Multiple pathways to homothallism in closely related yeast lineages in the Basidiomycota

Alexandra Cabrita[1], Márcia David-Palma[1*], Patrícia H. Brito[1], Joseph Heitman[2], Marco A. Coelho[1*] and Paula Gonçalves[1#]

[#] Corresponding author: pmz@fct.unl.pt

1. Applied Molecular Biosciences Unit-UCIBIO, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

2. Department of Molecular Genetics and Microbiology, Duke University, Duke University Medical Center, Durham, NC 27710, USA.

Running title: Homothallism in *Cystofilobasidium*

Abstract

Sexual reproduction in fungi relies on proteins with well-known functions encoded at the mating type (*MAT*) loci. In the Basidiomycota, *MAT* loci are often bipartite, the *P/R* locus encoding pheromone precursors and pheromone receptors and the *HD* locus encoding heterodimerizing homeodomain transcription factors (Hd1 and Hd2). The interplay between different alleles of these genes within a single species determines a variety of sexual systems and patterns, which may result in the emergence of hundreds of compatible mating types. However, a minority of species are homothallic, reproducing sexually without the need for a compatible partner.

Here we examine the organization and function of the *MAT* loci of *Cystofilobasidium capitatum,* a species in the order Cystofilobasidiales, which is unusually rich in homothallic species. For this, we determined *MAT* gene content and organization in *C. capitatum* and found that it resembled a mating type of the closely related heterothallic species *Cystofilobasidium ferigula.* To explain the homothallic sexual reproduction observed in *C. capitatum* we examined HD -protein interactions in the two individual *Cystofilobasidium* species and determined *C. capitatum MAT* gene expression both in the natural setting and upon heterologous expression in *Phaffia rhodozyma,* a homothallic species belonging to a clade sister to the *Cystofilobasidium.* We conclude that the molecular basis for homothallism in *C. capitatum* appears to be distinct from that previously established for *P. rhodozyma.* Unlike the latter species, homothallism in *C. capitatum* may involve constitutive activation or dispensability of the pheromone receptor and the functional replacement of the usual Hd1/Hd2 heterodimer by an Hd2 homodimer. Overall, our results suggest that homothallism evolved multiple times in the Cystofilobasidiales, underpinned by diverse molecular mechanisms.

Importance

Sexual reproduction is important for the biology of eukaryotes because it strongly impacts the dynamics of genetic variation. In fungi, although sexual reproduction is usually associated with the fusion between

[*] Present address: Department of Molecular Genetics and Microbiology, Duke University, Duke University Medical Center, Durham, NC 27710, USA.

cells belonging to different individuals (heterothallism), sometimes a single individual is capable of completing the sexual cycle alone (homothallism). Homothallic species are unusually common in a fungal lineage named Cystofilobasidiales. Here we studied the genetic bases of homothallism in one species in this lineage, *Cystofilobasidium capitatum*, and found it to be different in several aspects when compared to another homothallic species, *Phaffia rhodozyma*, belonging to the genus most closely related to *Cystofilobasidium*. Our results strongly suggest that homothallism evolved independently in *Phaffia* and *Cystofilobasidium*, lending support to the idea that transitions between heterothallism and homothallism are not as infrequent as previously thought. Our work also helps to establish the Cystofilobasidiales as a model lineage in which to study these transitions.

Introduction

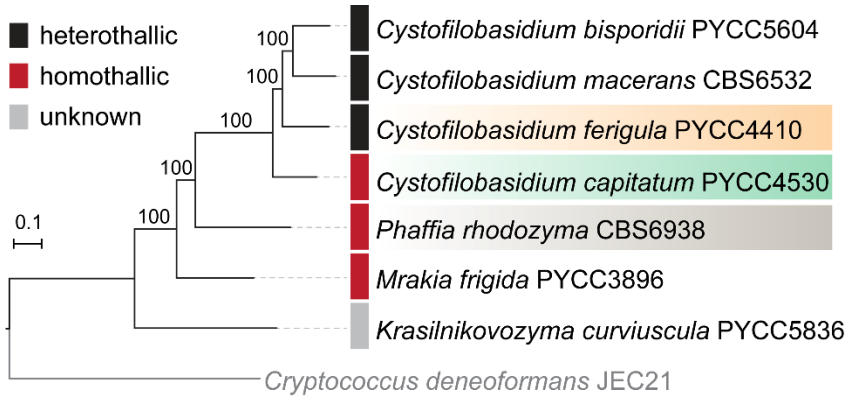
In Fungi, as in all eukaryotes, sexual reproduction is widespread, and some of the underlying mechanisms and molecular components are conserved even among distant lineages. The specific molecular pathways involved and the recognition systems responsible for triggering sexual reproduction may nonetheless vary greatly (1). Generally, sexual reproduction occurs through mating of two haploid individuals of the same species, possessing distinct mating types (2, 3). Mating types are defined by specific regions of the genome, the mating type (*MAT*) loci, which encode proteins responsible for triggering the major pathways leading to sexual development. Distinct mating types differ in the genetic content of the *MAT* loci. In Basidiomycetes, mating-type determination relies on two *MAT* loci (named *P/R* and *HD*) that encode two different classes of proteins (4–6). The *HD* locus contains two divergently transcribed genes encoding homeodomain transcription factors and the *P/R* locus comprises pheromone receptor and pheromone precursor encoding genes - *STE3* and *MFA,* respectively (4, 5, 7). In the Ascomycota, only one *MAT* locus encoding transcription factors is required to determine mating-type identity. Hence, the participation of receptors and pheromones in the determination of mating type is found only in the Basidiomycota.

The *HD* and *P/R MAT* loci in Basidiomycota can be either genetically linked or unlinked in the genome of a given species. If these loci are unlinked, they may segregate independently during meiosis, leading to the generation of four distinct mating types within the haploid progeny, which is the hallmark of the tetrapolar breeding system (4, 5, 7). The bipolar breeding system results from linkage of the two *MAT* loci and yields only two mating types in the haploid progeny (4, 5, 7). Bipolar mating also takes place if one of the two (unlinked) *MAT* loci loses its function in determining mating-type identity, which has been reported occasionally for the *P/R* locus (4, 5, 7).

Although heterothallism involving the fusion between compatible mating types as a requirement for sexual reproduction is more common, in some fungal species individuals are universally compatible, meaning that they can undergo sexual reproduction with any other individual, or even with or by itself (3, 7, 8). In basidiomycetes few homothallic organisms have been found and even fewer have had the genetic basis of their homothallism elucidated (1, 3, 7, 8). Two relevant examples of basidiomycete yeasts with distinct molecular strategies that result in a homothallic sexual mode of reproduction are the human pathogenic yeast *Cryptococcus deneoformans* (9, 10) and *Phaffia rhodozyma* (11).

77 *Cryptococcus deneoformans*, exhibits a particular form of homothallism named unisexual reproduction,
78 where cells of the same mating type can either fuse or undergo endoreplication of the entire genome
79 resulting in a diploid nucleus, develop hyphae, basidia and give rise to four haploid spores that are
80 products of a meiotic event. Interestingly, the key genes for heterothallic reproduction like *MFA*, *STE3*
81 and even *HD1* (*SXI1)* and *HD2* (*SXI2*) appear to be dispensable for unisexual reproduction of some *C.*
82 *deneoformans* strains (10, 12, 13).
83 More recently, the genetic basis of homothallism was dissected in *Phaffia rhodozyma* in our
84 laboratory (11). This astaxanthin-producing basidiomycetous yeast (14) belongs to the order
85 Cystofilobasidiales, a lineage with an unusually high proportion of homothallic species (15–20), but
86 comprising also heterothallic species and others for which sexual reproduction has not yet been
87 observed. In *P. rhodozyma*, deletion mutants were used to show that the two pairs of *STE3* and *MFA*
88 genes, and the single *HD1/HD2* gene pair present in the genome, are all required for robust sexual
89 reproduction (11). The two pairs of pheromone and pheromone receptors turned out to have reciprocal
90 compatibility, a single compatible *STE3* and *MFA* pair being sufficient for sexual reproduction. This is
91 what might be expected if the *P/R* loci of two putative mating types were present in the same genome,
92 in accord with the definition of primary homothallism. The only pair of Hd1 and Hd2 proteins encoded in
93 the genome is also essential for sexual development, but the mode of action is not fully understood (11).
94 This is because Hd1 and Hd2 are usually expected to heterodimerize to form functional transcription
95 factors, but proteins encoded in the same *HD* locus do not normally form dimers. This imposes
96 heterozygosity at the *HD* locus, with dimerization occurring only between proteins encoded by different
97 gene pairs, as a condition for sexual development. In line with this, the Hd1 and Hd2 proteins of *P.*
98 *rhodozyma* that are encoded in the same gene pair, do not appear to interact strongly, leading to the
99 tentative conclusion that a weak interaction between the two proteins might suffice for function (11).
100 Therefore, the *HD* locus configuration in *P. rhodozyma* is not fully aligned with the concept of primary
101 homothallism, where the presence of two distinct pairs of *HD* genes supporting strong cross dimerization
102 between Hd1 and Hd2 proteins would be expected.
103 Here, we examine in detail the content and function of the *MAT* loci of *Cystofilobasidium*
104 *capitatum*, a homothallic species closely related to *P. rhodozyma*, in order to understand if there are
105 common features between the molecular basis of homothallism in both species. We aim to shed some
106 light on the diversity of molecular mechanisms through which homothallism can occur in the phylum
107 Basidiomycota and improve the understanding of the evolution of mating patterns in the
108 Cystofilobasidiales.
109
110
111
112 **RESULTS**
113 ***MAT* loci in *Cystofilobasidium capitatum* and *Cystofilobasidium ferigula***
114 *Cystofilobasidium capitatum* and *Cystofilobasidium ferigula* belong to the order
115 Cystofilobasidiales which also comprises the genus *Phaffia*, for which two new species were recently
116 described (21), in addition to six other genera (22). The phylogenetic relationships within the order were

117 only recently clarified by a comprehensive genome-based phylogeny (21). This order contains a
118 strikingly large number of homothallic species (approximately one third of those described so far; **Table**
119 **S1**), a sexual mode uncommon in the Basidiomycota (3). In particular, *Phaffia* is composed entirely of
120 homothallic species (21) while *Cystofilobasidium* comprises species with a variety of sexual properties,
121 with *C. capitatum* and *C. intermedium* (23) representing the only two fully homothallic species in the
122 genus. The remaining species of *Cystofilobasidium* are heterothallic except for *C. macerans*, which
123 comprises strains exhibiting diverse sexual patterns (heterothallic, homothallic, and asexual) and *C.*
124 *alribaticum* for which no sexual reproduction has been observed under the conditions tested (23). The
125 genome-based phylogeny published for the Cystofilobasidiales (21) robustly supported *Phaffia* and
126 *Cystofilobasidium* as sister genera, a relationship that is recapitulated in the phylogeny shown in **Fig. 1,**
127 where a more limited number of species were included.



128

129 **Fig. 1.** Maximum likelihood phylogeny reconstructed from the concatenated protein alignments of 1147
130 single-copy genes shared across the studied taxa and the outgroup represented by *Cryptococcus*
131 *deneoformans*. Branch support was assessed by 1000 replicates of ultrafast bootstrap approximation
132 (UFBoot) with NNI optimization, and branch lengths are given in number of substitutions per site.
133

134 The availability of draft genomes for several *Cystofilobasidium* species allowed us to examine
135 the *MAT* loci of *C. capitatum* as well as that of the heterothallic species *C. ferigula* (**Fig. 2** and **Fig. S1**)
136 and to compare these with *MAT* loci in *P. rhodozyma* concerning gene content and organization (24).
137 In the genome of *C. capitatum* PYCC 4530 a single pair of divergently transcribed homeodomain
138 transcription factors (*HD1* and *HD2*) was found, as well as a single pheromone receptor gene (*STE3*)
139 flanked by two identical pheromone precursor genes (*MFA1a* and *MFA1b*; **Fig 2**). These two sets of
140 genes are located on different scaffolds and encode proteins with similar lengths to their counterparts
141 in *P. rhodozyma* (**Fig. 2A** and **Fig. 2B**). Both predicted Hd proteins have homeobox domains and nuclear
142 localization signals (NLS; **Fig.S2**). However, the *MFA* genes in *C. capitatum* encode a 41 amino acid
143 pheromone precursor protein where no site for N-terminal processing could be identified (**Fig. 2C**), which
144 may compromise the formation of a mature pheromone (25). Furthermore, analysis of gene synteny
145 conservation between *Phaffia* and *C. capitatum* indicates that the *P/R* locus of *C. capitatum* may be
146 restricted to the region containing the pheromones/receptors (highlighted in yellow in **Fig. 2B**) and the
147 *HD* locus most likely includes only the *HD1* and *HD2* genes, as observed in many other basidiomycetes
148 (11, 26, 27).

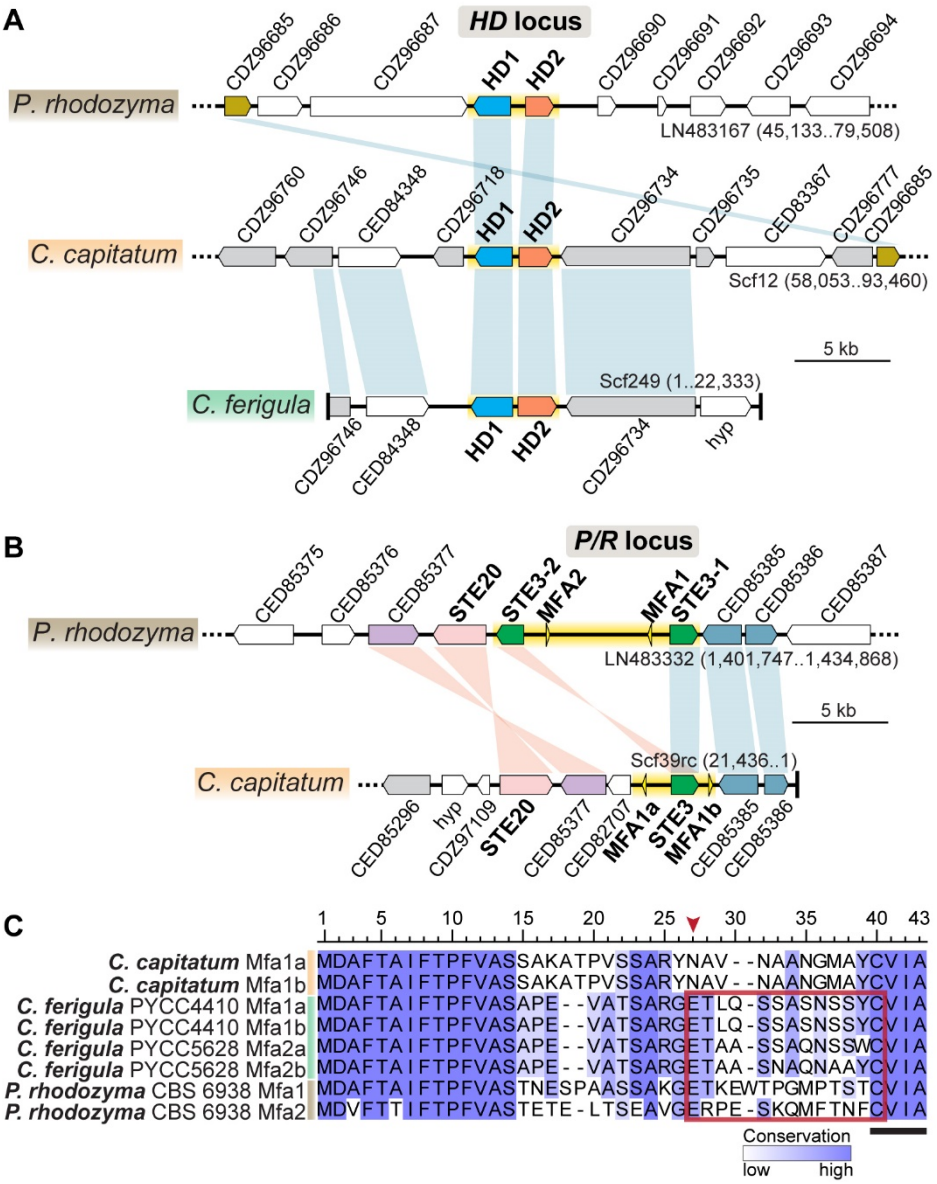


Fig. 2. Gene content and organization of the (A) *HD* and (B) *P/R* mating type loci in *C. capitatum* and *C. ferigula* (PYCC 4410), compared to *P. rhodozyma*. Genes are depicted as arrows denoting the direction of transcription. The genomic regions corresponding to the putative *MAT* loci are highlighted in yellow. Vertical bars connect orthologs that are in the same (blue) or inverted (pink) orientation. Non-syntenic genes are shown in white, and genes in grey in *C. capitatum* and *C. ferigula* are found scattered in the corresponding *P/R*- or *HD*-containing scaffold in *P. rhodozyma*, suggesting low level of synteny conservation beyond the core *MAT* regions. Gene names are based on top BLASTp hits in the *P. rhodozyma* genome and "hyp" denotes hypothetical proteins. The *P/R* locus of *C. ferigula* was omitted because the available genome assemblies are too fragmented in these regions to allow for a comparison. (C) Sequence alignment of pheromone precursors with amino acid positions colored in a blue gradient according to conservation. Sequences proposed as the putative mature pheromones are outlined by a red box, while those resembling the CAAX motif required for farnesylation are underlined. In *C. capitatum* the absence of a conserved position for N-terminal processing (the two charged amino acids indicated by a red arrowhead), precludes the prediction of a mature pheromone sequence.

For *C. ferigula,* the genomes of two compatible mating types were obtained (PYCC 4410 and PYCC 5628). Findings concerning *MAT* gene content were similar to *C. capitatum* (**Fig. 2** and **Fig. S1**), and in line with the genetic composition typically found in haploid mating types of basidiomycetes. In *C. ferigula*, two genes encoding pheromone precursors were also found in each of the mating types, but in

strain PYCC 5628 the two genes seem to encode slightly different mature pheromones. Because these genes are found in very small contigs in the current genome assemblies, it was not possible to determine their position in relation to the *STE3* gene nor the exact number of copies in the genome. It is conceivable that additional pheromones genes may become apparent when more complete assemblies of *C. capitatum* and *C. ferigula* genomes are available. As in *C. capitatum*, the *HD1/HD2* and the *STE3* genes were also found on different scaffolds in both *C. ferigula* strains, but the higher level of fragmentation of the current assemblies precludes a precise determination of the length and configuration of the *P/R* and *HD MAT* loci.

While heterothallic species are expected to harbour at least two different mating types with distinct alleles of *MAT* genes, this does not necessarily apply to homothallic species because there is no requirement for an operational self/nonself-recognition system in this case. To assess allele diversity of *MAT* genes in the *Cystofilobasidium* species under study, we obtained *MAT* gene sequences for as many strains as possible for both species. For *C. ferigula* two clearly distinct *STE3* alleles (sharing ~51% amino acid sequence identity) could be recognized among the five strains examined, as expected for a heterothallic species (**Fig. 3A**). For *C. capitatum,* although several different alleles could be identified in the 11 strains examined, they were much less divergent (sharing ~98% amino acid sequence identity) than *STE3* alleles known to encode proteins with different specificities, like those of *P. rhodozyma* (~50% sequence identity; (11)) and *C. ferigula* (**Fig. 3A**).

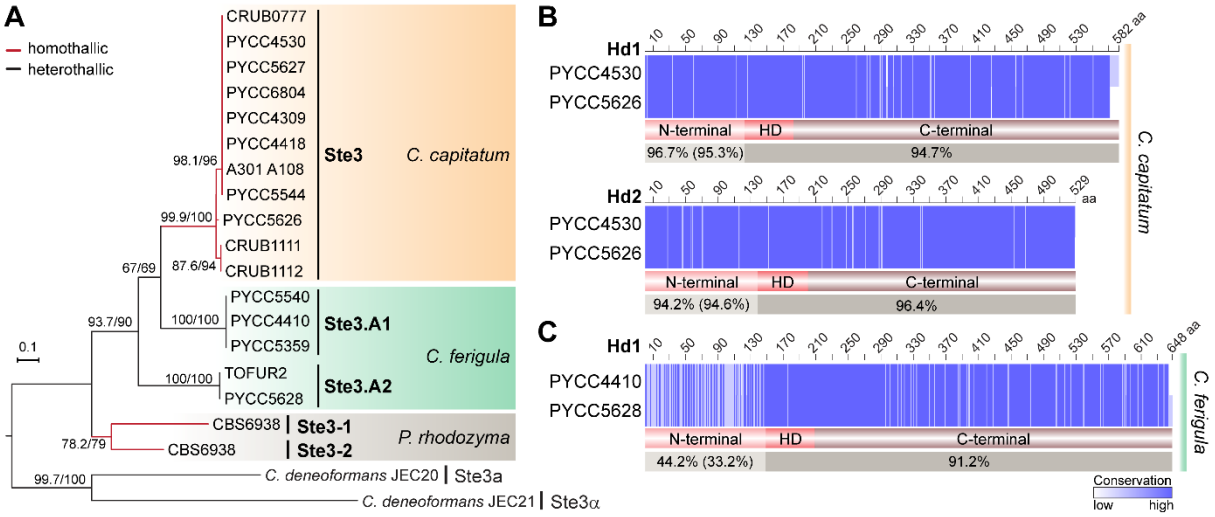


Fig. 3. Sequence diversity of *STE3* and *HD* mating type genes in *C. capitatum* and *C. ferigula*. (A) Maximum likelihood phylogeny of pheromone receptors (Ste3) obtained from various strains of *C. capitatum* (homothallic) and *C. ferigula* (heterothallic), along with the previously characterized Ste3 sequences of *P. rhodozyma* (homothallic) and *C. deneoformans* (heterothallic). The tree was inferred with the LG+F+G4 model of amino acid substitution and was rooted in the midpoint. Branch support values separated by a slash were assessed by 10,000 replicates of both the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) and the ultrafast bootstrap approximation (UFBoot). Compared to the other species, the low sequence divergence of Ste3 sequences among *C. capitatum* strains suggest the absence of functionally distinct, mating-type specific, alleles in this species. (B) and (C) Sequence alignments of the *HD1* and *HD2* gene products of *C. capitatum* and *C. ferigula*. Sequence identity between each pair of Hd1 and Hd2 proteins is given for the variable (N-termini) and conserved (homeodomain and C-termini) regions with amino acid positions colored in a blue gradient according to conservation. The comparison between Hd2 proteins of *C. ferigula* was not performed because the *HD2* gene of PYCC 5628 is fragmented in the current genome assembly. Values in brackets for the N-

202 terminal regions are the average identity as calculated from different allele products (see **Fig. S3C** for
203 details on number of alleles). In *C. capitatum*, variable amino acid positions are evenly distributed
204 throughout the length of the Hd1 and Hd2 proteins. In contrast, the N-terminal region of Hd1 in *C. ferigula*
205 is comparatively more variable, as commonly observed in Hd1 and Hd2 proteins of other heterothallic
206 basidiomycetes (5, 7, 12, 26).
207
208 The comparison of *C. capitatum* Hd proteins suggest the existence of three Hd variants in the
209 species with a degree of divergence between them that is lower than observed for functionally different
210 alleles of heterothallic species (**Fig. S3B and S3C)**. Moreover, for the two proteins that could be
211 examined over their entire length, the differences in the sequences are distributed homogeneously
212 throughout the protein (**Fig. 3B**), like observed for *P. rhodozyma* (11). This is unlike the divergence
213 observed in Hd1 sequences of *C. ferigula* and of other heterothallic basidiomycetes, which is more
214 extensive and concentrated in the N-terminal region responsible for self/nonself-recognition (**Fig. 3C**).
215 *C. ferigula* was reported to be bipolar, as only two different mating types had been identified so
216 far (16). This would mean that each *STE3* allele is expected to be linked to a single *HD* allele, defining
217 two mating types. However, as shown in **Fig. S3C**, our analysis uncovered three *HD* alleles, instead of
218 the expected two, and the *HD.B1* allele appears associated with the two receptors in different strains,
219 whereas the *STE3.A1* allele is associated with *HD.B1* and *HD.B3*. From these observations it seems
220 more likely that *C. ferigula* may have a tetrapolar mating system, in contrast to previous assumptions
221 (16).
222
223 **Involvement of the Hd proteins of *C. capitatum* in the homothallic sexual cycle**
224 In heterothallic basidiomycetes, the Hd1 and Hd2 proteins encoded in the *HD* locus control the
225 later stages of sexual reproduction through heterodimerization of non-allelic Hd1 and Hd2 proteins
226 brought together by cell fusion (4, 12, 26). Previous studies in our laboratory concerning the molecular
227 mechanisms of sexual reproduction of *P. rhodozyma* revealed that both the Hd1 and Hd2 proteins are
228 required for normal sporulation, and likely act through heterodimerization despite the weak nature of
229 their interaction (11).
230 To understand if *C. capitatum* resembled *P. rhodozyma* in this respect, the yeast two-hybrid
231 assay was employed to assess the ability of the Hd1 and Hd2 proteins of *C. capitatum* to interact with
232 each other. *HD1* and *HD2* cDNAs were isolated from *C. capitatum* strain PYCC 5626 and utilized for
233 the construction of the Gal4 fusion genes for this assay. The results of the assay, presented in **Fig. 4**,
234 are consistent with a complete absence of interaction between the Hd1 and Hd2 proteins of *C.*
235 *capitatum*, unlike the results for *P. rhodozyma* (11). Notably, a strong homodimerization was detected
236 for the Hd2 protein from this species. (**Fig. 4**). Weak homodimerizations of Hd proteins were also
237 previously observed for *P. rhodozyma* (11). For *C. ferigula* a strong heterodimerization of the Hd1 and
238 Hd2 proteins derived from strains of different mating types (PYCC 5628 and PYCC 4410, respectively)
239 was noted (**Fig. 4**), in line with observations in other heterothallic basidiomycete species (12, 26).
240 Consistently, beta-galactosidase activity resulting from activation of the *lacZ* reporter gene was similar
241 to the positive control both for heterodimerization of Hd proteins from *C. ferigula* and for
242 homodimerization of Hd2 in *C. capitatum* (**Fig. 4B**). Homodimerization of Hd1 in *C. capitatum* could not

be tested, because the construct of the fusion protein between Hd1 and the Gal4 DNA binding domain could not be stably expressed in the pertinent *S. cerevisiae* strain.

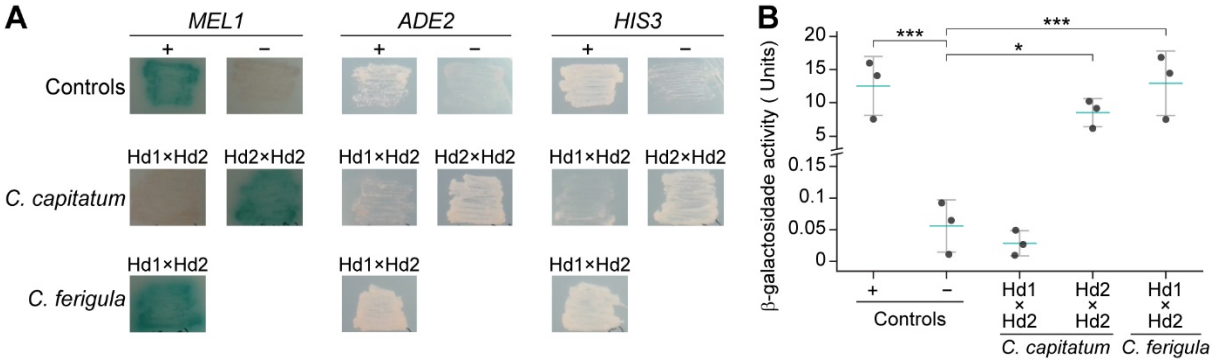


Fig. 4. Results of the Yeast Two-Hybrid assay for Hd proteins of *C. capitatum* and *C. ferigula*. (A) Qualitative results concerning the dimerization of Hd1 and Hd2 of *C. capitatum* and of *C. ferigula*, as well as homodimerization of Hd2 of *C. capitatum*, through the activation of the reporter genes *MEL1*, *ADE2*, and *HIS3*. In *C. ferigula* the Hd1 protein sequence is derived from strain PYCC 5628 and the Hd2 protein is derived from strain PYCC 4410. (B) Quantitative results of the β-Galactosidase assay of interactions shown in panel A. The interaction between a fusion protein containing the Gal4 activation domain fused to the SV40 large T-antigen and a fusion between Gal4 binding domain and p53 was used as positive control, while the negative control employed the same Gal4 activation domain fusion in combination with a fusion between Gal4 binding domain and lamin. Plasmids encoding the positive and negative control proteins were provided with the Matchmaker Gold Yeast Two-Hybrid System, by Takara Bio USA.

The *HD* locus of *C. capitatum* partially complements a *P. rhodozyma HD* deletion mutant

The results obtained in the yeast two-hybrid assay suggest that homodimerization of Hd2 may play a role in homothallic sexual development in *C. capitatum,* possibly functionally replacing the usual Hd1/Hd2 heterodimer. To investigate this, we set out to assess function of the *C. capitatum HD* locus by heterologous expression in a *P. rhodozyma HDΔ* mutant.

Integration of the complete *HD* locus of *C. capitatum* strain PYCC 4530 in the rDNA locus of the *P. rhodozyma HDΔ* mutant (construct *HDΔ+HD1/HD2-PYCC4530*) resulted in a very weak but consistent recovery of sporulation (**Table 1, Table S2 and Table S3**). Because no interaction between the Hd1 and Hd2 proteins could be detected in the yeast two-hybrid assay but instead strong homodimerization of the Hd2 protein was observed, we subsequently decided to assess whether expression of *HD2* alone was sufficient to restore sexual development of the *P. rhodozyma HDΔ* mutant. To this end, a construct containing in addition to the intergenic region only a residual (344 bp) 5´portion of the *HD1* gene (excluding the homeodomain; *HDΔ+HD2-PYCC4530*) was used to transform the *P. rhodozyma HDΔ* mutant. Sporulation levels like those observed upon transformation of the complete *C. capitatum HD* locus were observed (**Table 1, Table S2 and Table S3**).

Table 1. Complementation of *P. rhodozyma MAT* loci mutant strains with *C. capitatum* genes. Sporulation patterns were scored qualitatively using the criteria explained as given in the key.

Strain	*P. rhodozyma* WT	*HD*Δ	*HD*Δ*+HD2-PYCC4530*	*HD*Δ*+HD1/HD2-PYCC4530*	*P/R*Δ	*P/R*Δ*+STE3/MFA-PYCC4530*
Sporulation phenotype	+++	–	+	+	–	–

(–): < 5 basidia; (+): 5-30 basidia; (+++) > 1000 basidia.

The *HD1 and HD2* genes of *C. capitatum* are differently expressed during growth in sporulation conditions and in *P. rhodozyma* mutants

To substantiate the hypothesis that Hd2 might be the sole important player in the *HD* locus of *C. capitatum*, we compared the expression levels of *HD1* and *HD2* in *C. capitatum* PYCC 5626, in sporulation conditions, and in the *P. rhodozyma* mutant containing the complete *HD* locus of *C. capitatum* PYCC 4530 (construct *HDΔ+HD1/HD2-PYCC4530*). The results, depicted in **Fig. 5**, show that of the *HD* gene pair, only the *HD2* gene seems to be substantially expressed in the *C. capitatum* strain. Heterologous expression of *HD2* in the *P. rhodozyma* mutant, although much lower than in *C. capitatum*, can be easily detected, while heterologous expression of *HD1* seems to be only vestigial, because it was impossible to measure in one of the biological replicates and was very low in the other.

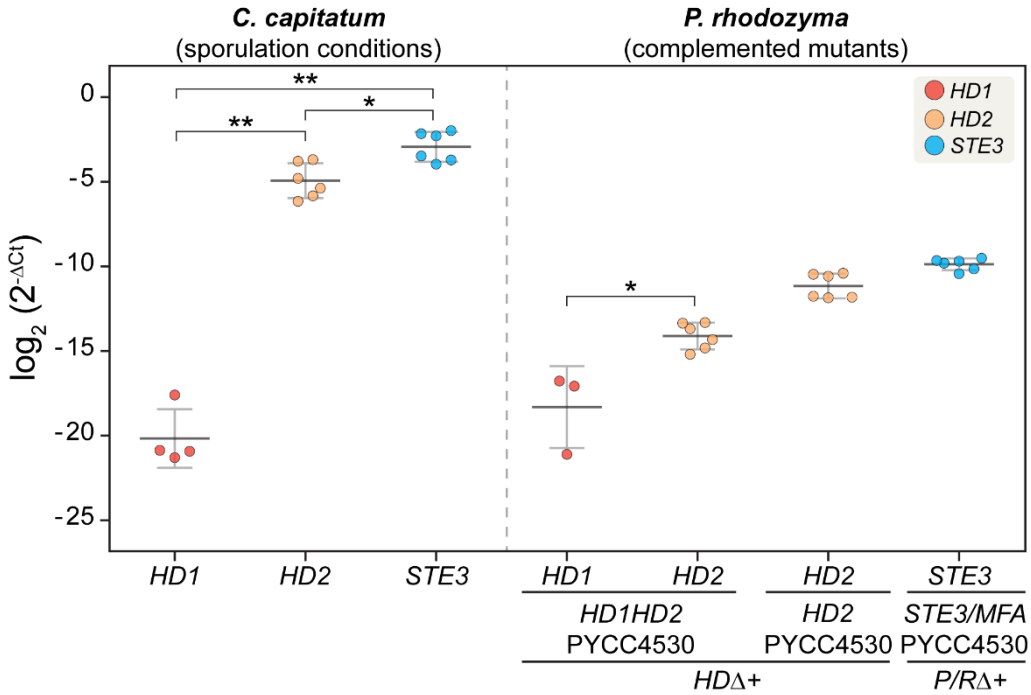


Fig. 5. Real time quantitative PCR results of *MAT* gene expression in *C. capitatum* under sporulation conditions and in *P. rhodozyma* complemented mutants. Expression of *MAT* genes is given as $\log_2$ fold difference relative to the expression of actin in each strain. These results are derived from two biological replicates. Significant differences between the expression of different genes were calculated with Mann-Whitney test (*p<0.05, **p<0.01).

305 ***P/R* locus of *C. capitatum* does not restore sporulation in a *P. rhodozyma* cognate deletion**
306 **mutant**
307 Because Hd function could be assessed using heterologous expression, a *P. rhodozyma*
308 deletion mutant of both *P/R* clusters (*P/RΔ*) was similarly used as host for the *P/R* locus of *C. capitatum*
309 strain PYCC 4530, encompassing the *STE3* and *MFA1*a genes and the respective native promotor
310 regions (*P/RΔ+STE3/MFA-PYCC4530*). However, integration of the *P/R* locus of *C. capitatum* in the
311 rDNA of the *P. rhodozyma P/RΔ* mutant failed to restore sporulation (**Table 1 and Table S2**), although
312 the *C. capitatum STE3* gene was expressed in *P. rhodozyma* (**Fig. 5**). Interestingly, *STE3* is the most
313 expressed among *MAT* genes of *C. capitatum* PYCC 5626 (**Fig. 5**), suggesting it may have a role in
314 sexual reproduction despite our inability to observe it in the heterologous setting.
315
316
317 **DISCUSSION**
318 The main aim of this study was to ascertain to which extent common features could be found
319 between the molecular bases of homothallism in different species of the Cystofilobasidiales, a lineage
320 in the Basidiomycota particularly rich in species exhibiting this uncommon sexual behaviour. The
321 molecular basis of homothallism was previously dissected in the genetically tractable Cystofilobasidiales
322 species *P. rhodozyma* (11). Here, we characterized the *MAT* loci of a second homothallic species
323 belonging to a sister genus, *C. capitatum,* by examining the structure of the loci in the available genome
324 of strain PYCC 4530 and by comparing *MAT* gene sequences in a total of 11 *C. capitatum* strains. The
325 most striking difference between the *MAT* loci in *P. rhodozyma* and *C. capitatum* is the presence in the
326 latter species of a single pheromone receptor gene, instead of the two distinct and functionally
327 complementary sets of pheromone receptor and pheromone precursor genes found in *P. rhodozyma*.
328 Therefore, the *C. capitatum MAT* gene content resembles a haploid mating type of a heterothallic
329 species (4), while that of *P. rhodozyma* is reminiscent of a fusion between two mating types (11).
330 However, like in *P. rhodozyma*, no evidence for functionally distinct variants (alleles) of *MAT* genes that
331 might form different mating types were found in *C. capitatum,* consistent with its homothallic behaviour.
332 We subsequently devised various experimental approaches to try to bring to light functional features of
333 *MAT* genes in *C. capitatum.* In these experimental approaches two *C. capitatum* strains were used: in
334 addition to sequenced strain PYCC 4530, strain PYCC 5626 was employed for the isolation of cDNAs
335 because unlike other strains, it readily sporulates in different experimental conditions.
336 *P. rhodozyma* can be described as a primary homothallic basidiomycete, with reciprocal
337 compatibility between the pheromones and receptors of the two *P/R* clusters and with a weak
338 heterodimerization between the only pair of Hd1 and Hd2 proteins being responsible for triggering of its
339 sexual cycle, with possible hints at homodimerization events (11). Therefore, in *P. rhodozyma*, the *P/R*
340 system seems to work similarly to what might be expected for a heterothallic mating type, while the *HD*
341 locus is apparently operating through weak dimerization between adjacently encoded Hd1 and Hd2.
342 This dimerization normally occurs only between proteins encoded by different *HD* alleles. Hence, the
343 weak interaction in *P. rhodozyma* may have evolved to support sexual development in the absence of
344 a second *HD* allele. Because the *MAT* loci in *C. capitatum* have the same gene content as haploid

345 mating types of heterothallic species, in this case both the *P/R* and the *HD* loci probably underwent
346 changes in their mode of operation to permit homothallism. For the *HD* locus, the complete absence of
347 interaction between the Hd1 and Hd2 proteins, the formation of Hd2 homodimers, complementation
348 results of the *P. rhodozyma HD∆* mutant and the very low expression of the *HD1* gene suggest that the
349 *HD* locus is relevant for sporulation but also that regulatory functions normally fulfilled by the Hd1/Hd2
350 heterodimer may have been taken over by the Hd2 homodimer. This has, to our knowledge, not been
351 reported before for any other basidiomycete, although some evidence for homodimerization of Hd
352 proteins had already been observed in *P. rhodozyma* (11). This hypothesis for the mode of action of the
353 *HD* locus predicts that Hd1 might be dispensable. In some heterothallic basidiomycetes, it has been
354 shown that the Hd1 and Hd2 proteins hold different functional domains that are essential for function of
355 the transcription regulation complex (28–30). One such species is the mushroom *Coprinopsis cinerea*,
356 where it has been shown that Hd1 contributes with the NLS region, allowing the heterodimer to be
357 transported into the nucleus, while the homeodomain of Hd2 is required for binding of the complex to
358 DNA (28, 31). Hence, in this case, formation of a heterodimer is required for function. In *C. capitatum*,
359 Hd2 possesses both an NLS and a homeodomain so in that perspective it seems possible that Hd2 may
360 indeed play a role in the homothallic life cycle of *C. capitatum* that is sufficient and independent of Hd1.
361 It is possible that this hypothesized function of Hd2 evolved recently in this species, which could explain
362 the fact that the *HD1* is not pseudogenized. Alternatively, the Hd1 may have acquired a function
363 unrelated to sexual reproduction, which might entail its expression under different conditions, and justify
364 its maintenance in the genome.

365 Introduction of the *P/R* locus it the *P. rhodozyma P/R∆* mutant failed to restore sporulation of
366 the cognate mutant, even though expression of the *STE3* gene in the heterologous setting could be
367 demonstrated. The explanations for this are presently unclear, since the ability of Ste3-like receptors to
368 activate heterologously a sexual development pathway over a much larger phylogenetic distance was
369 previously demonstrated (32). As to the mode of action of the receptor in its normal setting, the possibility
370 that it could be constitutively active should be considered, in line with the fact that the predicted
371 pheromone precursor genes lack the N-terminal processing signals, likely precluding the formation of
372 an active pheromone. Mutations within the pheromone receptor that can lead to constitutively active
373 receptors have been previously reported, resulting in a bypass of the pheromone for sexual reproduction
374 (25, 32–35). Mutations in the pheromone that allow it to activate the receptor encoded in the same *P/R*
375 cluster are also known (25, 36). In fact, mutations that lead to self-activation or self-compatibility have
376 been proposed to form the basis for the transition from a tetrapolar to a bipolar mating system, or even
377 from a heterothallic to a homothallic mating behaviour (3, 5) for some species. If the pheromone system
378 does not operate normally in *C. capitatum*, this may be a reason for the lack of complementation of the
379 *P. rhodozyma P/R*Δ mutant by the *C. capitatum P/R* locus.

380 Hence, taken together, our results suggest that *P. rhodozyma* and *C. capitatum* attained
381 homothallic breeding systems through different mechanisms, which is consistent with the hypothesis
382 that the *P. rhodozyma* mating system arose at the origin of the genus, possibly by fusion of *P/R* loci of
383 compatible mating types and the adaptation of the HD dimerization system. *C. capitatum*, on the other
384 hand, has an extant makeup of its *MAT* loci that suggests it may have evolved from a heterothallic

385 mating type, in which the Hd2 homodimer acquired a prominent role. If a mature pheromone is formed
386 despite the absence of an N-terminal processing signal, it might be that the pheromone-receptor pair in
387 *C. capitatum* have become self-activating. On the other hand, if no mature pheromone is formed, the
388 pheromone receptor gene might have bypassed the pheromone requirement and is now constitutively
389 active. Whatever the case may be, the *P/R* system still likely has a role in some aspects of the sexual
390 development, which in agreement with the fact that *STE3* gene seems to be highly expressed. Although
391 we cannot discard the hypothesis that the receptor is fulfilling a role unrelated to sexual reproduction,
392 and while such receptors have been reported (37), they were located in all instances outside the *P/R*
393 locus.
394 The different particularities of homothallism in the two Cystofilobasidiales species studied so far
395 are suggestive of remarkable levels of plasticity in the evolution of sexual reproduction in this order. It
396 will be interesting to conduct similar studies in other homothallic species of this order, which would allow
397 us to get a more complete insight in the array of mechanisms involved as well as possible genomic
398 rearrangement that may have been involved in the transitions between heterothallic and homothallic
399 species. Having uncovered *P. rhodozyma* as a viable host for heterologous expression, opens the
400 possibility of assessing the functionality of other *MAT* proteins from uncharacterised species in this
401 order.
402
403 **MATERIALS AND METHODS**
404 **Strains and culture conditions**
405 *P. rhodozyma*, *C. capitatum* and *C. ferigula* strains (**Table S4**) were routinely grown in YPD
406 medium (2% Peptone, 1% Yeast Extract, 2% Glucose, 2% Agar) at 17 to 20ºC. For the preparation of
407 electrocompetent cells, *P. rhodozyma* strains were grown in YPD liquid medium at 20ºC, and
408 transformants were incubated at 17ºC in selective medium, consisting of YPD plates supplemented with
409 the appropriate antifungal drugs (100 µg/ml of geneticin and/or 100 µg/ml of zeocin and/or 100 µg/ml of
410 hygromycin B).
411 *Escherichia coli* strain DH5α (Gibco-BRL, Carlsbad, CA, USA) was used as a cloning host and
412 was grown at 37ºC in LB medium (1% NaCl, 1% Tryptone, 0.5% Yeast Extract and 2% Agar for solid
413 medium) supplemented with ampicillin at 100 µg/ml when appropriate.
414
415 **DNA extraction, genome sequencing and assembly**
416 Genomic DNA of *C. ferigula* PYCC5628 was extracted from single cell-derived cultures using
417 the ZR Fungal/Bacterial DNA MiniPrep kit (ZYMO Research). DNA samples were quantified using Qubit
418 2.0. Genome sequencing was carried out by commercial providers, at the Genomics Unit of Instituto
419 Gulbenkian de Ciência, and at the Sequencing and Genomic Technologies Core Facility of the Duke
420 Center for Genomic and Computational Biology. Two short insert-size libraries (~500 bp) were prepared
421 with the Nextera Kit and subsequently sequenced using the Illumina MiSeq and HiSeq2500 systems to
422 generate paired 151- and 300-nt reads, respectively. After adaptor clipping using Trimmomatic (v0.36)
423 the two sets of reads were assembled with SPAdes (v3.11.1) (38) (with parameters: "--careful" to reduce
424 the number of mismatches and short indels in the final assembly, and the k-mer sizes: 21, 33, 55, 77,

425 99, 127, automatically selected based on read length). Genome assembly quality was assessed by the
426 QUAST (v.5.0.2) (39), and gene models were predicted ab initio using Augustus (40) trained on
427 *Cryptococcus neoformans*. Genome sequencing data generated, and final genome assembly statistics
428 are given in **Table S5**.
429
430 **Identification of mating type genes and synteny analyses**
431 Scaffolds containing *MAT* genes, namely the homeodomain transcription factors (*HD1*/*HD2*)
432 and the mating pheromones (*MFA*) and receptors (*STE3*), were identified by BLASTP or TBLASTN in
433 the genome assemblies of *C. capitatum* PYCC 4530 (BioProject: PRJNA371778) and *C. ferigula* PYCC
434 4410 (BioProject: PRJNA371793), and in the newly obtained assembly of *C. ferigula* PYCC 5628
435 (BioProject: PRJNA371809). Well-annotated *P. rhodozyma MAT* proteins (11) were used as search
436 query. The retrieved *MAT* genes were manually reannotated if required and analysed further: (i) the
437 transmembrane regions in the pheromone receptor protein were predicted by HMMTOP software (41);
438 (ii) the Homeodomain regions in Hd1 and Hd2 proteins were predicted by InterPro server (42) and
439 compared to the previously characterized homeodomain proteins in Pfam database; (iii) nuclear
440 localization signals (NLS) and coiled-coil motifs were identified in the complete Hd1 and Hd2 sequences
441 using, respectively, the SeqNLS (with a 0.8 cut-off) (43) and Jpred4 (44) (see **Figs. S1B** and **S1C** and
442 **Fig. S2**). Synteny between *MAT* regions of different strains and species was based on bidirectional
443 BLAST analyses of the corresponding predicted proteins. The short pheromone precursor genes in the
444 genomes of *C. ferigula* and *C. capitatum* were identified manually as they usually fail automatic
445 detection.
446
447 **Species and MAT gene phylogenies**
448 A phylogenetic analysis representing major lineages within Cystofilobasidiales was inferred on
449 a concatenated protein dataset of single copy core genes of four *Cystofilobasidium* species, *Phaffia*
450 *rhodozyma* CBS 6938, *Mrakia frigida* PYCC 3896, *Krasilnikovozyma curviuscula* PYCC 5836, and the
451 outgroup *Cryptococcus deneoformans* JEC21 (**Table S**4). Orthologous clusters were inferred with all-
452 against-all BLASTP (NCBI Blast-2.2) searches and the Markov cluster algorithm (OrthoMCL v1.4; (45))
453 with inflation factor (F) of 1.5, and minimum pairwise sequence alignment coverage of 50% implemented
454 in GET_HOMOLOGUES package (46). Clusters present in single copy in all analyzed genomes were
455 retained, aligned with MAFFT v7.407 using the G-INS-I method and default parameter values (47),
456 trimmed with BMGE v1.12 using the amino acid option (48), and finally concatenated into a single data
457 set. The species phylogeny was inferred with IQ-TREE v1.6.12 (49) using maximum-likelihood (ML)
458 inference under a LG+F+I+G4 model of sequence evolution. ModelFinder (50) was used to determine
459 best-fit model according to Bayesian Information Criterion (BIC) and branch support was estimated
460 using ultrafast bootstrap approximation (UFBoot) with NNI optimization (51), both implemented in IQ-
461 TREE package.
462 To analyse the *MAT* gene content across strains of *C. capitatum* and *C. ferigula*, protein
463 sequences of the *HD1*, *HD2* and STE3 genes were retrieved from the genome assemblies and aligned
464 separately. Conserved regions were used to design primers to amplify the corresponding genomic

465 regions across the available strains of each species (**Table S6**). These regions include a ~ 870-bp region
466 of the STE3 gene, and a ~1.5-kb fragment encompassing the 5′ end and intergenic regions of the HD1
467 and HD2 genes (**Table S6 and Fig. S3**). Genomic DNA was extracted through a standard Phenol-
468 Chloroform method (52) and the regions of interest were PCR-amplified, purified using Illustra GFX PCR
469 DNA and Gel Band Purification Kit (GE Healthcare Life Sciences), and then sequenced by Sanger
470 Sequencing, at STABVida (Portugal). For phylogenetic analysis of *MAT* genes, amino acid or nucleotide
471 sequences were individually aligned with MAFFT v7.310 (47) using the L-INS-i strategy (--localpair --
472 maxiterate 1000) and poorly aligned regions were trimmed with TrimAl (--gappyout) (53). The resulting
473 alignments were input to IQ-TREE v.1.6.5 (49) ML phylogenies using best-fit models automatically
474 determined by ModelFinder (50) (parameter: -m MFP). The exact model employed in tree reconstruction
475 is given in the respective figure legends. Branch support values were obtained from 10,000 replicates
476 of both UFBoot (51) and the nonparametric variant of the approximate likelihood ratio test (SH-aLRT)
477 (54). In addition, the option "-bnni" was employed to minimize the risk of overestimating branch supports
478 with UFBoot when in presence of severe model violations. The resulting phylogenies were midpoint
479 rooted and graphically visualized with iTOL v5.5.1 (55).
480
481
482 **Yeast Two-Hybrid assay**
483 To assess the interaction between the Hd1 and Hd2 proteins, the Matchmaker Gold Yeast Two-
484 Hybrid System kit (Takara) was used. Coding DNA sequences of the pertinent *HD* genes were cloned
485 into a pGADT7 and a pGBKT7 plasmids (Takara) containing the Activation and DNA binding domains
486 of Gal4, respectively, to obtain the required fusion proteins.
487 Synthetic genes were designed using the coding DNA sequences of the *HD1* and *HD2* genes
488 of *C. ferigula* and were synthesised at Eurofins Genomics (Germany). The *HD1* and *HD2* gene
489 sequences of strains PYCC 4410 and PYCC 5628 from *C. ferigula* respectively were adapted to the *S.*
490 *cerevisiae* codon usage (**Fig. S4**). The synthetic genes were obtained as inserts of pEX-A258 plasmids.
491 cDNAs of *HD1* and *HD2* from *C. capitatum* were obtained from total RNA isolated from strain PYCC
492 5626, briefly as follows. Strain PYCC 5626 was cultivated in GSA medium (0.2% Glucose, 0.2%
493 Soytone) in 10% of the flask volume, for 8 days at 20ºC and 90 rpm (Sartorius Certomat IS incubator),
494 until inspection under the microscope revealed the presence of teliospores (16). RNA extraction was
495 performed using the *ZR Fungal/Bacterial RNA MiniPrep kit* (by ZYMO Research), with a single step of
496 in-column DNase I Digestion to free the RNA samples of genomic DNA. cDNA was synthesized from
497 total mRNA using Maxima H Minus Reverse Transcriptase (by Thermo Scientific) and oligo (dT)20 as
498 primer, and synthesis of the second DNA strand was performed using specific primers for the complete
499 *HD1* and *HD2* genes (**Table S6**). The fragments corresponding to the *HD1* and *HD2* coding DNA
500 sequences of strain PYCC 5626 were sequenced by Sanger Sequencing, at STABVida (Portugal). The
501 protein sequences of Hd1 and Hd2 of strain PYCC 5626 were aligned with those of strain PYCC 4530
502 (**Fig. 3.B**) using the software MUSCLE (implemented in the software Unipro UGENE v1.30.0 (56)) and
503 level of intraspecific variability calculated.

504 The *HD1* and *HD2* complete synthetic genes from *C. ferigula* and cDNA fragments from *C.*
505 *capitatum* were amplified using primers that contained 40 bp tails at their 5' ends that correspond to the
506 flanking regions of the Multiple Cloning Sites present in pGADT7 and in pGBKT7 (**Table S6**). Plasmid
507 pGADT7 was then linearized at the Multiple Cloning Site by digestion with Cla I (Thermo Scientific),
508 while pGBKT7 was linearized by digestion with Pst I (Thermo Scientific). *S. cerevisiae* strains Y187 and
509 Y2HGold were transformed with inserts and linearized vectors using the transformation method
510 described in the Yeastmaker Yeast Transformation System 2 protocol. Transformants were selected in
511 appropriate media (Yeast Nitrogen Base without amino acids, with appropriate supplements (**Table S7**))
512 at 30ºC.
513 Mating of haploid *S. cerevisiae* strains was performed by incubating a single colony from each
514 of the two haploid transformants to be mated in 200 µl of YPD medium at 30ºC, for 24 hours, at 250
515 rpm. After incubation, the cells were recovered and thoroughly washed with distilled sterile water and
516 plated on appropriate selective media (**Table S7**).
517 To test the activation of the *ADE2* and *HIS3* reporter genes, haploid transformants (**Fig. S5**) and
518 diploids were plated on appropriate selective media without adenine nor histidine, respectively, while to
519 test the activation of *MEL1* the haploid transformants and diploid derivatives were plated on appropriate
520 selective medium supplemented with X-α-Gal (Takara Bio) at a final concentration of 40 µg/ml.
521 To quantify the activation of the *LacZ* reporter gene, a β-galactosidase activity assay was performed,
522 using o-nitrophenyl β-D-galactopyranoside (ONPG) as a substrate. The assay was performed as
523 described in the Yeast Protocols Handbook (Takara Bio). All reactions were performed in triplicate, so
524 that they could be stopped at 3 different points in time (after 2 hours, 6 hours, and 24 hours), by the
525 addition of 0.4 mL of 1M Na_2CO_3 to each suspension. Raw data concerning these assays is shown in
526 **Table S8.**
527
528 **Construction of recombinant plasmids and gene deletion cassettes**
529 For the construction of *P. rhodozyma* mutants, recombinant plasmids and gene deletion
530 cassettes were constructed, as follows. Primer sequences (**Table S6**) were based on available genome
531 sequences of *P. rhodozyma* strain CBS 6938 (NCBI project PRJEB6925 (57)) and *C. capitatum* strain
532 PYCC 4530 (Bioproject PRJNA371774). Plasmids used for the constructions were pJET1.2/blunt
533 (Thermo Scientific), pPR2TN containing a geneticin resistance cassette (58) and pBS-HYG (59)
534 containing a hygromycin resistance cassette. All fragments used for cloning and deletion cassettes were
535 amplified by PCR using Phusion High-Fidelity DNA polymerase (Thermo Scientific) and the amplified
536 products were purified using either *Illustra GFX PCR DNA and Gel Band Purification Kit* (GE Healthcare)
537 or *GeneJET Gel Extraction Kit* (Thermo Scientific). Constructions were performed using standard
538 molecular cloning methods (60) and *E. coli* strain DH5α as host.
539
540 **Transformation of *P. rhodozyma***
541 *P. rhodozyma* strains (CBS 6938 and mutants) (**Table S4**) were transformed by electroporation,
542 with the linearized recombinant plasmids or deletion cassettes, as previously described by Visser *et al*,
543 2005 (61), reducing the amount of DNA used to 2 ug for the transformations of the complemented *P.*

544 *rhodozyma* mutants. The electroporation conditions (Gene Pulser II Electroporation System, Bio-Rad)
545 consisted of an internal resistance of 1000 Ω, an electric pulse of 0.8 kV, a capacitance of 25 μF,
546 resulting in a pulse ranging from 18 to 20 ms (62, 63). The cells recovered subsequently in YPD liquid
547 medium, for at least 2.5 hours at 17 ºC before being plated on YPD medium supplemented with the
548 appropriate antifungal drug and incubated at 17ºC. The genotypes of the transformants were determined
549 as previously described (11, 64).
550
551 ***P. rhodozyma* sporulation assays**
552 To determine the ability of the *P. rhodozyma* mutants to sporulate, sporulation assays were
553 performed, where the strains were incubated in DWR solid medium with 0.5% of ribitol (0.5% Ribitol and
554 2.5% Agar), as previously described (11). Each sporulation assay was conducted on 3 plates containing
555 DWR+0.5% ribitol. On each plate, 10 colonies of each strain to be tested were spotted. Different strains
556 were employed in each assay as indicated in **Table S2**, but in all assays the *P. rhodozyma* wild type
557 was used as a positive sporulation control. Colonies were examined under the microscope after 10 and
558 20 days of incubation at 18ºC, and sporulation patterns were scored qualitatively. The numbers of
559 basidia counted in experiment E8 for the complemented mutants concerning the study of the *HD* locus
560 of *C. capitatum* after 20 days of incubation are listed in **Table S3**. In all cases the entire colony was
561 submitted to microscopic observation.
562
563 **Real-time quantitative PCR to assess expression of the *MAT* genes**
564 Total RNA was extracted from a sporulating culture of *C. capitatum* strain PYCC 5626.
565 Sporulation was induced by incubation in GSA liquid medium (0.2% Glucose, 0.2% Soytone) in 10% the
566 volume of the flask, at 17ºC, without agitation, until microscopic inspection revealed hyphae and
567 teliospores, the latter being thick-walled resting spores from which basidia arise (16). Total RNA
568 extraction was performed using the *ZR Fungal/Bacterial RNA MiniPrep kit* (by ZYMO Research). *P.*
569 *rhodozyma* complemented mutants were grown in YPD liquid medium to an OD_{600nm} of 1.0. The cultures
570 were then collected and frozen at -80ºC for 1 hour before proceeding with total RNA extraction through
571 a standard Trizol method. In-column DNase I digestion to free the RNA samples of genomic DNA using
572 the *RNA Clean & Concentrator kit* (by ZYMO Research) was used for all samples, and absence of gDNA
573 was verified by PCR. cDNA was synthesized from total mRNA using Maxima H Minus Reverse
574 Transcriptase (Thermo Scientific) and oligo (dT)20 as primer. Real-Time PCR was performed using the
575 *SensiFAST SYBR No-ROX kit* (by Bioline, London), with 20μl reactions, in a Rotor-Gene 6000 Corbett
576 apparatus. The reaction parameters consisted of an initial denaturation step at 95ºC, for 2 min, followed
577 by 40 cycles of 95ºC for 5 seconds, 57ºC for 10 seconds and 72ºC for 20 seconds. Two biological
578 replicates were performed, with triplicates performed for each reaction (**Table S9**). Relative expression
579 of the *MAT* genes was calculated using the $2^{-\Delta Ct}$ method, where $\Delta Ct = Ct_{test} - Ct_{reference}$, as described by
580 Livak and Schmittgen (2001) (65), and expression values were represented as the $\log_2 (2^{-\Delta Ct})$. Mann
581 Whitney tests were performed to determine if the differences in expression of genes within each strain
582 were statistically different.
583

ACKNOWLEDGEMENTS

This work was supported by UCIBIO-Unidade de Ciências Biomoleculares Aplicadas, which is financed by portuguese funds from Fundação para a Ciência e Tecnologia, Ministério da Ciência, Tecnologia e Ensino Superior (FCT/MCTES; https://www.fct. pt/) grant UID/Multi/04378/2019 and FCT/MCTES grants PTDC/BIA-GEN/112799/2009 (to P.G.), SFRH/BPD/79198/2011 (M.C.) and SFRH/BD/81895/2011 (M.D.P.). This work also benefited from the support of the INCD computing infrastructure funded by FCT and FEDER under the project 01/SAICT/2016 nº 022153.

This work was also supported by NIH/NIAID R37 Award Ai39115-23 and R01 award AI50113-16. J.H. is co-director and fellow for CIFAR program Fungal Kingdom: Threats & Opportunities.

References

1. Lee SC, Ni M, Li W, Shertz C, Heitman J. 2010. The Evolution of Sex: a Perspective from the Fungal Kingdom. Microbiol Mol Biol Rev 74:298–340.
2. Stanton BC, Hull CM. 2007. Mating-Type Locus Control of Cell Identity, p. 59–73. *In* Heitman, J, Kronstad, JW, Taylor, JW, Casselton, LA (eds.), Sex in Fungi: Molecular Determination and Evolutionary Implications. ASM Press, Washington, DC.
3. Lin X, Heitman J. 2007. Mechanisms of Homothallism in Fungi and Transitions between Heterothallism and Homothallism, p. 35–57. *In* Heitman, J, Kronstad, JW, Taylor, JW, Casselton, LA (eds.), Sex in Fungi: Molecular Determination and Evolutionary Implications. ASM Press, Washington, D.C.
4. Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal Sex: The Basidiomycota. Microbiol Spectr 5:FUNK-0046-2016.
5. Fraser JA, Hsueh Y-P, Findley KM, Heitman J. 2007. Evolution of the Mating-Type Locus: The Basidiomycetes, p. 19–34. *In* Heitman, J, Kronstad, JW, Taylor, JW, Casselton, LA (eds.), Sex in Fungi: Molecular Determination and Evolutionary Implications. ASM Press, Washington, DC.
6. Raudaskoski M, Kothe E. 2010. Basidiomycete Mating Type Genes and Pheromone Signaling. Eukaryot Cell 9:847–859.
7. Kües U, James TY, Heitman J. 2011. Mating Type in Basidiomycetes: Unipolar, Bipolar, and Tetrapolar Patterns of Sexuality, p. 97–160. *In* Pöggeler, S, Wöstemeyer, J (eds.), The Mycota, Evolution of Fungi and Fungal-Like Organisms, VolumeXIV. Springer, Berlin Heidelberg.

617 8. Wilson AM, Wilken PM, van der Nest MA, Steenkamp ET, Wingfield MJ, Wingfield BD.
618 2015. Homothallism: an umbrella term for describing diverse sexual behaviours. IMA
619 Fungus 6:207–214.
620 9. Lin X, Hull CM, Heitman J. 2005. Sexual reproduction between partners of the same
621 mating type in *Cryptococcus neoformans*. Nature 434:1017–1021.
622 10. Wang L, Lin X. 2011. Mechanisms of unisexual mating in *Cryptococcus neoformans*.
623 Fungal Genet Biol 48:651-660.
624 11. David-Palma M, Sampaio JP, Gonçalves P. 2016. Genetic Dissection of Sexual
625 Reproduction in a Primary Homothallic Basidiomycete. PLoS Genet 12: e1006110.
626 12. Hull CM, Boily M, Heitman J. 2005. Sex-Specific Homeodomain Proteins Sxi1α and
627 Sxi2a Coordinately Regulate Sexual Development in *Cryptococcus neoformans*.
628 Eukaryot Cell 4:526–535.
629 13. Gyawali R, Zhao Y, Lin J, Fan Y, Xu X, Upadhyay S, Lin X. 2017. Pheromone independent
630 unisexual development in *Cryptococcus neoformans*. PLOS Genet 13:e1006772.
631 14. Barredo J, García-Estrada C, Kosalkova K, Barreiro C. 2017. Biosynthesis of Astaxanthin
632 as a Main Carotenoid in the Heterobasidiomycetous Yeast *Xanthophyllomyces*
633 *dendrorhous*. J Fungi 3:44.
634 15. Kurtzman CP, Fell JW, Boekhout T. 2011. The Yeasts, a Taxonomic Study, 5th ed.
635 Elsevier.
636 16. Sampaio JP. 2011. *Cystofilobasidium* Oberwinkler & Bandoni (1983), p. 1423–1432. *In*
637 Kurtzman, CP, Fell, JW, Boekhout, T (eds.), The Yeasts, a Taxonomic Study, 5th ed.
638 Elsevier.
639 17. Fell JW, Johnson EA, Scorzetti G. 2011. *Xanthophyllomyces* Golubev (1995), p. 1595–
640 1599. *In* Kurtzman, CP, Fell, JW, Boekhout, T (eds.), The Yeasts, a Taxonomic Study,
641 5th ed. Elsevier.
642 18. Fell JW. 2011. *Mrakia* Y. Yamada & Komagata (1987), p. 1503–1510. *In* Kurtzman, CP,
643 Fell, JW, Boekhout, T (eds.), The Yeasts, a Taxonomic Study, 5th ed. Elsevier.
644 19. Sampaio JP. 2011. *Tausonia* Bab'eva (1998), p. 1999–2001. *In* Kurtzman, CP, Fell, JW,
645 Boekhout, T (eds.), The Yeasts, a Taxonomic Study, 5th ed. Elsevier.
646 20. Takashima M, Nakase T. 2011. *Udeniomyces* Nakase & Takematsu (1992), p. 2063–
647 2068. *In* Kurtzman, CP, Fell, JW, Boekhout, T (eds.), The Yeasts, a Taxonomic Study,
648 5th ed. Elsevier.

649 21. David-Palma M, Libkind D, Brito PH, Silva M, Bellora N, Coelho MA, Heitman J,
650 Gonçalves P, Sampaio JP. 2020. The Untapped Australasian Diversity of Astaxanthin-
651 Producing Yeasts with Biotechnological Potential - *Phaffia australis* sp. nov. and
652 *Phaffia tasmanica* sp. nov. Microorganisms (submitted).
653 22. Kachalkin A V, Turchetti B, Inácio J, Carvalho C, Mašínová T, Pontes A, Röhl O,
654 Glushakova AM, Akulov A, Baldrian P, Begerow D, Buzzini P, Sampaio JP, Yurkov AM.
655 2019. Rare and undersampled dimorphic basidiomycetes. Mycol Prog 18:945–971.
656 23. Pontes A, Röhl O, Carvalho C, Maldonado C, Yurkov AM, Paulo Sampaio J. 2016.
657 *Cystofilobasidium intermedium* sp. nov. and *Cystofilobasidium alribaticum* f.a. sp.
658 nov., isolated from Mediterranean forest soils. Int J Syst Evol Microbiol 66:1058–1062.
659 24. Bellora N, Moliné M, David-Palma M, Coelho MA, Hittinger CT, Sampaio JP, Gonçalves
660 P, Libkind D. 2016. Comparative genomics provides new insights into the diversity,
661 physiology, and sexuality of the only industrially exploited tremellomycete: *Phaffia*
662 *rhodozyma*. BMC Genomics 17:901.
663 25. Kües U. 2015. From two to many: Multiple mating types in Basidiomycetes. Fungal
664 Biol Rev 29:126–166.
665 26. Kämper J, Reichmann M, Romeis T, Bölker M, Kahmann R. 1995. Multiallelic
666 Recognition: Nonself-Dependent Dimerization of the bE and bW Homeodomain
667 Proteins in *Ustilago maydis*. Cell 81:73–83.
668 27. Maia TM, Lopes ST, Almeida JMGCF, Rosa LH, Sampaio JP, Gonçalves P, Coelho MA.
669 2015. Evolution of Mating Systems in Basidiomycetes and the Genetic Architecture
670 Underlying Mating-Type Determination in the Yeast *Leucosporidium scottii*. Genetics
671 201:75–89.
672 28. Asante-Owusu RN, Banham AH, Böhnert HU, Mellor EJC, Casselton LA. 1996.
673 Heterodimerization between two classes of homeodomain proteins in the mushroom
674 *Coprinus cinereus* brings together potential DNA-binding and activation domains.
675 Gene 172:25–31.
676 29. Schlesinger R, Kahmann R, Kämper J. 1997. The homeodomains of the heterodimeric
677 bE and bW proteins of *Ustilago maydis* are both critical for function. Mol Gen Genet
678 254:514–519.
679 30. Van Diepen LTA, Olson A, Ihrmark K, Stenlid J, James TY. 2013. Extensive trans-specific
680 polymorphism at the mating type locus of the root decay fungus *Heterobasidion*. Mol

681 Biol Evol 30:2286–2301.
682 31. Kües U, Göttgens B, Stratmann R, Richardson W V, O'Shea SF, Casselton L a. 1994. A
683 chimeric homeodomain protein causes self-compatibility and constitutive sexual
684 development in the mushroom *Coprinus cinereus*. EMBO J 13:4054–4059.
685 32. Hsueh Y-P, Xue C, Heitman J. 2009. A constitutively active GPCR governs morphogenic
686 transitions in *Cryptococcus neoformans*. EMBO J 28:1220–1233.
687 33. Olesnicky NS, Brown AJ, Dowell SJ, Casselton LA. 1999. A constitutively active G-
688 protein-coupled receptor causes mating self-compatibility in the mushroom *Coprinus*.
689 EMBO J 18:2756–2763.
690 34. Boone C, Davis NG, Sprague GF. 1993. Mutations that alter the third cytoplasmic loop
691 of the a-factor receptor lead to a constitutive and hypersensitive phenotype. Proc Natl
692 Acad Sci 90:9921–9925.
693 35. Fowler TJ, Mitton MF, Vaillancourt LJ, Raper CA. 2001. Changes in Mate Recognition
694 Through Alterations of Pheromones and Receptors in the Multisexual Mushroom
695 Fungus *Schizophyllum commune*. Genetics 158:1491–1503.
696 36. Riquelme M, Challen MP, Casselton LA, Brown AJ. 2005. The Origin of Multiple *B*
697 Mating Specificities in *Coprinus cinereus*. Genetics 170:1105–1119.
698 37. Díaz-Valderrama JR, Aime MC. 2016. The cacao pathogen *Moniliophthora roreri*
699 (Marasmiaceae) possesses biallelic *A* and *B* mating loci but reproduces clonally.
700 Heredity 116:491–501.
701 38. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
702 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler G,
703 Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its
704 applications to single-cell sequencing. J Comput Biol 19:455–477.
705 39. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome
706 assembly evaluation with QUAST-LG. Bioinformatics 34:i142–i150.
707 40. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically
708 mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics 24:637–
709 644.
710 41. Tusnády GE, Simon I. 2001. The HMMTOP transmembrane topology prediction server.
711 Bioinformatics 17:849–850.
712 42. Mitchell A, Chang H, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C,

713 McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong
714 S, Bateman A, Punta M, Attwood T, Sigrist C, Redaschi N, Rivoire C, Xenarios I, Kahn D,
715 Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale D, Wu C, Orengo
716 C, Sillitoe I, Mi H, Thomas P, Finn R. 2015. The InterPro protein families database: the
717 classification resource after 15 years. Nucleic Acids Res 43:D213-221.
718 43. Lin J rong, Hu J. 2013. SeqNLS: nuclear localization signal prediction based on frequent
719 pattern mining and linear motif scoring. PLoS One 8:e76864.
720 44. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary
721 structure prediction server. Nucleic Acids Res 43:W389–W394.
722 45. Li L, Stoeckert Jr CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for
723 eukaryotic genomes. Genome Res 13:2178–2189.
724 46. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software
725 package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol
726 79:7696–7701.
727 47. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
728 Improvements in Performance and Usability. Mol Biol Evol 30:772–780.
729 48. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a
730 new software for selection of phylogenetic informative regions from multiple
731 sequence alignments. BMC Evol Biol 10:210.
732 49. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
733 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol
734 32:268–274.
735 50. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
736 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods
737 14:587–589.
738 51. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2:
739 Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522.
740 52. Gonçalves P, Valério E, Correia C, de Almeida JMGCF, Sampaio JP. 2011. Evidence for
741 divergent evolution of growth temperature preference in sympatric *Saccharomyces*
742 species. PLoS One 6:e20739.
743 53. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated
744 alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–

745 1973.
746 54. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New
747 algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the
748 performance of PhyML 3.0. Syst Biol 59:307–321.
749 55. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
750 developments. Nucleic Acids Res 47:W256–W259.
751 56. Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, German
752 Grehov OG, Kandrov D, Rasputin K, Syabro M, Tleukenov T. 2012. Unipro UGENE: A
753 unified bioinformatics toolkit. Bioinformatics 28:1166–1167.
754 57. Sharma R, Gassel S, Steiger S, Xia X, Bauer R, Sandmann G, Thines M. 2015. The
755 genome of the basal agaricomycete *Xanthophyllomyces dendrorhous* provides insights
756 into the organization of its acetyl-CoA derived pathways and the evolution of
757 Agaricomycotina. BMC Genomics 16:233.
758 58. Verdoes JC, Krubasik P, Sandmann G, Van Ooyen AJJ. 1999. Isolation and functional
759 characterisation of a novel type of carotenoid biosynthetic gene from
760 *Xanthophyllomyces dendrorhous*. Mol Gen Genet 262:453-461.
761 59. Niklitschek M, Alcaíno J, Barahona S, Sepúlveda D, Lozano C, Carmona M, Marcoleta
762 A, Martínez C, Lodato P, Baeza M, Cifuentes V. 2008. Genomic organization of the
763 structural genes controlling the astaxanthin biosynthesis pathway of
764 *Xanthophyllomyces dendrorhous*. Biol Res 41:93-108.
765 60. Sambrook J, Russell D. 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold
766 Spring Harbor Laboratory Press, New York.
767 61. Visser H, Sandmann G, Verdoes JC. 2005. Xanthophylls in Fungi. Metabolic Engineering
768 of the Astaxanthin Biosynthetic Pathway in *Xanthophyllomyces dendrorhous.*
769 Microbial Processes and Products.
770 62. Rodríguez-Sáiz M, Godio RP, Álvarez V, De La Fuente JL, Martín JF, Barredo JL. 2009.
771 The NADP-dependent glutamate dehydrogenase gene from the astaxanthin producer
772 *Xanthophyllomyces dendrorhous*: Use of its promoter for controlled gene expression.
773 Mol Biotechnol 41:165-172.
774 63. Wery J, Verdoes JC, Van Ooyen AJJ. 1998. Efficient transformation of the astaxanthin-
775 producing yeast *Phaffia rhodozyma*. Biotechnol Tech 12:399-405.
776 64. Lin G, Bultman J, Johnson EA, Fell JW. 2012. Genetic manipulation of

Xanthophyllomyces dendrorhous and *Phaffia rhodozyma*. Methods Mol Biol 898:235–249.

65. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402–408.

Supplementary Figures

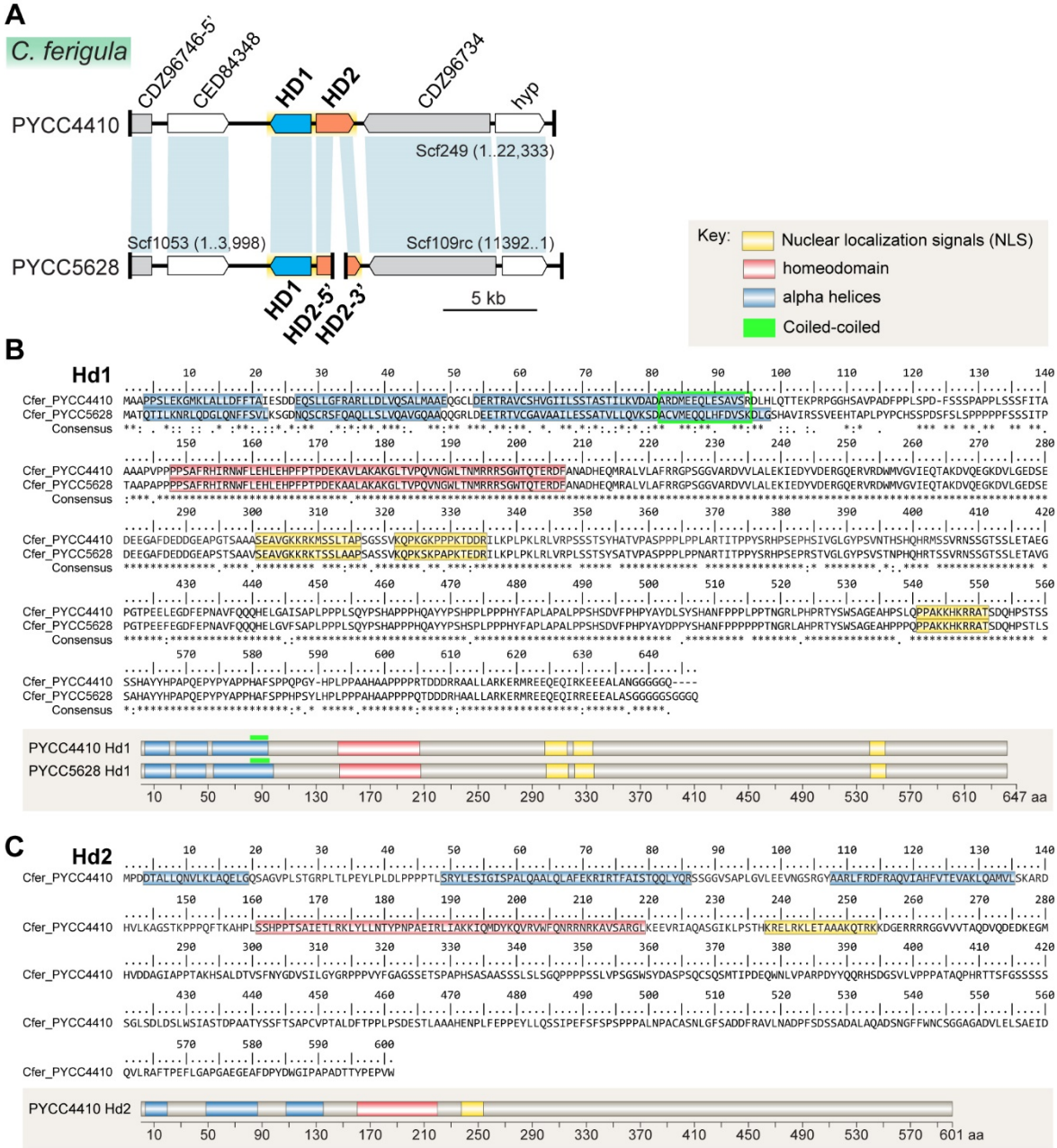


Fig. S1. Analysis of the *HD* locus of *C. ferigula*. (A) Genomic organization of the *HD1* and *HD2* genes in *C. ferigula* PYCC 4410 and PYCC 5628. The *HD2* gene in strain PYCC 5628 is fragmented and localizes, as depicted, at the ends of two different scaffolds. However, comparative analysis indicates this region is syntenic in both *C. ferigula* strains. (B) Hd1 amino acid sequences of strains PYCC 4410 and PYCC 5628 shown as a sequence alignment (on the top) and as a schematic representation (on the bottom). (C) Hd2 amino acid sequence of *C. ferigula* PYCC 4410. In panels B and C, typical protein secondary structure features are highlighted according to the key on the top right.

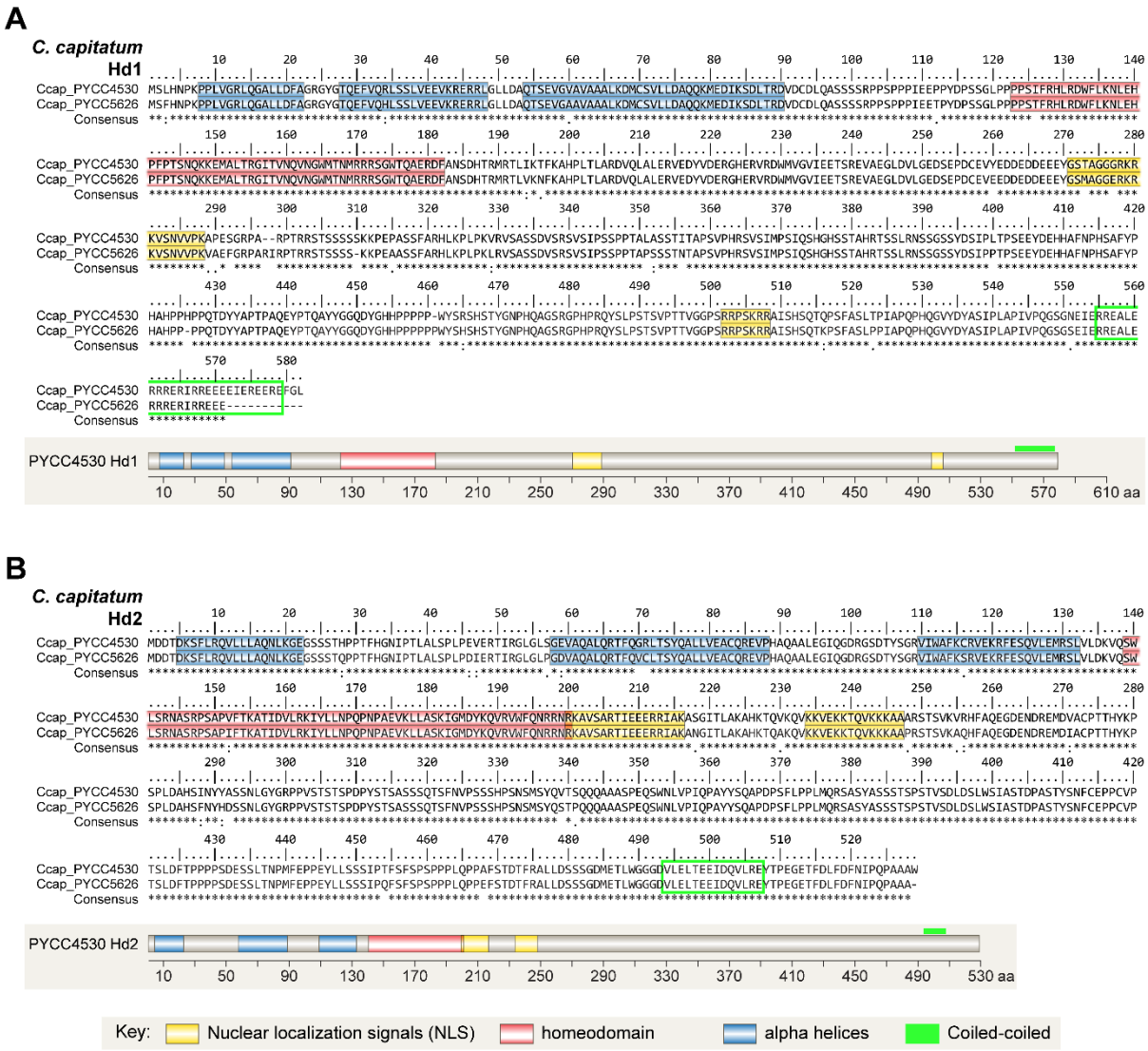


Fig. S2. Sequence alignment of the (A) *HD1* and (B) *HD2* gene products from *C. capitatum* PYCC 4530 and PYCC 5626. In both panels, the amino acid sequence alignment is shown on the top and a schematic representation is shown below. Typical protein secondary structure features are highlighted according to the key on the bottom.

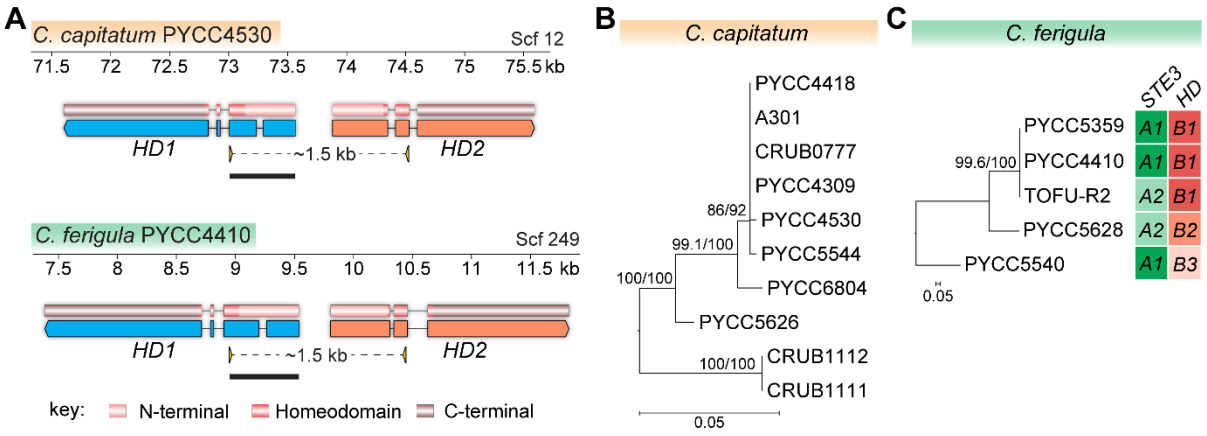


Fig. S3. Determination of different alleles of the *HD* loci in *C. capitatum* and *C. ferigula*. (A) A ~1.5-kb-long genomic region spanning the homeodomain, the 5' end region, and the common intergenic region of the *HD1* and *HD2* genes was PCR-amplified and sequenced from available 10 strains of *C. capitatum* and 5 strains of *C. ferigula*. Primers locations are shown as yellow arrowheads below the genes. (B) and (C) Maximum likelihood phylogenies inferred from the 5' end of the *HD1* gene (region underlined in panel A) of *C. capitatum* and *C. ferigula* strains using the best-fit substitution models HKY+F and K2P, respectively. Note that the branch length of the two trees is quite different; the divergence of the *C. ferigula HD1* alleles is much higher than that observed in *C. capitatum*. The trees are rooted in the midpoint and branch support values separated by a slash were assessed by 10,000 replicates of both Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) and the ultrafast bootstrap approximation (UFBoot). The molecular mating type assigned for each of the analyzed strains of *C. ferigula* provides evidence that this species has a biallelic *P/R* locus and a multiallelic *HD* locus, which are genetically unlinked (the *B1* allele appears associated with both *A1* and *A2* alleles).

***C. ferigula* PYCC5628 *HD1* synthetic gene sequence**

ATGGCAACCCAAACTATTCTCAAAAACCGACTTCAGGACGGGCTGCAGAATTTCTTCAGTGTACTAAAAAGCGGTGAC
AACCAGTCATGTCGAAGTTTCCAGGCACAACTTCTTTCACTAGTCCAGGCTGTTGGACAAGCAGCACAGCAGGGGCG
GTTGGATGAGGAAACCAGAACGGTATGTGGTGCTGTAGCAGCCATACTTGAATCGAGCGCCACTGTCCTTCTTCAGG
TTAAGTCTGACGCCTGTGTTATGGAACAGCAGCTGCATTTTGATGTTAGCAAAGATCTAGGCTCGCATGCCGTGATAAG
GTCGTCCGTGGAAGAACACACGGCGCCATTGCCCTACCCTTGCCACTCCTCGCCTGACTCGTTTAGTTTAAGCCCAC
CGCCCCCGCCATTCTCCTCTAGCATCACACCAACGGCGGCACCCGCTCCGCCTCCACCATCTGCTTTTCGCCACATC
AGAAACTGGTTTTTGGAGCATCTGGAACACCCGTTCCCTACACCTGACGAAAAAGCCGCGTTAGCCAAAGCAAAAGG
TTTAACTGTTCCACAAGTTAACGGTTGGCTAACAAATATGAGGCGCAGGTCGGGTTGGACTCAAACAGAAAGGGATTT
CGCCAACGCAGATCATGAACAAATGAGAGCGCTCGTGCTAGCCTTTCGGAGAGGCCCTTCCGGTGGCGTAGCCAGA
GATGTCGTGTTGGCCCTAGAAAAAATAGAAGACTATGTGGATGAAAGGGGACAAGAAAGAGTCAGAGATTGGATGGTA
GGTGTGATCGAGCAAACCGCTAAAGACGTACAAGAAGGAAAGGACGTTCTGGGTGAGGATTCCGAGGATGAAGAAG
GAGCATTTGATGAGGATGACGGAGAAGCGCCCAGTACGAGTGCGGCCGTCAGCGAAGCTGTGGGGAAGAAACGTAA
AACAAGTTCTTTGGCTGCACCATCCGCCAGCTCAGTCAAACAACCTAAATCTAAACCAGCGCCAAAGACTGAAGATCG
TATTCTGAAGCCGTTGCCGAAGTTGAGGCTCGTAAGGCCTTTGTCATCAACTAGTTACTCTGCAACTGTTCCCGCCAG
TCCGCCACCATTGCCCCCAAATGCGAGGACCATTACCCCGCCCTACAGTCGTCACCCAAGTGAGCCTAGAAGTACTG
TTGGCTTGGGGTACCCATCTGTTTCAACTAATCCTCATCAACATAGAACCTCTTCTGTCAGAAATTCATCAGGAACATCA
TCACTAGAAACCGCCGTTGGACCGGGAACTCCGGAAGAATTCGAAGGCGACTTCGAGCCAAATGCTGTTTTTCAACA
ACAACATGAGTTAGGCGTTTTTTCAGCTCCCTTACCGCCTCCTTTGTCCCAATACCCCTCTCACGCCCCCCCCCCCCA
TCAAGCATATTACCCTTCCCACTCACCATTGCCTCCCCCCCACTACTTTGCTCCTTTGGCTCCCGCATTACCACCTTCC
CATTCCGATGTTTTTCCTCACCCATACGCATATGATCCACCTTATTCCCACGCTAATTTTCCACCACCTCCTCCACCAAC
TAATGGCCGTTTAGCACACCCTCGTACATATTCTTGGTCAGCTGGTGAAGCTCATCCTCCACCACAACCACCAGCTAA
GAAGCATAAGAGAAGAGCTACATCTGATCAACATCCTTCTACATTATCATCTGCACATGCTTATTATCATCCAGCACCAC
AAGAACCTTATCCATATGCTCCACCTCATGCTTTTTCTCCTCCTCATCCATCTTATTTACATCCATTACCTCCACCTGCTC
ATGCTGCACCACCACCACCACAAACAGATGATGATAGACATGCAGCTTTATTAGCTAGAAAGGAAAGAATGAGAGAAG
AGCAAGAACAAATTAGAAGAGAGGAGGAAGCTTTAGCTTCTGGTGGTGGTGGTGGTTCTGGTGGTGGTCAATAA

***C. ferigula* PYCC4410 *HD2* synthetic gene sequence**

ATGCCGGATGATACTGCTCTTCTGCAGAATGTTTTGAAGCTTGCACAGGAGTTAGGTCAGTCTGCTGGAGTCCCGCTC
AGTACAGGAAGACCACTTACCTTACCGGAGTACCTCCCGCTGGATCTCCCACCTCCCCCCACCTTGTCGAGATATCTA
GAGTCGATTGGGATTTCACCGGCTTTGCAAGCAGCTCTGCAACTTGCATTCGAAAAAAGAATCCGAACATTTGCCATAT
CAACTCAACAATTGTATCAGCGATCGTCTGGGGGAGTGAGTGCACCATTGGGTGTCTTGGAGGAGGTCAATGGGTCT
AGGGGCTACGCGGCTAGATTGTTCCGTGACTTCCGTGCTCAAGTGATAGCCCACTTTGTAACCGAAGTTGCCAAATTG
CAAGCCATGGTTTTAAGCAAGGCAAGGGACCATGTACTGAAGGCTGGATCCACTAAACCGCCTCCCCAGTTCACTAAA
GCTCACCCATTGAGCAGTCATCCTCCCACTAGCGCTATAGAAACACTGCGGAAGCTGTACTTGTTAAATACTTACCCTA
ACCCTGCAGAAATACGTTTGATAGCTAAAAAAATCCAGATGGATTACAAACAGGTGAGGGTCTGGTTTCAAAACCGTAG
GAACAGAAAAGCTGTCTCCGCACGCGGACTAAAGGAAGAAGTAAGGATCGCACAGGCTTCCGGCATCAAATTACCGT
CCACGCACAAGCGTGAGTTGAGAAAACTTGAGACAGCCGCTGCAAAGCAAACTAGGAAGAAGGATGGCGAGAGAAG
ACGCAGAGGAGGCGTAGTCGTCACAGCTCAGGATGTACAAGATGAAGACAAAGAAGGCATGCATGTGGACGATGCC
GGCATCGCTCCTCCGACAGCCAAACATAGCGCTCTAGACACGGTATCCTTCAATTATGGGGATGTGTCAATTCTTGGTT
ATGGCAGACCACCTCCCGTTTATTTCGGAGCGGGAAGTTCAGAGACATCGCCAGCACCCCACAGTGCTAGTGCTGCT
AGCTCCTCTCTATCGTTGTCTGGTCAACCCCCTCCCCCATCATCCTTAGTGCCCAGTGGTTCTTGGAGCTATGATGCTT
CCCCCTCCCAATGTTCTCAATCGATGACTATTCCAGATGAACAATGGAATCTTGTTCCAGCTAGACCAGACTACTATCAA
CAAAGACATTCTGACGGTTCTGTTTTAGTTCCACCTCCAGCAACGGCACAACCTCATAGAACGACCTCATTCGGTAGC
AGCTCAAGTTCATCAGGTCTTTCGGATCTAGATTCCTTGTGGTCAATTGCAAGTACCGATCCAGCGGCGACTTACTCCT
CTTTCACATCAGCTCCTTGTGTTCCTACTGCTTTAGACTTTACCCCTCCTTTGCCTTCTGATGAAAGTACTTTGGCGGC
TGCACATGAAAACCCTCTATTCGAACCACCAGAATATTTATTACAATCTTCAATTCCTGAATTCAGTTTTTCTCCATCTCC
TCCACCAGCCCTAAATCCAGCATGTGCCTCTAATCTAGGTTTTTCTGCAGATGATTTTAGAGCCGTTTTAAACGCCGAT
CCTTTTTCAGACTCTTCTGCCGACGCGTTAGCGCAAGCTGACTCTAACGGTTTTTTTTGGAATTGCTCAGGTGGTGCG
GGTGCAGACGTTTTAGAATTATCAGCCGAAATTGATCAAGTTTTAAGAGCATTTACACCAGAATTTTTAGGTGCTCCAG
GTGCCGAAGGTGAAGCATTTGATCCATATGATTGGGGTATTCCAGCCCCAGCAGATACAACCTATCCAGAACCAGTTT
GGTAA

Fig. S4. DNA sequences of the designed synthetic *HD1* and *HD2* genes of *C. ferigula* PYCC 5628 and PYCC 4410, respectively. Sequences correspond to the coding DNA sequences of the *HD1* and *HD2* genes of PYCC 5628 and PYCC 4410, respectively, optimized for the codon usage of *S. cerevisiae*.

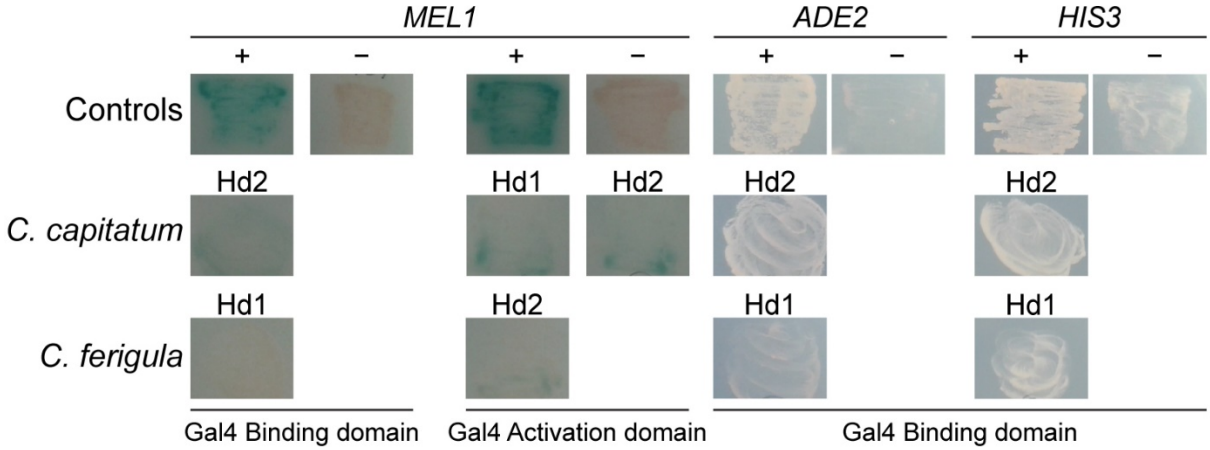


Fig. S5 Activation of the reporter genes in the Yeast Two-Hybrid Assay by the individual fusion proteins. This control consists in assessing activation of reporter genes in haploid transformants carrying each of the fusion proteins to be tested. Absence of this so-called autoactivation shows that activation of reporter genes in diploid strains relies on the presence of both interacting partners.