Supplementary File 1). For these species high quality genomes are available and therefore, overall genome quality should not affect our obtained results by means of incomplete genome sequences, and thus the possibility of absent protein sequences.

To determine if there was a group of proteins systematically present in a high number of species, we plotted the overall prevalence (%) across the 200 strains/species for each protein. We also plotted the overall prevalence of the two main fungal lineages Basidiomycota and Ascomycota separately in order to make sure that the proteins were present in high prevalence in both major clades. We used a 70% cut-off value in both clades as criteria for high prevalence and call the proteins that meet these criteria the conserved core of polarization proteins. We based this cut-off value on the observation that there is a gap in prevalence between 45% and 70% for the full dataset of 200 strains/species, dividing proteins into two groups. We also plotted the difference in prevalence between the Ascomycota and the Basidiomycota, to depict proteins that are particularly prevalent in either group.

Statistical analyses

We tested for a potential correlation between our obtained pattern of matches (i.e., the total number of 'true matches' per strain/species) and genome quality for which we used two assembly statistics. We obtained the number of contigs and N50 of contigs (i.e., length of contigs constituting 50% of the bases in the assembly), the number of scaffolds and the associated N50 for each of the genomes of the 200 selected strains/species from the European Nucleotide Archive (http://www.ebi.ac.uk/ena; Leinonen et al. 2010). Data was tested for normality with D'Agostino & Pearson omnibus normality tests as implemented in GraphPad Prism version 5.0 for Mac OS X (GraphPad Software, La Jolla California USA, www.graphpad.com). Correlations were tested with a spearman's rank correlations as implemented in GraphPad Prism.

We also tested if genome quality affected our obtained results of the full protein matrix by directly comparing the observed number of proteins per species clade-wise for four major lineages to the ones of the same lineages in the reduced matrix of 36 species. We used the Microsporidia, Saccharomycotina and Pezizomycotina lineages because these clades are also represented by at least 4 individuals in the reduced matrix. We divided the Saccharomycotina clade into two subclades because we observed distinctive lineage-specific patterns, such as gains/losses and distinctions in similarity scores, dividing them in two clades. Saccharomycotina clade 1 represents the Saccharomycetaceae family and the Saccharomycotina clade 2 includes the Debaryomycetaceae, Pichiaceae and Phaffomycetaceae families. Data was tested for normality with D'Agostino & Pearson omnibus normality tests as implemented in GraphPad Prism. We performed a Kruskal-Wallis test comparing the medians of the two matrices per defined lineage.

To test hypotheses about the cause(s) of differences observed in prevalence between the core and non-core proteins, we tested if these groups of proteins differed in number of genetic interactions, number of physical interactions and overall abundance of the proteins, as

279 indicative for gene expression level. Data for the number of interactions of the 34 proteins
280 with other members in the network in *S. cerevisiae* were obtained from SGD. We gathered
281 information about protein abundance (molecules/cells) in *S. cerevisiae* from Kulak *et al.*
282 (Kulak et al. 2014). Data was tested for normality with D'Agostino & Pearson omnibus
283 normality tests as implemented in GraphPad Prism. We also included the prevalence in the
284 analyses. We performed, a Kruskal-Wallis test (for interactions) and a Mann Whitney tests
285 (for prevalence and protein abundance) with GraphPad Prism.
286
287 *Multiple Factor Analysis*
288 Because we expected multiple continuous and categorical variables to potentially co-vary and
289 correlate with the observed number of proteins, we performed Multiple Factor Analysis
290 (MFA) on a dataset of the 200 strains/species. We included the following variables: proteins
291 (i.e., the total numbers of proteins as observed in the full protein matrix), genome quality (i.e.,
292 the number of Contigs and N50 of Contigs), Lineage (i.e., the main retrieved phylogenetic
293 clades: Microsporidia, Cryptomycota, Wallemiomycetes, Pucciniomycotina,
294 Ustilaginomycotina, Agaricomycotina, Taphrinomycotina, Saccharomycotina (clades 1 and 2,
295 as defined above), Pezizomycotina), genetic distance (in respect to the reference *S.*
296 *cerevisiae*), and lifestyle (i.e., unicellular, yeast, filamentous, dimorphic yeast-filamentous,
297 dimorphic yeast-pseudohyphal, and trimorphic). We calculated the genetic distance between
298 the examined strains/species and the reference. We used the concatenated amino acid
299 sequences of the 86 proteins from the phylogenetic analyses (158854 aa; see above) and
300 calculated the genetic distance by using the JTT model of amino acid evolution in MEGA 7:
301 Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets (Kumar et al.
302 2016). We obtained the lifestyle information from the Fungal Databases of the CBS-KNAW
303 Fungal Biodiversity Centre (http://www.cbs.knaw.nl/) and literature (Nagy et al. 2014;

304 Bastidas & Heitman 2009; Gauthier 2015). We performed the MFA with the FactoMineR R
305 package version 1.33 (Lê et al. 2008) package in R version 3.3.2 (R Core Team 2014) under
306 Rcmdr version 2.3-2 (Fox & Bouchet-Valat; Fox 2005; 2016). We used 3 quantitative groups:
307 genetic distance, genome quality (i.e., 2 variables) and proteins, and two qualitative groups;
308 lifestyle and lineage. Continuous variables were scaled and standard settings were used. We
309 first checked the eigenvalues for the first ten dimensions to determine the appropriate number
310 of dimensions to consider. In particular we checked for a drop in decline in variance (i.e.,
311 broken stick method; Jackson 1993). Length and directions of continuous variables were
312 plotted onto the first two dimensions and were visually checked. Partial axes for the fist first
313 three dimensions were visually checked. The five groups were plotted onto the first two
314 dimensions. We plotted individuals onto the first two dimensions and color-coded them
315 according to lineage.
316

Results

Phylogeny of 200 fungal strains/species is highly supported

In order to examine protein network evolution of fungal polarity establishment, we first estimated the phylogenetic relationship for our focal species. We inferred the phylogeny by means of the approximately maximum likelihood method on 86 homologous non-polarization proteins that these 200 strains/species have in common (see methods for details; total alignment length: 158,856 amino acids (aa; Figure 2)). Our phylogenetic tree is well-resolved and highly-supported and includes four main monophyletic phyla: the club fungi and relatives (Basidiomycota); sac fungi (Ascomycota); and the two basal phyla Cryptomycota and Microsporidia. Within the Basidiomycota, we found 100% support for the monophyletic subphyla Ustilaginomycotina, Pucciniomycotina, Agaricomycotina and Wallemiomycetes. Our observation of a basal position of the Wallemiomycetes species is in accordance with previous findings (Matheny et al. 2006; Zalar et al. 2005; Hibbett et al. 2007), although its exact position within the Basidiomycota phylum, and in relation to the Agaricomycotina in specific, has been varying (Matheny et al. 2006; Padamsee et al. 2012). Within the Ascomycota, we found full support for the monophyly of the Taphrinomycotina, Saccharomycotina and the Pezizomycotina, consistent with previous findings (Hibbett et al. 2007; Schoch et al. 2009). A discussion on relationships of deeper braches and clades is, however, beyond the scope of this work, although we observed almost exclusively high support values (i.e., >0.9) for the deeper branches and many relationships agree with previously published work (e.g., Saccharomycotina lineage; Shen et al. 2016).

Genome incompleteness has little impact on the protein matrix

We constructed a protein matrix consisting of the 34 polarization proteins and 200 strains/species based on our reciprocal BLAST approach (see methods for details). This way

we can both determine the presence of a protein in the examined strains/species and, when present, the level of sequence similarity in respect to *S. cerevisiae*.

Our approach resulted in a detailed protein matrix indicating the presence, level of divergence in respect to *S. cerevisiae*, and absence of the 34 polarization proteins in the examined strains/species (Figure 2). To determine if the quality of the genomic resources had an effect on the obtained results, we tested if there was a correlation between the total number of contigs, the N50 of contigs (i.e., length of contigs constituting 50% of the bases in the assembly) and the total number of proteins we obtained per strain/species. We observed that the quality of the 200 selected strains/species' genomic resources was highly variable (Supplementary File 1). Contig N50 range from 6 kb - 9 Mb. Genomic resources with short N50 may suffer from missing data such as missing exons and/or gaps (English et al. 2012), which could include, or result in, missing loci. A recent survey of >200 fungal genomes indicates that potentially only 40% reach the set cut-off for representative completeness (Cisse & Stajich). We found weak but significant correlations between the total number of obtained proteins per species and the number of contigs (Spearman rho = -0.29, P-value < 1.0×10^{-4}), and the contig N50 ($R = 0.32$, P-value < 1.0×10^{-4}; Figure 3). Thus, we potentially have cases of false negatives, or mismatches, in our protein matrix. In order to examine this pattern in more detail and to diminish the potential effect of missing proteins on our reciprocal BLAST analysis, we selected 36 strains/species with the most complete genomes (i.e., the number of scaffolds roughly equals the number of chromosomes) and constructed a reduced matrix (Figure 4). We then compared the overall number of observed proteins in the full matrix with the reduced matrix for all members of four lineages: Microsporidia, two Saccharomycotina clades, and the Pezizomycotina (Supplementary File 2). We did not observe significant differences in the number of proteins between the complete matrix and the reduced matrix (P-value > 0.05 for Microsporidia, Saccharomycotina I, Saccharomycotina II,

Pezizomycotina), indicating that the potential effect of missing data in our full matrix is, if any, minimal. Next, we examine the reduced matrix for patterns of protein conservation and divergence and compare them to the original matrix in order to test their generality.

The polarization protein network is dynamic and has a conserved core

To examine the fungal polarization network for patterns of protein network evolution, we screened the protein matrices for variation in protein prevalence (i.e., the overall number of species a protein is present in), levels of protein divergence and the repertoire of proteins (i.e., the composition of polarization network per species). We observed variation in the polarization network at different levels and magnitudes. We observed great variation at the level of amino acid similarity between different proteins, across different strains/species and between different lineages (Figure 4, Figure 2). For instance, Cdc42 is present with high levels of similarity in all species except in the Microsporidia, where it is not present in any of the examined species. Cla4, Ste20 and Cdc24 are found throughout the phylogeny as well, but their similarity scores vary greatly across species. These patterns were observed both in full and reduced matrices.

Next we examined the variability of the composition of the polarization protein repertoire between strains/species. We assessed the overall number of different protein combinations and the total number of unique repertoires across the reduced matrix (36 strains/species) and the full matrix (200 strains/species). We observed nearly identical fractions for the total number of different protein repertoires for the two matrices (i.e., 28/36 and 155/200). This indicates that the composition of repertoire is highly variable, with many different observed combinations of proteins. For both matrices the overall fraction of unique repertoires (i.e., protein repertoires observed in a single species) was also very similar, 0.64 (reduced matrix) and 0.67 (full matrix). We thus find that the majority of species are

392 characterized by a unique set of polarization proteins not found in other species. Interestingly,
393 we observed several specific combinations multiple times. For instance, we observed the
394 same pattern for all 7 Microsporidia species (Iqg1 and Ubi4). We observed most cases of
395 repeated repertoires in the species-rich and well-covered lineages Pezizomycotina (113
396 species in the full matrix) and Saccharomycotina lineages (20 species in the reduced matrix)
397 (Supplementary File 3). These combinations include prevalent and functionally diverse
398 proteins such as Rdi1, Bem1/2/3, Bni1, Axl2, Cla4, Ste20, Sec3/4/15, and Msb1/3.

399 To examine the overall prevalence of each protein across the 200 strains/species in
400 more detail, we screened the full matrix (Figure 5). We observed 23 proteins that were present
401 in ≥ 70% of all examined species (e.g., Iqg1), seven proteins are more commonly found in the
402 Basidiomycota (e.g., Bem2), and four proteins restricted to the Ascomycota (e.g., the paralogs
403 Msb3, Gic1 and Gic2). Most proteins are highest prevalent in the Ascomycota (Figure 5 top
404 panel). We observed a perceived threshold at ~60% prevalence for proteins across all species
405 examined that clearly divided the dataset (Figure 5). We found 11 proteins that are present in
406 less than 45% of the 200 strains/species, while the other 23 proteins are present in at least
407 70%. We used this 70% mark as cut-off value to determine conserved proteins (i.e., based on
408 prevalence) in both Ascomycota and Basidiomycota, individually. Using this, we excluded
409 those proteins that are more or less Ascomycota-specific, such as Nrp1. We observed the
410 following 16 proteins in high prevalence in both Ascomycota and Basidiomycota: Cla4, Axl2,
411 Rho3, Boi2, Ste20, Sec4, Bni1, Bem3, Spa2, Cdc24, Sec15, Ubi4, Bem1, Cdc42, Rdi1 and
412 Iqg1. We called these proteins the conserved core of polarization across fungi. We found this
413 full repertoire of core proteins in 66 out of 200 strains/species and >95% of all strains/species
414 had a protein network consisting of 12 or more core proteins (Supplementary File 3).
415
416 *Core proteins have higher protein abundance but not more interactions.*

As we observed a group of proteins at high prevalence across clades, we tested if there is a correlation between this conserved core of proteins and factors known to influence protein (network) evolution, such as number of protein-protein interactions and expression levels. We thus tested the following hypothesis: core proteins are conserved because they are either functionally important and/or because they are present in high quantity.

Core proteins had higher prevalence than non-core proteins (Figure 6A). We found no significant difference in the number of either genetic or physical interactions (based on observations in *S. cerevisiae*) between the core proteins and the non-core proteins (Figure 6B). We did find a barely significant difference in protein abundance (as measured as molecules per cell in *S. cerevisiae*; Kulak et al. 2014) between the core proteins and non-core proteins ($p=0.03$; Figure 6B). Core proteins have higher protein abundance than non-core proteins.

Lineage, lifestyle and genetic distance co-vary with protein network size

In order to characterize the 200 strains/species by different factors that could influence protein network evolution and in specific could correlate with the observed patterns of differences in overall number of polarization proteins, we performed Multiple Factor Analysis. We considered the following factors: size of protein repertoire (i.e., the total number of proteins we detected with the Reciprocal BLAST search per strain/species), lifestyle, lineage, genome quality (i.e., the number of contigs and the N50) and genetic distance to the reference *S. cerevisiae*, based on 86 shared proteins.

To determine the adequate number of dimensions to screen, we used the broken stick method (Jackson 1993). We found a drop in variance after the third dimension (Supplementary File 4), therefore we only considered the first three dimensions. Dimension 1 is constructed based on four groups: lineage (contribution is 27.44%), lifestyle (24.72%),

442 genetic distance (23.39%) and proteins (23.23%). Dimension 2 and 3 are both based on
443 lineage (44.16%; 50.61%) and lifestyle (43.87%; 46.31%). Together these three dimensions
444 accounted for 40.70% of the variance in the data. Dimension 1 explained 20.28%, dimension
445 2 12.37% and dimension 3 8.06% (Figure 7D). We did not find a substantial (>0.5)
446 contribution of genome quality, indicating that the number of contigs and/or N50 of contigs
447 did not explain the variation in the protein repertoire and other examined factors.
448 Supplementary File 4B shows that lifestyle and lineage vary together and that they vary
449 together with protein repertoire and genetic distance. Supplementary Files 4C & 4D indicate
450 that protein repertoire and genetic distance only correlate with dimension 1, while lineage and
451 lifestyle also correlate with dimensions 2 and 3.

452 We plotted the 200 strains/species onto the first two dimensions to visually examine if
453 they cluster to specific patterns based on e.g., morphology or descent (Figure 7A). As four
454 factors mainly contributed to the axes, various patterns were observed. Overall, prevalence
455 (proteins) declined horizontally (from left to right), with an increase in genetic distance.
456 Lineage and lifestyle contributed to both axes, and a clear distinction could be made between
457 three main clouds of individuals according to their morphology: the top left corner represents
458 the yeast-like fungi, the top right corner represents the unicellular non-yeast-like fungi, and
459 the lower center represents the filamentous fungi. The yeast-like fungi mainly constituted out
460 of the yeast and dimorphic (Y/P) lifestyles, while the lineages that mostly consist of the
461 filamentous lifestyle were found in the bottom center cloud (Figure 7C). We observed a split
462 in the Ustilaginomycotina species, as two species are yeast-like and two species are
463 filamentous. Interestingly, the Ascomycotan and mostly filamentous Pezizomycotina
464 clustered together with the Basidiomycotan individuals in the filamentous group. This
465 observation is not in line with the phylogenetic relationships between these clades, but it does
466 indicate that variation in the factor lifestyle is more similar for the Pezizomycotina and the

Basidiomycotan species, as opposed to the Pezizomycotina and other Ascomycotan species. We observed similar patterns when we examined these factors individually versus the protein repertoire (Figure 7E-G). Our results suggest that the Basidiomycota and the filamentous Ascomycotan species have a smaller repertoire than the yeast-like Ascomycota, and that the two basal lineages have an even smaller repertoire (Figure 7E). Lastly, we did observe a pattern of decreasing number of polarization proteins with greater genetic distance to our reference *S. cerevisiae* (Figure 7G).

Discussion

Here we assess evolution of the fungal polarization network at high phylogenetic resolution to examine patterns of lineage-specific, independently recurrent and/or conserved patterns of protein network composition along with levels of protein divergence. We observe that the fungal polarization protein network is characterized by both strong protein conservation *and* variation in protein prevalence and sequence similarity. Our results indicate that while certain proteins are nearly always needed potentially for specific functions (i.e., functional conservation), various other functional steps seem to be fulfilled by a variable repertoire, indicating flexibility in the network composition. Below, we discus these observations in context of protein network dynamics, functionality of the protein network, and potential causal factors of protein network evolution.

The fungal polarization network is highly variable at lineage-specific levels

It is clear that protein network evolution has a variety of outcomes, such as network expansion/reductions, interaction effects and protein divergence (Schüler & Bornberg-Bauer 2010; Voordeckers et al. 2015), brought forward by e.g., gen(om)e duplication, selection on protein function or structure and drift (Pál et al. 2006). We found that most proteins of the polarization network have high levels of divergence in amino acid sequence across fungi and that the specific build up of the protein repertoire per strains/species is highly variable. We find both variation at large phylogenetic distances, such as between subphyla, and between strains/species of the same clade. This indicates that, although the polarization network is involved in fundamental cellular functions across organisms, the network, as we know that in *S. cerevisiae*, is not a conserved entity. Work based on the first available fungal genomes reveal remarkable levels of divergence (Galagan et al. 2005), with even less than 50% similarity in amino acid sequence in comparisons of Ascomycotan species (Dean et al. 2005).

499 Screening these genomes for networks reveal that especially regulatory pathways are
500 recurrently characterized by substantial levels of variation, in that elements can be gained or
501 lost over time (Tanay et al. 2005; Muñoz et al. 2016; Tuch et al. 2008; Habib et al. 2012). Our
502 work provides further support for the eminent finding that proteomes and networks constantly
503 change (Coulombe-Huntington & Xia 2017), not only in Ascomycota as previously shown,
504 but also in Basidiomycota and basal lineages as Cryptomycota and Microsporidia.

505 The substantial levels of variation that we observed in the polarization network could
506 be caused by the remarkable differences in how fungal species polarize and grow (e.g.,
507 isotropic, (a)symmetric). In fact, we do find a clear clustering of yeast-like fungi, non-yeast
508 like unicellular fungi and filamentous fungi in our MFA plot. While budding yeast polarizes
509 in a switch-like way, filamentous species are characterized by continuous hyphal growth and
510 thus need a constant state of polarity. Differences at the protein levels between species with
511 differences in polarization/growth mode have also been described. The Rho GTPase Rac1 has
512 partly overlapping functions with Cdc42 in regulating polarization in a variety of filamentous
513 species (Banuett et al. 2008), but not in *S. cerevisiae*. We believe that the observed levels of
514 variation have a strong functional component.

515 To what extent does this high variability of the protein network affect functionality?
516 As functional studies are not available for the majority of examined species, we made use of
517 the functional classification of proteins of *S. cerevisiae* (see Figure 1). We found that over
518 90% of examined strains/species have at least one protein present from all nine defined
519 functional groups. This could imply that the overall functional pathway of polarity
520 establishment, by means of the regulation of Cdc42, is similar across the fungal tree.
521 Exceptions to this overall pattern of consensus are observed throughout the tree (e.g.,
522 Microsporidia, *Penicillium rubens*, *Suillus luteus*, *Serpula lacrymans*, lacking e.g., Cdc42

effectors and/or GEF proteins). Further functional exploration of protein networks in additional non-model species is needed to determine the level of orthology of these networks.

Variation in polarization network; from stark reductions to lineage-specific additions

We found high levels of lineage-specific patterns, of which various patterns coinciding with monophyletic clades. For instance, Msb3 and Axl2 are individually lost in various Pezizomycotina clades (Figure 2). The protein matrix also showed very similar patterns for the four Taphrinomycotina species, which are quite dissimilar from the other Ascomycota clades. These species have very distinct ecologies, as they are the only fission yeast species. We do observe further cases of lineage-specific patterns in clades with distinct ecologies.

We found that nearly all examined polarization proteins were absent in the Microsporidia (Figure 2), including most of the conserved core. The only proteins that we observed are Iqg1 and Ubi4. Interestingly, we did not observe this pattern in the other basal phylum, the Cryptomycota. We believe that our observation is a true lineage-specific loss in the Microsporidia, as the majority of the polarization proteins (29 out of 34 proteins) are found in non-fungal eukaryotes, such as animals, amoeba and/or plants (see Supplementary File 5). The genomes of the parasitic Microsporidia are known to be highly condensed and lack other essential proteins, such as MAP kinases and proteins involved in stress response (Miranda-Saavedra et al. 2007; Peyretaillade et al. 2011; Vivarès et al. 2002). This strong reduction in the proteome is hypothesized as an adaptation to their parasitic life style. It is currently not understood which proteins play a role in polarized cell growth in this genus.

In contrast to the strong reduction in the Microsporidia, we observed lineage-specific gain of polarization proteins in the budding yeast species Saccharomycetaceae. Bem4, Gic1 and Gic2 are all restricted to this clade (Figure 2). Various causes can be involved. Genome-wide comparisons across the eukaryote tree have identified an increase in proteins domains in

548 the lineage towards the Ascomycota (Zmasek & Godzik 2011). Furthermore, a whole genome
549 duplication occurred in the *Saccharomyces* lineage after the divergence from the
550 *Kluyveromyces* lineage, and has resulted in many duplicated genes (i.e., paralogs) and
551 instances of accelerated evolution (Kellis et al. 2004; Wolfe & Shields 1997). Our results
552 indicate that different processes have resulted in a myriad of lineage-specific patterns across
553 the fungal tree.
554
555 *The conserved core of polarization; functional versus structural conservation*
556 We observe a group of sixteen core proteins that are recurrently present in the vast majority of
557 examined species (Figure 5). Interestingly, the 16 core proteins cover all functional groups
558 from Cdc42 regulators and effectors to proteins involved in cytokinesis and exocytosis
559 (Figure 1). Even though this group consists of the most prevalent polarization proteins, it does
560 not represent the absolute minimal system needed for polarization. In fact, the majority of
561 species does not have the full set of core proteins (i.e., the complete core is present in 66 out
562 of the 200 strains/species), which can be seen as another indicator of high uniqueness of
563 structural constitution of the polarization network across fungi. Different strains/species might
564 achieve functional conservation of the core functions of the network by having different
565 combinations of core proteins. In fact, we observed 12 or more core proteins (i.e., 75%) in
566 191 strains/species (i.e., 95,5%). These results suggest that functional conservation of the
567 polarization network is high, but that structural conservation, in the sense of network
568 composition, of the individual proteins varies across the fungal strains/species.
569 Various protein characteristics have been elucidated that are (in part) responsible for
570 protein network conservation, such as position within the network, whether the proteins are
571 essential and the number of interactions (Liu et al. 2015; Giaever et al. 2002). We observed
572 high proportions of essential proteins (6 out of 7 essential proteins are core proteins) and short

single domain proteins (5 out of 10) for the core proteins (Figure 2). Selection is thought to be strong on these classes of proteins, because of their crucial functions and long protein domains (Pál et al. 2006; Buljan & Bateman 2009). These functional characteristics are based on studies in *S. cerevisiae* and could be less relevant in other species. We did not find significant differences in the number of genetic and physical interactions between the conserved core proteins and the non-core proteins. Interestingly, Cdc42, Bem1, Cdc24 and Cla4 were among the conserved core and have the most interactions with the other proteins. These proteins also take central parts in the polarization network, as key regulator (Cdc42; Etienne-Manneville 2004; Park & Bi 2007; Johnson 1999), scaffolding for protein complex (Cdc24 and Bem1; Butty et al. 2002), and signal transducing (Cla4; Johnson 1999). At the same time, we do find a low number of physical and genetic interactions for the core proteins, Rdi1, Axl2, Rho3. Our results indicate that factors like if a protein is essential and the number of interactions, only partially explain the conservation of core proteins. We did find a significant effect of protein abundance on the conservation of core proteins. This observation supports the hypothesis that conserved proteins are generally more expressed in a cell, as discussed previously (Wall et al. 2005; Drummond 2005).

Link to causal factors for variation in the polarization protein network

Here we aim to uncover potential causal factors influencing protein network evolution, in the sense of network size. Our MFA results show that the factors lifestyles, lineage and genetic distance co-vary with the size of the protein repertoire. These results indicate that the evolutionary background, adaptation to specific lifestyles (i.e., yeast-like, unicellular and filamentous) and evolutionary time, and thus an indirect measure of genetic drift, of a given species influence their polarization repertoire size. The examined factors do not explain all variation observed in the data, as 59% is undefined, indicative of missing causal factors. The

598 discovery of, and interplay of, causal factors of adaptation and differentiation between species
599 has gained much and long-term attention in the literature (Kimura 1967; Haldane 1927; Orr
600 2005; Masel 2011; Futuyma 2009). The long-lasting history of population genetics has shown
601 that genetic variation, and thus sequence similarity and ultimately presence/absence of
602 proteins, is caused by the interplay of mutation, natural selection, drift and gene flow, with
603 descent and thus the heritable characteristics as the starting conditions. It is clear that not all
604 these potential causal factors are incorporated in our study, mainly due to the scale of our
605 study and the unavailability of the particular data for our focal species. Even though our
606 analysis does not examine all factors that are likely to have played a role during the protein
607 network evolution, we identified several factors that are, in part, responsible for the complex
608 and highly dynamic polarization protein network evolution of fungi. Further expansion of
609 experimental datasets and development of reliable large-scale comparative tools, should aid a
610 better assessment of empirical data in light of the available theoretical models to study the full
611 scope of real life protein network evolution.
612
613 Our study characterizes the fungal polarization protein network as highly dynamic
614 across species, and we identify gene expression level, lineage, lifestyle and genetic drift, as
615 factors correlating with the observed patterns of variation and adaptation. Our results provide
616 further evidence that protein networks are often characterized by shared (ancient) conserved
617 components as well as taxa-specific components that are variable between even closely
618 related species. Our work sheds new light on the level and intensity of protein network
619 evolution across broad and deep phylogenetic levels.

Acknowledgements

We want to thank the members of the Laan Lab for providing critical feedback during earlier phases of this project. Thanks to W.KG. Daalman for advice on the statistical analyses. We would like to express our gratitude to the participants of the Gordon Research Conference on Cellular & Molecular Fungal Biology 2016 for constructive feedback on this project. Special thanks to Q. A. Justman, P. J. Boyton and E. M. Hyland for providing valuable comments on earlier drafts of this manuscript and to S.W.M. Pelders for assisting with writing python code. This work was supported by the Netherlands Organization for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program.

Competing interests

The authors declare to have no competing interests.

Supplementary Files

Supplementary File 1. Proteins, Strains & Species information

Supplementary File 2. Protein Matrix Comparison 200 versus 36sp

Comparison of number of observed proteins between the full dataset of 200 strains/species and the 36 selected species with highest genome quality. Depicted are scatterplots for the following four lineages: Microsporidia (far left, squares), Saccharomycotina 1 representing the Saccharomycetaceae family (center left, circles), Saccharomycotina 2 representing e.g., the Debaryomycetaceae family (center right, triangles), Pezizomycotina (far right, diamonds). For each lineage a scatterplot of the full dataset (200 strains/species; right scatterplots) and the reduced dataset is depicted (36 species; left scatterplots). Lineages are color-coded as in Figure 4A. Medians (shown as grey lines) do not differ between the two datasets for all four lineages.

645 Supplementary File 3. Repeated Combinations Proteins

646 Supplementary File 4. MFA analysis

647 (A) Broken stick method. (B) Groups. (C) Correlation map of variables. (D) Partial

648 axes.

649 Supplementary File 5. Eukaryote outgroups

650

651 **Figure Legends**

652 Fig. 1. The central part of budding yeast's polarization protein network. (A) Polarity
653 establishment and subsequent budding takes place at one side of the budding yeast cell
654 (cartoon). Insert: schematic overview of the 34 selected proteins and their functional
655 groupings based on SGD; http://www.yeastgenome.org/, (Drees et al. 2001; Chang &
656 Peter 2003; Martin & Arkowitz 2014; Madhani 2007). Cdc42 cycles between an
657 active membrane-bound state (GTP) and an inactive cytosolic state (GDP). Depicted
658 are the Cdc42 regulators and effectors, the Bem1/Cdc42 protein complex and several
659 downstream steps. Nrp1 has a presumed function in polarity establishment (see Laan
660 et al. 2015), Ubi4 has a described epistatic interaction with Cdc42 (BioGRID;
661 thebiogrid.org). (B) Matrix of the genetic (in red) and physical (in black) interactions
662 between the 34 selected polarization proteins based on SGD protein data. Proteins are
663 color coded with the functional groupings from the (A) panel. Protein Abundance
664 following Kulak et al. 2014 is displayed in the most right panel. Note that for Gic1
665 and Ubi4 no expression data was available.

666 Fig. 2. Phylogenetic relationships between the 200 fungal strains/species and the protein
667 matrix for the 34 selected polarization proteins. The phylogeny is based on 86 protein
668 sequences (158856 aa) and we used the approximately maximum likelihood method
669 and the JTT model of amino acid evolution. Support values are almost exclusively
670 above 0.9, except when shown on the tree (12 instances). The tree includes the phyla:
671 Microsporidia (in orange), Cryptomycota (in yellow), Basidiomycota (in blue) and the
672 Ascomycota (in purple). Subphyla are shades of the same phylum color. Phylogenetic
673 relationships greatly follow known relationships (Matheny et al. 2006; Zalar et al.
674 2005; Hibbett et al. 2007; Schoch et al. 2009; Shen et al. 2016). The protein matrix
675 displays the similarity scores of the reciprocal BLAST search. White fields represent

676 no match of the query protein in the respective PDB; black field represent a match
677 with 100% similarity score; grey fields represent a match with <100% similarity score.
678 Proteins are ordered and color coded following Figure 1A. Essential proteins (in *S.*
679 *cerevisiae*), paralogs and short single domain proteins are labeled with green, pink and
680 yellow bullets at the bottom of the matrix. Various recurrent and lineage-specific
681 patterns, as discussed in the main text, are highlighted by red outlines. Genome quality
682 as in number of contigs is shown in the most right column, followed by the life styles
683 of the fungal species (cartoons: yeast-like (orange), non-yeast-like unicellular (green),
684 pseudohyphal (light brown), filamentous (dark brown)).

685 Fig. 3. Correlation between genome quality and number of retrieved proteins. The top
686 panel (Count) shows the distribution of strains/species for the number of retrieved
687 proteins. The center panel shows a statistically significant positive correlation between
688 the N50 of contigs of the genome and the number of retrieved proteins. One data point
689 ($x=11$, $y=9.5x10^6$) was omitted from the plot for clarity of the plot. The bottom panel
690 shows statistically significant negative correlation between the number of contigs in
691 the genome and the number of retrieved proteins. One data point ($x=12$, $y=25607$) was
692 omitted from the plot for clarity of the plot.

693 Fig. 4. Protein matrix of the 36 species with highest genome quality. The matrix displays
694 the similarity scores of the reciprocal BLAST search for species of the Microsporidia
695 (in orange), Basidiomycota (in blue) and the Ascomycota (in purple). The cladogram
696 on the left represents the phylogenetic relationship between strains/strains based on
697 Figure 2 (note: the branch lengths are fixed and do not represent amino acid
698 substitutions). Proteins are ordered and color coded following Figure 1A. Essential
699 proteins (in *S. cerevisiae*), paralogs and short single domain proteins are labeled with

green, pink and yellow bullets at the bottom of the matrix. The life styles of the species are depicted on the right column.

Fig. 5. Polarization proteins prevalence. Prevalence of the 34 polarization proteins for all examined fungal species (black circles), the Basidiomycota species (blue circles) and the Ascomycota species (purple circles). Proteins are ordered based on their overall prevalence in all examined strains/species. The 70% criterion is marked by a horizontal red dotted line. Shading in the bottom part reflects grouping of proteins with < 70% prevalence in all strains/species (light grey; left), proteins with prevalence > 70%, but prevalence in the Basidiomycota is not in all cases >70% (grey; center), proteins with >70% prevalence in all examined groups (i.e., core proteins; dark grey; right). Core proteins in the middle group are marked by an asterisk (*). Difference in prevalence between the Ascomycota and Basidiomycota is presented in the top panel (pink diamonds).

Fig. 6. Comparison between the core and non-core proteins. (A) Significant difference in the observed prevalence of the core and non-core proteins (P-value < 0.0001). (B) Number of genetic interactions (in red) and physical interactions (in black) between the 34 examined polarization proteins. No difference was observed between the core and non-core proteins in the number of genetic or physical interactions. Data was obtained from SGD protein data. (C) Significant difference in protein abundance between the two groups. Core proteins have higher protein abundance (P-value = 0.03). Note that data for Gic1 (non-core) and Ubi4 (core) was unavailable. Core proteins are depicted as circles, while non-core proteins are depicted as squares. Grey lines depict medians.

Fig. 7. Multiple Factor Analysis and correlations. (A) Multiple Factor Analysis of the number of polarization proteins, lineage, lifestyle, genomic quality and genetic

distance. The 200 strains/species are plotted and color-coded according to their phylogenetic lineage as in Figure 2. Dimension 1 explains 20.28% of the observed variation and the following four factors constitute to its construction (in order of importance): lineage, lifestyle, genetic distance, number of observed proteins. Dimension 2 explains 12.37% of the variation in the data and is based on the variables lineage and lifestyle. Main areas occupied by specific lineages are labeled accordingly for clarity. A clear distinction can be made between yeast-like fungi (left top corner), filamentous fungi (center lower part) and unicellular non-yeast like fungi (right top corner). (B) Cartoon depicting the topology of the major clades. Note that the two Saccharomycotina lineages are shown together. The length of branches do not represent observed branch lengths. See Figure 2 for full phylogeny. (C) The distribution of lifestyles (in percentages) for the ten different phylogenetic lineages. The number of strains/species per lineages is given. Lifestyles are color-coded as in legend at the bottom right of the figure. The 200 strains/species are classified as unicellular, yeast, filamentous, dimorphic (either yeast/filamentous or yeast/pseudohyphal) and trimorphic following Figure 2. (D) Pie plot depicting the percentage of variation explained by the three main dimensions. The three dimensions account for 41% of the observed variation, leaving 59% undefined. (E) The number of observed proteins in ten different phylogenetic lineages. Groups are color-coded per lineage as in Figure 2. Medians are given as grey lines. (F) The number observed proteins in the different lifestyles. Grey lines represent medians. (G) The number of observed proteins plotted versus the genetic distance (in respect to *S. cerevisiae*). Strains/species are color-coded according to their lifestyle morphology.

References

Banuett F, Quintanilla RH Jr, Reynaga-Peña CG. 2008. The machinery for cell polarity, cell morphogenesis, and the cytoskeleton in the Basidiomycete fungus Ustilago maydis—A survey of the genome sequence. Fungal Genetics and Biology. 45:S3–S14. doi: 10.1016/j.fgb.2008.05.012.

Bastidas RJ, Heitman J. 2009. Trimorphic stepping stones pave the way to fungal virulence. Proc. Natl. Acad. Sci. U.S.A. 106:351–352. doi: 10.1073/pnas.0811994106.

Bi E, Park HO. 2012. Cell Polarization and Cytokinesis in Budding Yeast. Genetics. 191:347–387. doi: 10.1534/genetics.111.132886.

Brown JL, Jaquenoud M, Gulli MP, Chant J, Peter M. 1997. Novel Cdc42-binding proteins Gic1 and Gic2 control cell polarity in yeast. Genes & Development. 11:2972–2982.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. Nat Meth. 12:59–60. doi: 10.1093/nar/gkq275.

Buljan M, Bateman A. 2009. The evolution of protein domain families. Biochm. Soc. Trans. 37:751–755. doi: 10.1042/BST0370751.

Butty A-C et al. 2002. A positive feedback loop stabilizes the guanine-nucleotide exchange factor Cdc24 at sites of polarization. EMBO J. 21:1565–1576. doi: 10.1093/emboj/21.7.1565.

Chang F, Peter M. 2003. Yeasts make their mark. Nat Cell Biol. 5:294–299. doi: 10.1038/ncb0403-294.

Chant J. 1999. Cell polarity in yeast. Annu. Rev. Cell Dev. Biol. 15:365–391. doi: 10.1146/annurev.cellbio.15.1.365.

Cherry JM et al. 2011. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Research. 40:D700–D705. doi: 10.1093/nar/gkr1029.

Cisse OH, Stajich JE. FGMP: assessing fungal genome completeness and gene content. doi: 10.1101/049619.

Coulombe-Huntington J, Xia Y. 2017. Network Centrality Analysis in Fungi Reveals Complex Regulation of Lost and Gained Genes. PLoS ONE. 12:e0169459. doi: 10.1371/journal.pone.0169459.s001.

Dean RA et al. 2005. The genome sequence of the rice blast fungus Magnaporthe grisea. Nature. 434:980–986. doi: 10.1038/nature03449.

Diepeveen ET, Daalman WK-G, Laan L. Evolutionary dynamics in the fungal polarization network, a mechanistic perspective. Biophysical Reviews. under review.

Drees BL et al. 2001. A protein interaction map for cell polarity development. Journal of Cell Biology. 154:549–571. doi: 10.1083/jcb.200104057.

Drummond DA. 2005. A Single Determinant Dominates the Rate of Yeast Protein Evolution. Molecular Biology and Evolution. 23:327–337. doi: 10.1093/molbev/msj038.

785 Ebersberger I et al. 2012. A Consistent Phylogenetic Backbone for the Fungi. Molecular
786 Biology and Evolution. 29:1319–1334. doi: 10.1093/molbev/msr285.

787 English AC et al. 2012. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS
788 Long-Read Sequencing Technology. PLoS ONE. 7:e47768. doi:
789 10.1371/journal.pone.0047768.t003.

790 Etienne-Manneville S. 2004. Cdc42 - the centre of polarity. Journal of Cell Science.
791 117:1291–1300. doi: 10.1242/jcs.01115.

792 Evlampiev K, Isambert H. 2008. Conservation and topology of protein interaction networks
793 under duplication-divergence evolution. Proc. Natl. Acad. Sci. U.S.A. 105:9863–9868. doi:
794 10.1073/pnas.0804119105.

795 Evlampiev K, Isambert H. 2007. Modeling protein network evolution under genome
796 duplication and domain shuffling. BMC Syst Biol. 1:49. doi: 10.1186/1752-0509-1-49.

797 Fox J. 2005. The R commander: a basic-statistics graphical user interface to R. J Stat Softw.
798 14. doi: 10.18637/jss.v014.i09.

799 Fox J. 2016. *Using the R Commander: A Point-and-Click Interface or R*. Chapman and
800 Hall/CRC.

801 Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the
802 protein interaction network. Science. 296:750–752. doi: 10.1126/science.1068696.

803 Futuyma DJ. 2009. Evolution. Sinauer Associates, Sunderland, MA pp. 279–301.

804 Galagan JE, Henn MR, Ma L-J, Cuomo CA, Birren B. 2005. Genomics of the fungal
805 kingdom: insights into eukaryotic biology. Genome Research. 15:1620–1631. doi:
806 10.1101/gr.3767105.

807 Gauthier GM. 2015. Dimorphism in Fungal Pathogens of Mammals, Plants, and Insects.
808 PLoS Pathog. 11:e1004608. doi: 10.1371/journal.ppat.1004608.t001.

809 Giaever G et al. 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature.
810 418:387–391. doi: 10.1038/nature00935.

811 Gladieux P et al. 2014. Fungal evolutionary genomics provides insight into the mechanisms
812 of adaptive divergence in eukaryotes. Molecular Ecology. 23:753–773. doi:
813 10.1111/mec.12631.

814 Goryachev AB, Pokhilko AV. 2008. Dynamics of Cdc42 network embodies a Turing-type
815 mechanism of yeast cell polarity. FEBS Letters. 582:1437–1443. doi:
816 10.1016/j.febslet.2008.03.029.

817 Habib N, Wapinski I, Margalit H, Regev A, Friedman N. 2012. A functional selection model
818 explains evolutionary robustness despite plasticity in regulatory networks. Molecular Systems
819 Biology. 8:619. doi: 10.1038/msb.2012.50.

820 Haldane JBS. 1927. A mathematical theory of natural and artificial selection, Part V:
821 Selection and Mutation. Mathematical Proceedings of the Cambridge Philosophical Society.

822 23:838–844. doi: 10.1017/S0305004100015644.

823 Hibbett DS et al. 2007. A higher-level phylogenetic classification of the Fungi. Mycological
824 Research. 111:509–547. doi: 10.1016/j.mycres.2007.03.004.

825 Hirsh AE, Fraser HB. 2001. Protein dispensability and rate of evolution. Nature. 411:1046–
826 1049. doi: 10.1038/35082561.

827 Huynen MA, Dandekar T, Bork P. 1999. Variation and evolution of the citric-acid cycle: a
828 genomic perspective. Trends Microbiol. 7:281–291.

829 Irazoqui JE, Gladfelter AS, Lew DJ. 2003. Scaffold-mediated symmetry breaking by Cdc42p.
830 Nat Cell Biol. 5:1062–1070. doi: 10.1083/jcb.142.5.1301.

831 Jackson DA. 1993. Stopping Rules in Principal Components Analysis: a Comparison of
832 Heuristic and Statistical Approaches. Ecology. 74:2204–2214. doi: 10.2307/1939574.

833 Johnson DI. 1999. Cdc42: an essential Rho-type GTPase controlling eukaryotic cell polarity.
834 Microbiology and Molecular Biology Reviews. 63:54–105.

835 Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices
836 from protein sequences. Comput. Appl. Biosci. 8:275–282.

837 Jordan IK, Wolf YI, Koonin EV. 2003. No simple dependence between protein evolution rate
838 and the number of protein-protein interactions: only the most prolific interactors tend to
839 evolve slowly. BMC Evolutionary Biology. 3:1.

840 Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome
841 duplication in the yeast Saccharomyces cerevisiae. Nature. 428:617-24. doi:
842 10.1038/nature02424.

843 Kersey PJ et al. 2016. Ensembl Genomes 2016: more genomes, more complexity. Nucleic
844 Acids Research. 44:D574–D580. doi: 10.1093/nar/gkv1209.

845 Kim PM, Lu LJ, Xia Y, Gerstein MB. 2006. Relating three-dimensional structures to protein
846 networks provides evolutionary insights. Science. 314:1938–1941. doi:
847 10.1126/science.1136174.

848 Kimura M. 1967. On the evolutionary adjustment of spontaneous mutation rates. Genet. Res.
849 9:23–34. doi: 10.1017/S0016672300010284.

850 Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. 2014. Minimal, encapsulated proteomic-
851 sample processing applied to copy-number estimation in eukaryotic cells. Nat Meth. 11:319–
852 324. doi: 10.1038/nmeth.2834.

853 Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
854 Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. 33:1870–1874. doi:
855 10.1093/molbev/msw054.

856 Laan L, Koschwanez JH, Murray AW. 2015. Evolutionary adaptation after crippling cell
857 polarization follows reproducible trajectories. Elife. 4. doi: 10.7554/eLife.09638.

858 Leinonen R et al. 2010. The European Nucleotide Archive. Nucleic Acids Research. 39:D28–
859 D31. doi: 10.1093/nar/gkq967.

860 Lê S, Josse J, Husson F. 2008. FactoMineR: an R package for multivariate analysis. J. Stat.
861 Soft. 25. doi: 10.18637/jss.v025.i01.

862 Liang Z, Xu M, Teng M, Niu L. 2006. Comparison of protein interaction network reveals
863 species conservation and divergence. BMC Bioinformatics. 7:457. doi: 10.1186/1471-2105-7-
864 457.

865 Liu G et al. 2015. Gene Essentiality Is a Quantitative Property Linked to Cellular
866 Evolvability. Cell. 163:1388–1399. doi: 10.1016/j.cell.2015.10.069.

867 Lucking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT. 2009. Fungi evolved right on
868 track. Mycologia. 101:810–822. doi: 10.3852/09-016.

869 Madhani H. 2007. *From a to alpha: Yeast as a Model for Cellular Differentiation*. Cold
870 Spring Harbor Laboratory Press.

871 Martin SG. 2015. Spontaneous cell polarization: Feedback control of Cdc42 GTPase breaks
872 cellular symmetry. Bioessays. 37:1193–1201. doi: 10.1002/bies.201500077.

873 Martin SG, Arkowitz RA. 2014. Cell polarization in budding and fission yeasts. FEMS
874 Microbiol Rev. 38:228–253. doi: 10.1111/1574-6976.12055.

875 Masel J. 2011. Genetic drift. Curr. Biol. 21:R837–R838. doi: 10.1016/j.cub.2011.08.007.

876 Matheny PB, Gossmann JA, Zalar P, Kumar TKA, Hibbett DS. 2006. Resolving the
877 phylogenetic position of the Wallemiomycetes: an enigmatic major lineage of Basidiomycota.
878 Can. J. Bot. 84:1794–1805. doi: 10.1139/b06-128.

879 Miranda-Saavedra D et al. 2007. The complement of protein kinases of the microsporidium
880 Encephalitozoon cuniculi in relation to those of Saccharomyces cerevisiae and
881 Schizosaccharomyces pombe. BMC Genomics. 8:309. doi: 10.1186/1471-2164-8-309.

882 Mody A, Weiner J, Ramanathan S. 2009. Modularity of MAP kinases allows deformation of
883 their signalling pathways. Nat Cell Biol. 11:484–491. doi: 10.1038/ncb1856.

884 Mueller GM, Schmit JP. 2007. Fungal biodiversity: what do we know? What can we predict?
885 Biodivers Conserv. 16:1–5. doi: 10.1007/s10531-006-9117-7.

886 Muñoz A, Santos Muñoz D, Zimin A, Yorke JA. 2016. Evolution of transcriptional networks
887 in yeast: alternative teams of transcriptional factors for different species. BMC Genomics.
888 17:826. doi: 10.1186/s12864-016-3102-7.

889 Nagy LG et al. 2014. Latent homology and convergent regulatory evolution underlies the
890 repeated emergence of yeasts. Nature Communications. 5:4471. doi: 10.1038/ncomms5471.

891 Neumann RS, Kumar S, Shalchian-Tabriz K. 2014. BLAST output visualization in the new
892 sequencing era. Brief Bioinf. 15:484–503. doi: 10.1093/bib/bbt009.

893 O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R. 2005. Fungal Community

894 Analysis by Large-Scale Sequencing of Environmental Samples. Applied and Environmental
895 Microbiology. 71:5544–5550. doi: 10.1128/AEM.71.9.5544-5550.2005.

896 Orr HA. 2005. The genetic theory of adaptation: a brief history. Nat Rev Genet. 6:119–127.
897 doi: 10.1038/nrg1523.

898 Padamsee M et al. 2012. The genome of the xerotolerant mold Wallemia sebi reveals
899 adaptations to osmotic stress and suggests cryptic sexual reproduction. Fungal Genetics and
900 Biology. 49:217–226. doi: 10.1016/j.fgb.2012.01.007.

901 Papin JA, Hunter T, Palsson BO, Subramaniam S. 2005. Reconstruction of cellular signalling
902 networks and analysis of their properties. Nat Rev Mol Cell Biol. 6:99–111. doi:
903 10.1038/nrm1570.

904 Park HO, Bi E. 2007. Central Roles of Small GTPases in the Development of Cell Polarity in
905 Yeast and Beyond. Microbiology and Molecular Biology Reviews. 71:48–96. doi:
906 10.1128/MMBR.00028-06.

907 Pawson T, Nash P. 2003. Assembly of cell regulatory systems through protein interaction
908 domains. Science. 300:445–452. doi: 10.1126/science.1083653.

909 Pál C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. Nat Rev Genet.
910 7:337–348. doi: 10.1038/nrg1838.

911 Peyretaillade E et al. 2011. Extreme reduction and compaction of microsporidian genomes.
912 Research in Microbiology. 162:598–606. doi: 10.1016/j.resmic.2011.03.004.

913 Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing Large Minimum Evolution Trees
914 with Profiles instead of a Distance Matrix. Molecular Biology and Evolution. 26:1641–1650.
915 doi: 10.1093/molbev/msp077.

916 Pruyne D, Bretscher A. 2000. Polarization of cell growth in yeast. I. Establishment and
917 maintenance of polarity states. Journal of Cell Science. 113:365–375.

918 Pruyne D, Legesse-Miller A, Gao L, Dong Y, Bretscher A. 2004. Mechanisms of polarized
919 growth and organelle segregation in yeast. Annu. Rev. Cell Dev. Biol. 20:559–591. doi:
920 10.1146/annurev.cellbio.20.010403.103108.

921 Rhind N et al. 2011. Comparative Functional Genomics of the Fission Yeasts. Science.
922 332:930–936. doi: 10.1126/science.1203357.

923 Richman TJ et al. 2004. Analysis of cell-cycle specific localization of the Rdi1p RhoGDI and
924 the structural determinants required for Cdc42p membrane localization and clustering at sites
925 of polarized growth. Current Genetics. 45:339–349. doi: 10.1007/s00294-004-0505-9.

926 R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation
927 for Statistical Computing: Vienna, Austria http://www.R-project.org/.

928 Schoch CL et al. 2009. The Ascomycota Tree of Life: A Phylum-wide Phylogeny Clarifies
929 the Origin and Evolution of Fundamental Reproductive and Ecological Traits. Systematic
930 Biology. 58:224–239. doi: 10.1093/sysbio/syp020.

931 Schüler A, Bornberg-Bauer E. 2010. The Evolution of Protein Interaction Networks. In:
932 Methods in Molecular Biology Vol. 696 Humana Press: Totowa, NJ pp. 273–289. doi:
933 10.1007/978-1-60761-987-1_17.

934 Sharma KK. 2016. Fungal genome sequencing: basic biology to biotechnology. Critical
935 Reviews in Biotechnology. 36:743–59. doi: 10.1186/1471-2164-14-274.

936 Shen XX et al. 2016. Reconstructing the backbone of the Saccharomycotina yeast phylogeny
937 using genome-scale data. G3: Genes|Genomes|Genetics. 6:3927–3939. doi:
938 10.1543/g3.116.034744.

939 Sievers F et al. 2011. Fast, scalable generation of high-quality protein multiple sequence
940 alignments using Clustal Omega. Molecular Systems Biology. 7:539. doi:
941 10.1038/msb.2011.75.

942 Stajich JE et al. 2011. FungiDB: an integrated functional genomics database for fungi.
943 Nucleic Acids Research. 40:D675–D681. doi: 10.1093/nar/gkr918.

944 Tanay A, Regev A, Shamir R. 2005. Conservation and evolvability in regulatory networks:
945 The evolution of ribosomal regulation in yeast. Proc. Natl. Acad. Sci. U.S.A. 102:7203–7208.
946 doi: 10.1073/pnas.0502521102.

947 Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD. 2008. The evolution of
948 combinatorial gene regulation in fungi. Plos Biol. 6:352–364. doi:
949 10.1371/journal.pbio.0060038.

950 Vivarès CP, Gouy M, Thomarat F, Méténier G. 2002. Functional and evolutionary analysis of
951 a eukaryotic parasitic genome. Current Opinion in Microbiology. 5:499–505.

952 Vleugel M, Hoogendoorn E, Snel B, Kops GJPL. 2012. Evolution and Function of the Mitotic
953 Checkpoint. Developmental Cell. 23:239–250. doi: 10.1016/j.devcel.2012.06.013.

954 Voordeckers K, Pougach K, Verstrepen KJ. 2015. How do regulatory networks evolve and
955 expand throughout evolution? Current Opinion in Biotechnology. 34:180–188. doi:
956 10.1016/j.copbio.2015.02.001.

957 Wall DP et al. 2005. Functional genomic analysis of the rates of protein evolution. Proc. Natl.
958 Acad. Sci. U.S.A. 102:5483–5488. doi: 10.1073/pnas.0501761102.

959 Wedlich-Soldner R, Altschuler S, Wu L, Li R. 2003. Spontaneous cell polarization through
960 actomyosin-based delivery of the Cdc42 GTPase. Science. 299:1231–1235. doi:
961 10.1126/science.1080944.

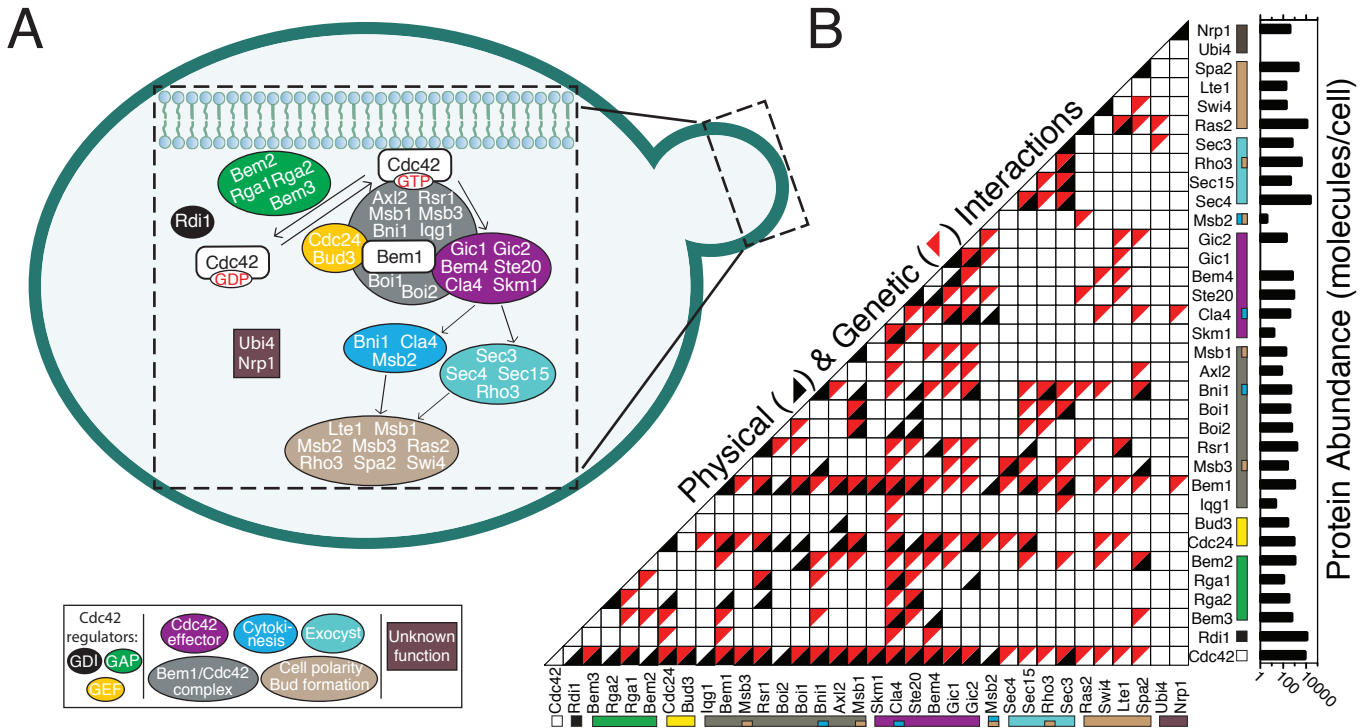
962 Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire
963 yeast genome. Nature. 387:708–713. doi: 10.1038/42711.

964 Zalar P, Sybren de Hoog G, Schroers H-J, Frank JM, Gunde-Cimerman N. 2005. Taxonomy
965 and phylogeny of the xerophilic genus Wallemia (Wallemiomycetes and Wallemiales, cl. et
966 ord. nov.). Antonie Van Leeuwenhoek. 87:311–328. doi: 10.1007/s10482-004-6783-x.

967 Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. Nat Rev
968 Genet. 16:409–420. doi: 10.1038/nrg3950.

969 Zmasek CM, Godzik A. 2011. Strong functional patterns in the evolution of eukaryotic
970 genomes revealed by the reconstruction of ancestral protein domain repertoires. Genome Biol.
971 12:R4. doi: 10.1186/gb-2011-12-1-r4.

972


A
Bem2
Rga1 Rga2
Bem3
Cdc42
GTP
Axl2 Rsr1
Msb1 Msb3
Bni1 Iqg1
Rdi1
Cdc24
Bud3
Bem1
Gic1 Gic2
Bem4 Ste20
Cla4 Skm1
Cdc42
GDP
Boi1
Boi2
Ubi4
Nrp1
Bni1 Cla4
Msb2
Sec3
Sec4 Sec15
Rho3
Lte1 Msb1
Msb2 Msb3 Ras2
Rho3 Spa2 Swi4
B
Physical (◢) & Genetic (◤) Interactions
Nrp1
Ubi4
Spa2
Lte1
Swi4
Ras2
Sec3
Rho3
Sec15
Sec4
Msb2
Gic2
Gic1
Bem4
Ste20
Cla4
Skm1
Msb1
Axl2
Bni1
Boi1
Boi2
Rsr1
Msb3
Bem1
Iqg1
Bud3
Cdc24
Bem2
Rga1
Rga2
Bem3
Rdi1
Cdc42
Protein Abundance (molecules/cell)
1
100
10000
Cdc42 regulators:
GDI
GAP
GEF
Cdc42 effector
Cytoki-nesis
Exocyst
Bem1/Cdc42 complex
Cell polarity Bud formation
Unknown function

Ascomycota

Pezizomycotina

Saccharomycotina

Taphrinomycotina

Basidiomycota

Agaricomycotina

Ustilaginomycotina

Pucciniomycotina

Wallemiomycetes

Cryptomycota

Microsporidia

0.14

0.02

0.21

0.39

0.81

0.86

0.53

0.73

WGD

0.55

0.33

0.61

0.05

0.2

1.00

0.75

0.50

0.25

0.00

Essential in S. cerevisiae

Paralog

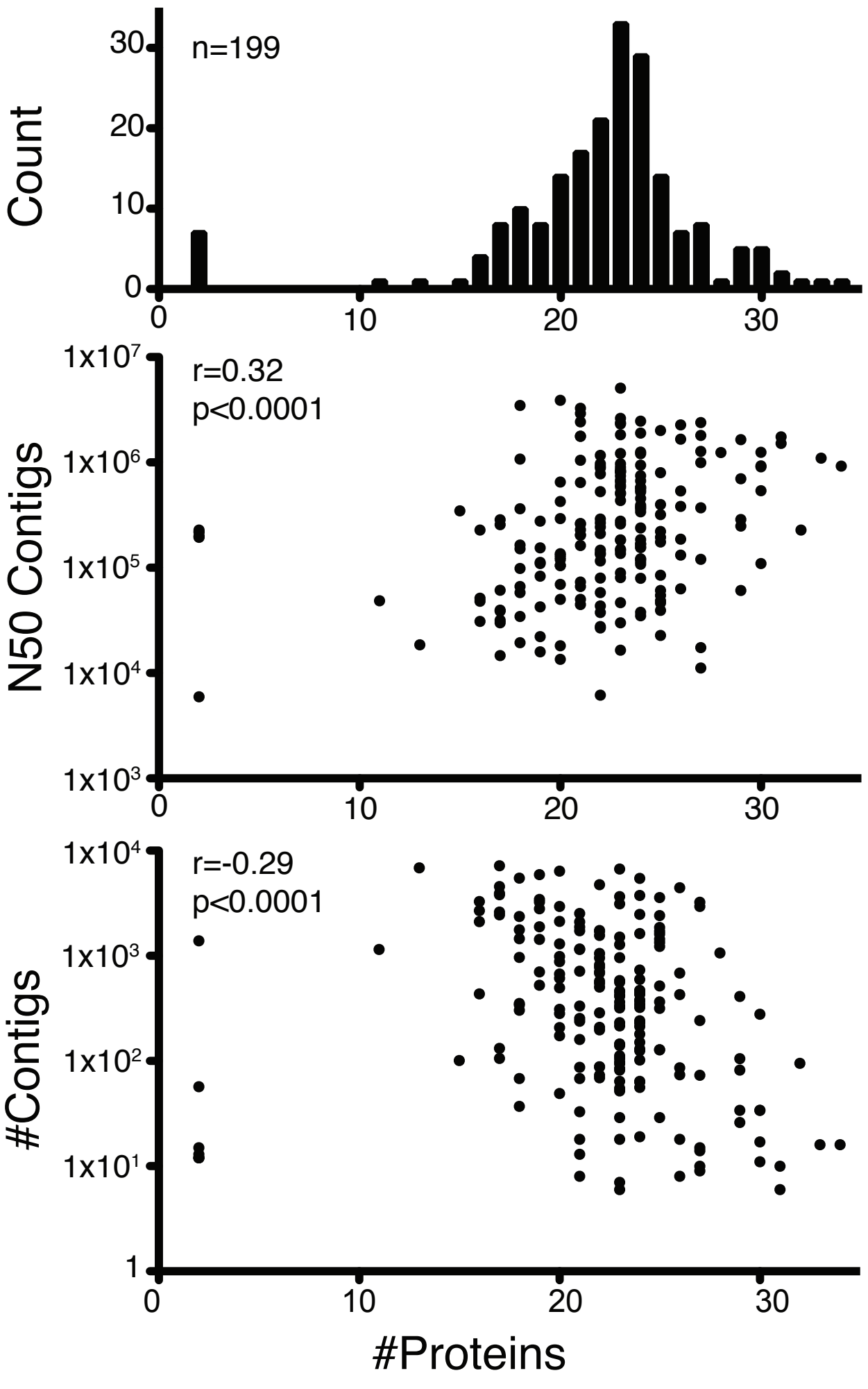
Short single domain protein

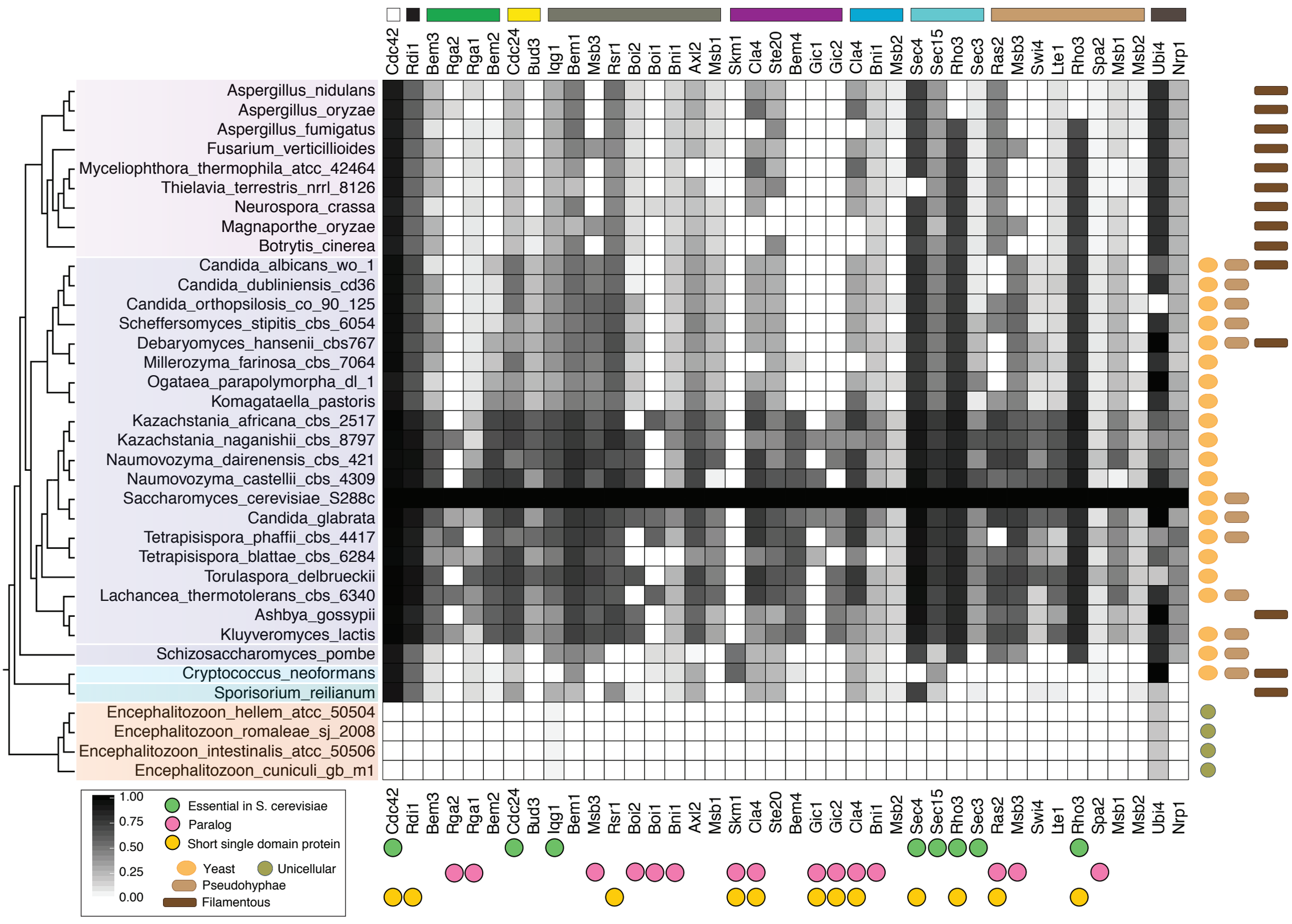
Unicellular

Yeast

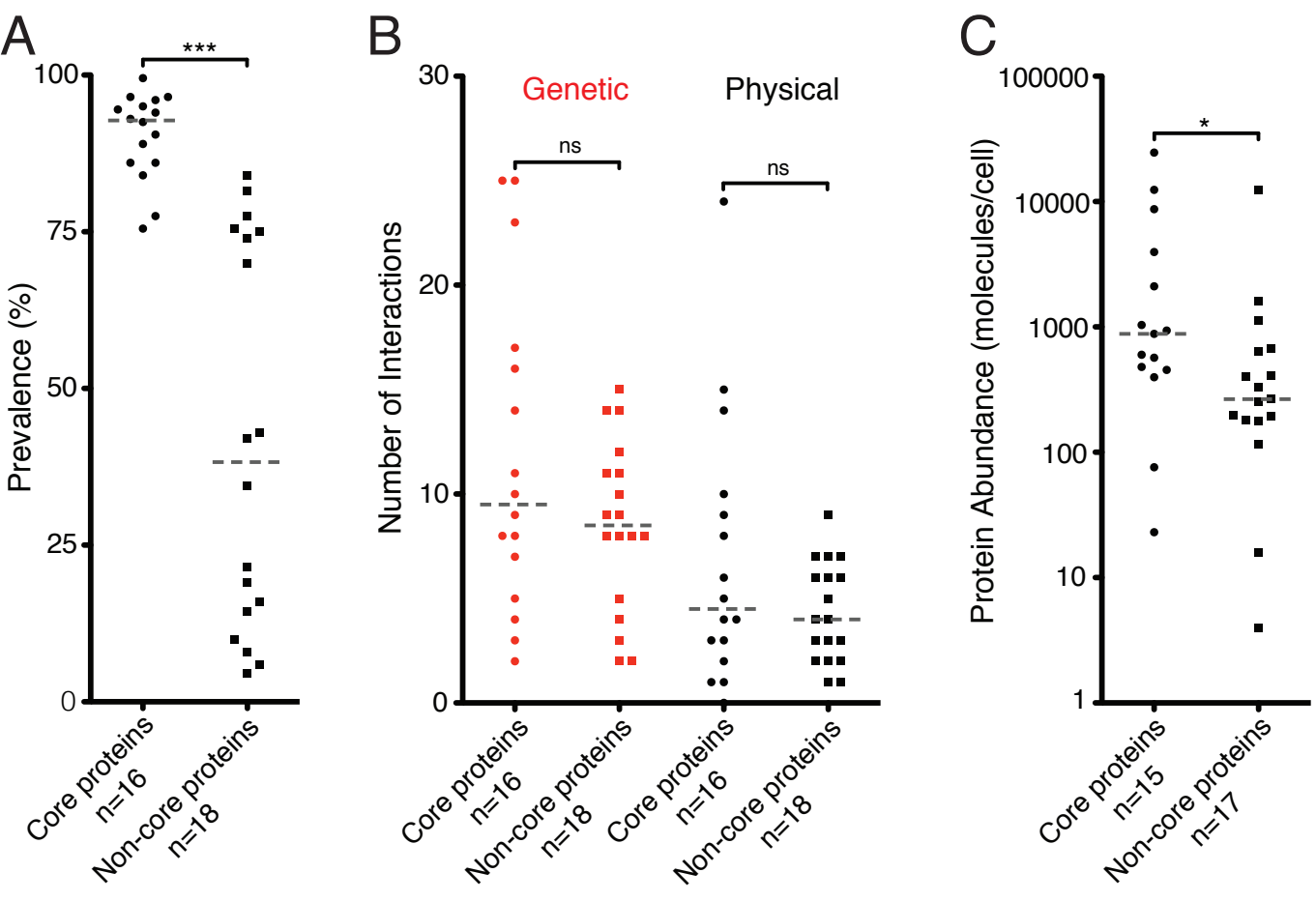
Pseudohyphae

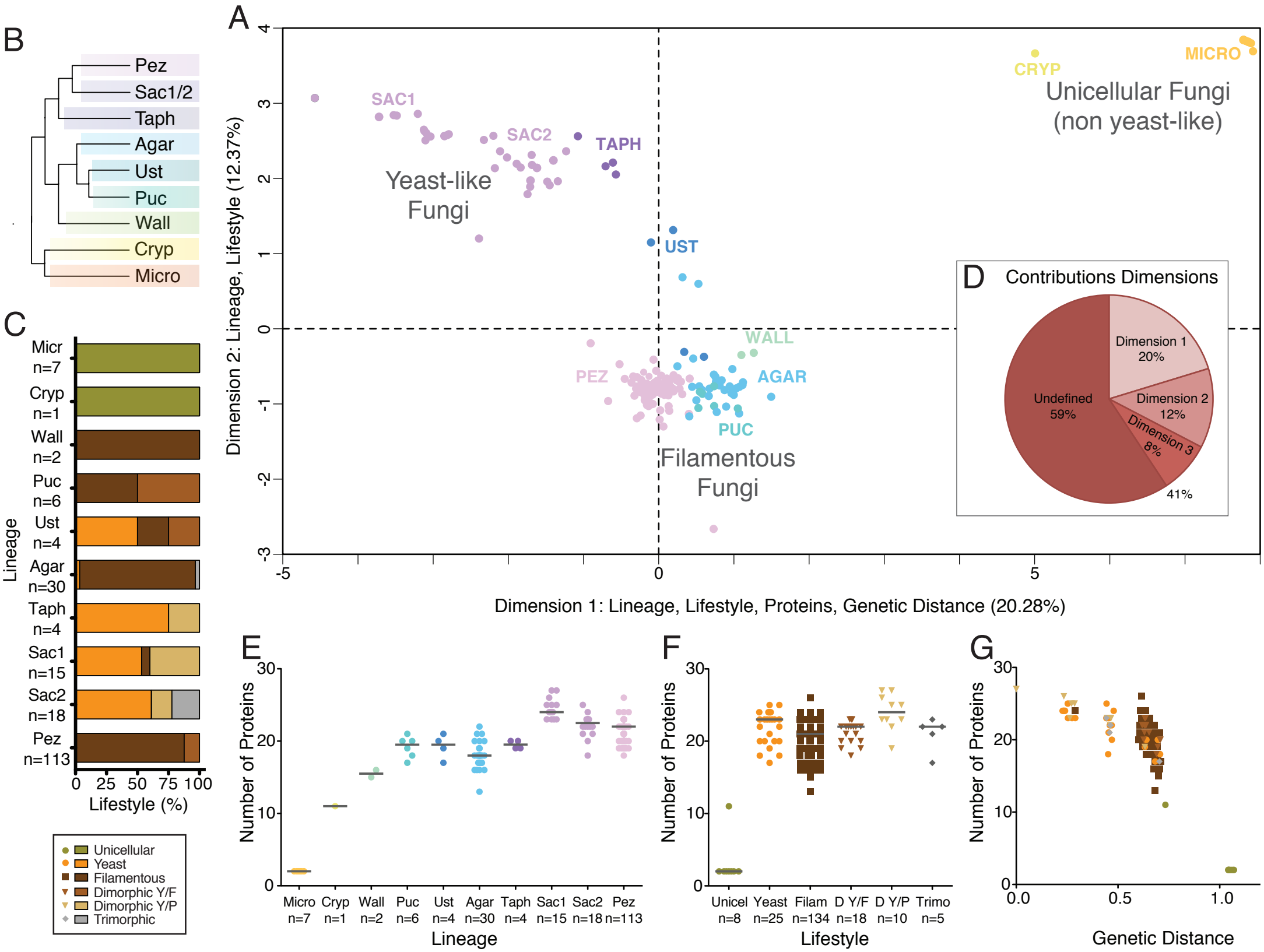
Filamentous


n=199
Count
30
20
10
0
0
10
20
30
r=0.32
p<0.0001
N50 Contigs
1x10^7
1x10^6
1x10^5
1x10^4
1x10^3
r=-0.29
p<0.0001
#Contigs
1x10^4
1x10^3
1x10^2
1x10^1
1
#Proteins


Cdc42
Rdi1
Bem3
Rga2
Rga1
Bem2
Cdc24
Bud3
Iqg1
Bem1
Msb3
Rsr1
Boi2
Boi1
Bni1
Axl2
Msb1
Skm1
Cla4
Ste20
Bem4
Gic1
Gic2
Cla4
Bni1
Msb2
Sec4
Sec15
Rho3
Sec3
Ras2
Msb3
Swi4
Lte1
Rho3
Spa2
Msb1
Msb2
Ubi4
Nrp1
Aspergillus_nidulans
Aspergillus_oryzae
Aspergillus_fumigatus
Fusarium_verticillioides
Myceliophthora_thermophila_atcc_42464
Thielavia_terrestris_nrrl_8126
Neurospora_crassa
Magnaporthe_oryzae
Botrytis_cinerea
Candida_albicans_wo_1
Candida_dubliniensis_cd36
Candida_orthopsilosis_co_90_125
Scheffersomyces_stipitis_cbs_6054
Debaryomyces_hansenii_cbs767
Millerozyma_farinosa_cbs_7064
Ogataea_parapolymorpha_dl_1
Komagataella_pastoris
Kazachstania_africana_cbs_2517
Kazachstania_naganishii_cbs_8797
Naumovozyma_dairenensis_cbs_421
Naumovozyma_castellii_cbs_4309
Saccharomyces_cerevisiae_S288c
Candida_glabrata
Tetrapisispora_phaffii_cbs_4417
Tetrapisispora_blattae_cbs_6284
Torulaspora_delbrueckii
Lachancea_thermotolerans_cbs_6340
Ashbya_gossypii
Kluyveromyces_lactis
Schizosaccharomyces_pombe
Cryptococcus_neoformans
Sporisorium_reilianum
Encephalitozoon_hellem_atcc_50504
Encephalitozoon_romaleae_sj_2008
Encephalitozoon_intestinalis_atcc_50506
Encephalitozoon_cuniculi_gb_m1
1.00
0.75
0.50
0.25
0.00
Essential in S. cerevisiae
Paralog
Short single domain protein
Yeast
Unicellular
Pseudohyphae
Filamentous

dF = A(●) - B(●)
100
50
0
-50
Ascomycota-specific
Prevalence (%)
100
80
60
40
20
0
Non-core proteins
* Core proteins
Fungi
Ascomycota
Basidiomycota
Gic1
Skm1
Gic2
Bem4
Rga2
Swi4
Boi1
Bud3
Msb3
Bem2
Msb2
Rsr1
Nrp1
Ras2
Msb1
Cla4
Axl2
Lte1
Rga1
Rho3
Sec3
Boi2
Ste20
Sec4
Bni1
Bem3
Spa2
Cdc24
Sec15
Ubi4
Bem1
Cdc42
Rdi1
Iqg1


A
***
100
75
50
25
0
Prevalence (%)
Core proteins n=16
Non-core proteins n=18
B
Genetic
Physical
ns
ns
30
20
10
0
Number of Interactions
Core proteins n=16
Non-core proteins n=18
Core proteins n=16
Non-core proteins n=18
C
*
100000
10000
1000
100
10
1
Protein Abundance (molecules/cell)
Core proteins n=15
Non-core proteins n=17


A
B
Pez
Sac1/2
Taph
Agar
Ust
Puc
Wall
Cryp
Micro
C
Micr n=7
Cryp n=1
Wall n=2
Puc n=6
Ust n=4
Agar n=30
Taph n=4
Sac1 n=15
Sac2 n=18
Pez n=113
Lineage
0 25 50 75 100
Lifestyle (%)
Unicellular
Yeast
Filamentous
Dimorphic Y/F
Dimorphic Y/P
Trimorphic
SAC1
SAC2
TAPH
UST
WALL
AGAR
PEZ
PUC
CRYP
MICRO
Yeast-like Fungi
Unicellular Fungi (non yeast-like)
Filamentous Fungi
Dimension 2: Lineage, Lifestyle (12.37%)
Dimension 1: Lineage, Lifestyle, Proteins, Genetic Distance (20.28%)
D
Contributions Dimensions
Dimension 1 20%
Dimension 2 12%
Dimension 3 8%
Undefined 59%
41%
E
Number of Proteins
Micro n=7
Cryp n=1
Wall n=2
Puc n=6
Ust n=4
Agar n=30
Taph n=4
Sac1 n=15
Sac2 n=18
Pez n=113
Lineage
F
Number of Proteins
Unicel n=8
Yeast n=25
Filam n=134
D Y/F n=18
D Y/P n=10
Trimo n=5
Lifestyle
G
Number of Proteins
Genetic Distance