Title: Gene regulatory network plasticity predates a switch in function of a conserved transcription regulator

Authors: Isabel Nocedal[1,2], Eugenio Mancera[1,3], Alexander D. Johnson[1]*

Affiliations:

[1]Department of Microbiology and Immunology, Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA.

[2]Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

[3]Present address: Departamento de Ingeniería Genética, CINVESTAV Unidad Irapuato, Irapuaro, México

*Correspondence to: ajohnson@cgl.ucsf.edu

ABSTRACT

The rewiring of gene regulatory networks can generate phenotypic novelty. It remains an open question, however, how the large number of connections needed to form a novel network arise over evolutionary time. Here we address this question using the network controlled by the fungal transcription regulator Ndt80. This conserved protein has undergone a dramatic switch in function—from an ancestral role regulating sporulation to a derived role regulating biofilm formation. This switch in function corresponded to a large-scale rewiring of the genes regulated by Ndt80. However, we demonstrate that the Ndt80-target gene connections were undergoing extensive rewiring prior to the switch in Ndt80's regulatory function. We propose that extensive drift in the Ndt80 regulon allowed for the exploration of alternative network structures without a loss of ancestral function, thereby facilitating the formation of a network with a new function.

INTRODUCTION

The emergence of novel traits has long fascinated evolutionary biologists, with many intriguing examples observed across the tree of life (Pigliucci, 2012; Rieppel, 2001; Shubin, Tabin, & Carroll, 2009; G. P. Wagner & Lynch, 2010). Although the rewiring of gene regulatory networks over evolutionary time is recognized as a key source of variation responsible for the modification of complex phenotypes (Carroll, 2005; Davidson & Erwin, 2006; Hao Li & Johnson, 2010; Wray, 2007), we still lack an understanding of how new, large gene regulatory networks controlling novel phenotypes arise over evolutionary time. Typically, gene regulatory networks are composed of "master" transcription regulators and many downstream "target" genes, whose expression is controlled by these regulators in response to environmental or developmental signals. It has been proposed that new gene regulatory networks arise from a combination of *de novo* genes as well as conserved genes that have undergone changes in regulation (G. P. Wagner & Lynch, 2010). Yet it remains unclear how a large number of genes (whether old or new) can be brought together to form a new network.

To address these questions, we examined the evolutionary history of Ndt80, a sequence-specific DNA-binding protein that is deeply conserved across a large group of fungal species encompassing approximately 300 million years of diversity. In most of these species, Ndt80 controls meiosis and sporulation, the coupled processes that form a portion of the fungal sexual cycle (Chu & Herskowitz, 1998; Xu, Ajimura, Padmore, Klein, & Kleckner, 1995). However, in the narrow lineage leading to the human fungal pathogen species *Candida albicans*, Ndt80 acquired a new role as a master regulator of the gene regulatory network that controls the formation of biofilms, multicellular communities of surface-associated cells (Figure 1A) (Nobile

et al., 2012). This newly evolved trait enables *C. albicans* to persist on mucosal surfaces and on implanted medical devices (Bonhomme & D'enfert, 2013; Kojic & Darouiche, 2004; Nobile et al., 2012) and is responsible for many of the disease-causing properties of *C. albicans* (Nobile & Johnson, 2015). The entire biofilm gene regulatory network in *C. albicans* is complex—Ndt80 alone controls the expression of hundreds of genes involved in biofilm formation— but it appears to have evolved relatively recently (Nobile et al., 2012). In this paper, we infer that approximately 20 to 100 million years ago (Taylor & Berbee, 2006), Ndt80's role changed from regulating sporulation and meiosis (its ancestral role) to regulating biofilm formation. We rigorously exclude a model in which the Ndt80 network remained relatively static until a sudden change occurred in the *Candida* clade when Ndt80's overall function changed from sporulation to biofilm formation. Instead, we found that the Ndt80 regulon was continuously undergoing significant rewiring in all lineages examined, even in those where the overall function of Ndt80 remained unchanged. Indeed, at the resolution of our experiments, the extent of Ndt80 rewiring was approximately the same whether its function had changed or not.

We propose that the inherent flexibility of the Ndt80 regulon facilitated the exploration of new regulatory networks, allowing it to reach positions in "network space" that allowed for the evolution of a novel phenotype. This idea is analogous to examples of protein evolution in which mutational space can be sampled by drift without compromising the ancestral function of the protein. Such neutral excursions can then be exploited if they provide, as a by-product, a new function, for example weak catalysis of an alternative reaction by an enzyme (Aharoni et al., 2004; Bridgham, Carroll, & Thornton, 2006; Coyle, Flores, & Lim, 2013; Khersonsky & Tawfik, 2010; A. Wagner, 2005b; 2005a). Our work provides an empirical, case study of an analogous

exploitation of regulatory network plasticity in the evolution of novel transcription regulator function.

It has been well established that transcription regulator-target gene interactions change over evolutionary timescales. For example, the regulators controlling the ribosomal genes and the **a**-specific genes in fungi have changed over several hundred million years of evolution (Baker, Booth, Sorrells, & Johnson, 2012; Ihmels et al., 2005; Lavoie et al., 2010; Tanay, Regev, & Shamir, 2005; Tsong, Tuch, Li, & Johnson, 2006) (Sorrells, Booth, Tuch, & Johnson, 2015). Likewise, transcription regulators (such as Mcm1 and Ste12) have been shown to undergo extensive rewiring of their regulatory targets over similar timescales (Borneman et al., 2007; Tuch, Galgoczy, Hernday, Li, & Johnson, 2008). Such rewiring is not restricted to fungi; for example enhancers in *Drosophila* have been shown to undergo rapid changes in the nature and arrangement of their *cis*-regulatory sequences (Frankel, Wang, & Stern, 2012; Gompel, Prud'homme, Wittkopp, Kassner, & Carroll, 2005; Ludwig, Bergman, Patel, & Kreitman, 2000; Swanson, Schwimmer, & Barolo, 2011). This high degree of network plasticity seems to be a general feature of many regulatory networks. Based on the case study of Ndt80, we propose that this plasticity can be captured to produce new, complex regulatory networks.

RESULTS

Ndt80 target genes differ markedly between *S. cerevisiae* and *C. albicans*

As described above, the transcription factor Ndt80 is required for sporulation and meiosis in *S. cerevisiae* and biofilm formation in *C. albicans* (Figure 1A). In principle, the difference in

phenotype could be due to a difference in the identity of the genes regulated by Ndt80 between these species; alternatively, Ndt80 could regulate the same gene set in both species and the difference in phenotype could be due to something else, for example, Ndt80 target genes acquiring new enzymatic functions or Ndt80 itself being regulated differently in the two species.

To distinguish between these models, we identified the genes directly regulated by Ndt80 in both *S. cerevisiae* and *C. albicans* using chromatin immunoprecipitation of epitope-tagged Ndt80 followed by high throughput sequencing (ChIP-Seq). To minimize the effect of species-specific differences in the levels of Ndt80 expression (and to capture as much of the network as possible) the promoters of Ndt80 in both species were replaced with high-expression constitutive promoters (Materials and Methods). It has been well established that, even using proper controls, chromatin immunoprecipitation produces false positives in addition to valid instances of binding (Fan & Struhl, 2009; Teytelman, Thurtle, Rine, & Van Oudenaarden, 2013). To eliminate spurious signals and identify bona fide instances of Ndt80 binding, we employed four different criteria of increasing stringency (Figure 1C). First, we simply identified genes with significant peaks of ChIP enrichment in their intergenic regions; this is our least stringent criteria. Second, we filtered this set to include only those ChIP enrichment peaks that contained an Ndt80 *cis*-regulatory motif in the intergenic region (this motif is discussed in more detail below). Third, of those ChIP peaks that contained an Ndt80 motif, we further refined the set of Ndt80 targets by requiring that the motif also be present in the orthologous intergenic region of two very closely related species, as this greatly increases the likelihood that the Ndt80 binding site was maintained by selection. (For *S. cerevisiae*, we used *Saccharomyces mikatae* and *Saccharomyces kudriavzevii*, for *C. albicans* we used *Candida tropicalis* and *Candida dubliniensis* (Byrne & Wolfe, 2005; Lohse et al., 2013)). Lastly, we used gene expression data

to restrict Ndt80 targets to ChIP peaks with an Ndt80 motif where the gene also exhibited a significant change in mRNA expression when Ndt80 was deleted (Chu et al., 1998; Nobile et al., 2012).

Each of the four criteria were used to identify and compare the genes regulated by Ndt80 between *S. cerevisiae* and *C. albicans*. While the number of genes identified as Ndt80 targets differs depending on the criteria used, all four methods produce the same overall conclusion: Ndt80 regulates many genes in each species, with very little overlap between them (Figure 1D). By all methods, fewer than 13% of the targets of Ndt80 in these two species are shared; with the most stringent criteria this drops to 3.4% (Figure 1D). To ensure that constitutive overexpression of Ndt80 was not responsible for this large difference in Ndt80 targets, we also performed ChIP-Seq on Ndt80 in both species under the control of the endogenous promoter. While many fewer regions are bound than with constitutive Ndt80 expression, we similarly find very little overlap in the Ndt80 targets in *S. cerevisiae* and *C. albicans* (less than 12%, Figure 1—figure supplement 1). We note that experimental noise is not sufficient to explain the difference in targets, as the differences we observe between biological replicates accounts for only a small fraction of the difference observed between *S. cerevisiae* and *C. albicans* (Figure 1—figure supplement 4).

These results all lead to the conclusion that Ndt80 regulates distinct sets of genes in *S. cerevisiae* and *C. albicans*. Most of these genes have 1:1 orthologs in both species, although there are also a smaller number of species-specific genes. If we exclude the species-specific genes from our analysis we still find relatively little overlap in Ndt80 targets (4-21% targets shared, depending on target identification criteria used, Figure 1—figure supplement 2), showing that gene gains and losses alone cannot account for the change in Ndt80 targets between *S. cerevisiae* and *C. albicans*. Despite the significant differences in Ndt80 targets, however, the

147 *cis*-regulatory sequence bound by Ndt80 in both species is highly conserved ((Jolly, Chin,
148 Herskowitz, & Li, 2005; Nobile et al., 2012), Figure 1B). This observation indicates that the
149 Ndt80 regulon has been significantly rewired between *S. cerevisiae* and *C. albicans* without a
150 change in the DNA-binding specificity of Ndt80.
151
152 **Divergence in Ndt80 targets not a result of Ndt80 gene duplication**
153 While the *NDT80* genes in *S. cerevisiae* and *C. albicans* discussed above are
154 homologous, *C. albicans* has an additional paralog of Ndt80, resulting from a gene duplication
155 (Figure 2A, (Sellam et al., 2010)). The Ndt80 homolog whose targets we identified above will
156 be referred to as Ndt80B, while the additional paralog will be referred to as Ndt80A (Figure 2—
157 figure supplement 1). While Ndt80B has been shown to be required for biofilm formation in *C.*
158 *albicans* (Nobile et al., 2012), the regulatory function of Ndt80A is unknown (deletion of
159 Ndt80A has no effect on biofilm formation (Figure 4H, (Nobile et al., 2012)) or any other
160 phenotype tested). To test the possibility that the difference in targets between Ndt80 in *S.*
161 *cerevisiae* and Ndt80B in *C. albicans* is simply a consequence of this gene duplication, we
162 identified the regulatory targets of Ndt80A in *C. albicans* by ChIP-Seq under control of the same
163 constitutive promoter used for Ndt80B. We found that the targets regulated by Ndt80A represent
164 a small subset of the targets of Ndt80B (9-11%, Figure 1—figure supplement 3); that is, Ndt80A
165 binds to many fewer genomic regions, but all of these regions are also bound by Ndt80B. If we
166 compare the targets of Ndt80 in *S. cerevisiae* to the targets of Ndt80A in *C. albicans*, we find
167 that less than 4% of the overall targets are shared, regardless of target identification criteria
168 (Figure 1—figure supplement 3), demonstrating that Ndt80A's targets are no more similar to the
169 targets of Ndt80 in *S. cerevisiae* than that of Ndt80B. These experiments demonstrate that both

of the Ndt80 paralogs in *C. albicans* have undergone significant rewiring since *S. cerevisiae* and *C. albicans* diverged, indicating that the Ndt80 gene duplication was not directly responsible for the divergence in targets between these two species.

Ndt80 target genes also differ between *S. cerevisiae*, *K. lactis*, *P. pastoris*, *S. stipitis*, and *C. albicans*

To reconstruct a timeline of the evolution of the Ndt80 gene regulatory network, we conducted ChIP-Seq experiments in three additional species that branch from the common ancestor of *S. cerevisiae* and *C. albicans* at highly informative points: *Kluyveromyces lactis, Pichia pastoris,* and *Scheffersomyces stipitis* (Figure 2A). In each case, we used the strategy of expressing Ndt80 under the control of a strong constitutive promoter to minimize expression differences between species. In *C. albicans* and *S. stipitis,* targets were identified for the two Ndt80 paralogs separately, and found to be highly overlapping (Figure 1—figure supplement 3, Supplementary File 1). Thus, the union of Ndt80 paralog targets were taken to represent all Ndt80 targets in that species.

We first investigated whether the regulatory targets of Ndt80 in each of these species more closely resemble the targets in *S. cerevisiae* or the targets in *C. albicans*. Ndt80 targets were identified in all species based on the criteria of ChIP enrichment plus Ndt80 motif presence (Figure 1C), and compared to the targets of Ndt80 in *S. cerevisiae* and *C. albicans*. Counter to our initial expectation, we found that a very large number of Ndt80 targets in each species are not targets of Ndt80 in either *S. cerevisiae* or *C. albicans* (50% of all targets in *K. lactis*, 70% in *P. pastoris*, and 48% in *S. stipitis,* Figure 2B). Thus, the Ndt80 regulons in these three species do not closely resemble those of either *S. cerevisiae* or *C. albicans*; instead, each have acquired a

distinctive set of Ndt80 target genes. If we repeat this analysis using alternative criteria for identifying Ndt80 targets (Figure 1C), this conclusion still holds (Figure 2—figure supplement 2), indicating that it is robust to the stringency of Ndt80 target identification.

Consistent with the idea that Ndt80 has acquired a unique set of targets in each species, only ten genes are targets of Ndt80 in all five species tested (Figure 2C). While this is more than expected strictly by chance considering the number of regions bound in each species ($p < 10^{-5}$, Figure 2—figure supplement 3), this number pales in comparison to the 3,261 genes that are Ndt80 targets in just one of the five species (Figure 2D). Even if we consider only genes with 1:1 orthologs across all five species, there are more targets bound in only one species (1,041) than bound in two or more species (772) (Figure 2D). In short, the Ndt80 regulon differs extensively between each of the five species tested.

Ndt80 binding differences are largely determined by the gain and loss of *cis*-regulatory sites

Because the Ndt80-target gene connections differ significantly between species, we considered whether these differences were due primarily to changes in the Ndt80 protein itself, or to changes in the distribution of its *cis*-regulatory motif across the different genomes. Although the *cis*-regulatory sequence recognized by Ndt80 is conserved across these species (Figure 2—figure supplement 4), changes in the Ndt80 protein itself that alter, for example, protein-protein interactions with other transcription regulators, could account for many of the differences between species. The sequence of the Ndt80 protein does differ considerably across the species tested, with only 55% similarity in the DNA-binding domain between *S. cerevisiae*

and *C. albicans* (Sellam et al., 2010), and 35% similarity in the proteins overall (Figure 2—figure supplement 1).

To test whether these changes in protein sequence were primarily responsible for the observed differences in Ndt80 target genes between species, we expressed the *P. pastoris NDT80* gene in *S. cerevisiae* under the control of the same constitutive promoter used for ChIP-Seq of the endogenous *S. cerevisiae NDT80*, and carried out ChIP-Seq on the *P. pastoris NDT80* (Figure 3A). If we compare Ndt80 ChIP enrichment genome-wide in this experiment to Ndt80 enrichment for the native Ndt80 in each species, we find a much stronger correlation between genomic binding of the heterologous *P. pastoris* Ndt80 and the native *S. cerevisiae* Ndt80 (Figure 3B) than between the heterologous *P. pastoris* Ndt80 and the native *P. pastoris* Ndt80 (Figure 3C). Similarly, we find more shared target genes between the different Ndt80 proteins expressed in *S. cerevisiae* (75 genes, Figure 3—figure supplement 1) than between the same *P. pastoris* Ndt80 protein expressed in different species (4 genes, Figure 3—figure supplement 1). These results demonstrate that changes in the Ndt80 protein sequence across species have had only a minor effect on the Ndt80-target gene connections; instead, the connections are predominantly determined by the distribution of *cis*-regulatory sequences across the different genomes.

Although it seems likely that gains and losses of Ndt80 *cis*-regulatory sequences are largely responsible, we cannot formally exclude the contributions of changes in *cis*-regulatory sequences for other (thus far, unidentified) proteins that might help recruit Ndt80 to DNA. These observations on the key role of *cis*-regulatory change do not mean that changes in the protein are unimportant; indeed, the heterologous *P. pastoris* Ndt80 does not complement the sporulation defect observed in an *ndt80* deletion in *S. cerevisiae* (data not shown), nor does it

occupy all the genome positions characteristic of the *S. cerevisiae* protein. Our results do show, however, that the heterologous *P. pastoris* protein does occupy a substantial fraction of the *S. cerevisiae*-specific positions along the genome. This observation strongly supports the conclusion that the majority of the differences in the Ndt80 regulons across species are due to gains and losses of *cis*-regulatory sequences.

Ndt80 is required for sporulation in *K. lactis*, *P. pastoris*, and *C. lusitaniae*

To determine whether the rewiring of the Ndt80 gene regulatory network corresponds to the changes in the overall function of Ndt80, we determined the phenotype of an *ndt80* deletion mutant in three species that branch between *S. cerevisiae* and *C. albicans*: *K. lactis*, *P. pastoris,* and *C. lusitaniae* (Figure 2A). We first tested whether this deletion had an effect on sporulation and found, in *K. lactis* and *P. pastoris,* that the *ndt80* deletion mutant is completely deficient in its ability to sporulate (Figure 4A and B). C. *lusitaniae*, like *C. albicans*, has two paralogs of Ndt80. Deletion of one paralog (Ndt80A) results in a complete deficiency in sporulation, while deletion of the other paralog (Ndt80B) results in a significant reduction in sporulation efficiency (Figure 4C). In contrast to these four species, *C. albicans* has never been observed to undergo sporulation and meiosis, relying instead on an alternate parasexual cycle (Bennett & Johnson, 2003; Butler et al., 2009). Given that Ndt80 is required for sporulation in *S. cerevisiae* (Chu & Herskowitz, 1998; Xu et al., 1995), *K. lactis, P. pastoris*, and *C. lusitaniae*, and given the phylogenetic relationship between these species (Figure 2A), the most parsimonious model is that, in the ancestor of these species, Ndt80 regulated sporulation.

We next investigated the requirement for Ndt80 in biofilm formation in these species. In *C. albicans*, biofilms consist of thick structures of different cell types (yeast-form and hyphae)

that can form both *in vitro* and *in vivo* (Andes et al., 2004)). Deletion of *NDT80B* in *C. albicans* causes a severe defect in surface adherence and biofilm formation (Figure 4H). In contrast, in *S. cerevisiae*, *K. lactis*, *P. pastoris*, and *C. lusitaniae*, the *ndt80* mutants and wild-type are comparable in their ability to adhere to a solid surface (Figure 4D - G). It is worth noting that *S. cerevisiae*, *K. lactis* and *P. pastoris* form only a thin layer of yeast-form cells on the solid surface rather than the type of thick, multicellular biofilms characteristic of those formed in *C. albicans* while *C. lusitaniae* forms a biofilm somewhat thinner than that of *C. albicans*. These results highlight the specialized type of biofilm produced by *C. albicans* and the pivotal role played by Ndt80 in its evolution.

Taken together, these results indicate that Ndt80 regulated sporulation in the shared ancestor of *S. cerevisiae* and *C. albicans*, and that it gained a role in biofilm formation along the *C. albicans* lineage, after it branched off from *C. lusitaniae* (Figure 2A). We believe this scenario is more likely than that of the next most parsimonious model, which holds that Ndt80 regulated both sporulation and biofilm formation in the shared ancestor, as this would require at least three independent losses of biofilm regulation. Given that Ndt80 regulates hundreds of genes in each species, a single gain of Ndt80 function seem more plausible than three independent losses.

Combining this phenotypic data with the Ndt80 target identification previously discussed, we can begin to pinpoint the importance of individual genes in the overall switch in Ndt80 function. Only ten genes are targets of Ndt80 in all five species tested (*S. cerevisiae*, *K. lactis, P. pastoris*, *S. stipitis*, and *C. albicans*) (Figure 2C). This suggests that a very limited set of genes may have been repurposed from sporulation to biofilm formation. Notably, five of these genes (*CLN2*, *CLG1*, *GLN1*, *GDH2*, *RPS7a*) are strongly up-regulated in biofilm formation (Fox et al.,

2015). If we examine the ten shared genes based on their known functions in *S. cerevisiae*, three are involved in cell cycle regulation (*CLG1*, *CLN2*, *YOX1*), four are metabolic enzymes (*GDH2*, *GLN1*, *RKI1*, *SGA1*), two are proteins associated with the ribosome (*RPS7a*, *TMA10*), and one is involved in regulation of cell wall biosynthesis (*USV1*) (Figure 2C). The protein Usv1 is particularly intriguing, as its ortholog in *C. albicans*, Bcr1, is one of the other master regulators of biofilm formation (Nobile et al., 2012; Nobile & Mitchell, 2005). This observation suggests that two of the master regulators of biofilm formation, Ndt80 and Bcr1, may have had an ancestral regulatory relationship that was co-opted in the evolution of the biofilm regulatory network. In summary, while there are small vestiges of the sporulation network present in the *C. albicans* biofilm network, the great majority of connections are new.

The transcription network that controls meiosis and sporulation changes extensively

Because the targets of Ndt80 appear to have rewired significantly during the switch from sporulation regulation to biofilm regulation, we tested whether Ndt80 targets are more similar among species with a conserved Ndt80 phenotype than among species with a diverged Ndt80 phenotype. If so, this would suggest that the dramatic rewiring of targets was associated with the switch in Ndt80 regulatory function. We focused on the four species for which we know both the Ndt80 phenotype and the binding targets: *S. cerevisiae*, *K. lactis*, *P. pastoris*, and *C. albicans*. For every two-species pair, we compared the Ndt80 targets, considering only genes with 1:1 orthologs for both species. We find relatively little overlap in Ndt80 targets for any of these comparisons (9-14%), but more importantly, we do not find a correlation between target overlap and conservation of overall Ndt80 function (Figure 5A). For example, a larger fraction of targets (14%) are shared between *S. cerevisiae* and *C. albicans* than between *S. cerevisiae* and

307 *P. pastoris* (9.2%). If we take into account the different divergence times for each of these two-
308 species comparisons (normalizing to the divergence time between *S. cerevisiae* and *K. lactis)*, we
309 still find that conservation of overall Ndt80 function does not correlate with increased
310 conservation of Ndt80 targets (Figure 5B). This pattern also holds if we use more stringent
311 criteria to identify Ndt80 targets in each species (Figure 5—figure supplement 1). Overall, these
312 results show that the targets of Ndt80 change extensively, even while Ndt80's conserved role in
313 regulating sporulation is maintained. While we cannot entirely rule out the possibility that Ndt80
314 has another, unknown regulatory function in one or more of these species that accounts for the
315 significant difference in Ndt80 targets, we know that *S. cerevisiae* Ndt80 is produced specifically
316 during sporulation (Chu & Herskowitz, 1998). Similarly, by performing transcription profiling
317 we find that in *K. lactis* and *P. pastoris,* Ndt80 is not expressed in mitotic cells (Supplementary
318 File 1), consistent with the idea that Ndt80 is primarily needed during sporulation and is unlikely
319 to have additional regulatory roles.
320 The rewiring of Ndt80 targets across species in which Ndt80 is known to regulate
321 sporulation indicates that only a small number of Ndt80 targets must be maintained in order for
322 the sporulation network to be functional. Although there are several possible explanations for
323 how Ndt80 could have maintained a role in sporulation despite the near complete turnover of its
324 target genes, we tested perhaps the most intriguing hypothesis: the genes required for sporulation
325 and meiosis are themselves changing across species. To test this hypothesis, we measured global
326 gene expression during sporulation and mitotic growth in *K. lactis* and *P. pastoris* (Materials and
327 Methods). We compared the genes up-regulated during sporulation in these species to those
328 upregulated during sporulation in *S. cerevisiae* (Chu et al., 1998) and found extensive differences
329 in the genes activated during sporulation across these species (Figure 6A). Only 25 genes show

sporulation-activation across *S. cerevisiae*, *K. lactis*, and *P. pastoris*, while 577 genes are activated uniquely in only one of the three species. Of the 65 genes known to be required for sporulation in *S. cerevisiae* that also exhibit sporulation-specific activation, only 9 are activated during sporulation in both *K. lactis* and *P. pastoris*, with 21 of these genes missing altogether from the genomes of one or both of these species (Figure 6—figure supplement 2). We also measured global gene expression in an *ndt80* deletion mutant in both *K. lactis* and *P. pastoris* under sporulation conditions, and found that the genes whose expression depends on Ndt80 during sporulation also differ significantly across species, with only seven genes showing Ndt80-dependent expression in all three species (Figure 6B, Figure 6—figure supplement 1). We conclude that there are significant differences in sporulation-specific gene expression among *S. cerevisiae*, *K. lactis*, and *P. pastoris*, and that this contributes significantly to the large differences observed in Ndt80 targets across species.

DISCUSSION

Using experiments performed in six different extant yeast species, we have deduced that the transcription regulator Ndt80 underwent extensive rewiring of its regulatory connections before it became incorporated into a newly evolving regulatory network. This incorporation, which occurred on the lineage leading to *C. albicans*, resulted in an overall switch in Ndt80's function from an ancestral role in regulating sporulation and meiosis to a derived role in regulating biofilm formation. We have shown that this switch in function was not accompanied by a change in the DNA-binding specificity of Ndt80 (Figure 1B, Figure 2—figure supplement 4); rather, a change in the distribution of the conserved Ndt80 *cis*-regulatory sequence across

353 hundreds of genes in the genome largely accounts for the different functions of Ndt80 in *S.*
354 *cerevisiae* and *C. albicans*.

355 A priori, we considered two likely explanations for the change in Ndt80 function. First,
356 we hypothesized that a duplication of Ndt80 in the *C. albicans* clade could have allowed the
357 ancestral function (sporulation) to be conserved in one paralog while the other acquired a new
358 function (biofilm regulation). However, the similarity in function of Ndt80 paralogs in *C.*
359 *lusitaniae* (Figure 4C and F) as well as the similarity in targets of Ndt80 paralogs in *C. albicans*
360 (Figure 1—figure supplement 3) rule out this model. A second plausible model held that a
361 sudden change in Ndt80-target gene connections occurred along the *C. albicans* lineage, perhaps
362 triggered by the loss of meiosis in this clade, relaxing constraints on Ndt80 target gene
363 connections and allowing old connections to rapidly break and new connections to rapidly form.
364 However, we also ruled out this model by showing that similarly high rates of Ndt80 rewiring
365 occurred in all phylogenetic branches examined, even those where the function of Ndt80 remains
366 conserved in sporulation (Figure 5A and B). Rather than a sudden shift in regulation, our results
367 indicate continuous flexibility in the regulon of Ndt80, allowing it to sample many
368 configurations even while retaining its ancestral function. In support of this idea, we showed
369 that genes up-regulated during sporulation (by Ndt80 and other regulators of sporulation) vary
370 considerably across different species (Figure 6), indicating that many different network
371 configurations can support sporulation. Consistent with this idea, it has previously been noted
372 that several key genes required for meiosis and sporulation in *S. cerevisiae* (*IME1*, *ZIP2*, *SPO13*)
373 are missing from the genomes of species such as *C. lusitaniae* that are known to undergo meiosis
374 and sporulation (Bennett & Johnson, 2003; Sherwood & Bennett, 2009). These observations,
375 taken together, indicate that there is no "universal" solution to meiosis and sporulation in terms

of both the genes required for this process and the regulatory network itself. Instead, many different networks can apparently orchestrate this ancient process.

This work illustrates how a transcription regulator with hundreds of ancestral connections can nonetheless completely shift its cellular function. We propose that the continuous exploration of regulatory connections available to Ndt80—while still maintaining its role in sporulation—facilitated it reaching a network configuration that supported a new role in biofilm regulation. Computational models had previously led to this view, namely that gene regulatory networks are inherently flexible, allowing for the exploration of many configurations that maintain an overall function (Barve & Wagner, 2013; A. Wagner, 2005b). This exploration allows networks to sample many points in "network space", some of which may be only a few changes away from generating a novel output. We believe that Ndt80 represents a tangible example of this idea: we propose that it is only through the inherent flexibility of the Ndt80 connections that the network—while still maintaining sporulation and meiosis—could reach a position where the initial transition to regulating biofilm formation would be a simple one.

Although other transcription regulators may be subject to stronger evolutionary constraints than Ndt80, there is ample evidence that many transcription networks, like that of Ndt80, can drift through new configurations while preserving their transcriptional output (Baker et al., 2012; Hare, Peterson, Iyer, Meier, & Eisen, 2008; Lavoie et al., 2010; Swanson et al., 2011; Tanay et al., 2005; Tsong et al., 2006; Villar et al., 2015). Although the overall rate of network rewiring has not been extensively documented, it seems likely that the rewiring of the Ndt80 regulon is not unusually rapid when compared to at least some other fungal regulators (Borneman et al., 2007; Tuch et al., 2008). While much regulatory rewiring likely occurs through drift with no obvious shift in phenotype, we propose that continuous rewiring also

predisposes regulators to a shift in function. Such extreme, inherent flexibility in gene regulatory networks, such as that exhibited by Ndt80, seems likely to be a general model to explain how complex regulatory networks can evolve to produce novel phenotypes.

MATERIALS AND METHODS

Media

All strains were grown in YPD at 30°C unless otherwise noted. For galactose induction, strains were grown in SRaffinose + 1.7% galactose. For sporulation, *S. cerevisiae* strains were grown in liquid YPA (2% peptone, 2% potassium acetate, 1% yeast extract) and incubated at room temperature for 20-30 hours. For *K. lactis*, saturated liquid cultures in YPD were spotted onto SPO plates (1% potassium acetate, 2% agar + amino acids) and incubated at room temperature for 3 days. For *P. pastoris*, cells growing on YPD plates were patched onto 0.5% sodium acetate, 1% potassium chloride, 1% glucose, 2% agar plates and incubated at room temperature for 3 days. For *C. lusitaniae*, cells growing on YPD plates were patched onto PDA plates (0.37% potato dextrose + 1.45% agar) and incubated at room temperature for 4 days. To quantify sporulation efficiency, three technical replicates (the same strain, grown up and plated on sporulation media independently) were performed and 200 cells were counted for each sample by DIC microscopy. In addition, a biological replicate (an independently-generated deletion strain) was assayed to confirm the phenotype of the first biological replicate. Nuclear staining using DAPI (*K. lactis*, *P. pastoris*) and Hoecsht (*C. lusitaniae*) was used to verify spore formation in each experiment.

Strain construction

Strains used in this study are listed in Supplementary File 2A and primers used for strain construction are listed in Supplementary File 2B. Gene disruption cassettes for Ndt80 deletions were constructed by fusion PCR. For *K. lactis*, two cassettes were constructed, one with a *URA3*

marker, one with a KanMX marker, each with 700 bp homology flanking the markers on either side. These constructs were transformed, sequentially, into a wild-type diploid strain. For *P. pastoris* and *C. lusitaniae*, a split marker approach was used and two constructs were generated for each gene disruption. For *P. pastoris*, a Hygromycin resistance marker and 700 bp flanking homology were used, and the split markers were transformed into two haploid strains of complementary mating types. These strains were then mated to form a diploid *ndt80* deletion (ploidy was verified by FACS). For *C. lusitaniae*, a NAT resistance marker was used along with 1 kb of flanking homology. The split markers were transformed into a wild-type α-cell, that was then mated with a wild-type **a**-cell, sporulated, and *ndt80* deletion **a**-cells were picked and verified by PCR. These were then mated with the *ndt80* deletion α-cell to generate a diploid *ndt80* deletion strain.

Tagged strains for ChIP were generated using a 13x C-terminal Myc tag (Longtine et al., 1998) inserted into the genome at the C terminus of the endogenous Ndt80 protein sequence. For *S. cerevisiae*, long primers were used to amplify the Myc tag fused to a KAN marker from pFA6a-13Myc- kanMX6 (Longtine et al., 1998) and this construct was integrated into a wild-type W303 strain. For *K. lactis*, split marker constructs were generated with 700 bp flanking homology fused to the Myc-KAN cassette and transformed into a wild-type diploid strain. For *P. pastoris*, split marker constructs were generated with 700 bp flanking homology fused to the Myc-NAT construct amplified from pADH34 (Hernday, Noble, Mitrovich, & Johnson, 2010). In *S. stipitis*, homologous recombination was found to be very inefficient, and thus constructs were randomly integrated into the genome. Constructs containing a NAT marker upstream of an *E. gossypii* Tef1 promoter upstream of *S. stipitis* Ndt80 fused to a 13x C-terminal Myc tag with a Sat1

terminator were generated by PCR for both Ndt80A and Ndt80B and transformed into a wild-type strain. Two independent isolates were generated for each paralog and tested independently by ChIP-Seq to verify that the location of integration did not affect Ndt80 function or genomic binding. To test binding of *P. pastoris* Ndt80 in *S. cerevisiae*, fusion PCR was used to generate a construct with a Hygromycin resistance marker fused to the Gal1 promoter from *S. cerevisiae* upstream of *P. pastoris* Ndt80 fused to a 13x C-terminal Myc tag with a Sat1 terminator. This was fused to homology to *URA3* and integrated in an *S. cerevisiae* wild-type at the *URA3* locus. All Myc-tagged Ndt80 strains were tested for their ability to sporulate to ensure the tag did not interfere with endogenous Ndt80 function.

To generate strains with highly expressed Ndt80, different constitutive promoters were used, and constructs generated by fusion PCR to integrate these upstream of Ndt80. In *S. cerevisiae*, pGal1 from *S. cerevisiae* was used; in *K. lactis* pGal1 from *K. lactis* was used; in *P. pastoris* and *S. stipitis* pTef1 from E. gossypii was used; in *C. albicans* pTDH3 from *C. albicans* was used. In all cases, RT-qPCR was performed on the resulting strains as well as a wild-type strain to ensure that the constitutive promoters were indeed driving high Ndt80 expression.

S. cerevisiae and *C. albicans* were transformed using standard lithium acetate protocols. Electroporation protocols were used for *K. lactis* (Booth, Tuch, & Johnson, 2010), *P. pastoris* (Cregg et al., 2009), and *S. stipitis*. The protocol for *S. stipitis* was adapted from (Cregg et al., 2009), with cells harvested at an OD of 1.3-1.5.

Genome sequences, gene annotations, and orthology mapping

Genome sequences and gene annotations for *S. mikatae*, *S. kudriavzevii*, *K. lactis*, *E. gossypii*, and *E. cymbalariae* were downloaded from the Yeast Gene Order Browser (YGOB) (Byrne &

Wolfe, 2005). Genome sequences and gene annotations for *S. cerevisiae*, *C. lusitaniae*, *S. stipitis*, *C. dubliniensis*, *C. tropicalis*, and *C. albicans* were downloaded from the Candida Gene Order Browser (CGOB) (Maguire et al., 2013). The genome sequence of *P. pastoris* strain CB7435 (Küberl et al., 2011) was downloaded from NCBI (accession number: PRJEA62483), and gene annotations were generated using the Yeast Genome Annotation Pipeline (Proux-Wéra, Armisén, Byrne, & Wolfe, 2012). Gene annotations for YGOB, CGOB, and YGAP also include synteny-based orthology mapping, which was used to comparie genes across species.

Chromatin immunoprecipitation and high throughput sequencing

For all ChIP experiments with endogenous or pTef1 promoters driving Ndt80 expression, samples were isolated from log-phase cultures grown in YPD and chromatin immunoprecipitation was performed as previously described (Hernday et al., 2010) using an anti-Myc monoclonal antibody (RRID: AB_2536303). Libraries were prepared using NEBNext Multiplex Kit for Illumina as previously described (Sorrells et al., 2015) and sequenced on an Illumina HiSeq 4000. For the mid-sporulation ChIP in *S. cerevisiae*, cells were induced to sporulate as previously described (Chu & Herskowitz, 1998; Hepworth, Friesen, & Segall, 1998), and samples were isolated after 16 hours in sporulation media. For ChIP of strains with pGal1 driving Ndt80 expression, strains were grown overnight in SRaffinose, diluted back and grown until log-phase in SRaffinose, then grown in SRaffinose + 1.7% galactose for 5 hours.

For each ChIP-Seq experiment, a control experiment was performed using a matched strain missing the C-terminal Myc tag. Two biological replicates (independently grown single-colonies

of the same strain) were performed for both the tagged strain and the control untagged strain to identify regions bound by Ndt80 for that strain and condition.

RNA sequencing in *P. pastoris* design and analysis

RNA expression was measured in *P. pastoris* using RNA-Seq. Samples were isolated from log-phase cultures in YPD or from mid-sporulation cultures after 30 hours. Total RNA was isolated using the Ambion RiboPure kit, and mRNA was isolated using the Oligotex mRNA Mini kit. Purified mRNA was then concentrated using the Zymo RNA Clean and Concentrator kit. Sequencing libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit and sequenced on an Illumina HiSeq 4000. Three technical replicates (independently-grown single colonies of the same strain) were performed for each strain.

Reads were aligned to the genome using TopHat (Trapnell, Pachter, & Salzberg, 2009). Transcripts were then assembled, abundance was estimated, and differential expression was detected using Cufflinks (Trapnell et al., 2012).

Full genome expression microarray in *K. lactis*

Custom-designed *K. lactis* microarrays were used to measure differential expression of genes in sporulation vs. mitotic growth and in an *ndt80* deletion vs. wild-type in sporulation media as previously described (Booth et al., 2010). Samples were isolated from log-phase cultures in YPD or from mid-sporulation cultures after 21 hours. Two technical replicates (independently-grown single colonies of the same strain) of each experiment were performed and the average of these replicates was taken.

Comparing gene expression across *S. cerevisiae, K. lactis*, and *P. pastoris*

To identify genes up-regulated in sporulation (Figure 6—figure supplement 1), different thresholds were used to compare gene expression in a wild-type strain grown in YPD to a wild-type strain grown in sporulation media. For *S. cerevisiae*, this threshold was a 2.5-fold expression increase in sporulation vs. mitotic growth, in a published dataset (Chu et al., 1998). For *K. lactis*, the threshold was a 1.5-fold expression increase in sporulation vs. mitotic growth in expression microarrays (described in detail above). For *P. pastoris*, the threshold was a statistically significant increase in sporulation vs. mitotic growth, as determined by analysis using Cuffdiff (Trapnell et al., 2012).

Similarly, to identify genes exhibiting Ndt80-specific expression in sporulation (Figure 6—figure supplement 1), expression was compared between a wild-type strain grown in sporulation media and an *ndt80* deletion strain grown in sporulation media. Different thresholds were applied to identify genes showing Ndt80-dependent expression: at least 2-fold expression increase in WT vs. *ndt80* mutant for *S. cerevisiae*, at least 1.5-fold expression increase in in WT vs. *ndt80 mutant* for *K. lactis*, and a statistically significant increase (Trapnell et al., 2012) in *P. pastoris*.

Biofilm experiments

Biofilms were grown and imaged by confocal scanning laser microscopy (CSLM) similar to previously described (Mancera, Porman, Cuomo, Bennett, & Johnson, 2015). In brief, silicon squares were pre-incubated overnight in bovine serum at 30^{o}C, washed with phosphate-buffered

539 saline (PBS), and then submerged in Spider media containing 1% glucose instead of mannitol.
540 Then cells from a culture grown overnight in YEPD at 30°C were added to the silicone squares
541 upon reaching an OD_{600} of 0.5. The squares were incubated for 90 minutes at 30°C shaking at
542 200 rpm for cell adherence. After adherence the squares were washed with PBS and transferred
543 to fresh Spider 1% glucose and incubated for 48 hours at 30°C shaking at 200 rpm. For CSLM,
544 the biofilms grown on the silicon squares were stained with concanavalin A Alexa Fluor 594
545 conjugate (50 μg/ml) and visualized using a Nikon Eclipse C1si upright spectral imaging
546 confocal microscope and a 40x/0.80W Nikon objective.
547
548 **Identifying regions of Ndt80 binding**
549 Sequencing reads were aligned to the genome using Bowtie (Langmead, Trapnell, Pop, &
550 Salzberg, 2009). Alignments were converted for visualization using SAMtools (Heng Li et al.,
551 2009). Peaks and fold enrichment values were generated using MACS2 (Zhang et al., 2008)
552 with peak shift sizes generated using SPP (Kharchenko, Tolstorukov, & Park, 2008). Peaks were
553 assigned to genes if any portion of the peak overlapped with the intergenic region upstream of
554 that gene using MochiView (Homann & Johnson, 2010). For the heterologous *P. pastoris* Ndt80
555 in *S. cerevisiae* ChIP, fold enrichment values were mapped to genes by taking the maximum fold
556 enrichment value for each intergenic region and mapping that to neighboring genes, also using
557 MochiView.
558
559 Two technical replicates were performed for each ChIP-Seq experiment. Genes with Ndt80
560 peaks were identified for each replicate, and only genes with peaks in both replicates were
561 considered as bona fide targets. For fold enrichment per gene calculations, the average of the

562 maximum fold enrichment per intergenic region in each replicate was taken as the fold
563 enrichment for neighboring genes.
564
565 Ndt80 targets were identified using several different criteria (Figure 1C). For Criteria 1, a gene
566 was considered to be bound by Ndt80 if a peak was present in the intergenic upstream of that
567 gene. For Criteria 2, a gene was considered to be bound by Ndt80 if a peak was present in
568 upstream intergenic region and a consensus motif (CACAAA) was also present in that intergenic
569 region (we did not require the motif to overlap the peak location). For Criteria 3, a gene was
570 considered to be bound by Ndt80 if a peak was present in the upstream intergenic region, and a
571 consensus motif was present in the intergenic of that species as well as intergenic regions
572 upstream of orthologous genes in two closely related species. For *S. cerevisiae*, *S. mikatae* and
573 *S. kudriavzevii* were used in this analysis; for *K. lactis*, *E. gossypii* and *E. cymbalariae* were
574 used; and for *C. albicans*, *C. tropicalis* and *C. dubliniensis* were used (Figure 2—figure
575 supplement 1). For Criteria 4, a gene was considered to be bound by Ndt80 is a peak was present
576 in the upstream intergenic region, a consensus motif was also present in the upstream intergenic
577 region, and the gene exhibited Ndt80-dependent expression. In *S. cerevisiae*, this was defined as
578 at least a two-fold change in expression in an *ndt80* deletion strain (up or down) compared with a
579 wild-type strain during middle-meiosis (Chu et al., 1998). In *K. lactis*, this was defined as at
580 least a 1.5-fold change in expression in the average of two replicates of an *ndt80* deletion
581 compared to wild-type, as measured by expression array. In *P. pastoris*, this was defined as a
582 statistically significant change in expression, as measured by Cufflinks (Trapnell et al., 2012)
583 between three replicates of an *ndt80* deletion strain and three replicates of a wild-type strain. In

584 *C. albicans*, this was defined as at least a 1.5-fold change in expression between an *ndt80*
585 deletion and wild-type (Nobile et al., 2012).
586
587 **DNA *cis*-regulatory sequence motif discovery and enrichment**
588 Ndt80 motifs were generated *de novo* for each ChIP-Seq experiment using DREME (Bailey,
589 2011). The union of peak locations in two technical replicates for each experiment were
590 submitted to DREME, with a random set of genomic sequences of the same length as a negative
591 control. The motif most closely matching the known Ndt80 binding site from *S. cerevisiae* (Jolly
592 et al., 2005) and *C. albicans* (Nobile et al., 2012) was recorded along with the accompanying e-
593 value (Figure 2—figure supplement 4).
594
595 A consensus Ndt80 motif across all species was generated using the union of replicate peak
596 locations for all experiments performed with highly-expressed Ndt80. For *C. albicans* and *S.*
597 *stipitis*, the intersection of replicate peaks was taken for each Ndt80 paralog, and then the union
598 of paralog peak locations were taken to represent Ndt80 binding in that species. Peak locations
599 for all five species tested were submitted to DREME, using a shuffled set of the same sequences
600 as a negative control. The resulting motif (Figure 2—figure supplement 4) was trimmed by one
601 base pair to give the high-confidence "consensus motif" using for identifying high-confidence
602 Ndt80 binding: CACAAA.
603
604 To determine enrichment of the consensus motif in binding locations for each ChIP-Seq
605 experiment, compared to a random genomic background, a Fisher's one-tailed exact test was
606 performed. The number of motifs in the peak locations for an experiment was compared to the

number of motifs in a set of randomly generated genomic sequences of the same length to generate a p-value representing enrichment (Figure 2—figure supplement 4).

Phylogenetic tree building

A phylogenetic tree of relevant species was constructed as previously described (Lohse et al., 2013). Protein sequences for 73 orthologs present in a single copy in all species were concatenated and aligned using MUSCLE (Edgar, 2004) and a tree was constructed using PHYML (Guindon et al., 2010). Two outgroup species, *N. crassa* and *A. nidulans* were used in the building of the tree to improve root placement, but were omitted from the tree image (Figure 2A) for simplicity.

A phylogenetic tree of Ndt80 protein sequences (Figure 2—figure supplement 1) was similarly constructed by aligning all Ndt80 protein sequences using MUSCLE (Edgar, 2004) and using PHYML to build a tree (Guindon et al., 2010). Similarity matrix of protein sequences (Figure 2—figure supplement 1) was generated by MUSCLE.

Estimating sequence conservation between species

To determine overall sequence conservation between two species (Figure 6B, Figure 5—figure supplement 1), branch lengths from the phylogenetic species tree were used (Figure 2A). The additive branch lengths separating two species from the most recent common ancestor were calculated for each two-species comparison, and then this was normalized to the value for *S. cerevisiae* and *K. lactis* (arbitrarily set to 1). This number was then multiplied by the fraction of

Ndt80 targets conserved for any given two-species comparison to get a divergence time-corrected value.

Simulation of overlap in Ndt80 ChIP-Seq targets across all species

In order to determine whether the 10 genes shown to be Ndt80 targets in all five species represented more than expected by chance (based on the number of targets in each species), we performed 100,000 simulations. In each simulation, genes were randomly selected from the pool of 3,171 genes with orthologs in all five species, to represent the Ndt80 targets in all five species. For *S. cerevisiae*, 424 genes were selected, for *K. lactis* 269, for *P. pastoris* 300, for *S. stipitis* 958, and for *C. albicans* 989, as these are the number of mappable targets of Ndt80 in each species according to Criteria 2 (Figure 1C). For 100,000 simulations, the median number of genes found in all five sets was zero, and the maximum was five. Ten or more genes were not found in any round of the simulation (p-value < 0.00001).

Data deposition

ChIP-Seq and RNA-Seq data has been deposited to the NCBI Gene Expression Omnibus (GEO) repository under accession numbers GSE90660 (*S. cerevisiae* mitotic ChIP-Seq), GSE90661 (*S. cerevisiae* meiotic ChIP-Seq), GSE90662 (*K. lactis* ChIP-Seq), GSE90663 (*P. pastoris* ChIP-Seq), GSE90664 (*S. stipitis* ChIP-Seq), GSE90665 (*C. albicans* ChIP-Seq), and GSE92766 (RNA-Seq).

ACKNOWLEDGEMENTS

We thank C. Dalal, S. Coyle, C. Baker, C. Britton, S. Singh-Babak, and L. Pack for valuable comments on the manuscript; T. Sorrells and C. Nobile for important technical contributions and comments on the manuscript; E. Chow and the UCSF CAT for expert advice; and R. Bennett and R. Sherwood for strains and protocols. This work was supported by grant R01 GM037049 from the National Institutes of Health. E.M. was supported by the Human Frontier Science Program and UC-MEXUS.

COMPETING INTERESTS

The authors declare that no competing interests exist.

REFERENCES

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C., & Tawfik, D. S. (2004). The “evolvability” of promiscuous protein functions. *Nature Genetics*, 1–4. http://doi.org/10.1038/ng1482

Andes, D., Nett, J., Oschel, P., Albrecht, R., Marchillo, K., & Pitula, A. (2004). Development and Characterization of an In Vivo Central Venous Catheter Candida albicans Biofilm Model. *Infection and Immunity*, *72*(10), 6023–6031. http://doi.org/10.1128/IAI.72.10.6023-6031.2004

Bailey, T. L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, *27*(12), 1653–1659. http://doi.org/10.1093/bioinformatics/btr261

Baker, C. R., Booth, L. N., Sorrells, T. R., & Johnson, A. D. (2012). Protein Modularity, Cooperative Binding, and Hybrid Regulatory States Underlie Transcriptional Network Diversification. *Cell*, *151*(1), 80–95. http://doi.org/10.1016/j.cell.2012.08.018

Barve, A., & Wagner, A. (2013). A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, *500*(7461), 203–206. http://doi.org/10.1038/nature12301

Bennett, R. J., & Johnson, A. D. (2003). Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. *The EMBO Journal*, *22*(10), 2505–2515. http://doi.org/10.1093/emboj/cdg235

Bonhomme, J., & D'enfert, C. (2013). Candida albicans biofilms: building a heterogeneous, drug-tolerant environment. *Current Opinion in Microbiology*, *16*(4), 398–403. http://doi.org/10.1016/j.mib.2013.03.007

Booth, L. N., Tuch, B. B., & Johnson, A. D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature*, *468*(7326), 959–963. http://doi.org/10.1038/nature09560

Borneman, A. R., Gianoulis, T. A., Zhang, Z. D., Yu, H., Rozowsky, J., Seringhaus, M. R., et al. (2007). Divergence of Transcription Factor Binding Sites Across Related Yeast Species. *Science (New York, NY)*, *317*(5839), 815–819. http://doi.org/10.1126/science.1140748

Bridgham, J. T., Carroll, S. M., & Thornton, J. W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. *Science (New York, NY)*, *312*(5770), 97–101. http://doi.org/10.1126/science.1123348

Butler, G., Rasmussen, M. D., Lin, M. F., Santos, M. A. S., Sakthikumar, S., Munro, C. A., et al. (2009). Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature*, *459*(7247), 657–662. http://doi.org/10.1038/nature08064

Byrne, K. P., & Wolfe, K. H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, *15*(10), 1456–1461. http://doi.org/10.1101/gr.3672305

Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biology*, *3*(7), e245. http://doi.org/10.1371/journal.pbio.0030245

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research*, *40*(Database issue), D700–5. http://doi.org/10.1093/nar/gkr1029

Chu, S., & Herskowitz, I. (1998). Gametogenesis in yeast is regulated by a transcriptional cascade dependent on Ndt80. *Molecular Cell*, *1*(5), 685–696.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., & Herskowitz, I.

(1998). The transcriptional program of sporulation in budding yeast. *Science (New York, NY)*, *282*(5389), 699–705.
Coyle, S. M., Flores, J., & Lim, W. A. (2013). Exploitation of Latent Allostery Enables the Evolution of New Modes of MAP Kinase Regulation. *Cell*, *154*(4), 875–887. http://doi.org/10.1016/j.cell.2013.07.019
Cregg, J. M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., & Chappell, T. (2009). Expression in the yeast Pichia pastoris. *Methods in Enzymology*, *463*, 169–189. http://doi.org/10.1016/S0076-6879(09)63013-5
Davidson, E. H., & Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science (New York, NY)*, *311*(5762), 796–800. http://doi.org/10.1126/science.1113832
Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. http://doi.org/10.1093/nar/gkh340
Fan, X., & Struhl, K. (2009). Where Does Mediator Bind In Vivo? *PloS One*, *4*(4), e5029–6. http://doi.org/10.1371/journal.pone.0005029
Fox, E. P., Bui, C. K., Nett, J. E., Hartooni, N., Mui, M. C., Andes, D. R., et al. (2015). An expanded regulatory network temporally controls Candida albicansbiofilm formation. *Molecular Microbiology*, *96*(6), 1226–1239. http://doi.org/10.1111/mmi.13002
Frankel, N., Wang, S., & Stern, D. L. (2012). Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(51), 20975–20979. http://doi.org/10.1073/pnas.1207715109
Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., & Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature*, *433*(7025), 481–487. http://doi.org/10.1038/nature03235
Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. http://doi.org/10.1093/sysbio/syq010
Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R., & Eisen, M. B. (2008). Sepsid even-skipped Enhancers Are Functionally Conserved in Drosophila Despite Lack of Sequence Conservation. *PLoS Genetics*, *4*(6), e1000106–13. http://doi.org/10.1371/journal.pgen.1000106
Hepworth, S. R., Friesen, H., & Segall, J. (1998). NDT80 and the meiotic recombination checkpoint regulate expression of middle sporulation-specific genes in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *18*(10), 5750–5761.
Hernday, A. D., Noble, S. M., Mitrovich, Q. M., & Johnson, A. D. (2010). Methods in Enzymology, *470*, 737–758. http://doi.org/10.1016/S0076-6879(10)70031-8
Homann, O. R., & Johnson, A. D. (2010). MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biology*, *8*, 49. http://doi.org/10.1186/1741-7007-8-49
Hulsen, T., de Vlieg, J., & Alkema, W. (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, *9*(1), 488–6. http://doi.org/10.1186/1471-2164-9-488
Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., & Barkai, N. (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science (New York, NY)*, *309*(5736), 938–940. http://doi.org/10.1126/science.1113833

Jolly, E. R., Chin, C.-S., Herskowitz, I., & Li, H. (2005). Genome-wide identification of the regulatory targets of a transcription factor using biochemical characterization and computational genomic analysis. *BMC Bioinformatics*, *6*, 275. http://doi.org/10.1186/1471-2105-6-275

Kharchenko, P. V., Tolstorukov, M. Y., & Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology*, *26*(12), 1351–1359. http://doi.org/10.1038/nbt.1508

Khersonsky, O., & Tawfik, D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry*, *79*, 471–505. http://doi.org/10.1146/annurev-biochem-030409-143718

Kojic, E. M., & Darouiche, R. O. (2004). Candida Infections of Medical Devices. *Clinical Microbiology Reviews*, *17*(2), 255–267. http://doi.org/10.1128/CMR.17.2.255-267.2004

Küberl, A., Schneider, J., Thallinger, G. G., Anderl, I., Wibberg, D., Hajek, T., et al. (2011). High-quality genome sequence of Pichia pastoris CBS7435. *Journal of Biotechnology*, *154*(4), 312–320. http://doi.org/10.1016/j.jbiotec.2011.04.014

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. http://doi.org/10.1186/gb-2009-10-3-r25

Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., & Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biology*, *8*(3), e1000329. http://doi.org/10.1371/journal.pbio.1000329

Li, Hao, & Johnson, A. D. (2010). Evolution of transcription networks - lessons from yeasts. *Current Biology : CB*, *20*(17), R746–53. http://doi.org/10.1016/j.cub.2010.06.056

Li, Heng, Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. http://doi.org/10.1093/bioinformatics/btp352

Lohse, M. B., Hernday, A. D., Fordyce, P. M., Noiman, L., Sorrells, T. R., Hanson-Smith, V., et al. (2013). Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(19), 7660–7665. http://doi.org/10.1073/pnas.1221734110

Longtine, M. S., McKenzie, A., Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., et al. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast (Chichester, England)*, *14*(10), 953–961. http://doi.org/10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U

Ludwig, M. Z., Bergman, C., Patel, N. H., & Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*, *403*(6769), 564–567. http://doi.org/10.1038/35000615

Maguire, S. L., Oheigeartaigh, S. S., Byrne, K. P., Schroder, M. S., O'gaora, P., Wolfe, K. H., & Butler, G. (2013). Comparative Genome Analysis and Gene Finding in Candida Species Using CGOB. *Molecular Biology and Evolution*, *30*(6), 1281–1291. http://doi.org/10.1093/molbev/mst042

Mancera, E., Porman, A. M., Cuomo, C. A., Bennett, R. J., & Johnson, A. D. (2015). Finding a Missing Gene: EFG1 Regulates Morphogenesis in Candida tropicalis. *G3 (Bethesda, Md.)*, *5*(5), 849–856. http://doi.org/10.1534/g3.115.017566

Nobile, C. J., & Johnson, A. D. (2015). Candida albicans Biofilms and Human Disease. *Annual*

Review of Microbiology, *69*(1), 71–92. http://doi.org/10.1146/annurev-micro-091014-104330

Nobile, C. J., & Mitchell, A. P. (2005). Regulation of cell-surface genes and biofilm formation by the C. albicans transcription factor Bcr1p. *Current Biology : CB*, *15*(12), 1150–1155. http://doi.org/10.1016/j.cub.2005.05.047

Nobile, C. J., Fox, E. P., Nett, J. E., Sorrells, T. R., Mitrovich, Q. M., Hernday, A. D., et al. (2012). A Recently Evolved Transcriptional Network Controls Biofilm Development in Candida albicans. *Cell*, *148*(1), 126–138. http://doi.org/10.1016/j.cell.2011.10.048

Pigliucci, M. (2012). What, if Anything, Is an Evolutionary Novelty? *Philosophy of Science*, *75*(5), 887–898.

Proux-Wéra, E., Armisén, D., Byrne, K. P., & Wolfe, K. H. (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*, *13*(1), 237. http://doi.org/10.1186/1471-2105-13-237

Rieppel, O. (2001). Turtles as hopeful monsters. *BioEssays*, *23*(11), 987–991. http://doi.org/10.1002/bies.1143

Sellam, A., Askew, C., Epp, E., Tebbji, F., Mullick, A., Whiteway, M., & Nantel, A. (2010). Role of transcription factor CaNdt80p in cell separation, hyphal growth, and virulence in Candida albicans. *Eukaryotic Cell*, *9*(4), 634–644. http://doi.org/10.1128/EC.00325-09

Sherwood, R. K., & Bennett, R. J. (2009). Fungal meiosis and parasexual reproduction—lessons from pathogenic yeast. *Current Opinion in Microbiology*, *12*(6), 599–607. http://doi.org/10.1016/j.mib.2009.09.005

Shubin, N., Tabin, C., & Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature*, *457*(7231), 818–823. http://doi.org/10.1038/nature07891

Sorrells, T. R., Booth, L. N., Tuch, B. B., & Johnson, A. D. (2015). Intersecting transcription networks constrain gene regulatory evolution. *Nature*, *523*(7560), 361–365. http://doi.org/10.1038/nature14613

Swanson, C. I., Schwimmer, D. B., & Barolo, S. (2011). Rapid Evolutionary Rewiring of a Structurally Constrained Eye Enhancer. *Current Biology*, *21*(14), 1186–1196. http://doi.org/10.1016/j.cub.2011.05.056

Tanay, A., Regev, A., & Shamir, R. (2005). Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7203–7208. http://doi.org/10.1073/pnas.0502521102

Taylor, J. W., & Berbee, M. L. (2006). Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia*, *98*(6), 838–849.

Teytelman, L., Thurtle, D. M., Rine, J., & Van Oudenaarden, A. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(46), 18602–18607. http://doi.org/10.1073/pnas.1316064110

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2012). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. http://doi.org/10.1038/nbt.2450

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111. http://doi.org/10.1093/bioinformatics/btp120

Tsong, A. E., Tuch, B. B., Li, H., & Johnson, A. D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature*, *443*(7110), 415–420.

http://doi.org/10.1038/nature05099
Tuch, B. B., Galgoczy, D. J., Hernday, A. D., Li, H., & Johnson, A. D. (2008). The evolution of combinatorial gene regulation in fungi. *PLoS Biology*, *6*(2), e38. http://doi.org/10.1371/journal.pbio.0060038
Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., et al. (2015). Enhancer Evolution across 20 Mammalian Species. *Cell*, *160*(3), 554–566. http://doi.org/10.1016/j.cell.2015.01.006
Wagner, A. (2005a). Robustness and Evolvability in Living Systems (pp. 1–13). Princeton University Press.
Wagner, A. (2005b). Robustness, evolvability, and neutrality. *FEBS Letters*, *579*(8), 1772–1778. http://doi.org/10.1016/j.febslet.2005.01.063
Wagner, G. P., & Lynch, V. J. (2010). Evolutionary novelties. *Current Biology : CB*, *20*(2), R48–52. http://doi.org/10.1016/j.cub.2009.11.010
Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, *8*(3), 206–216. http://doi.org/10.1038/nrg2063
Xu, L., Ajimura, M., Padmore, R., Klein, C., & Kleckner, N. (1995). NDT80, a meiosis-specific gene required for exit from pachytene in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *15*(12), 6572–6581. http://doi.org/10.1128/MCB.15.12.6572
Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. http://doi.org/10.1186/gb-2008-9-9-r137

FIGURE CAPTIONS

Figure 1: Ndt80 target genes differ between *S. cerevisiae* and *C. albi*cans.

(A) Diagram of sporulation in *S. cerevisiae* and biofilm formation in *C. albicans*. Scale bars represent 5 μm. (B) The *cis*-regulatory motif most highly enriched at locations of Ndt80 ChIP binding in *S. cerevisiae* and *C. albicans*. Motifs were generated independently for each species. The *S. cerevisiae* Ndt80 motif determined *de novo* in this study closely matches that determined previously for Ndt80 (Chu & Herskowitz, 1998; Jolly et al., 2005; Nobile et al., 2012). (C) Diagram of the four criteria used to identify Ndt80 regulatory targets. Criteria 1: significant ChIP-Seq enrichment in the intergenic region upstream of a gene relative to untagged control experiments. Criteria 2: ChIP-Seq enrichment and the presence of an Ndt80 motif in the intergenic region. Criteria 3: ChIP-Seq enrichment with the Ndt80 motif present in the intergenic region and also in orthologous intergenic regions of two very closely related species, suggesting the motif has been maintained by selection. Criteria 4: ChIP-Seq enrichment with the Ndt80 motif present in the intergenic region and Ndt80-dependent expression of the nearby gene, indicating that expression of the gene is under Ndt80 control. (D) Overlap in targets of *S. cerevisiae* Ndt80 (red) and *C. albicans* Ndt80B (blue), using the four different criteria from (C) to identify targets, when Ndt80 is highly expressed in each species (Materials and Methods). Venn diagrams are roughly area-proportional (Hulsen, de Vlieg, & Alkema, 2008). Overlap in targets of *S. cerevisiae* Ndt80 and *C. albicans* Ndt80 when each is endogenously expressed is shown in Figure 1—figure supplement 1. Overlap in targets considering only 1:1 orthologs between the species is shown in Figure 1—figure supplement 2. Overlap of *C. albicans* Ndt80 paralogs with each other and with *S. cerevisiae* Ndt80 shown in Figure 1—figure supplement 3.

Figure 2: Ndt80 target genes differ across species descending from the *S. cerevisiae - C. albicans* common ancestor.

(A) Phylogenetic tree of the species investigated, inferred from protein sequences of 73 highly conserved genes (Lohse et al., 2013) with the scale representing the number of substitutions per site. The likely position of the *NDT80* gene duplication is indicated with a green circle. Ndt80 protein tree shown in Figure 2—figure supplement 1. (B-D) Ndt80 targets identified by Criteria 2 (Figure 1C). (B) The proportion of Ndt80 targets in *K. lactis*, *P. pastoris*, and *S. stipitis* that are shared with the Ndt80 targets in *S. cerevisiae*, *C. albicans*, or both. The proportion of targets shared, using all four different criteria for identifying targets (Figure 1C) shown in Figure 2—figure supplement 2. (C) Histogram of all Ndt80 targets in the five species tested (*S. cerevisiae, K. lactis, P. pastoris, S. stipitis,* and *C. albicans*) according to the number of species in which that gene is a target. All target genes (black) and those with 1:1 orthologs across all five species (dashed) are shown. The overlap in Ndt80 targets in all five species was simulated, assuming the target genes are randomly selected, and is shown in Figure 2—figure supplement 3. (D) List of genes bound by Ndt80 in all five species tested with functional annotations (color-coded) from *S. cerevisiae,* as described in the text (SGD, (Cherry et al., 2012)). For (B –D), the union of Ndt80A and Ndt80B bound genes were used for *S. stipitis* and *C. albicans*. The *cis*-regulatory site bound by Ndt80 in each species shown in Figure 2—figure supplement 4.

Figure 3: *P. pastoris* Ndt80 binds to regions bound by *S. cerevisiae* Ndt80 when expressed in *S. cerevisiae.*

(A) Diagram of strains used to generate data in (B) and (C): (1) tagged native Ndt80 in *S. cerevisiae*, (2) tagged native Ndt80 in *P. pastoris*, and (3) tagged heterologous *P. pastoris* Ndt80 expressed in *S. cerevisiae*. (B and C) Comparisons of the maximum ChIP-Seq fold enrichment for the intergenic regions of all genes with 1:1 orthologs in *S. cerevisiae* and *P. pastoris*, with Spearman's rank correlation coefficient shown (p-value = 7.2×10^{-117} for (B), 1.02×10^{-296} for (C)). Numbers in axis labels correspond to simplified diagrams in (A). (B) *S. cerevisiae* Ndt80 vs. *P. pastoris* Ndt80 expressed in *S. cerevisiae*. (C) *P. pastoris* Ndt80 expressed in *P. pastoris* vs. *P. pastoris* Ndt80 expressed in *S. cerevisiae*. A comparison of Ndt80 targets identified in each experiment using Criteria 1 shown in Figure 3—figure supplement 1.

Figure 4: Ndt80 is required for sporulation, but is dispensable for biofilm formation, in *K. lactis*, *P. pastoris*, and *C. lusitaniae.*

(A-C) Light microscope images of genetically matched wild-type and *ndt80* deletion strains (Stars indicate diploid cells that have undergone sporulation) and quantification of the percent of cells exhibiting spores, as measured by microscopy (200 cells counted for each strain). (D-H) Confocal scanning laser microscopy images of biofilm formation for genetically matched wild-type and *ndt80* deletion strains. Top view of biofilm shown above side view for each, with scale bars representing 25 μm.

Figure 5: The targets of Ndt80 differ among species even with a conserved Ndt80 phenotype.

(A) Percent of all Ndt80-bound genes (identified using Criteria 2) shared between any two species tested; for this analysis, only 1:1 orthologs in each two-species comparison were considered. (B) Percent of all Ndt80-bound genes shared between any two species, normalized to the genome-wide substitutions per site (*S. cerevisiae – K. lactis* set to 1). Species comparisons between two species with a conserved Ndt80 phenotype shown with black outline for (A) and (B). Percent overlap and normalized percent overlap for targets identified using all four criteria (Figure 1C) shown in Figure 5—figure supplement 1. (C) Cladogram of species for which Ndt80 phenotype and target genes have been identified, with simplified cartoons representing Ndt80 ChIP-Seq targets in each species. The grey circle in the center represents Ndt80, while the smaller circles represent genes present in all four species, with each small circle representing ~100 genes. The arrows indicate that Ndt80 binds to those genes in that species.

Figure 6: Genes induced during sporulation differ significantly across *S. cerevisiae*, *K. lactis*, and *P. pastoris*.

(A) Gene expression in meiotic growth compared to mitotic growth. Genes with significant upregulation in meiosis in at least one species shown. (B) Gene expression in a wild-type strain compared with an *ndt80* Δ/Δ strain for all three species. Genes with significant upregulation in wild-type in at least one species shown. For (A) and (B), each line represents a single gene with the color of the line representing the ratio of expression in the two conditions. Numbers below represent the number of genes shared in all three species (left), in two species (middle), or in just one species (right). Data for *S. cerevisiae* from (Chu et al., 1998), data for *K. lactis* and *P. pastoris* generated in this study. A comparison of the 25 shared sporulation genes (A) and the 7 shared Ndt80-dependent genes (B) shown in Figure 6—figure supplement 1.

SUPPLEMENTAL FIGURE CAPTIONS

Figure 1—figure supplement 1: Ndt80 targets in *S. cerevisiae* and *C. albicans* when Ndt80 is endogenously expressed.

Ndt80 targets identified for *S. cerevisiae* Ndt80 (red) and *C. albicans* Ndt80B (blue), using four different criteria to identify targets (Figure 1C). Ndt80 expression in each species is driven by the native promoter (*S. cerevisiae* ChIP performed during sporulation, *C. albicans* ChIP performed during mitotic growth). As the sizes of the regulons are dramatically different between the species, the percent of targets shared is shown for each species individually. The black numbers above the venn diagrams represent the number of overlapping targets.

Figure 1—figure supplement 2: Ndt80 targets in *S. cerevisiae* and *C. albicans* considering only 1:1 orthologs between species.

Ndt80 targets identified for *S. cerevisiae* Ndt80 (red) and *C. albicans* Ndt80B (blue), using four different criteria to identify targets (Figure 1C). Ndt80 expression in each species is driven by a high expression promoter, and ChIP experiments were performed as in Figure 1. Only genes with 1:1 orthologs between *S. cerevisiae* and *C. albicans* are shown.

Figure 1—figure supplement 3: Ndt80 targets for paralogs Ndt80A and Ndt80B in *C. albicans*.

(A) Ndt80 targets identified for Ndt80A (light blue) and Ndt80B (dark blue) in *C. albicans* using four different criteria (Figure 1C). (B) Targets of Ndt80A in *C. albicans* compared to the targets of Ndt80 in *S. cerevisiae.*

Figure 1—figure supplement 4: Fraction of Ndt80 targets shared between biological replicates and between *S. cerevisiae* and *C. albicans*

Fraction of Ndt80 ChIP-Seq targets shared between two biological replicates of *S. cerevisiae* (red), two biological replicates of *C. albicans* (blue), and between the intersections of the replicates for each species (dashed), using four different criteria (Figure 1C). Ndt80 expression in each species is driven by a high expression promoter, and ChIP experiments were performed as in Figure 1. Only genes with 1:1 orthologs between *S. cerevisiae* and *C. albicans* are shown.

Figure 2—figure supplement 1: Similarity of Ndt80 protein sequences across many yeast species.

(A) Phylogenetic tree of Ndt80 proteins in diverse ascomycetes, built using MUSCLE (Edgar, 2004) and PHYML (Guindon et al., 2010). Scale bar indicates substitutions per site. (B) List of genes used in the phylogenetic tree, and in the manuscript. (C) Similarity matrix of Ndt80 proteins for species discussed in this work, calculated using MUSCLE.

Figure 2—figure supplement 2: Ndt80 targets in *K. lactis*, *P. pastoris*, and *S. stipitis* categorized according to overlap with *S. cerevisiae* and *C. albicans*.

The proportion of Ndt80 targets in *K. lactis*, *P. pastoris*, and *S. stipitis* shared with Ndt80 in *S. cerevisiae* and *C. albicans*. Different criteria for Ndt80 target identification (Figure 1C) were used to identify targets in *K. lactis*, *P. pastoris*, and *S. stipitis*, where possible. Criteria 3 could not be used for *P. pastoris* and *S. stipitis*, as it requires two closely related species with mapped orthologs, while Criteria 4 could not be used for *S. stipitis* as expression profiling was not performed in this species.

Figure 2—figure supplement 3: Histogram of simulated overlap in Ndt80 targets across five species.

Number of shared Ndt80 targets shown for 100,000 simulations of ChIP experiments. Genes were randomly selected from the set of genes with orthologs in all five species (3,171 in total) to make target sets for each species corresponding to the number of actual ChIP targets we observed (using Criteria 2 from Figure 1C; 424 for *S. cerevisiae*, 269 for *K. lactis*, 300 for *P. pastoris*, 958 for *S. stipitis*, and 989 for *C. albicans*). Histogram shows the number of genes found as targets in all five species in 100,000 simulations, with the number of runs shown above each bar; five genes was the maximum overlap found.

Figure 2—figure supplement 4: Ndt80 *cis*-regulatory motif in ChIP-Seq experiments.

(A) *De novo* motifs were generated for Ndt80 for each ChIP-Seq experiment performed, using DREME (Bailey, 2011), with the union of replicate peak locations compared to a random set of genomic sequences of the same lengths, for each species. The e-values generated by DREME represent the enrichment p-value multiplied by the number of candidate motifs tested. Shown are the motifs most closely resembling the known Ndt80 binding motif in *S. cerevisiae* and *C. albicans* (Jolly et al., 2005) (Nobile et al., 2012). (B) A composite motif for all species was generated, using peak locations for Ndt80 ChIP-Seq in all five species tested. For species with two Ndt80 paralogs, the intersection of replicates was taken and the union of the paralogs was taken to represent all binding locations and submitted to DREME. These sequences were scrambled as a background for the analysis. From this analysis, a "consensus motif" sequence was inferred to use for later analyses. (C) Enrichment of the consensus motif (CACAAA) for each ChIP-Seq experiment. Motifs were counted in the peak locations compared to a random set

of genomic sequences, for each species, and a p-value was generated using a Fisher's one-tailed exact test.

Figure 3—figure supplement 1: Heterologous *P. pastoris* Ndt80 targets compared to native *P. pastoris* and *S. cerevisiae* Ndt80 targets.

(A and B) Ndt80 targets, defined by ChIP-Seq enrichment alone (Criteria 1), for *P. pastoris* Ndt80 when expressed in *S. cerevisiae*, compared to targets of Ndt80 in *S. cerevisiae* (A) and *P. pastoris* (B) when expressed endogenously. Genes that are Ndt80 targets in both *S. cerevisiae* and *P. pastoris* have been omitted from this analysis.

Figure 5—figure supplement 1: Percent overlap in Ndt80 targets across different species.

(A-D) Percent of all Ndt80-bound genes shared between any two species tested (left column) and the percent shared normalized to the genome-wide substitutions per site between species (right column) (*S. cerevisiae – K. lactis* set to 1). Targets identified using Criteria 1 (A), Criteria 2 (B), Criteria 3 (C), or Criteria 4 (D) (Figure 1C). For this analysis, only 1:1 orthologs in each two-species comparison were considered. Species comparisons between two species with a conserved Ndt80 phenotype shown with black outline.

Figure 6—figure supplement 1: Comparing sporulation-induced genes and Ndt80-dependent genes across *S. cerevisiae*, *K. lactis*, and *P. pastoris*.

Venn diagram comparison of shared sporulation genes (upregulated in all three species, Figure 6A with shared Ndt80-dependent genes (downregulated in *ndt80* deletion in all three species,

Figure 6B), with corresponding gene names from *S. cerevisiae* listed. Only genes with 1:1:1 orthologs in all three species were considered in this analysis.

Figure 6—figure supplement 2: Comparing patterns of meiotic expression for genes required for sporulation in *S. cerevisiae*

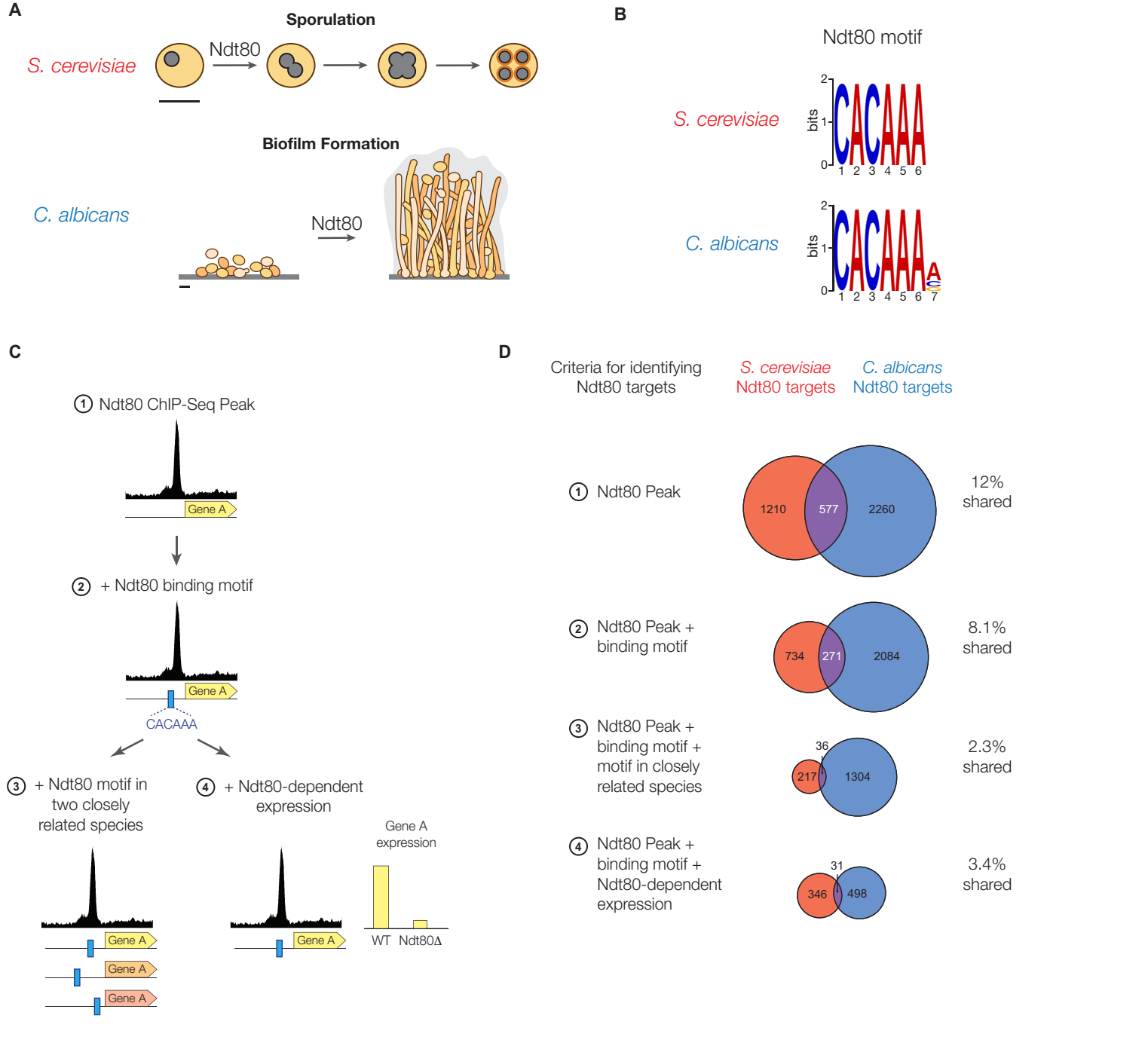
Table shows the 65 genes required for sporulation in *S. cerevisiae* (SGD, (Cherry et al., 2012)) that exhibit meiotic upregulation in *S. cerevisiae* ((Chu et al., 1998), Materials and Methods), along with meiotic expression data from *K. lactis* and *P. pastoris*. Yellow indicates the gene is present and up-regulated in meiotic vs. mitotic growth in that species (Materials and Methods), white indicates the gene is present but not significantly up-regulated, and grey indicates that an ortholog of the gene cannot be found in that species.

Supplementary File 1: Excel spreadsheet containing processed data for all ChIP-Seq, RNA-Seq, and expression array experiments performed.

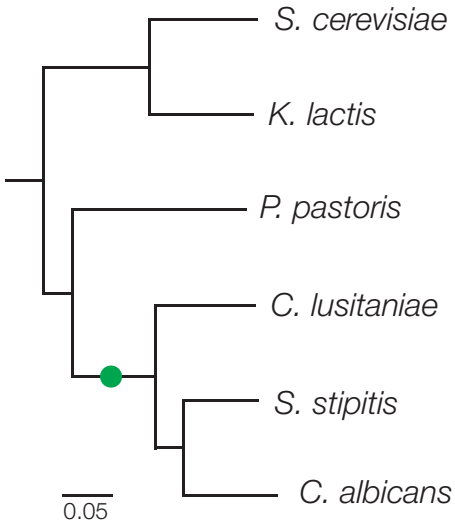
Rows correspond to groups of orthologous genes. Columns show results of all genomic experiments discussed here, colors correspond to experiments in different species. Cell value of "N/A" indicates no orthologous gene for that orthogroup in that species.

Supplementary File 2: Excel spreadsheet containing strains and primers used in this study.

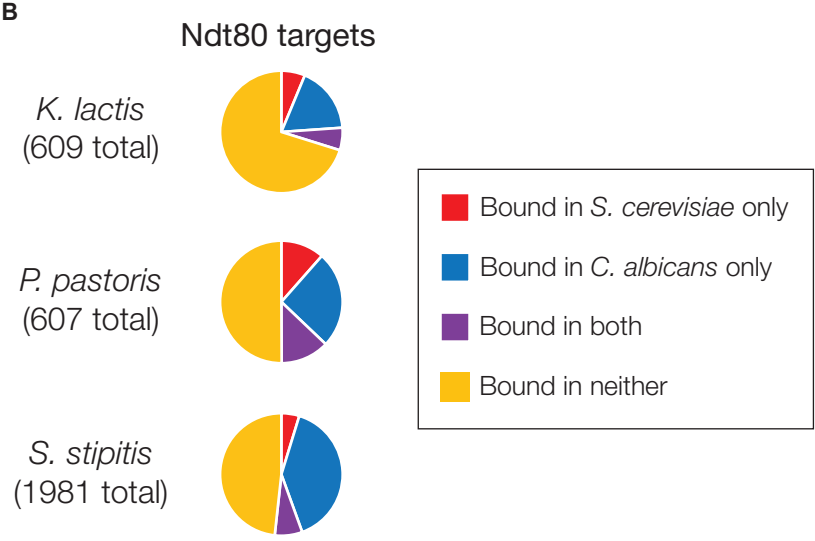
(A) Strains used in this study. (B) Primers used in this study.


A
Sporulation
S. cerevisiae
Ndt80
Biofilm Formation
C. albicans
Ndt80
B
Ndt80 motif
S. cerevisiae
bits
C. albicans
bits
C
1 Ndt80 ChIP-Seq Peak
Gene A
2 + Ndt80 binding motif
Gene A
CACAAA
3 + Ndt80 motif in two closely related species
Gene A
Gene A
Gene A
4 + Ndt80-dependent expression
Gene A
Gene A expression
WT Ndt80Δ
D
Criteria for identifying Ndt80 targets
S. cerevisiae Ndt80 targets
C. albicans Ndt80 targets
1 Ndt80 Peak
1210
577
2260
12% shared
2 Ndt80 Peak + binding motif
734
271
2084
8.1% shared
3 Ndt80 Peak + binding motif + motif in closely related species
36
217
1304
2.3% shared
4 Ndt80 Peak + binding motif + Ndt80-dependent expression
31
346
498
3.4% shared

A



B

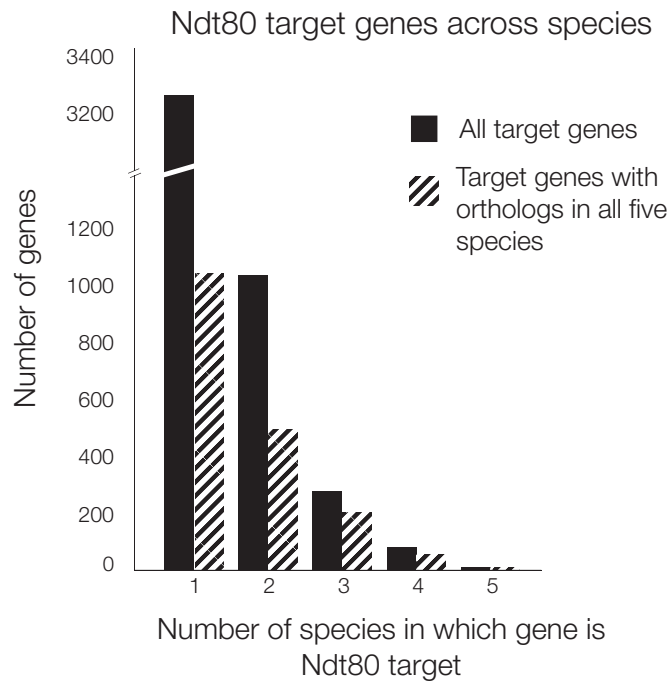


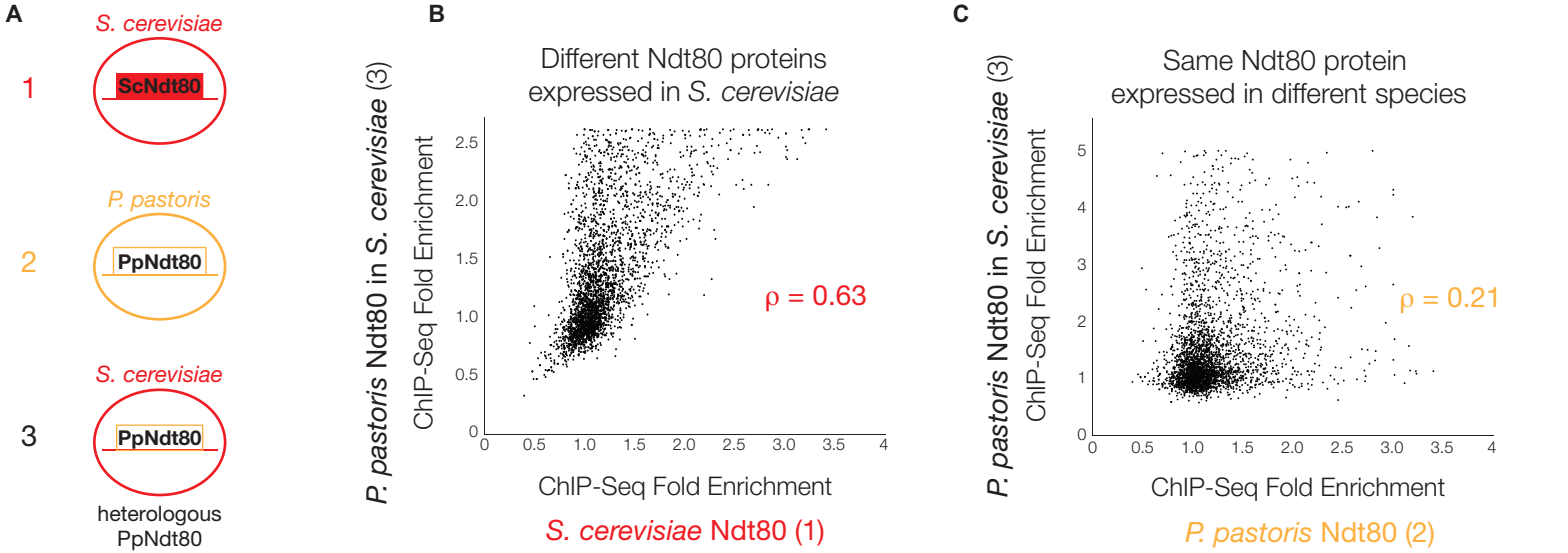
C

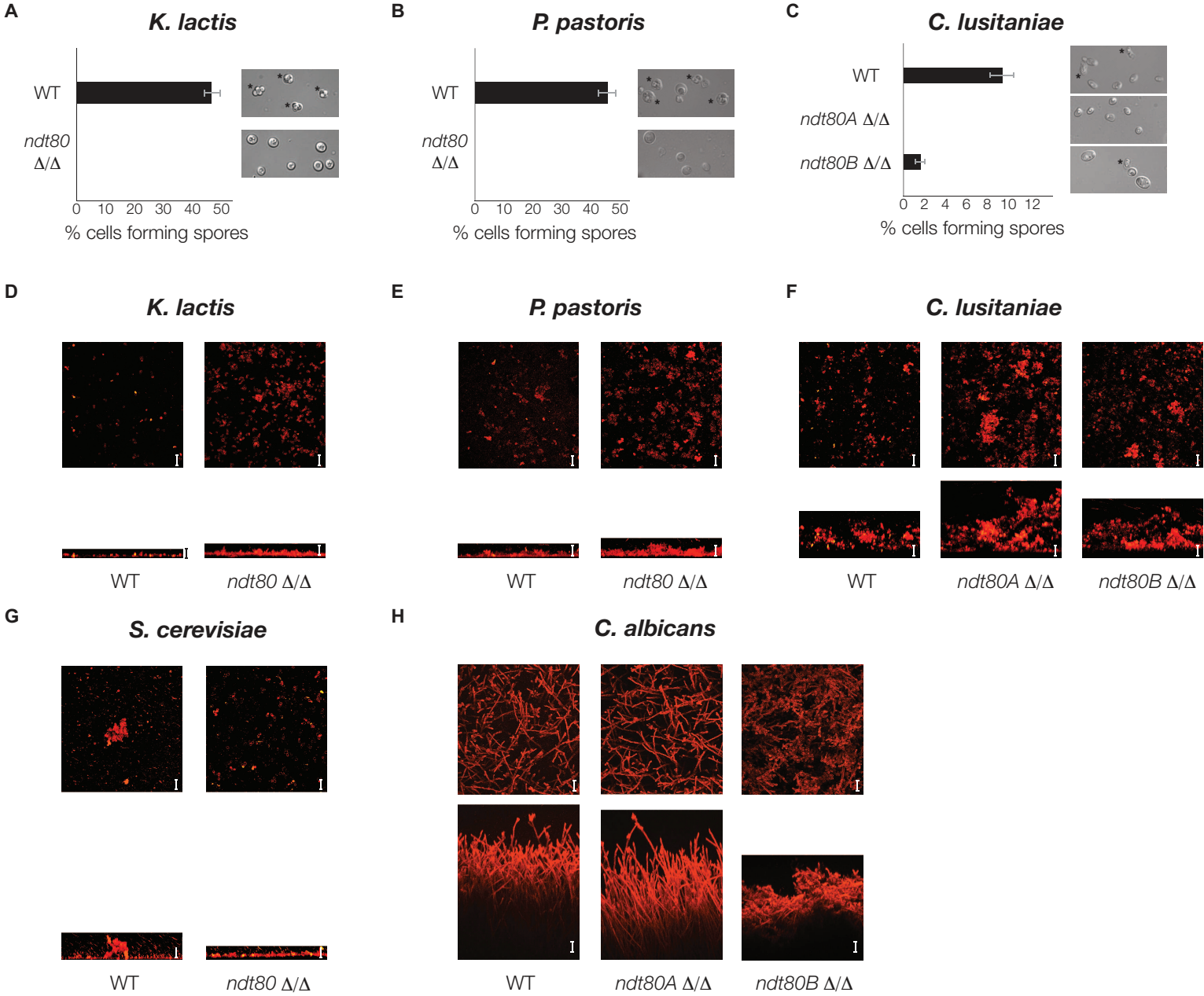
Conserved Ndt80 Target Genes

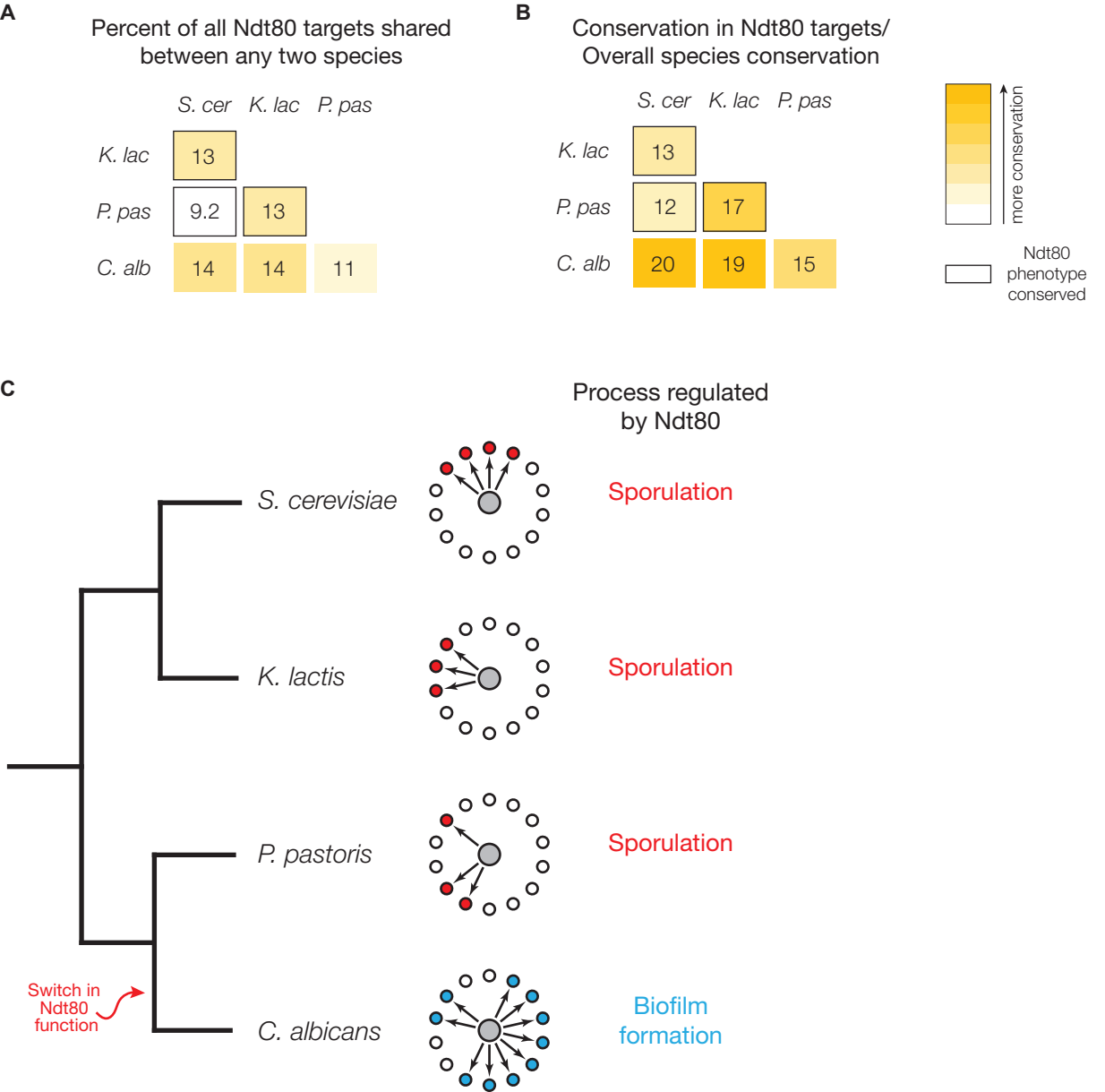
Gene	Function in *S. cerevisiae*
Cln2	G1 cyclin involved in regulation of the cell cycle
Clg1	Cyclin-like protein
Yox1	Transcriptional repressor; binds promoters of cell cycle-regulated genes
Rki	Ribose-5-phosphate ketol-isomerase
Gln1	Glutamine synthetase
Sga1	Intracellular sporulation-specific glucoamylase
Gdh2	NAD(+)-dependent glutamate dehydrogenase
Rps7a	Protein component of the small ribosomal subunit
Tma10	Protein of unknown function that associates with ribosomes
Usv1	Putative transcription factor; affects regulation of genes involved in cell wall biosynthesis

D

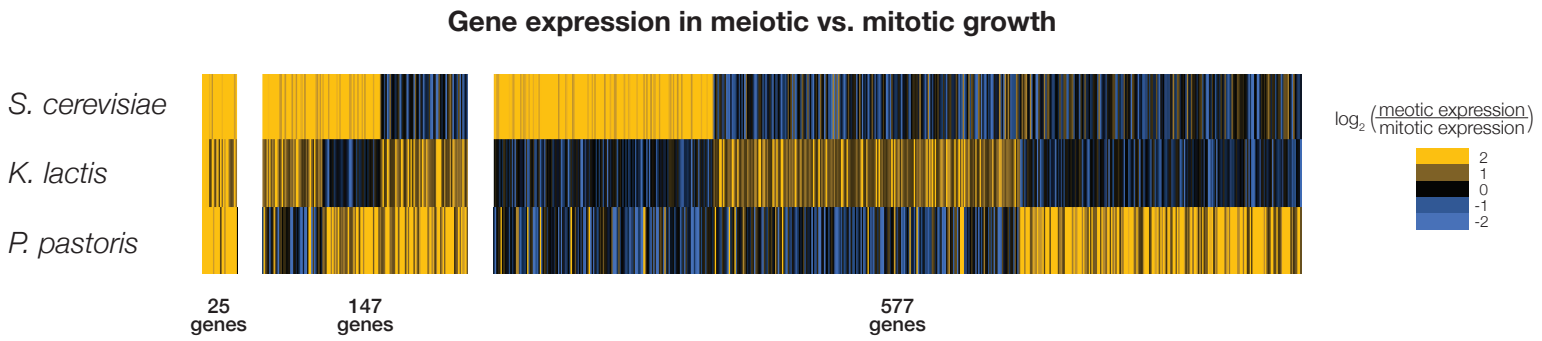



A
S. cerevisiae
1
ScNdt80
P. pastoris
2
PpNdt80
S. cerevisiae
3
PpNdt80
heterologous
PpNdt80
B
Different Ndt80 proteins
expressed in S. cerevisiae
P. pastoris Ndt80 in S. cerevisiae (3)
ChIP-Seq Fold Enrichment
ρ = 0.63
ChIP-Seq Fold Enrichment
S. cerevisiae Ndt80 (1)
C
Same Ndt80 protein
expressed in different species
P. pastoris Ndt80 in S. cerevisiae (3)
ChIP-Seq Fold Enrichment
ρ = 0.21
ChIP-Seq Fold Enrichment
P. pastoris Ndt80 (2)

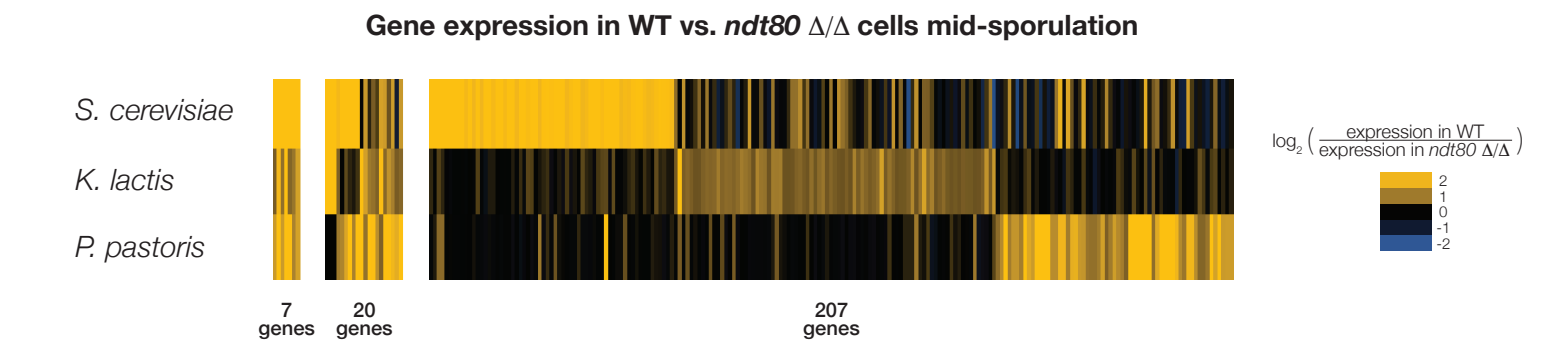

A
K. lactis
WT
ndt80 Δ/Δ
0 10 20 30 40 50
% cells forming spores
B
P. pastoris
WT
ndt80 Δ/Δ
0 10 20 30 40 50
% cells forming spores
C
C. lusitaniae
WT
ndt80A Δ/Δ
ndt80B Δ/Δ
0 2 4 6 8 10 12
% cells forming spores
D
K. lactis
WT
ndt80 Δ/Δ
E
P. pastoris
WT
ndt80 Δ/Δ
F
C. lusitaniae
WT
ndt80A Δ/Δ
ndt80B Δ/Δ
G
S. cerevisiae
WT
ndt80 Δ/Δ
H
C. albicans
WT
ndt80A Δ/Δ
ndt80B Δ/Δ


A
Percent of all Ndt80 targets shared between any two species
S. cer
K. lac
P. pas
K. lac
13
P. pas
9.2
13
C. alb
14
14
11
B
Conservation in Ndt80 targets/ Overall species conservation
S. cer
K. lac
P. pas
K. lac
13
P. pas
12
17
C. alb
20
19
15
more conservation
Ndt80 phenotype conserved
C
Process regulated by Ndt80
S. cerevisiae
Sporulation
K. lactis
Sporulation
P. pastoris
Sporulation
Switch in Ndt80 function
C. albicans
Biofilm formation

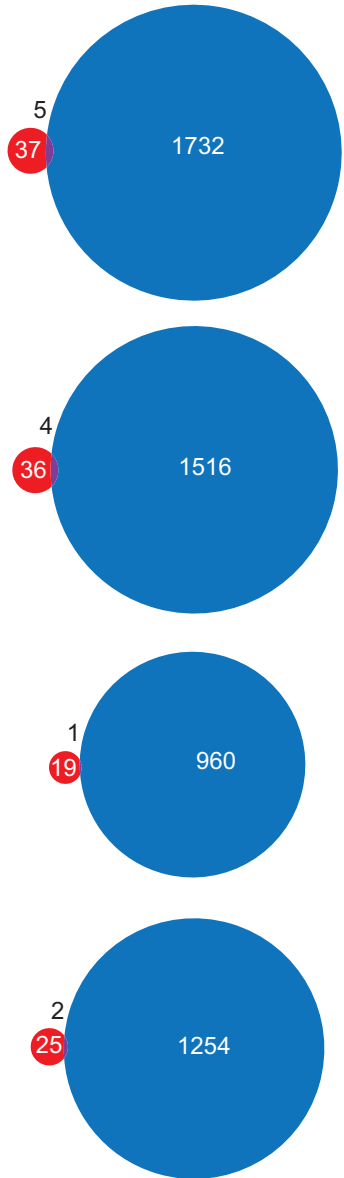
A


Gene expression in meiotic vs. mitotic growth
S. cerevisiae
K. lactis
P. pastoris
25 genes
147 genes
577 genes
log2 (meotic expression / mitotic expression)
2
1
0
-1
-2

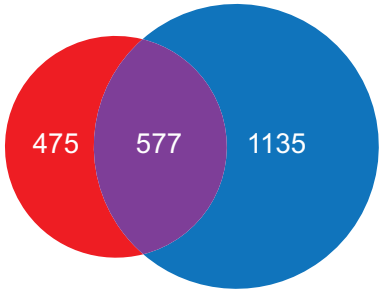
B

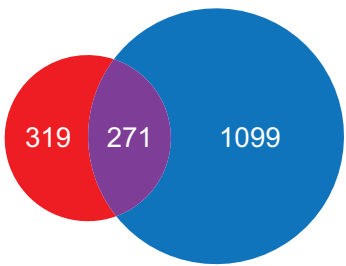

Gene expression in WT vs. ndt80 Δ/Δ cells mid-sporulation
S. cerevisiae
K. lactis
P. pastoris
7 genes
20 genes
207 genes
log2 (expression in WT / expression in ndt80 Δ/Δ)
2
1
0
-1
-2

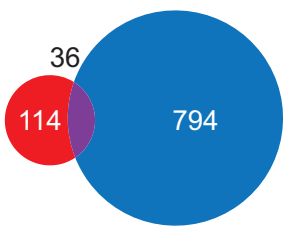
Criteria for identifying Ndt80 targets	*S. cerevisiae* Ndt80 targets	*C. albicans* Ndt80 targets	% *S. cerevisiae* Ndt80 targets shared	% *C. albicans* Ndt80 targets shared
① Ndt80 Peak	37	1732	12%	3.6%
② Ndt80 Peak + binding motif	36	1516	10%	0.26%
③ Ndt80 Peak + binding motif + motif in closely related species	19	960	5.0%	0.10%
④ Ndt80 Peak + binding motif + Ndt80-dependent expression	25	1254	7.4%	0.16%

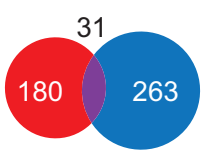


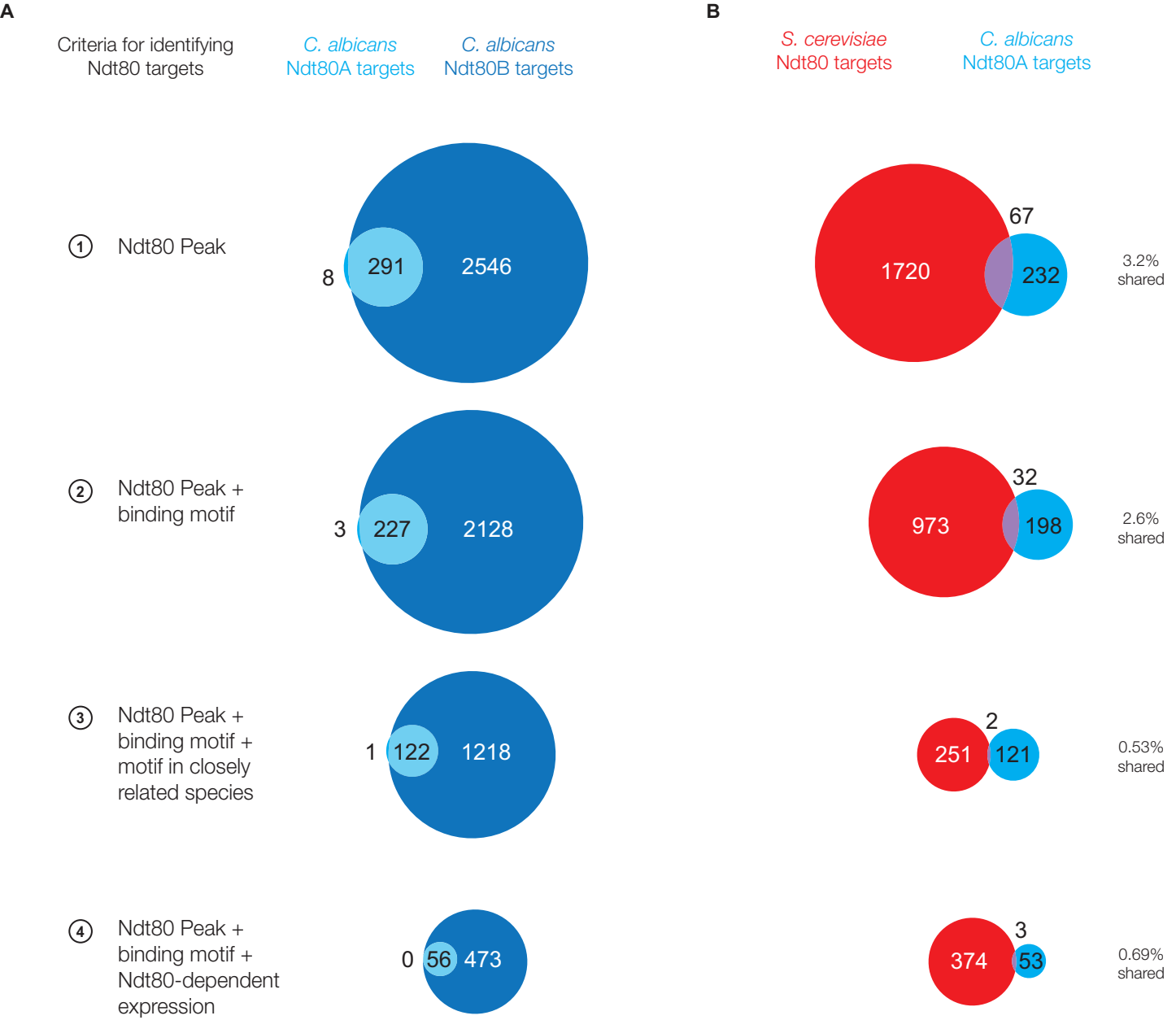
Criteria for identifying Ndt80 targets	*S. cerevisiae* Ndt80 targets	*C. albicans* Ndt80 targets	
① Ndt80 Peak	475	1135	21% shared
② Ndt80 Peak + binding motif	319	1099	14% shared
③ Ndt80 Peak + binding motif + motif in closely related species	114	794	3.7% shared
④ Ndt80 Peak + binding motif + Ndt80-dependent expression	180	263	6.1% shared

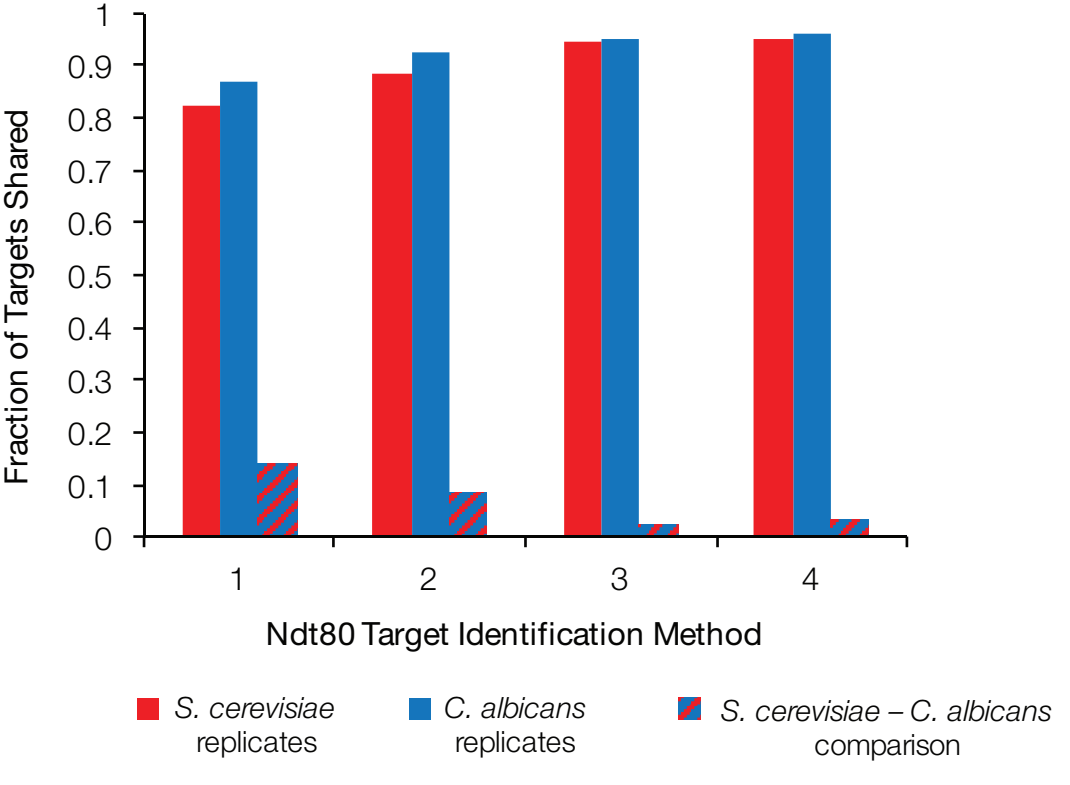




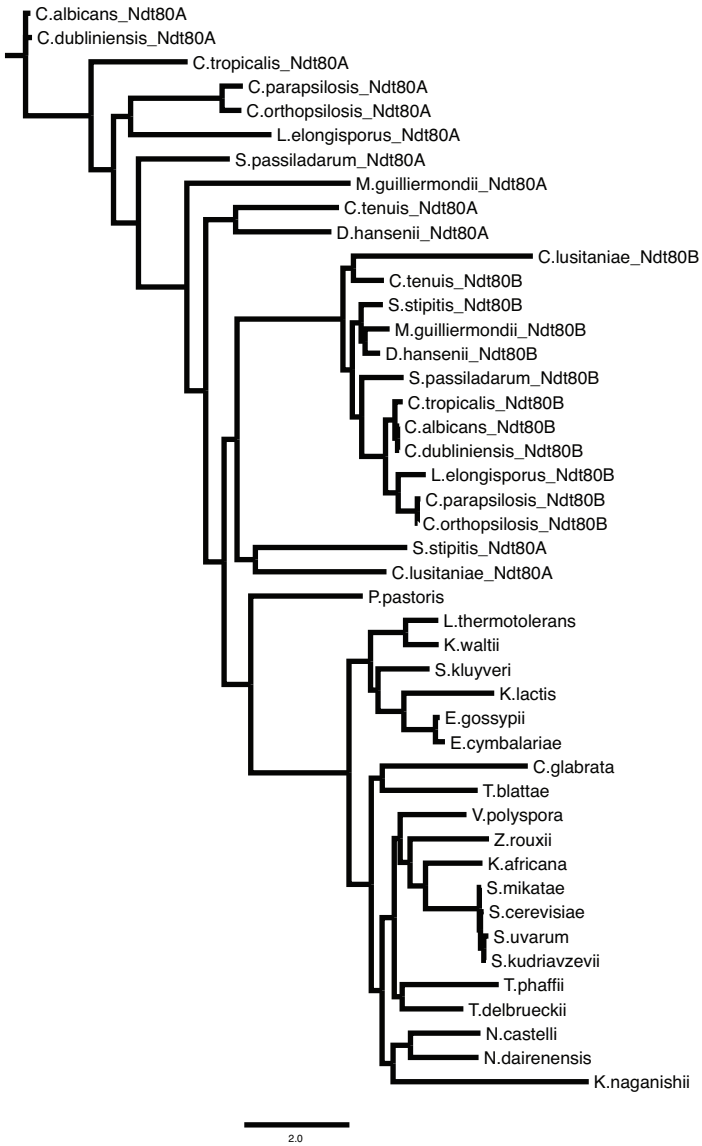





A
Criteria for identifying Ndt80 targets
C. albicans Ndt80A targets
C. albicans Ndt80B targets
1 Ndt80 Peak
8
291
2546
2 Ndt80 Peak + binding motif
3
227
2128
3 Ndt80 Peak + binding motif + motif in closely related species
1
122
1218
4 Ndt80 Peak + binding motif + Ndt80-dependent expression
0
56
473
B
S. cerevisiae Ndt80 targets
C. albicans Ndt80A targets
67
1720
232
3.2% shared
32
973
198
2.6% shared
2
251
121
0.53% shared
3
374
53
0.69% shared


Fraction of Targets Shared
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
Ndt80 Target Identification Method
S. cerevisiae replicates
C. albicans replicates
S. cerevisiae – C. albicans comparison

A

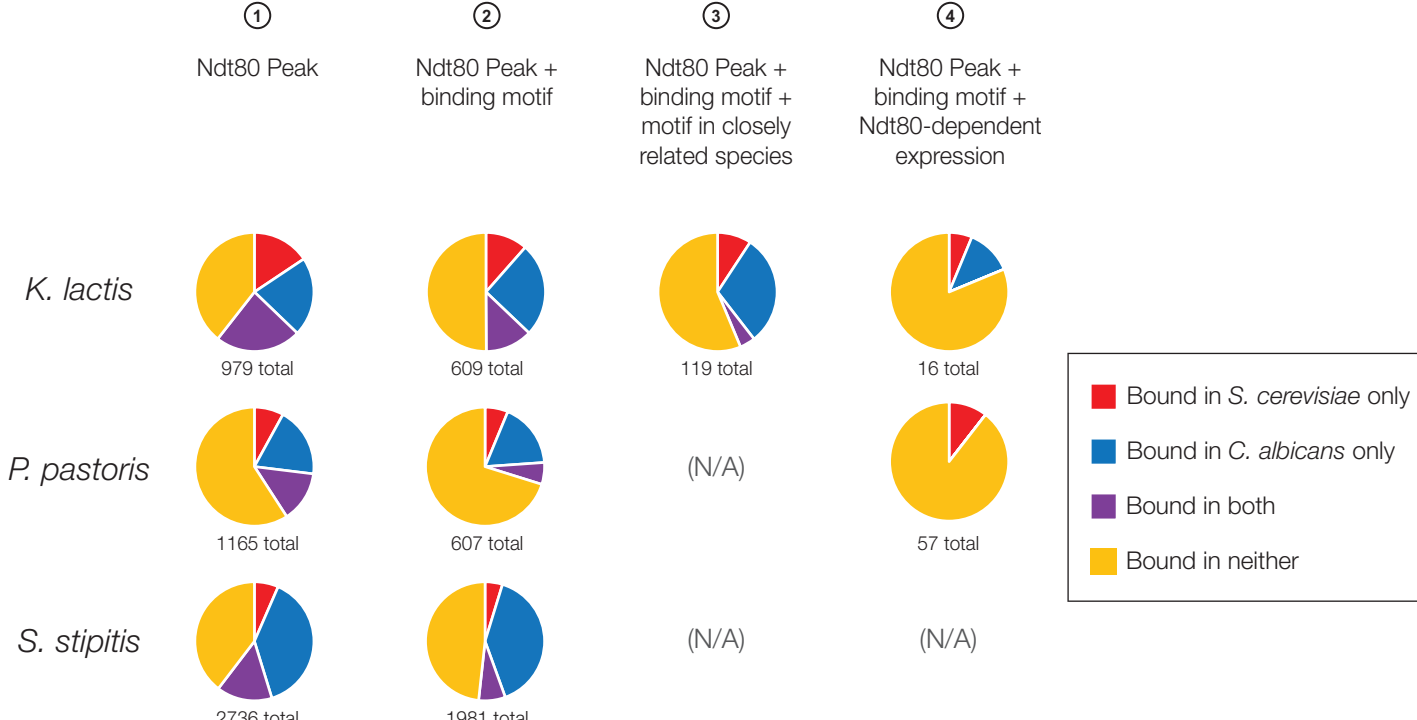


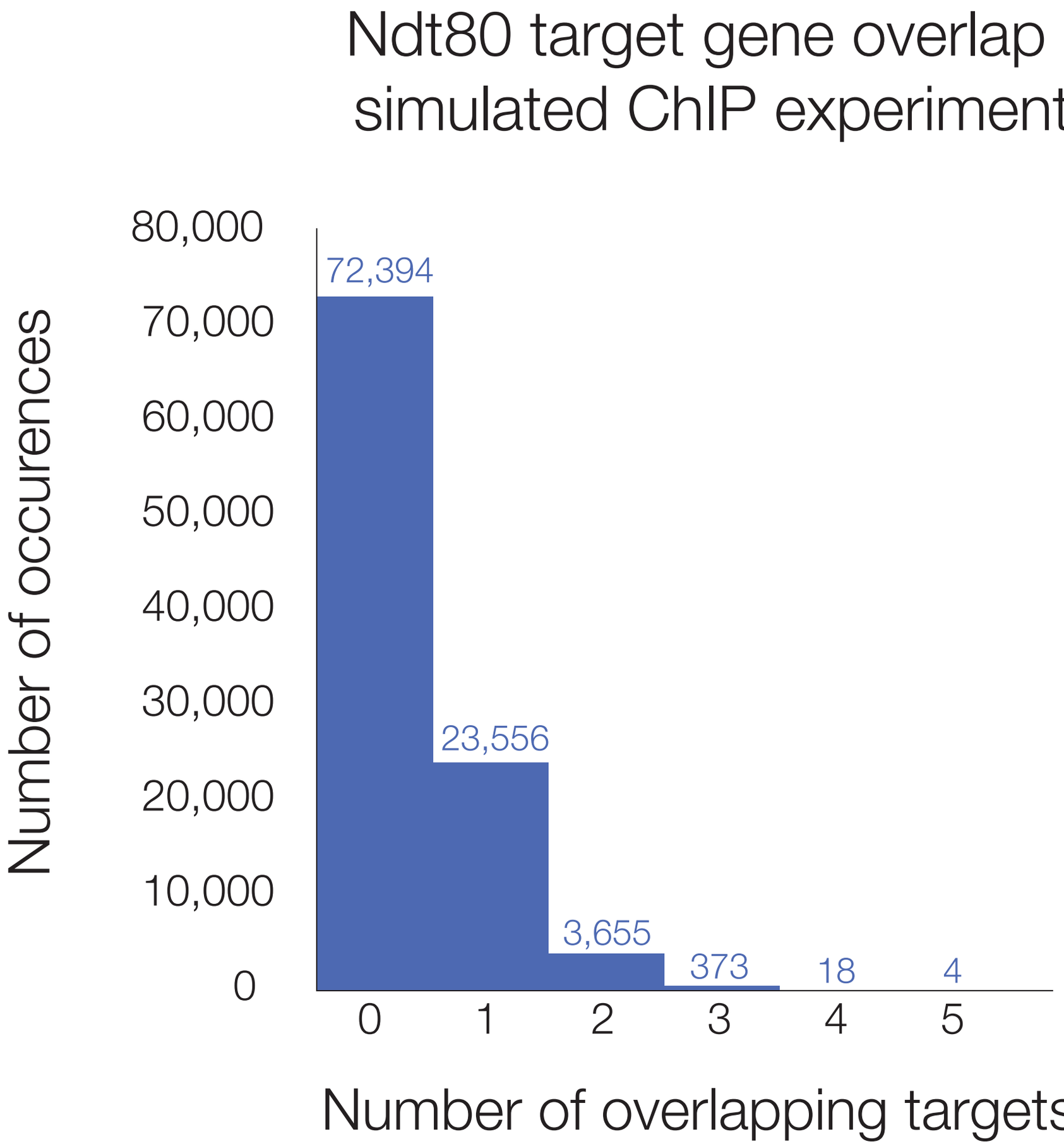
B

Species	Ndt80	Ndt80A	Ndt80B
S. cerevisiae	YHR124W	N/A	N/A
K. lactis	KLLA0F24420g	N/A	N/A
V. polyspora	Kpol_1048.8	N/A	N/A
T. phaffii	TPHA0I01370	N/A	N/A
T. blattae	TBLA0D00350	N/A	N/A
N. dairenensis	NDAI0B04400	N/A	N/A
N. castellii	NCAS0B07070	N/A	N/A
K. naganishii	KNAG0H03590	N/A	N/A
K. africana	KAFR0F00660	N/A	N/A
C. glabrata	CAGL0L13090g	N/A	N/A
S. uvarum	Suva_15.321	N/A	N/A
S. kudriavzevii	Skud_8.185	N/A	N/A
S. mikatae	Smik_8.202	N/A	N/A
Z. rouxii	ZYRO0G06160g	N/A	N/A
T. delbrueckii	TDEL0B04970	N/A	N/A
E. gossypii	AGR347W	N/A	N/A
E. cymbalariae	Ecym_8187	N/A	N/A
S. kluyveri	SAKL0E11330g	N/A	N/A
L. thermotolerans	KLTH0B05544g	N/A	N/A
K. waltii	Kwal_33.14699	N/A	N/A
C. albicans	N/A	orf19.513	orf19.2119
C. dubliniensis	N/A	CD36_29580	CD36_15120
C. tropicalis	N/A	CTRG_00389	CTRG_01097
C. parapsilosis	N/A	CPAR2_202090	CPAR2_213640
C. orthopsilosis	N/A	CORT0D02140	CORT0A12970
L. elongisporus	N/A	LELG_03337	LELG_01178
D. hansenii	N/A	DEHA2F21230g	DEHA2A07282g
S. stipitis	N/A	PICST_21174	PICST_81430
C. tenuis	N/A	CANTEDRAFT_111455	CANTEDRAFT_116020
S. passalidarum	N/A	SPAPADRAFT_134277	SPAPADRAFT_60902
M. guilliermondii	N/A	PGUG_00339	PGUG_02096
C. lusitaniae	N/A	CLUG_05634	CLUG_00404

C

	S.cerevisiae	*K.lactis*	*P.pastoris*	*C.lusitaniae Ndt80A*	*S.stipitis Ndt80A*	*C.albicans Ndt80A*	*C.lusitaniae Ndt80B*	*S.stipitis Ndt80B*	*C.albicans Ndt80B*
S.cerevisiae	100								
K.lactis	32.83	100							
P.pastoris	25.91	28.7	100						
C.lusitaniae Ndt80A	23.71	24.19	24.92	100					
S.stipitis Ndt80A	37.8	38.59	46.89	45.45	100				
C.albicans Ndt80A	24.06	25.5	30.4	25.11	42.98	100			
C.lustaniae Ndt80B	33.66	31.8	33.13	38.75	38.59	32.65	100		
S.stipitis Ndt80B	34.66	34.39	38.4	41.14	41.46	37.67	36.78	100	
C.albicans Ndt80B	35.05	32.29	35.09	40.46	39.84	35.08	38.66	54.15	100


①
Ndt80 Peak
②
Ndt80 Peak + binding motif
③
Ndt80 Peak + binding motif + motif in closely related species
④
Ndt80 Peak + binding motif + Ndt80-dependent expression
K. lactis
979 total
609 total
119 total
16 total
P. pastoris
1165 total
607 total
(N/A)
57 total
S. stipitis
2736 total
1981 total
(N/A)
(N/A)
Bound in S. cerevisiae only
Bound in C. albicans only
Bound in both
Bound in neither

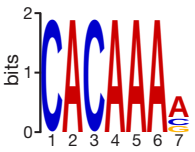

Ndt80 target gene overlap in simulated ChIP experiments
Number of occurences
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
72,394
23,556
3,655
373
18
4
0
1
2
3
4
5
Number of overlapping targets

A

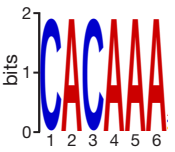
Species	Tagged protein	Tagged protein promoter	De novo motif	e-value
S. cerevisiae	Ndt80	pGal1		1.5E-257
S. cerevisiae	Ndt80	Endogenous		2.7E-047
K. lactis	Ndt80	pGal1		1.8E-010
P. pastoris	Ndt80	pTef1		1.3E-009
S. stipitis	Ndt80A	pTef1		1.1E-002
S. stipitis	Ndt80B	pTef1		2.5E-084
C. albicans	Ndt80A	pTDH3		6.6E-003
C. albicans	Ndt80B	pTDH3		1.5E-257
C. albicans	Ndt80B	Endogenous		2.9E-166

B

Composite species de novo motif



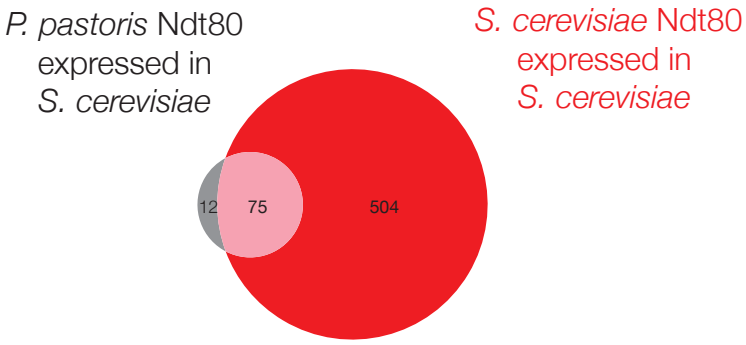
"Consensus motif"



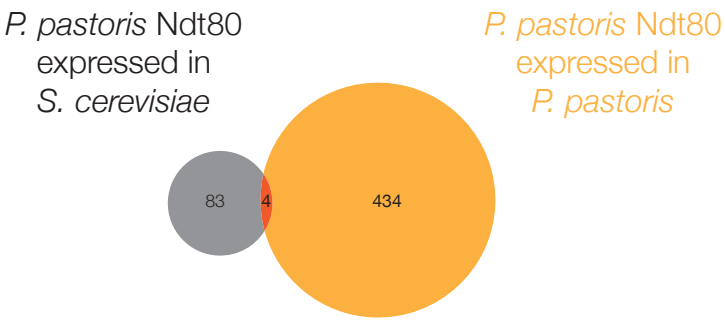
C

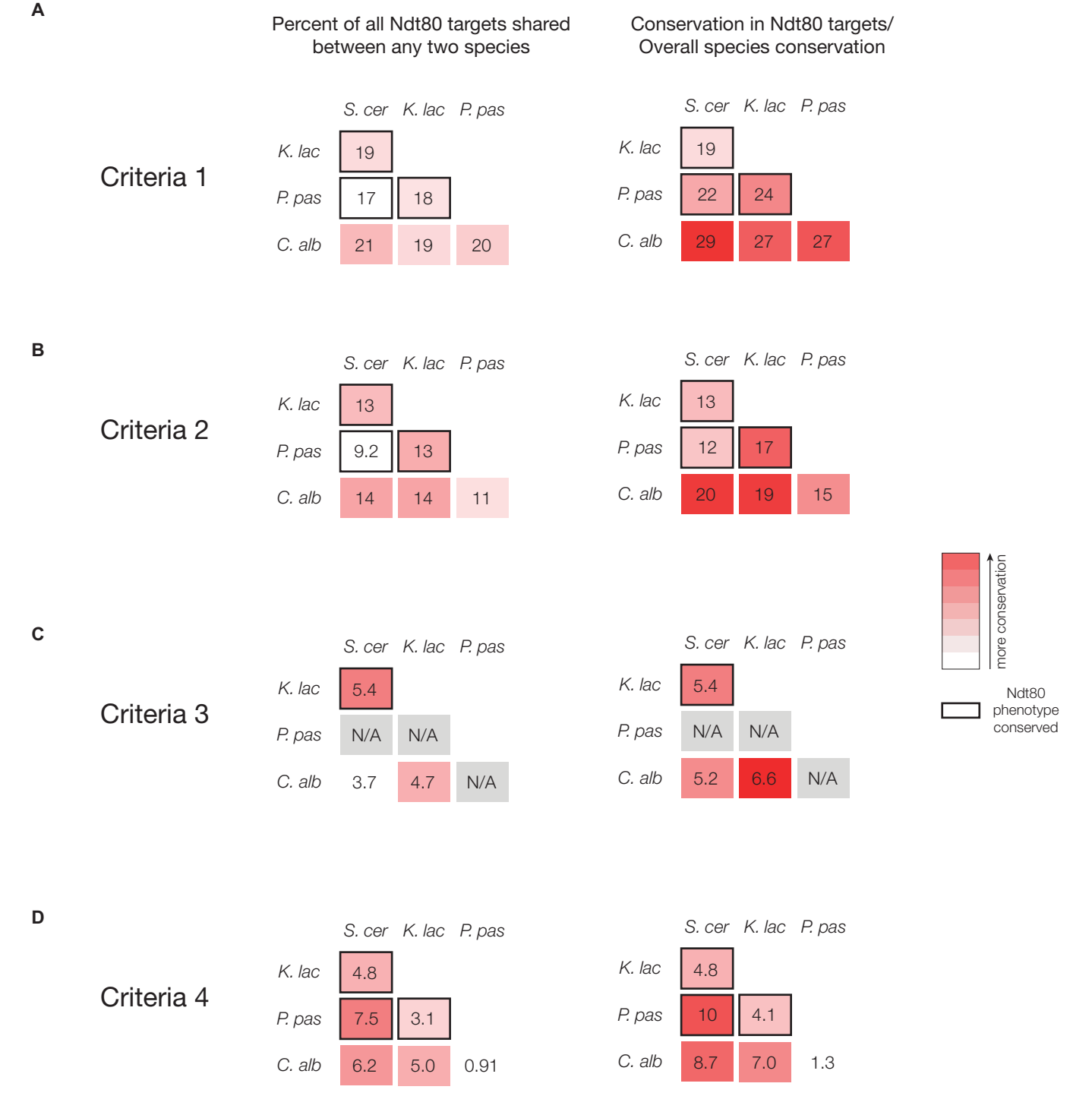
Species	Tagged protein	Tagged protein promoter	Total peaks	Peaks with consensus motif	Enrichment p-value
S. cerevisiae	Ndt80	pGal1	2767	1097	3.16E-124
S. cerevisiae	Ndt80	Endogenous	596	214	2.97E-20
K. lactis	Ndt80	pGal1	1036	246	1.22E-08
P. pastoris	Ndt80	pTef1	1242	274	3.27E-10
S. stipitis	Ndt80A	pTef1	772	131	1.89E-04
S. stipitis	Ndt80B	pTef1	3355	882	5.92E-53
C. albicans	Ndt80A	pTDH3	1461	277	6.46E-01
C. albicans	Ndt80B	pTDH3	4473	1858	1.58E-163
C. albicans	Ndt80B	Endogenous	2728	1155	4.86E-76

A


P. pastoris Ndt80
expressed in
S. cerevisiae
S. cerevisiae Ndt80
expressed in
S. cerevisiae
12
75
504

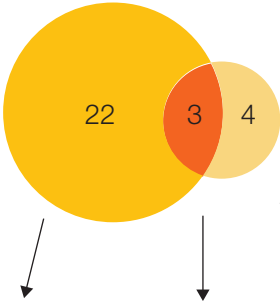
B


P. pastoris Ndt80
expressed in
S. cerevisiae
P. pastoris Ndt80
expressed in
P. pastoris
83
4
434


A
Percent of all Ndt80 targets shared between any two species
Conservation in Ndt80 targets/ Overall species conservation
Criteria 1
S. cer K. lac P. pas
K. lac 19
P. pas 17 18
C. alb 21 19 20
S. cer K. lac P. pas
K. lac 19
P. pas 22 24
C. alb 29 27 27
B
Criteria 2
S. cer K. lac P. pas
K. lac 13
P. pas 9.2 13
C. alb 14 14 11
S. cer K. lac P. pas
K. lac 13
P. pas 12 17
C. alb 20 19 15
C
Criteria 3
S. cer K. lac P. pas
K. lac 5.4
P. pas N/A N/A
C. alb 3.7 4.7 N/A
S. cer K. lac P. pas
K. lac 5.4
P. pas N/A N/A
C. alb 5.2 6.6 N/A
more conservation
Ndt80 phenotype conserved
D
Criteria 4
S. cer K. lac P. pas
K. lac 4.8
P. pas 7.5 3.1
C. alb 6.2 5.0 0.91
S. cer K. lac P. pas
K. lac 4.8
P. pas 10 4.1
C. alb 8.7 7.0 1.3

Shared sporulation genes

Shared Ndt80-dependent genes



ATG4
CAT2
DIT2
DPH1
ERF2
FKS3
IME2
INO1
MUM2
MUM3
NUS1
PFS1
RIM4
RRT8
SGA1
SIP4
SMC3
SMK1
SPS22
YAT2
YGL138C
YOR338W

GAS2
HYM1
MIP6

CDC5
CDC10
CDC3
SSP1

Gene	*K. lactis*	*P. pastoris*
ATG8		
MUM2		
SPO73		
AMA1		
RIM4		
SMC3		
IME2		
CDC16		
ATG4		
IME4		
SET1		
NDT80		
SPO75		
IRC19		
ATP10		
SSP2		
RTC6		
SPO1		
CDC10		
CDC14		
VAM7		
SSP1		
SMA2		
CDC5		
ATG3		
BUB3		
SWM1		
HOP2		
ATG1		
MND1		
SAE3		
SPO77		
OSW1		
CDC20		
SPO74		
YPT7		
CNM67		
PKH2		
CBC2		
HRR25		
GIP1		
ALG7		
SPO71		
ARG82		
PEP7		
BIM1		
YPT1		
SAE2		
BUD32		
CDC23		
SWI3		
SPO14		
PDS5		
SEC14		
STO1		
MET22		
VPS30		
PLC1		
YGR226C		
EMI2		
SPO20		
YMR158W-B		
SPO21		
MPC54		
SSO1		

Gene upregulated in meiotic vs. mitotic growth

Gene not upregulated in meiotic vs. mitotic growth

Gene not present in species