Multiple hybridization events punctuate the evolutionary trajectory of *Malassezia furfur*

Bart Theelen[a,§], Verónica Mixão[b,c,*,§], Giuseppe Ianiri[d], Joleen Goh Pei Zhen[e], Jan Dijksterhuis[a], Joseph Heitman[f], Thomas L. Dawson Jr.[e,g], Toni Gabaldón[b,c,h,#,+], Teun Boekhout[a,i,#,+]

[a] Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands

[b] Life Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain

[c] Mechanisms of Disease Programme, Institute for Research in Biomedicine (IRB), Barcelona, Spain

[d] Department of Agricultural, Environmental and Food Sciences, University of Molise, 86100 Campobasso, Italy

[e] Skin Research Institute of Singapore (SRIS), Agency for Science, Technology, and Research (A∗STAR), Singapore

[f] Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, United States of America

[g] Center for Cell Death, Injury and Regeneration, Departments of Drug Discovery and Biomedical Sciences and Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, South Carolina, USA

[h] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

[i] Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, The Netherlands

Running head: Hybridization in *Malassezia furfur*

#Address correspondence to Teun Boekhout (t.boekhout@wi.knaw.nl) and Toni Gabaldón (toni.gabaldon@bsc.es)

[*] Present address: Bioinformatics Unit, Infectious Diseases Department, National Institute of Health Dr. Ricardo Jorge, Av. Padre Cruz, 1649-016, Lisbon, Portugal

[§] Bart Theelen and Verónica Mixão contributed equally to this work. Author order was determined by alphabetical order.

[+] Toni Gabaldón and Teun Boekhout share senior authorship and correspondence.

Abstract: 209 words

Importance: 149 words

Text: 6823 words

Abstract

Malassezia species are important fungal skin commensals and are part of the normal microbiota of humans and other animals. However, under certain circumstances these fungi can also display a pathogenic behaviour. For example, *Malassezia furfur* is a common commensal of human skin, and yet is often responsible for skin disorders but also systemic infections. Comparative genomics analysis of *M. furfur* revealed that some isolates have a hybrid origin, similar to several other recently described hybrid fungal pathogens. Because hybrid species exhibit genomic plasticity that can impact phenotypes, we sought to elucidate the genomic evolution and phenotypic characteristics of *M. furfur* hybrids in comparison to their parental lineages. To this end, we performed a comparative genomics analysis between hybrid strains and their presumptive parental lineages, and assessed phenotypic characteristics. Our results provide evidence that at least two distinct hybridization events occurred between the same parental lineages, and that the parental strains may have originally been hybrids themselves. Analysis of the mating-type locus reveals that *M. furfur* has a pseudobipolar mating system, and provides evidence that hybridization involved sexual liaisons resulting in cell-cell fusion leading to a diploid/aneuploid state. This study provides new insights into the evolutionary trajectory of *M. furfur* and contributes valuable genomic resources for future pathogenicity studies.

Importance

Malassezia furfur is a common commensal member of human/animal microbiota that is also associated with several pathogenic states. Recent studies report involvement of

Malassezia species in Crohn's disease, a type of inflammatory bowel disease, pancreatic cancer progression, and exacerbation of Cystic fibrosis. A recent genomics analysis of *M. furfur* revealed the existence of hybrid isolates and identified their putative parental lineages. In this study, we explored the genomic and phenotypic features of these hybrids in comparison to their putative parental lineages. Our results revealed the existence of a pseudobipolar mating system in this species and showed evidence for the occurrence of multiple hybridization events in the evolutionary trajectory of *M. furfur*. These findings significantly advance our understanding of the evolution of this commensal microbe and are relevant for future studies exploring the role of hybridization in the adaptation to new niches or environments, including the emergence of pathogenicity.

Introduction (573 words)

Malassezia are basidiomycetous yeasts that are the dominant fungal component of the healthy human skin microbiome (1, 2). Nevertheless, they can also take on pathogenic roles in various skin disorders, and have been implicated in cases of infection associated with several comorbidities (3–12). Non-culture based sequencing methods revealed *Malassezia* presence in other ecological niches, including insects, nematodes, corals, sponges, deep-sea environments, and soils (13–15), suggesting that *Malassezia* are ecologically diverse. Interestingly, *Malassezia* are evolutionarily related to fungal plant pathogens (16, 17), potentially pointing towards an early host shift from plants to animals, facilitated by the loss of genes coding for proteins involved in the degradation of complex carbohydrates, and expansion of lipid hydrolases required to break down lipids available on animal skin (16, 17).

91 In *Basidiomycetes*, species with either a tetrapolar or bipolar mating system have been
92 described (18, 19). In a tetrapolar mating system, the pheromone and pheromone
93 receptor locus (P/R) and the homeodomain locus (HD) are unlinked, whereas in a
94 bipolar mating system both loci are linked and contiguous on the same chromosome
95 which precludes recombination. An intermediate -pseudobipolar- mating system was
96 described in three *Malassezia* species, namely *Malassezia globosa, Malassezia*
97 *sympodialis*, and *Malassezia yamatoensis*. In this configuration, the P/R and HD
98 mating-type loci are linked on the same chromosome but far enough apart from each
99 other that recombination can still occur (16, 20, 21). The presence of mating and
100 meiotic genes suggests that *Malassezia* are potentially capable of sexual reproduction,
101 but mating remains hitherto unobserved, and the genus is known to propagate
102 asexually through unipolar budding (16, 21, 22).
103
104 Our study focuses on *Malassezia furfur*, a species isolated from a wide variety of
105 hosts, from humans (with healthy skin, skin disorders, or bloodstream infections), to a
106 range of domestic and zoo animals (12, 15, 23–26). This species presents variable cell
107 sizes and shapes and different ploidies and karyotypes (27–29). Indeed, previous
108 studies reported two different karyotypes -one displaying more chromosomes (10-11
109 vs. 7-8) and a larger genome size (~14 Mbp vs. 8.5 Mbp) than the other- and a high
110 degree of genetic variation that uncovered a possible hybrid genotype (28, 30, 31).
111 These findings were later corroborated in a genomic analysis of members of the
112 *Malassezia* genus. This analysis revealed four *M. furfur* strains (CBS1878, CBS4172,
113 CBS7019, CBS7710) with double genome size due to gene duplication that possibly
114 originated from a hybridization event between members of the lineages of strains
115 CBS7982 and CBS14141 (syn. JPLK23)(16).

116

117 The coexistence of genomic material of two diverged lineages in a single cell is often
118 associated with hybrids' lack of viability. Nevertheless, several genomic mechanisms,
119 such as recombination, often lead to loss of heterozygosity (LOH), thus erasing some
120 of the genomic incompatibilities (32). In such a scenario, organisms carrying highly
121 plastic genomes and unique phenotypes may survive leading to new lineages able to
122 thrive in new environments (32–34). For instance, findings of multiple pathogenic
123 hybrids, such as *Candida albicans* or the *Cryptococcus gattii* / *Cryptococcus*
124 *neoformans* species complex, has led to the hypothesis that hybridization plays an
125 important role in the emergence of pathogenicity (30, 35–42).

126

127 Taking advantage of the existence of known candidate parental haploid lineages, this
128 study explored genomic and some general phenotypic features of *M. furfur* hybrids to
129 assess the impact of hybridization on their evolutionary path. This study reveals that
130 *M. furfur* has a pseudobipolar mating system and shows evidence that multiple
131 hybridization events punctuated the evolution of this species.

132

133 **Results and Discussion (4155 words)**

134 ***M. furfur* genomic patterns reveal the current existence of two hybrid lineages**

135 Previous AFLP and comparative genomics analyses suggested the existence of *M.*
136 *furfur* hybrid strains (16, 31). In this study, we further explored the genomic patterns
137 of these hybrids and of 17 additional strains presently classified as *M. furfur* (Table 1)
138 to assess how widespread hybridization is within this species. Following the original
139 observations of a hybrid genotype (31), we performed AFLP with two different
140 adaptor and primer set combinations for 22 *M. furfur* strains (Supplementary Table 1

141 and Materials and Methods for more details). These AFLP analyses confirmed that
142 the banding patterns of three previously identified hybrids (CBS1878, CBS4172,
143 CBS7019) are indeed a combination of those of the proposed parental lineages
144 (Figure 1) (16). Nevertheless, the hybrid strain CBS7019 did not cluster with the
145 other two, suggesting the existence of two putative hybrid clades (Figure 1). In total,
146 the AFLP clustering analysis identified four lineages: P1 (parental 1), which
147 corresponds to the parental CBS7982 lineage with five additional strains; P2, which
148 corresponds to the parental CBS14141 lineage and four additional strains; H1 (hybrid
149 1), which corresponds to CBS1878, CBS4172, and five additional hybrid strains; and
150 finally, H2, which harbors CBS7019 and three additional hybrid strains (Figure 1).
151
152 **[Figure 1]**
153
154 [Table 1]
155
156 To confirm the existence of two independent hybridization events, 13 strains (Table
157 1) comprising representative isolates of all four AFLP-determined lineages were
158 further compared in detail at the genomic level. Hybrid sequencing reads were
159 mapped to a combined reference including the P1 (CBS9595: 8.2 Mb, 8 scaffolds)
160 and P2 (CBS14141: 8.3 Mb, 9 scaffolds) genome assemblies (Supplementary file 1,
161 Table 2, Materials and Methods for more details) to identify the source and sequence
162 of the different sub-genomes of the hybrids (Figure 2a).
163
164 [Table 2]
165

166 AFLP-determined H1 and H2 lineages on average presented 6 SNPs/kb
167 (Supplementary Table 5), the majority of them (>85%) corresponding to P2 for all
168 hybrid strains, thus suggesting that CBS9595 is possibly closer to the actual parent 1
169 than CBS14141 is to the actual parent 2. Noteworthy, these SNPs were not
170 homogeneously distributed along the genome, but rather formed blocks of
171 homozygous variants (see Materials and Methods for more details on block
172 definition). Therefore, based on the assumption that strains originating from the same
173 hybridization event would share the same blocks of variants, we utilized these as
174 signatures of the hybrids' evolutionary past and compared them between the different
175 hybrids to confirm the number of hybridization events. Jaccard metrics revealed a
176 high block overlap between the strains identified by AFLP as H1 (>84.6%), and the
177 strains identified by AFLP as H2 (>91.8%), thus confirming the results of the AFLP
178 analysis. It is important to highlight that among the H1 strains, CBS4172 presented
179 the lowest similarity with its peers. While this strain revealed 84.6% and 84.9% block
180 overlap with CBS1878 and CBS9365, respectively, the latter two strains had an
181 overlap of 99%. The block overlap in strains of hybrid lineage H2 was more
182 homogeneous. The overlap of the high polymorphic regions between strains of H1
183 and H2 varied between 20.8% and 23.1%, thus suggesting independent origins.
184 Together, these results confirm that the hybrid lineages resulted from two independent
185 hybridization events between P1 and P2. These findings suggest an apparent
186 propensity of P1 and P2 to hybridize, and for their hybrids to survive.
187
188 [Figure 2]
189

190 **Genomic divergence and different mitochondria uncover the hybrid origin of P1**
191 **and P2**
192 Considering the possible ancestry of the highly polymorphic blocks in the two hybrid
193 lineages, we sought to determine their presence in the parental lineages. To this end,
194 the same methodology was followed as for the hybrids (see Materials and Methods),
195 and we verified that indeed, the strains of both P1 and P2 that were not used for
196 genome assembly also harbor highly polymorphic genomic regions. A comparison
197 between these blocks and the ones observed in the hybrid strains showed that the
198 highest overlap between H1 strains and P1 and P2 is 14.6% and 21%, respectively.
199 Similarly, comparing the regions of high variability in the hybrid lineage H2 with
200 what was detected in P1 and P2 revealed that the highest overlaps were 64% and
201 13%, respectively. This is in line with our previous observations in the hybrid
202 genomes and suggests that none of the strains sequenced for the P1 or P2 lineages
203 represents the direct parental strain of the hybrids, but rather their close relatives.
204
205 The origin of these blocks of high genomic variability was unknown, but as reported
206 by Mixão and Gabaldón (35) for *C. albicans*, two possible models could explain their
207 existence: i) continuous admixture between different strains, and ii) a hybrid ancestor
208 that experienced genomic recombination. A key aspect to distinguish these models is
209 the divergence between the reference genome and the observed sequence, which we
210 expect to vary between the different blocks in the first model, and to be similar among
211 them in the second one (35). Therefore, we next assessed the level of divergence
212 between these blocks and the reference genome in both lineages. For P2 strains, we
213 identified a clear single peak of sequence divergence in both CBS8735 and PM315
214 (Figure 2b), suggesting that the genomic material of all the blocks was acquired at a

215 single time-point, which in turns supports the existence of a hybrid ancestor. Our
216 estimations point to a current haplotype divergence between the strains of parental
217 lineage P2 of 4% to 4.5% (Supplementary Table 5). To determine if any sequenced
218 strain could be the alternative parent of P2 lineage, we performed a BLASTn search
219 in the NCBI genome database. The only hit obtained was the CBS14141 (the same
220 strain we use as the P2 reference) genome with 95.2% sequence similarity, a value
221 consistent with our estimations. This means that the alternative parent of the P2
222 lineage has not been sequenced thus far.
223
224 For the P1 lineage, this analysis was complicated by the limited number of
225 polymorphic blocks (min 78 - max 463 blocks), and their short size (average = 245
226 bp), because a variation in a single polymorphism can have a large impact on the
227 estimation of the sequence divergence. Indeed, despite the observed overlap between
228 the peaks of sequence divergence in the different P1 strains, we also detected multiple
229 sequence divergence peaks (Figure 2c). This did not allow exclusion of any of the
230 proposed models, and properly estimate the sequence divergence. Even so, we
231 performed a BLASTn search in the NCBI genome database using the longest blocks
232 (>500 bp), and we verified that their best hits always presented a sequence similarity
233 of ~93% with the CBS14141 genome (P2 lineage), thus suggesting that P1 also has a
234 hybrid origin.
235
236 We next focused on comparing the mitochondrial genome sequences of the different
237 *M. furfur* lineages. Genomic reads of each parental strain were mapped against
238 mitochondrial genomes of the respective strains. As reference mitochondria for the P2
239 lineages, we selected the CBS14141 mitochondrial sequence available in NCBI

240 (accession number: KY911086.1). As expected, all P2 strains shared the same
241 mitochondrial genome. In the P1 lineage, the publicly available mitochondrial
242 genome of CBS7982 (NCBI accession number: KY911085.1) was used as a reference
243 for mapping sequencing reads from P1 strains. Mapping results revealed the presence
244 of two different mitochondrial sequences in the P1 lineage with the mitochondrial
245 sequence of CBS7982 differing from that of CBS9595, CD866, and CBS9574
246 (Supplementary Figure 1). To compare both P1 mitochondrial types, a draft
247 mitochondrial genome assembly with 49kb was generated for CBS9595 (see
248 Materials and Methods). A BLASTn search with this mitochondrial genome against
249 the NCBI nr database revealed that its sequence is equally distant to that of the
250 mitochondrial genomes of CBS7982 (P1 lineage) and CBS14141 (P2 lineage), with a
251 similarity of 93.45%. These results are in line with our observations for the
252 polymorphic regions of the nuclear genome, further providing support that P1 results
253 from the cross of two diverged lineages, i.e. hybridization, and CBS7982 harbors the
254 mitochondrial genome of the alternative parent. A recent study assessing three
255 mitochondrial genomic loci of 43 *M. furfur* strains, found two rather divergent
256 mitochondrial clades for samples belonging to the P1 lineage (43), supporting our
257 results and suggesting a wider presence of two different mitochondria among P1
258 strains. When reads of the strains belonging to both hybrid lineages were mapped
259 against the mitochondrial reference genomes, it was observed that all hybrid strains
260 inherited the mitochondria from the P2 lineage.
261
262 Together, these analyses suggest that both parental lineages are not genetically "pure"
263 but are genomic mosaics likely resulting from a hybridization event between two
264 unknown lineages that are approximately 7% divergent in the case of P1, and 4%

265 divergent in the case of P2 (Figure 3). The observed small differences in the genetic
266 mosaics detected in the nuclear genomes in each of the parental lineages, suggest that
267 strains in each lineage (P1 and P2) diverged before some of the recombination events
268 occurred, a scenario which implies a non-haploid state of the strains at the time of
269 their respective divergence. Nevertheless, all analyses performed in this study support
270 a haploid state for all of them (Supplementary Figure 4,Figure 5), indicating
271 independent ploidy reduction.
272
273 [Figure 3]
274
275 ***M. furfur* hybrid lineages are mostly diploid with some LOH**
276 Karyotype data from this study and previous publications (44, 45) demonstrate
277 significant chromosomal variation in *M. furfur* with hybrid strains containing
278 additional chromosomal bands when compared to strains from their hypothesized
279 haploid parental lineages (Supplementary Figure 4a). Moreover, FACS ploidy
280 analysis for H1 strains displayed a DNA content between 1n and 2n suggesting
281 aneuploidy, whereas the DNA-content of both analyzed H2 strains is consistent with
282 assignment as diploids (Supplementary Figure 4b). Based on whole-genome
283 sequencing data analysis, the analyzed strains for both hybrid lineages generally seem
284 of diploid nature (Supplementary Figures 2 and 3), with a few exceptions, such as a
285 putative triplication of a chromosome from P2 in the hybrid strain CBS9365 (H1,
286 Supplementary Figure 2).
287
288 Considering the non-haploid state of these hybrid strains, we decided to look for LOH
289 events, an important feature to restore genomic stability through the erasure of one of

290 the haplotypes of the recombining region (32). In the case of the *M. furfur* hybrids
291 analyzed in this study, we estimated that LOH covers approximately 20% of the
292 genome (Supplementary Table 5). This is a low value when compared to other fungal
293 hybrids, such as *Candida* hybrids, where LOH has been estimated to cover at least
294 50% of the genome (35–38, 42). Considering that sequence divergence between the
295 parental lineages is higher in *M. furfur* (10%) than in *Candida* hybrids (~4%) (35–38,
296 42), we hypothesize that the lower occurrence of LOH in *M. furfur* may be related to
297 a lower number of potential recombining sites. Indeed, a recent study on *C.*
298 *neoformans* x *C. gattii* hybrids, which present 7% sequence divergence, has shown
299 that they experience fewer recombination events (39).
300
301 The direction of the LOH event, i.e. which allele is retained, varies from hybrid to
302 hybrid, or even from niche to niche according to the most advantageous phenotypes
303 (46). Therefore, it is important to understand the genomic alterations that the
304 coexistence of the parental sub-genomes may cause in the hybrid. While in some
305 hybrids there seems to be a tendency to retain the allele of a given parental lineage, in
306 others this appears to be a random process (32, 38, 42, 47, 48). In *M. furfur* H1 and
307 H2 hybrid lineages we found that 60% of the LOH tends to the allele of P2
308 (Supplementary Table 5). Although this result revealed a slight tendency to retain the
309 allele of P2 parental, we consider that it is not as robust and evident as in other
310 hybrids where >80% of the genome retains the allele of the same parent (47, 48).
311 Previous studies have shown that incompatibilities between the mitochondrial and the
312 nuclear genome may be one of the major sources of incompatibilities in hybrids (49,
313 50). Nevertheless, a previous study on *Saccharomyces* hybrids revealed that the
314 overall direction of LOH is not correlated with the direction of mitochondrial

315 inheritance (51). The fact that all *M. furfur* hybrid strains retained the mitochondria of
316 P2, is in accordance with the observed direction of LOH of the nuclear genome.
317 However, we are not certain if the direction of LOH and this observation have a
318 causal relation.
319

320 ***M. furfur* possesses the genetic machinery of a pseudo-bipolar mating system**

321 To understand the origin of the hybridization events leading to the H1 and H2
322 lineages, analysis of the mating-type genes of *M. furfur* was carried out (see
323 Supplementary file 2) for a detailed description of the results). The *MAT* genes of *M.*
324 *furfur* haploid strains were identified based on similarity searches with the *MAT* genes
325 of two opposite mating-types of *M. sympodialis* (strain CBS12506, *MAT a1b1*; strain
326 ATCC44340, *MAT a2b2*). Two *MAT a* and two *MAT b* loci were identified in the
327 haploid strains of *M. furfur*. The *MAT a* and *b* loci of the *M. furfur* haploid strains
328 analyzed are ~ 590 kb apart (Supplementary Table 6), suggesting a pseudo-bipolar
329 configuration. A representative comparison of the *MAT* loci of CBS14139 (*MAT*
330 *a1b1*) and CBS7982 (*MAT a2b2*), as model strains, is shown in Figure 4. The *M.*
331 *furfur MAT* structure reflects that of the *MAT* locus of *M. yamatoensis (16)* and differs
332 from the *M. globosa* and *M. sympodialis MAT* loci, for which *MAT a* and *MAT b* are
333 ~167 kb and ~140 kb apart, respectively (20, 22). This is consistent with a whole-
334 genome based clustering that groups *M. furfur* and *M. yamatoensis* in a separate
335 phylogenetic cluster within the *Malassezia* genus (16). In all *M. furfur* haploid strains
336 analyzed, the two *MAT a* locus genes are divergently oriented, corroborating previous
337 findings for other *Malassezia* species (16, 20, 22). Moreover, a comparison between
338 the *MAT a1* and *a2* loci revealed that the *Mfa* and *Pra* genes have an opposite
339 arrangement albeit being located in highly syntenic flanking regions (Figure 4). In

340 *Basidiomycetes*, the tetrapolar mating system is thought to be ancestral, and transition
341 from a tetrapolar to bipolar system has been linked to the evolution of pathogenicity
342 (52, 53). For example in the genus *Cryptococcus*, pathogenic species have a bipolar
343 mating system, whereas closely related non-pathogenic species such as *Cryptococcus*
344 *amylolentus* have a tetrapolar mating system (54). Based on findings for the red yeast
345 *Sporobolomyces salmonicolor* (cited as *Sporidiobolus salmonicolor*), the pseudo-
346 bipolar mating system was proposed to be a gradual stage in the transition from a
347 tetrapolar to bipolar system (53).
348
349 [Figure 4]
350
351 Hybrid genome searches with the *MAT* genes of haploid *M. furfur* identified two *MAT*
352 *a* loci and two additional *MAT b* loci in the *M. furfur* hybrid strains CBS1878 and
353 CBS7019. The sexual identity of the *b* loci among strains from all lineages was
354 assigned following sequence comparison and phylogenetic analysis of the predicted
355 proteins, resulting in *MAT b1*, *b2*, *b3*, and *b4* alleles, with *MAT b1* clustering with *b3*,
356 and *MAT b2* clustering with *b4* (Figure 5C). In particular, based on genome data,
357 strain CBS1878 was designated as *MAT a2a2b1b4,* strain CBS4172 as *MAT a2b1b4*,
358 and strain CBS7019 as *MAT a1a2b3b4*. Interestingly, CBS1878 did not contain *MAT*
359 *a1* but the *a2* locus was present in two copies, suggesting a duplication of this region
360 or LOH from an ancestral *a1a2* state to a derived *a2a2* configuration. In addition,
361 CBS4172 did not possess the *MAT a1* locus and only possessed one copy of the *MAT*
362 *a2* locus.
363

364 **A molecular assay for *M. furfur* mating type identifies mating-compatible strains**
365 **in the parental lineages**
366 Following these initial findings for the *M. furfur* mating-type regions, a PCR assay
367 was developed with primers that specifically amplify *MAT a1*, *MAT a2*, *MAT b1-b3*,
368 and *MAT b2-b4*. PCR results for the *MAT a1* and *a2*, and sequencing analysis of *MAT*
369 *b1-b3* and *b2-b4*, indicated additional mating-type configurations (Figure 5a-c and
370 Supplementary Table 7). Both P1 and P2 parental lineages contain strains with either
371 *MAT a1* or *MAT a2*, while all P1 strains are of the *MAT b2* type and all P2 strains are
372 of the *MAT b1* type (Figure 5a-c and Supplementary Table 7). The presence of *a1b1,*
373 *a2b1, a1b2* and *a2b2 MAT* combinations supports the hypothesis that recombination
374 occurs within this region, corroborating findings on the pseudo-bipolar *MAT* structure
375 of *M. sympodialis* and in contrast with the lack of recombination reported for *Ustilago*
376 *hordei*, which is bipolar (20). Our findings also highlight incompatibility at the B loci
377 within each parental lineage, despite compatibility being present at the *MAT a* locus,
378 which might explain why sexual reproduction could not be observed under laboratory
379 conditions. Attempts to cross compatible *MAT a* and *MAT b* strains belonging to the
380 P1 and P2 parental lineages were also carried out, but also in this case sexual
381 reproduction could not be observed. It seems likely that two individual mating events
382 between representatives of both parental lineages originally resulted in hybrid strains
383 that contained the combination *MAT a1a2b1b2*, yet our data showed that all H1-
384 strains lost the *MAT a1* copy and in some instances may have duplicated the *MAT a2*
385 allele (only confirmed for CBS1878, whereas CBS4172 only has one *MAT a2* copy).
386 Strains belonging to the H2 hybrid lineage still have both parental *MAT a* copies, with
387 the exception of CBS6093, that seems to have lost the parental *MAT a2* copy. In the
388 hybrids, two unique *MAT b* arrangements were observed: *MAT b4* in both hybrid

389 lineages, with similarity to *MAT b2* of the P1 strains; and *MAT b3* which is only
390 present in H2 strains, and is a phylogenetic sister of *MAT b1* of the P2 strains (Figure
391 5a-c and Supplementary Table 7). Strains of hybrid lineage H1 retained the *MAT b1*
392 copy. Considering that recombination in the *MAT* locus has previously been observed
393 in other hybrid lineages (38, 55, 56), and is associated with a possible restoration of
394 hybrid fertility (55, 56), we hypothesize that strains in lineages H1 and H2 may have
395 undergone genomic changes leading to re-establishment of a viable sexual state.
396
397 [Figure 5]
398
399
400
401 **Targeted sequencing of five nuclear loci and genomic data suggest P1 and P2**
402 **lineages may be two separate species**
403 The species *M. furfur* is represented by two neotype cultures, namely CBS1878 and
404 CBS7019, corresponding to the respective names *Malassezia furfur* and *Pityrosporum*
405 *ovale* (synonym of *M. furfur*), but these belong to the hybrid lineages H1 and H2,
406 respectively. Many species are described based on a limited number of nuclear DNA
407 loci, ribosomal loci ITS and LSU being two of the most frequently used taxonomic
408 markers in fungi, but these loci may not reflect the genetic heterogeneity of a yeast
409 strain sufficiently as is also shown in this study for the ITS of ribosomal DNA and
410 three protein coding genes (β-tubulin, chitin synthase (*CHS2*), and translation
411 elongation factor 1-α (EF1- α)) (Supplementary Figure 5). Phylogenetic analysis of
412 published sequencing data for the intergenic transcribed spacer (IGS) of the rDNA
413 resulted in separate clusters for each of the studied lineages, making it a potential

414 diagnostic tool for identifying hybrids belonging to H1 or H2 among genetically
415 uncharacterized *M. furfur* isolates (Supplementary Figure 5). As chromatograms of
416 hybrid strains for the protein coding genes possessed multiple sites with two different
417 nucleotide peaks representing both parental backgrounds, they were phased into two
418 sequences representing their respective parental copies (Supplementary Figure 5).
419 Based on ITS data, both hybrid lineages H1 and H2 cannot be distinguished from the
420 parental lineage P2, yet for the most part they contain genetic material from both P1
421 and P2. Type strains or neotype strains are reference strains, often used as
422 representatives for a species in specific functional assays, but as in this case they
423 represent hybrid lineages with a combined genomic content of two haploid parental
424 lineages, they may not be the best choice to serve as reference strains and should
425 probably be disqualified to serve as neotypes. At present, all strains considered in this
426 study are classified as *M. furfur*. However, based on sequence divergence between the
427 P1 and P2 lineages for the five analyzed nuclear genomic loci it is likely that P1 and
428 P2 may represent two closely related species. For example, ITS sequence similarity
429 between P1 strain CBS9595 and the P2 strain CBS14141 is 98% but their similarity
430 for protein coding genes is much lower: 97.3% for EF1- α, 96.4% for β-tubulin, and
431 87.6 % for *CHS2* respectively. According to a study that assessed more than 9 000
432 yeast isolates to establish species and genus thresholds for ribosomal ITS and LSU
433 regions, a species-threshold of 98.41% was presented for ITS (57), which supports the
434 assignment of these lineages into two species. Based on genome sequence alignments
435 for CBS9595 (P1) and CBS14141 (P2), an overall sequence similarity of 89 % was
436 estimated, further supporting that these strains represent two species.
437

438 **Hybrid lineages H1 and H2 show different morphologies but limited**
439 **differentiation based on traditional physiological properties**
440 At the phenotypic level, cells belonging to hybrid lineages differ in size and shape
441 from parental lineage cells (Figure 6, Supplementary Table 8), with hybrid cells being
442 thinner and more elongated. However, growth profiling experiments traditionally
443 used for *Malassezia* species identification, seem not to differ significantly between
444 hybrid and parental lineages (Supplementary Table 9). One noteworthy observation is
445 that all P2 strains were positive for β-glucosidase activity and strains from the other
446 lineages were all negative, except for one P1 strain (CBS9589) and for one H2-strain,
447 CBS6093. This latter strain also seems somewhat different from other H2-strains
448 based on mating type, AFLP-pattern, and Matrix-Assisted Laser Desorption
449 Ionization-time Of Flight (MALDI-TOF) mass spectrum (Supplementary Figure 6)
450 and would be an interesting candidate for further genomic exploration. Variable
451 results for Cremophor EL utilization and β-glucosidase activity have been described
452 previously for *M. furfur* isolates (12, 15, 23–26) and may be an expression of the
453 heterogeneity of the species. Analysis of mass spectra generated with a Bruker
454 MALDI Biotyper also illustrated the differences between P1 and P2 lineages and
455 formed separate dendrogram clusters for H1, H2, P1, and P2, with a few exceptions
456 (Supplementary Figure 6). The putative H2 strain CBS6093 clustered basal to clusters
457 for P2 and H2, supporting above mentioned deviating findings for that strain. In
458 addition, the MALDI-TOF mass spectrum of H1-strain CBS4172 clusters with the P2
459 strains. This divergent character is in line with a deviating PFGE chromosomal
460 banding pattern and this strain has lost *MAT* a1, and only has one copy of *MAT* a2.
461 These findings need to be repeated but confirm the current high heterogeneity present
462 in the species *M. furfur* that might be better interpreted as a species complex.

463

464 Of note, as has also previously been mentioned (30, 58), deep-seated isolates (e.g.
465 blood, urine, rectal swabs) seem to almost exclusively belong to the genotype
466 representing the P2 lineage. Based on AFLP, mating type analysis, and Sanger
467 sequencing data, the H1 hybrid lineage consists of seven strains, five originating from
468 animal skin and two from diseased human skin. Hybrid lineage H2 consists of four
469 strains, all isolated from human diseased skin, with the exception of CBS6093 for
470 which the origin is unknown. Leong and colleagues explored antifungal susceptibility
471 patterns of 26 *M. furfur* strains, including some strains also considered in this study.
472 Four hybrid strains (H1: CBS1878, CBS7019; H2:CBS6000, CBS600) showed
473 reduced susceptibility to certain azoles. This feature was not exclusive for strains
474 from hybrid lineages H1 and H2 but rather seemed linked to disease state
475 backgrounds of strains, and reduced azole susceptibility also included disease isolates
476 belonging to parental lineage P2 (59). Whether hybridization at the basis of any of the
477 lineages H1, H2, or P2 may have facilitated genetic changes driving this reduced
478 azole susceptibility, or whether this was the result from mere exposure to these drugs,
479 remains to be elucidated. Considering that five out of the seven H1-strains are derived
480 from animals, we hypothesize that the hybridization event for H1 may have facilitated
481 a host-shift event between humans and animals. This hypothesis should be tested in
482 future analyses including a larger sampling of strains.

483

484 [Figure 6]

485

486 In summary, this study identified two individual hybrid lineages H1 and H2, with P1
487 and P2 representing their parental lineages, although not the exact parental strains.

488 We propose that the diploid hybrid lineages H1 and H2 are the result of two separate
489 mating events between mating-compatible strains from the P1 and P2 groups.
490 Interestingly, our genome analysis shows that both haploid parental lineages were
491 themselves the result of prior hybridization events but became haploid before some of
492 their members mated again to form hybrid lineages H1 and H2. The various
493 hybridization events and subsequent further evolutionary changes have contributed to
494 additional genetic diversification in *M. furfur*, but also at the phenotypic level, the
495 hybrid strains differ from their parental lineages. The implications of possible
496 hybridization-driven changes in pathogenicity, or adaptation to new environments,
497 will drive further analysis and in depth examination comparing pathogenicity factors,
498 lipid production and utilization, and salient physiological and phenotypic facets.
499

500 **Materials and Methods (2095 words)**

501 **DNA extraction**

502 To extract DNA for whole-genome sequencing, cells were grown on modified Dixon
503 agar (mDA) (60) for 48 h at 30°C and harvested into 50 ml tubes. Yeast cells were
504 lysed, using the QIAGEN Genomic DNA Purification procedure for yeast samples
505 (Qiagen, Hilden, Germany), with minor modifications. Lyticase incubation was
506 performed for 2 h at 30°C, and RNase/Proteinase incubation was performed for 2 h at
507 55°C. Genomic DNA was purified using Genomic-tip 100/G prep columns, according
508 to the manufacturer's handbook. For AFLP, MLST, and the mating type assay, gDNA
509 extraction was performed following the CTAB method as described by (61) with the
510 following modification: DNA was purified in two steps- first with phenol-chloroform
511 and secondly with chloroform only.
512

513 **AFLP**

514 AFLP analysis was performed according to (62) with some modifications. Various
515 restriction enzymes and adaptor pairs were combined with multiple primer
516 combinations (Supplementary Table 1). A combined restriction ligation reaction was
517 performed for 2 h at 37°C. PCR products were diluted 200x after purification and
518 then combined with Orange600 size standard (Nimagen, Nijmegen, the Netherlands)
519 before fragment analysis on a 3730xl DNA Analyzer (Thermo Fisher Scientific,
520 Waltham, Massachusetts, USA). Data was imported and analyzed using Bionumerics
521 v.7.6.3 (Applied Maths, Sint-Martens-Latem, Belgium), and dendrograms were
522 created with UPGMA clustering and Pearson correlation coefficient. After
523 purification of PCR products, all reactions were diluted 200x.

524

525 **Whole-genome sequencing**

526 Thirteen samples were selected for whole-genome sequencing data analysis (Table 1).
527 For three of them (CBS7982, CBS14141 and CBS4172), we retrieved Illumina data
528 from SRA (Table 1) (16). For the remaining ones, whole-genome sequencing was
529 performed at the Genomics Unit from the Centre for Genomic Regulation (group 1:
530 CBS8735, PM315, CBS1878 and CBS7019) and at the Genome Institute Singapore
531 (group 2: CBS9595, CD866, CBS9574, CBS6000 and CBS6001). Except when
532 specified, the protocol was similar in both groups of strains. Libraries were prepared
533 using the NEBNext Ultra DNA Library Prep kit for Illumina (New England BioLabs,
534 United States) according to manufacturer's instructions. All reagents subsequently
535 mentioned are from the NEBNext Ultra DNA Library Prep kit for Illumina, if not
536 specified otherwise. 1 µg of gDNA was fragmented by ultrasonic acoustic energy in
537 Covaris to a size of ~600 bp in group 1 and ~300-400 bp in group 2. After shearing,

538 the ends of the DNA fragments were blunted with the End Prep Enzyme Mix, and
539 then NEBNext Adaptors for Illumina were ligated using the Blunt/TA Ligase Master
540 Mix. The adaptor-ligated DNA was cleaned-up using the MinElute PCR Purification
541 kit (Qiagen, Germany) and a further size selection step was performed using an
542 agarose gel. Size-selected DNA was then purified using the QIAgen Gel Extraction
543 Kit with MinElute columns (Qiagen) and library amplification was performed by
544 PCR with the NEBNext Q5 Hot Start 2X PCR Master Mix and index primers (12–15
545 cycles in group 1, 6 cycles in group 2). A purification step was done using AMPure
546 XP Beads (Agentcourt, United States). The final library was analyzed using Agilent
547 DNA 1000 chip (Agilent) to estimate the quantity and check size distribution, and it
548 was then quantified by qPCR using the KAPA Library Quantification Kit
549 (KapaBiosystems, United States) prior to amplification with Illumina's cBot.
550 Libraries were loaded and sequenced 2 x 125 on Illumina's HiSeq2500 for group 1
551 and 2 x 101 on Illumina's HiSeq2000 for group 2. Base-calling was performed using
552 Illumina pipeline software. De-convolution was performed using the CASAVA
553 software (Illumina, United States).
554
555 Four samples (CBS7982, CBS9595, CBS14141 and CBS1878) were additionally
556 sequenced with PacBio long-read sequencing strategy. In addition, previously
557 generated PacBio sequencing data was used for mating type analysis only, for five
558 samples (CBS14139, CBS8735, PM315, CBS4172, CBS9369 and CBS7019). All
559 samples were sequenced on the PacBio RSII platform (Pacific Biosciences, United
560 States), and, except for CBS9595, libraries were prepared in 2015 with the DNA
561 Template Prep Kit 3.0, and polymerase/template complexes were subsequently
562 formed using Polymerase Binding Kit P6 v2, and then sequenced with Sequencing

563 Reagent Kit 4.0. Sample CBS9595 was sequenced more recently, with the following
564 specifications: the library was prepared using the SMRTbell Template Prep Kit 1.0,
565 polymerase/template complexes were generated with DNA/Polymerase Binding Kit
566 P6 v2, and the sample was sequenced using DNA Sequencing Reagent Kit 4.0 v2
567 with 360 min runtime per SMRTcell.
568
569 ***De novo* genome assembly**
570 The genomes of the samples exclusively sequenced with PacBio (CBS14139,
571 CBS8735, PM315, CBS4172, CBS9369 and CBS7019) were assembled with HGAP3
572 within the SMRT portal of PacBio, SMRTanalysis v3.1, with standard settings. The
573 genomes of the four samples with short- and long-read sequencing libraries
574 (CBS7982, CBS9595, CBS14141 and CBS1878) were assembled with a pipeline that
575 combines short- and long-read assemblers. Briefly, Illumina reads were filtered and
576 trimmed with Trimmomatic v0.36 (63) and assembled with Platanus v1.2.4 (64).
577 PacBio reads were corrected with Canu (65) and assembled with DBG2OLC
578 (v20180222) (66) using Platanus assembly, MaSurCA v3.3.0 (67), and WTDBG2
579 v2.1 (68). Ragout v2.2 (69) was used for scaffolding using DBG2OLC, WTDBG2
580 and MaSurCA assemblies. Assembly correction was performed with Pilon v1.22 (70).
581 The assemblies' quality was assessed with Quast v4.5 (71) and K-mer Analysis
582 Toolkit v2.4.1 (KAT, (72)). The best assembly for each sample was chosen based on
583 N50, level of fragmentation and estimated assembly completeness by KAT (72).
584 Augustus Web-server (73, 74) was used for genome annotation, using *Malassezia*
585 *restricta* proteome as training set (accession number: GCA_003290485.1 (75)).
586 Predicted protein-coding genes completeness was assessed with BUSCO v4 using the
587 Basidiomycota database (76). Functional annotation was performed with eggNOG-

588 mapper web-server using the default settings (77).

589

590 **Read mapping and variant calling**

591 All paired-end Illumina libraries were inspected with FastQC v0.11.5

592 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed and filtered

593 with Trimmomatic v0.36 (63). Read mapping was performed with sppIDer pipeline

594 (78) using a combined reference including the genome assembly of CBS9595 (as

595 representative of P1 lineage, Supplementary File 1 for more details) and CBS14141

596 (as representative of P2 lineage, Supplementary File 1 for more details). To guarantee

597 the proper correspondence between the scaffolds of both parentals, we aligned both

598 genomes with the nucmer tool of MUMmer v3 (79). Average genome coverage for

599 each sample was estimated with Samtools v1.9 (80). Variant calling was performed

600 with HaploTypo v1.0.1 (81) selecting Freebayes v1.3.2 (82) as variant caller, and

601 using the default settings for the remaining parameters.

602

603 **Definition of blocks of high genomic variability**

604 We used the methodology developed and tested by (38) for LOH blocks definition.

605 Briefly, we used bedtools merge (83) with a distance of 100 bp to merge the

606 homozygous SNPs of each sample, and we set a minimum polymorphic region size of

607 100 bp. These blocks were compared among the different strains using bedtools

608 jaccard (83). The sequence divergence between the reference alleles and the allele of

609 the polymorphic regions was calculated by dividing the number of SNPs overlapping

610 such regions by the total number of base pairs covered by them.

611

612 **LOH blocks definition**

613 LOH block definition relied on read depth of coverage. Briefly, bedtools genomecov
614 (83) was used to determine the number of reads covering each position. Positions
615 covered by 0 reads were considered deleted, while those covered by 150% of the
616 average genome coverage were considered duplicated. Similarly to the procedure
617 developed by (38), we determined a minimum block size of 100 bp. Bedtools
618 intersect was used to determine the intersection of the deleted regions of P1 and the
619 duplicated regions of P2, and *vice-versa*. Only duplicated regions in one parent that
620 intersect a deleted region in the alternative one were considered as LOH. An
621 enrichment analysis of the genes overlapping LOH blocks was performed with
622 FatiGO (84).
623
624 **Mitochondrial genome assembly**
625 *De novo* genome assembly of CBS9595 mitochondrial genome was performed with
626 NOVOPlasty v2.7.2 (85), using CBS7982 *cox2* sequence as seed (accession number:
627 KY911085.1).
628
629 **Identification of *M. furfur* mating type region**
630 P/R (*MAT A*) and HD (*MAT B*) loci of *M. sympodialis* strains ATCC42132 (*MAT*
631 *a1b1*) and ATCC44340 (*MAT a2b2*) (20) were used as query for tBLASTn analysis
632 on the PacBio genomic assemblies of *M. furfur* haploid strains. The designation of *M.*
633 *furfur MAT A* loci was assigned following that of the closest *M. sympodialis* orthologs
634 based on the *E*-value of the tBLASTn outcome, whereas that of *MAT B* loci was
635 assigned according to phylogenetic clustering of the predicted concatenated HD
636 proteins (see below). In all cases, the *MAT* genes identified in *M. furfur* strains were
637 confirmed by reciprocal BLASTx on GenBank. The nomenclature of the *M. furfur*

638 *MAT* genes follows that of the closely-related Ustilaginomycotina: *mfa* is the
639 pheromone-encoding gene, *pra* is the pheromone receptor, *bE* and *bW* are the HD
640 transcription factors, followed by a number to distinguish from different alleles (18).
641
642 Open reading frames of the *MAT* genes were predicted by comparison with their
643 respective orthologs through BLASTx on GenBank, and using RNAseq data available
644 for *M. furfur* CBS14141 (Bioproject PRJNA741845 (59)). DNA sequences of the
645 *MAT* genes were aligned with MUSCLE (86) and their phylogenetic reconstruction
646 was performed with MEGA7 (87) using the maximum likelihood method [Tamura 3-
647 parameter model with Gamma distribution] and 100 bootstrap replications. Similarly,
648 translated HD proteins were predicted with ExPASy translate tool (88), and
649 concatenated bE-bW sequences were aligned with MUSCLE (86) and the respective
650 maximum likelihood tree [Jones-Tailor-Thornton (JTT), uniform rates] with 100
651 bootstrap replications was obtained with MEGA7 (87).
652
653 The identified genes of the *MAT a1, a2, b1,* and *b2* loci of the *M. furfur* haploid
654 strains were used as queries for BLASTn and tBLASTx analyses to identify the *MAT*
655 regions in the *M. furfur* hybrid strains CBS1878, CBS4172, and CBS7019. The *MAT*
656 *a1* and *MAT a2* designation followed that of the haploid strains used as input. For the
657 identified *MAT B* loci, the DNA sequences of *bE* and *bW* genes and their predicted
658 encoded proteins were aligned with MUSCLE (86) and then subjected to
659 phylogenetic analysis with MEGA7 (87) as described above (data not shown).
660
661 **Characterization of *M. furfur* mating type**

662 Primers for the amplification of *MAT a1, a2, b1,* and *b2* alleles were designed on the
663 basis of an alignment of all *MAT* loci of the available *M. furfur* parental and hybrid
664 strains derived from their genome sequences. The primers for *MAT A* loci specifically
665 amplified the *a1* or *a2* allele, while the primers for the *MAT B* loci amplified *b1* and
666 *b3,* or *b2* and *b4* (Supplementary Table 3). For the amplification of the *MAT A* loci,
667 ExTaq (Taqara) was used following the manufacturer's instructions with the following
668 PCR conditions: initial denaturation 2 min at 94˚C, followed by 33 cycles of 30 sec at
669 94˚C, 30 sec annealing at 58˚C, 1 min 15 sec extension at 72˚C, and a final extension
670 of 5 min at 72˚C. For the amplification of the *MAT b1-b3,* LATaq (Taqara)
671 supplemented with GC Buffer II was used following the manufacturer's instruction,
672 and PCR conditions were: initial denaturation 2 min at 94˚C, followed by 33 cycles of
673 30 sec at 94˚C, 30 sec annealing at 60˚C, 2 min extension at 72˚C, and a final
674 extension of 5 min at 72˚C. The *MAT b2-b4* alleles were amplified using the same
675 conditions as reported for *MAT b1-b3*, except for the use of LATaq (Taqara)
676 supplemented with GC Buffer I. *MAT B* loci are high in G+C (~65 %) and the use of
677 specific Taq polymerase was important for their successful amplification.
678
679 The amplified MAT b1-b3 and MaT b2-b4 alleles were then sequenced using the
680 primers used for amplification, and sequences were aligned using MEGAX with
681 MUSCLE and then subjected to phylogenetic analysis using the maximum likelihood
682 method [Tamura-Nei model] and 500 bootstrap replications (86, 89).
683
684 **Scanning Electron Microscopy**
685 Hybrid- and parent strains were cultivated on modified Dixon (mDixon) medium for
686 72 h and a loop of cells was suspended in water. The cells were briefly vortexed to

687 dislodge from each other. Droplets of 1, 2 and 3 µl were gently placed on mDixon
688 agar and dried for 1 h in a laminar flow cabinet to fix the cells onto the agar. After
689 pre-examination under a stereo microscope, small 4 x 4 mm selections with both
690 individual cells and cells grouped together were cut out using a surgical blade
691 (Swann-Morton, no. 11, Sheffield, UK) and glued on a copper sample cup with a
692 small droplet of frozen-tissue medium (KP-Cryoblock, Klinipath, Duiven, The
693 Netherlands) and subsequently snap-frozen in nitrogen slush, and transferred into an
694 Oxford CT1500 Cryostation connected to a JEOL 5600LV scanning electron
695 microscope (JEOL, Tokyo, Japan). Samples were sputter-coated (3 x 1 min) using a
696 gold target in the cryostation. Electron micrographs were taken at an acceleration
697 voltage of 5 kV.
698
699 **Data availability**
700 Sequencing data, genome assemblies and annotations are available at NCBI database
701 under the BioProject PRJNA732434.
702
703 **Acknowledgements**
704 The authors thank Bart Kraak for some exploratory PCR and microscopy work,
705 Simon Denil for helping with initial bioinformatics assessment of strain CBS1878,
706 Claudia Cafarchia for providing strain CD866 and Marina Marcet-Houben for all the
707 helpful discussions on the bioinformatics analyses. This work was supported by the
708 European Union's Horizon 2020 research and innovation programme under the Marie
709 Sklodowska-Curie grant agreement No H2020-MSCA-ITN-2014-642095. TG group
710 also acknowledges support from the Spanish Ministry of Economy, Industry, and
711 Competitiveness (MEIC) for the EMBL partnership, and grants 'Centro de Excelencia

712 Severo Ochoa 2013-2017' SEV-2012-0208, and BFU2015-67107 co-founded by
713 European Regional Development Fund (ERDF); from the CERCA Programme /
714 Generalitat de Catalunya; from the Catalan Research Agency (AGAUR) SGR857,
715 and grants from the European Union's Horizon 2020 research and innovation
716 programme under the grant agreement ERC-2016-724173. TG also receives support
717 from an INB Grant (PT17/0009/0023 - ISCIII-SGEFI/ERDF). G.I and J.H. were
718 supported by NIH/NIAID R37 award AI39115-24 and R01 award AI50113-16A1.
719 J.H. is fellow and co-director of the CIFAR program Fungal Kingdom: Threats and
720 Opportunities. TD was supported by the A*STAR Industry Alignment Fund
721 (H18/01/a0/016).

722 **References**

723 1. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D,
724 Nomicos E, Park M, NIH Intramural Sequencing Center Comparative
725 Sequencing Program, Kong HH, Segre JA. 2013. Topographic diversity of
726 fungal and bacterial communities in human skin. Nature 498:367–370.

727 2. Byrd AL, Belkaid Y, Segre JA. 2018. The human skin microbiome. Nat Rev
728 Microbiol 16:143–155.

729 3. Soret P, Vandenborght L-E, Francis F, Coron N, Enaud R, Avalos M,
730 Schaeverbeke T, Berger P, Fayon M, Thiebaut R, Delhaes L, Mucofong
731 Investigation Group. 2020. Respiratory mycobiome and suggestion of inter-
732 kingdom network during acute pulmonary exacerbation in cystic fibrosis. Sci
733 Rep 10:3589.

734 4. Limon JJ, Tang J, Li D, Wolf AJ, Michelsen KS, Funari V, Gargus M, Nguyen
735 C, Sharma P, Maymi VI, Iliev ID, Skalski JH, Brown J, Landers C, Borneman J,
736 Braun J, Targan SR, McGovern DPB, Underhill DM. 2019. Malassezia Is
737 Associated with Crohn's Disease and Exacerbates Colitis in Mouse Models. Cell
738 Host Microbe 25:377–388.e6.

739 5. Prohic A, Jovovic Sadikovic T, Krupalija-Fazlic M, Kuskunovic-Vlahovljak S.
740 2016. Malassezia species in healthy skin and in dermatological conditions. Int J
741 Dermatol 55:494–504.

742 6. LeibundGut-Landmann S, Dawson T. 2021. Malassezia: A Skin Commensal
743 Yeast Impacting Both Health And Disease. Frontiers Media SA.

744 7. Iatta R, Cafarchia C, Cuna T, Montagna O, Laforgia N, Gentile O, Rizzo A,
745 Boekhout T, Otranto D, Montagna MT. 2014. Bloodstream infections by
746 Malassezia and Candida species in critical care patients. Med Mycol 52:264–
747 269.

748 8. Ha CWY, Martin A, Sepich-Poore GD, Shi B, Wang Y, Gouin K, Humphrey G,
749 Sanders K, Ratnayake Y, Chan KSL, Hendrick G, Caldera JR, Arias C,
750 Moskowitz JE, Ho Sui SJ, Yang S, Underhill D, Brady MJ, Knott S, Kaihara K,
751 Steinbaugh MJ, Li H, McGovern DPB, Knight R, Fleshner P, Devkota S. 2020.
752 Translocation of Viable Gut Microbiota to Mesenteric Adipose Drives
753 Formation of Creeping Fat in Humans. Cell 183:666–683.e17.

754 9. Spatz M, Richard ML. 2020. Overview of the Potential Role of in Gut Health
755 and Disease. Front Cell Infect Microbiol 10:201.

756 10. Aykut B, Pushalkar S, Chen R, Li Q, Abengozar R, Kim JI, Shadaloey SA, Wu
757 D, Preiss P, Verma N, Guo Y, Saxena A, Vardhan M, Diskin B, Wang W,
758 Leinwand J, Kurz E, Kochen Rossi JA, Hundeyin M, Zambrinis C, Li X, Saxena
759 D, Miller G. 2019. The fungal mycobiome promotes pancreatic oncogenesis via
760 activation of MBL. Nature 574:264–267.

761 11. Saunte DML, Gaitanis G, Hay RJ. 2020. -Associated Skin Diseases, the Use of
762 Diagnostics and Treatment. Front Cell Infect Microbiol 10:112.

763 12. Rhimi W, Theelen B, Boekhout T, Otranto D, Cafarchia C. 2020. spp. Yeasts of
764 Emerging Concern in Fungemia. Front Cell Infect Microbiol 10:370.

765 13. Amend AS, Barshis DJ, Oliver TA. 2012. Coral-associated marine fungi form
766 novel lineages and heterogeneous assemblages. ISME J 6:1291–1301.

767 14. Amend A. 2014. From dandruff to deep-sea vents: Malassezia-like fungi are
768 ecologically hyper-diverse. PLoS Pathog 10:e1004277.

769 15. Theelen B, Cafarchia C, Gaitanis G, Bassukas ID, Boekhout T, Dawson TL Jr.
770 2018. Malassezia ecology, pathophysiology, and treatment. Med Mycol 56:S10–
771 S25.

772 16. Wu G, Zhao H, Li C, Rajapakse MP, Wong WC, Xu J, Saunders CW, Reeder
773 NL, Reilman RA, Scheynius A, Sun S, Billmyre BR, Li W, Averette AF,
774 Mieczkowski P, Heitman J, Theelen B, Schröder MS, De Sessions PF, Butler G,
775 Maurer-Stroh S, Boekhout T, Nagarajan N, Dawson TL Jr. 2015. Genus-Wide
776 Comparative Genomics of Malassezia Delineates Its Phylogeny, Physiology, and
777 Niche Adaptation on Human Skin. PLoS Genet 11:e1005614.

778 17. Wang Q-M, Theelen B, Groenewald M, Bai F-Y, Boekhout T. 2014.
779 Moniliellomycetes and Malasseziomycetes, two new classes in
780 Ustilaginomycotina. Persoonia 33:41–47.

781 18. Fraser JA, Heitman J. 2003. Fungal mating-type loci. Curr Biol 13:R792–5.

782 19. Raudaskoski M, Kothe E. 2010. Basidiomycete mating type genes and
783 pheromone signaling. Eukaryot Cell 9:847–859.

784 20. Gioti A, Nystedt B, Li W, Xu J, Andersson A, Averette AF, Münch K, Wang X,
785 Kappauf C, Kingsbury JM, Kraak B, Walker LA, Johansson HJ, Holm T, Lehtiö
786 J, Stajich JE, Mieczkowski P, Kahmann R, Kennell JC, Cardenas ME,
787 Lundeberg J, Saunders CW, Boekhout T, Dawson TL, Munro CA, de Groot
788 PWJ, Butler G, Heitman J, Scheynius A. 2013. Genomic insights into the atopic
789 eczema-associated skin commensal yeast Malassezia sympodialis. MBio

790 4:e00572–12.

791 21. Zhu Y, Engström PG, Tellgren-Roth C, Baudo CD, Kennell JC, Sun S, Billmyre
792 RB, Schröder MS, Andersson A, Holm T, Sigurgeirsson B, Wu G,
793 Sankaranarayanan SR, Siddharthan R, Sanyal K, Lundeberg J, Nystedt B,
794 Boekhout T, Dawson TL Jr, Heitman J, Scheynius A, Lehtiö J. 2017.
795 Proteogenomics produces comprehensive and highly accurate protein-coding
796 gene annotation in a complete genome assembly of Malassezia sympodialis.
797 Nucleic Acids Res 45:2629–2643.

798 22. Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, Kuramae EE, Kronstad JW,
799 Deangelis YM, Reeder NL, Johnstone KR, Leland M, Fieno AM, Begley WM,
800 Sun Y, Lacey MP, Chaudhary T, Keough T, Chu L, Sears R, Yuan B, Dawson
801 TL Jr. 2007. Dandruff-associated Malassezia genomes reveal convergent and
802 divergent virulence traits shared with plant and human fungal pathogens. Proc
803 Natl Acad Sci U S A 104:18730–18735.

804 23. Cafarchia C, Latrofa MS, Figueredo LA, da Silva Machado ML, Ferreiro L,
805 Guillot J, Boekhout T, Otranto D. 2011. Physiological and molecular
806 characterization of atypical lipid-dependent Malassezia yeasts from a dog with
807 skin lesions: adaptation to a new host? Med Mycol 49:365–374.

808 24. Zinkeviciene A, Norkunas V, Citavicius D. 2012. Atypical non-lipid-dependent
809 strains of Malassezia furfur. Open Life Sci 7:241–249.

810 25. González A, Sierra R, Cárdenas ME, Grajales A, Restrepo S, de García MCC,
811 Celis A. 2009. Physiological and Molecular Characterization of Atypical Isolates
812 of Malassezia furfur. Journal of Clinical Microbiology.

813 26. Puig L, Bragulat MR, Castellá G, Cabañes FJ. 2018. Phenotypic and genetic
814 diversity of Malassezia furfur from domestic and zoo animals. Med Mycol
815 56:941–949.

816 27. Mittag H. 1994. Fine structural investigation ofMalassezia furfur:I. Size and
817 shape of the yeast cells and a consideration of their ploidy. Mycoses.

818 28. Boekhout T, Kamp M, Guého E. 1998. Molecular typing of Malassezia species
819 with PFGE and RAPD. Med Mycol 36:365–372.

820 29. Guého E, Midgley G, Guillot J. 1996. The genus Malassezia with description of
821 four new species. Antonie Van Leeuwenhoek 69:337–355.

822 30. Theelen B, Silvestri M, Guého E, van Belkum A, Boekhout T. 2001.
823 Identification and typing of Malassezia yeasts using amplified fragment length
824 polymorphism (AFLP), random amplified polymorphic DNA (RAPD) and
825 denaturing gradient gel electrophoresis (DGGE). FEMS Yeast Res 1:79–86.

826 31. Theelen B, Batra R, Gupta A, Summerbell R, Boekhout T. 2004. Molecular
827 identification of Malassezia species at Spring Meeting, Dutch Association of
828 Medical Microbiology (Arnhem, The Netherlands).

829 32. Mixão V, Gabaldón T. 2018. Hybridization and emergence of virulence in
830 opportunistic human yeast pathogens. Yeast 35:5–20.

831 33. Gabaldón T. 2020. Hybridization and the origin of new yeast lineages. FEMS
832 Yeast Res 20.

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

835 35. Mixão V, Gabaldón T. 2020. Genomic evidence for a hybrid origin of the yeast
836 opportunistic pathogen Candida albicans. BMC Biol 18:48.

837 36. Mixão V, Saus E, Boekhout T, Gabaldón T. 2021. Extreme diversification driven
838 by parallel events of massive loss of heterozygosity in the hybrid lineage of
839 Candida albicans. Genetics 217.

840 37. Mixão V, Hansen AP, Saus E, Boekhout T, Lass-Florl C, Gabaldón T. 2019.
841 Whole-Genome Sequencing of the Opportunistic Yeast Pathogen Candida
842 inconspicua Uncovers Its Hybrid Origin. Frontiers in Genetics.

843 38. Pryszcz LP, Németh T, Saus E, Ksiezopolska E, Hegedűsová E, Nosek J, Wolfe
844 KH, Gacser A, Gabaldón T. 2015. The Genomic Aftermath of Hybridization in
845 the Opportunistic Pathogen Candida metapsilosis. PLoS Genet 11:e1005626.

846 39. Priest SJ, Coelho MA, Mixão V, Clancey SA, Xu Y, Sun S, Gabaldón T,
847 Heitman J. 2021. Factors enforcing the species boundary between the human
848 pathogens Cryptococcus neoformans and Cryptococcus deneoformans. PLoS
849 Genet 17:e1008871.

850 40. Samarasinghe H, Xu J. 2018. Hybrids and hybridization in the Cryptococcus
851 neoformans and Cryptococcus gattii species complexes. Infect Genet Evol
852 66:245–255.

853 41. Mixão V, Hegedűsová E, Saus E, Pryszcz LP, Cillingová A, Nosek J, Gabaldón
854 T. 2021. Genome analysis of Candida subhashii reveals its hybrid nature and
855 dual mitochondrial genome conformations. DNA Res 28.

856 42. Schröder MS, Martinez de San Vicente K, Prandini THR, Hammel S, Higgins

857 DG, Bagagli E, Wolfe KH, Butler G. 2016. Multiple Origins of the Pathogenic
858 Yeast Candida orthopsilosis by Separate Hybridizations between Two Parental
859 Species. PLoS Genet 12:e1006404.

860 43. Theelen B, Christinaki AC, Dawson TL, Boekhout T, Kouvelis VN. 2021.
861 Comparative analysis of Malassezia furfur mitogenomes and the development of
862 a mitochondria-based typing approach. FEMS Yeast Res
863 https://doi.org/10.1093/femsyr/foab051.

864 44. Boekhout T, Kurtzman CP, O'Donnell K, Smith MT. 1994. Phylogeny of the
865 yeast genera Hanseniaspora (anamorph Kloeckera), Dekkera (anamorph
866 Brettanomyces), and Eeniella as inferred from partial 26S ribosomal DNA
867 nucleotide sequences. Int J Syst Bacteriol 44:781–786.

868 45. Boekhout T, Bosboom RW. 1994. Karyotyping of Malassezia yeasts: Taxonomic
869 and epidemiological implications. Syst Appl Microbiol 17:146–153.

870 46. Morales L, Dujon B. 2012. Evolutionary role of interspecies hybridization and
871 genetic exchanges in yeasts. Microbiol Mol Biol Rev 76:721–739.

872 47. Louis VL, Despons L, Friedrich A, Martin T, Durrens P, Casarégola S,
873 Neuvéglise C, Fairhead C, Marck C, Cruz JA, Straub M-L, Kugler V, Sacerdot
874 C, Uzunov Z, Thierry A, Weiss S, Bleykasten C, De Montigny J, Jacques N,
875 Jung P, Lemaire M, Mallet S, Morel G, Richard G-F, Sarkar A, Savel G,
876 Schacherer J, Seret M-L, Talla E, Samson G, Jubin C, Poulain J, Vacherie B,
877 Barbe V, Pelletier E, Sherman DJ, Westhof E, Weissenbach J, Baret PV,
878 Wincker P, Gaillardin C, Dujon B, Souciet J-L. 2012. Pichia sorbitophila, an
879 Interspecies Yeast Hybrid, Reveals Early Steps of Genome Resolution After

880 Polyploidization. G3 2:299–311.

881 48. Pontes A, Čadež N, Gonçalves P, Sampaio JP. 2019. A Quasi-Domesticate Relic
882 Hybrid Population of × Adapted to Olive Brine. Front Genet 10:449.

883 49. Lee H-Y, Chou J-Y, Cheong L, Chang N-H, Yang S-Y, Leu J-Y. 2008.
884 Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility
885 between two yeast species. Cell 135:1065–1073.

886 50. Jhuang H-Y, Lee H-Y, Leu J-Y. 2017. Mitochondrial-nuclear co-evolution leads
887 to hybrid incompatibility through pentatricopeptide repeat proteins. EMBO Rep
888 18:87–101.

889 51. Langdon QK, Peris D, Baker EP, Opulente DA, Nguyen H-V, Bond U,
890 Gonçalves P, Sampaio JP, Libkind D, Hittinger CT. 2019. Fermentation
891 innovation through complex hybridization of wild and domesticated yeasts. Nat
892 Ecol Evol 3:1576–1586.

893 52. Bakkeren G, Kämper J, Schirawski J. 2008. Sex in smut fungi: Structure,
894 function and evolution of mating-type complexes. Fungal Genet Biol 45 Suppl
895 1:S15–21.

896 53. Coelho MA, Sampaio JP, Gonçalves P. 2010. A deviation from the bipolar-
897 tetrapolar mating paradigm in an early diverged basidiomycete. PLoS Genet 6.

898 54. Sun S, Yadav V, Billmyre RB, Cuomo CA, Nowrousian M, Wang L, Souciet J-
899 L, Boekhout T, Porcel B, Wincker P, Granek JA, Sanyal K, Heitman J. 2017.
900 Fungal genome and mating system transitions facilitated by chromosomal
901 translocations involving intercentromeric recombination. PLoS Biol

902 15:e2002527.

903 55. Watanabe J, Uehara K, Mogi Y, Tsukioka Y. 2017. Mechanism for Restoration
904 of Fertility in Hybrid Zygosaccharomyces rouxii Generated by Interspecies
905 Hybridization. Appl Environ Microbiol 83.

906 56. Ortiz-Merino RA, Kuanyshev N, Braun-Galleani S, Byrne KP, Porro D,
907 Branduardi P, Wolfe KH. 2017. Evolutionary restoration of fertility in an
908 interspecies hybrid yeast, by whole-genome duplication after a failed mating-
909 type switch. PLoS Biol 15:e2002128.

910 57. Vu D, Groenewald M, Szöke S, Cardinali G, Eberhardt U, Stielow B, de Vries
911 M, Verkleij GJM, Crous PW, Boekhout T, Robert V. 2016. DNA barcoding
912 analysis of more than 9 000 yeast isolates contributes to quantitative thresholds
913 for yeast species and genera delimitation. Stud Mycol 85:91–105.

914 58. Gupta AK, Boekhout T, Theelen B, Summerbell R, Batra R. 2004. Identification
915 and typing of Malassezia species by amplified fragment length polymorphism
916 and sequence analyses of the internal transcribed spacer and large-subunit
917 regions of ribosomal DNA. J Clin Microbiol 42:4253–4260.

918 59. Leong C, Chan JWK, Lee SM, Lam YI, Goh JPZ, Ianiri G, Dawson TL Jr. 2021.
919 Azole Resistance Mechanisms in Pathogenic Malassezia furfur. Antimicrob
920 Agents Chemother 65.

921 60. Guého-Kellermann E, Boekhout T, Begerow D. 2010. Biodiversity, Phylogeny
922 and Ultrastructure. Malassezia and the Skin.

923 61. O'Donnell K, Cigelnik E, Weber NS, Trappe JM. 1997. Phylogenetic

924 relationships among ascomycetous truffles and the true and false morels inferred
925 from 18S and 28S ribosomal DNA sequence analysis. Mycologia 89:48–65.

926 62. Lu X-L, Najafzadeh MJ, Dolatabadi S, Ran Y-P, van den Ende AHGG, Shen Y-
927 N, Li C-Y, Xi L-Y, Hao F, Zhang Q-Q, Li R-Y, Hu Z-M, Lu G-X, Wang J-J,
928 Drogari-Apiranthitou M, Klaassen C, Meis JF, Hagen F, Liu W-D, de Hoog GS.
929 2013. Taxonomy and epidemiology of Mucor irregularis , agent of chronic
930 cutaneous mucormycosis. Persoonia - Molecular Phylogeny and Evolution of
931 Fungi.

932 63. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for
933 Illumina sequence data. Bioinformatics 30:2114–2120.

934 64. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana
935 M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T,
936 Itoh T. 2014. Efficient de novo assembly of highly heterozygous genomes from
937 whole-genome shotgun short reads. Genome Res 24:1384–1395.

938 65. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017.
939 Canu: scalable and accurate long-read assembly via adaptive -mer weighting and
940 repeat separation. Genome Res 27:722–736.

941 66. Ye C, Hill CM, Wu S, Ruan J, Ma ZS. 2016. DBG2OLC: Efficient Assembly of
942 Large Genomes Using Long Erroneous Reads of the Third Generation
943 Sequencing Technologies. Sci Rep 6:31900.

944 67. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The
945 MaSuRCA genome assembler. Bioinformatics.

946 68. Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nature
947 Methods.

948 69. Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout--a reference-assisted
949 assembly tool for bacterial genomes. Bioinformatics.

950 70. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo
951 CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool
952 for comprehensive microbial variant detection and genome assembly
953 improvement. PLoS One 9:e112963.

954 71. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment
955 tool for genome assemblies. Bioinformatics.

956 72. Mapleson D, Accinelli GG, Kettleborough G, Wright J, Clavijo BJ. 2016. KAT:
957 a K-mer analysis toolkit to quality control NGS datasets and genome assemblies.
958 Bioinformatics.

959 73. Hoff KJ, Stanke M. 2013. WebAUGUSTUS--a web service for training
960 AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res 41:W123–
961 8.

962 74. Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction
963 in eukaryotes that allows user-defined constraints. Nucleic Acids Research.

964 75. Park M, Cho Y-J, Lee YW, Jung WH. 2017. Whole genome sequencing analysis
965 of the cutaneous pathogenic yeast Malassezia restricta and identification of the
966 major lipase expressed on the scalp of patients with dandruff. Mycoses 60:188–
967 197.

968 76. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome
969 Assembly and Annotation Completeness. Methods Mol Biol 1962:227–245.

970 77. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering
971 C, Bork P. 2017. Fast Genome-Wide Functional Annotation through Orthology
972 Assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122.

973 78. Langdon QK, Peris D, Kyle B, Hittinger CT. 2018. sppIDer: A Species
974 Identification Tool to Investigate Hybrid Genomes with High-Throughput
975 Sequencing. Mol Biol Evol 35:2835–2849.

976 79. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C,
977 Salzberg SL. 2004. Versatile and open software for comparing large genomes.
978 Genome Biol 5:R12.

979 80. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis
980 G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The
981 Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

982 81. Pegueroles C, Mixão V, Carreté L, Molina M, Gabaldón T. 2020. HaploTypo: a
983 variant-calling pipeline for phased genomes. Bioinformatics 36:2569–2571.

984 82. Garrison E MG. 2012. Haplotype-based variant detection from short-read
985 sequencing. arXiv preprint arXi:1207.3907.

986 83. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing
987 genomic features. Bioinformatics.

988 84. Al-Shahrour F, Diaz-Uriarte R, Dopazo J. 2004. FatiGO: a web tool for finding
989 significant associations of Gene Ontology terms with groups of genes.

990 Bioinformatics.

991 85. Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of
992 organelle genomes from whole genome data. Nucleic Acids Res 45:e18.

993 86. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy
994 and high throughput. Nucleic Acids Res 32:1792–1797.

995 87. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary
996 Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–
997 1874.

998 88. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003.
999 ExPASy: The proteomics server for in-depth protein knowledge and analysis.
1000 Nucleic Acids Res 31:3784–3788.

1001 89. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular
1002 Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol
1003 35:1547–1549.

1004

1005

1006 **Tables**

1007 Table 1. Strains used in this study with information about their lineage, source, country of isolation, NGS sequencing data analysis, usage for

1008 PFGE, AFLP, sanger sequencing, mating-type, FACS, LM, MALDI-TOF and physiology analyses.

Lineage	Strain	Source	Country	Data availability										
				NGS data	NGS strategy	PFGE	AFLP	Sanger sequencing 5 loci	Mating type	FACS	LM	MALDI-TOF	Physi ology
P1	CBS 9595	back (man)	Greece	PRJNA732434	Reference assembly P1	-	+	+	+	+	+	+	+
	CBS 9575	back (man)	Greece	-	-	-	+	+	+	-	+	+	-
	CBS 9574	back (man)	Greece	PRJNA732434	Read mapping	-	+	+	+	+	+	+	+
	CBS 9589	folds (man)	Greece	-	-	-	+	+	+	+	+	+	-
	CBS 7982	Skin of ear (man)	France	PRJNA286710* / PRJNA732434	Read mapping	+	+	+	+	+	+	+	-
	CD 866	Poodle	Brazil	PRJNA732434	Read mapping	-	+	+	+	+	+	+	+
	CBS 9369	Scalp of patient	Canada	-	-	+	-	+	+	-	+	+	+
P2	CBS 8735	bronchial wash	Canada	PRJNA732434	Read mapping	+	+	+	+	+	+	+	+
	CBS 14139	urine, human	France	-	-	+	+	+	+	+	+	+	+
	CBS 14141	catheter, blood	France	PRJNA286710* / PRJNA732434	Reference assembly P2	+	+	+	+	+	+	+	+
	PM 315	anal swab neonate	Germany	PRJNA732434	Read mapping	-	+	+	+	+	+	+	+
H1	CBS 1878 NT	Dandruff, man	Unknown	PRJNA732434	Read mapping	+	+	+	+	+	+	+	+
	CBS 4171	Ear of cow	Unknown	-	-	+	+	+	+	+	+	+	+
	CBS 4172	Skin of eland	Unknown	PRJNA286710*	Read mapping	+	+	+	+	+	+	+	+
	CBS 7969	Asian elephant	France	-	-	+	+	+	+	+	+	+	+

	CBS 9365	Elephant in zoo	France	PRJNA732434	Read mapping	+	+	+	+	+	+	+	+
	CBS 5101	skin scales, from PV	USA	-	-	-	+	+	+	-	+	+	-
	CBS 4162	Ear of pig	unknown	-	-	-	+	+	+	-	+	+	-
	CBS 6000	Dandruff	India	PRJNA732434	Read mapping	+	+	+	+	+	+	+	+
	CBS 6001	PV	India	PRJNA732434	Read mapping	+	+	+	+	-	+	+	-
H2	CBS 6093	Unknown	unknown	-	-	-	+	+	+	-	+	+	-
	CBS 7019 NT	PV on trunk of 15-year-old girl	Finland	PRJNA732434	Read mapping	+	+	+	+	+	+	+	+

H1 = hybrid lineage 1, H2 = hybrid lineage 2, P1 = Parental lineage 1, P2 = parental lineage 2, PFGE = Pulsed-Field Gel Electrophoresis, AFLP = Amplified Fragment Length Polymorphism. NT = neotype strain, + = analysis performed, - = analysis not performed, *Data from Wu et al. 2015.

1013 **Table 2.** Summary of assembly statistics for the best genome assembly obtained for

1014 CBS9595 and CBS14141.

	CBS9595	**CBS14141**
Assembly strategy	Ragout	DBG2OLC_sparse
Estimated genome size (Mb)*	7.8	8.3
Genome size (bp)	8 107 776	8 241 990
Contigs	8	9
Contigs > 50kb	8	8
N50 (bp)	1 622 862	1 642 932
GC (%)	64,48%	64,97%
Mapped reads (%)	95,50%	77,2%***
Protein-coding genes	4.376	4.441
Completeness (BUSCO)*	96.1%	95.4%
Completeness (KAT)**	99,90%	99,81%

1015 *BUSCO (Waterhouse et al. 2017), **KAT (Mapleson et al. 2017)

Multiple hybridization events punctuate the evolutionary trajectory of *Malassezia furfur*

Figures main text – illustration qualities may be reduced in this file, required quality files are available for publication.

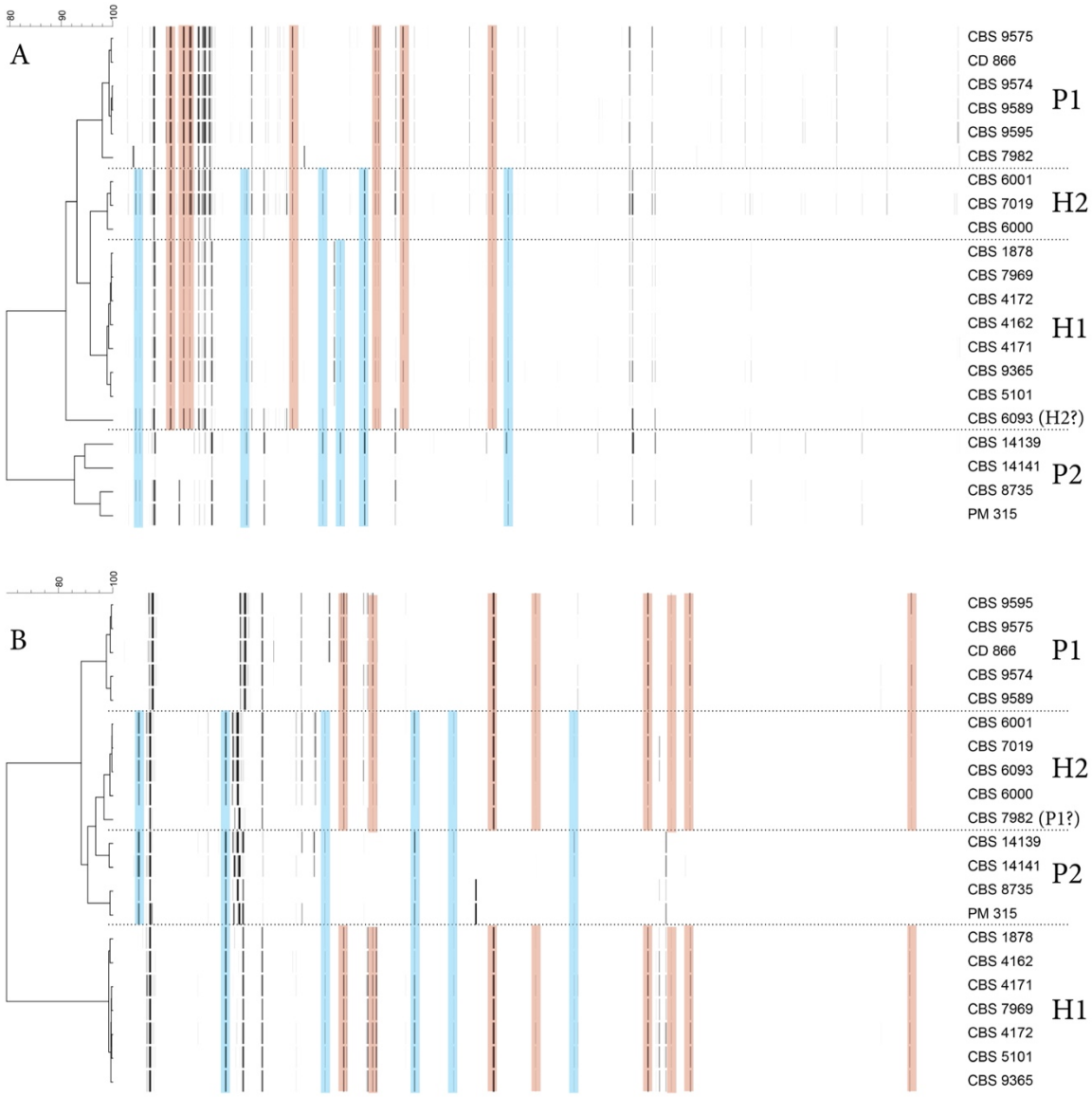

Figure 1. AFLP banding pattern-based Neighbor Joining trees for two different adaptor/primer combinations (A-B) (see Materials & Methods, and Supplementary Table 1). Both versions resulted from using different primer/adaptor pairs, reflecting different polymorphisms in the genomic DNA and thus resulting in some clustering variation for some strains. CBS6093 belongs to the H2 lineage based on dendrogram B, but clusters outside any of the other lineages in dendrogram A, suggesting genomic deviation from other H2 strains, a

finding also supported by mating type, β-glucosidase activity, and MALDI-TOF data. CBS7982 clusters together with other P1 strains as expected in dendrogram A, but clusters close to H2 strains in dendrogram B. Interestingly, CBS7982 was found to contain a mitochondrial sequence different from other P1 strains (CBS9595, CD866, and CBS9574, Supplementary Figure 1).

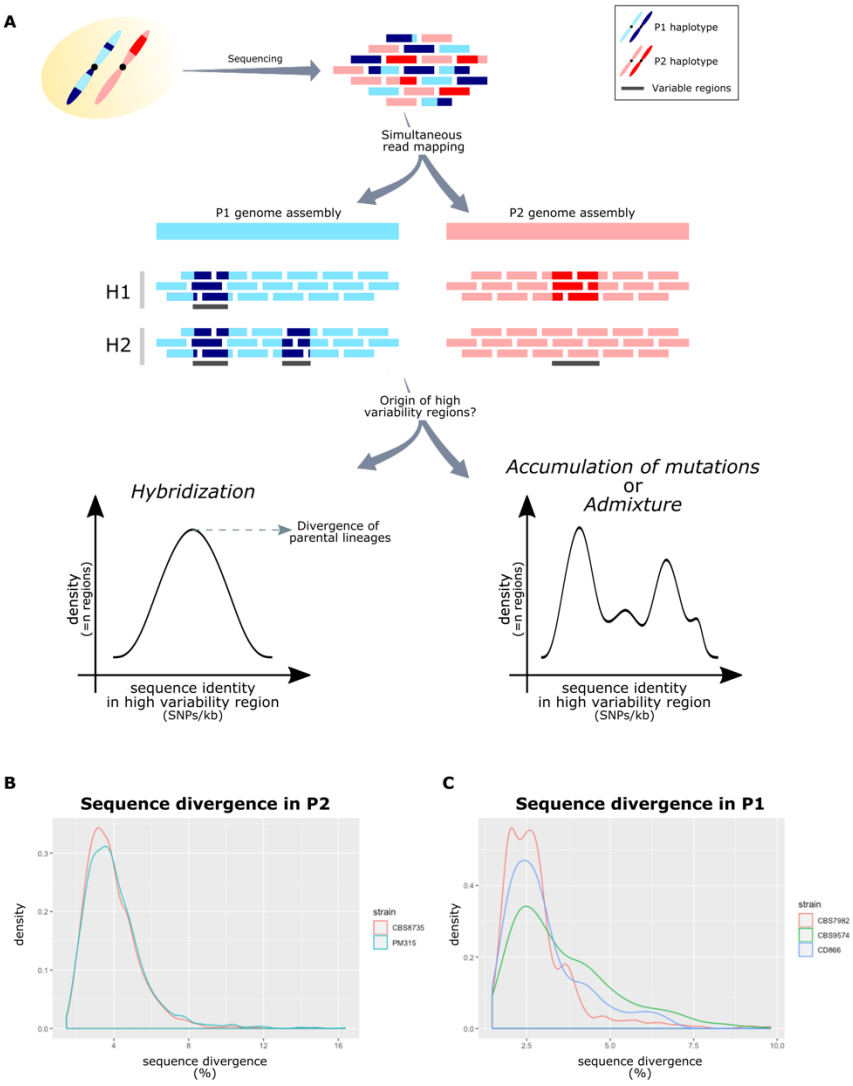

Figure 2. Analysis of the genomic patterns of hybrid genomes. A) Sequencing libraries of hybrid genomes were simultaneously mapped on the genome assemblies of the putative lineages. Differences in the patterns of genomic variability were used to determine the different hybrid lineages. Estimated sequence divergence in the blocks of genomic variability were used to determine the origin of such blocks: hybridization or admixture between different strains. **B)** Sequence divergence in the blocks of genomic variability of P2 lineages show a single density peak, suggesting a hybrid origin. **C)** Sequence divergence in the blocks of genomic variability of P1 lineages show multiple density peaks, with a single peak shared by all strains, not allowing the exclusion of any of the above-mentioned scenarios.

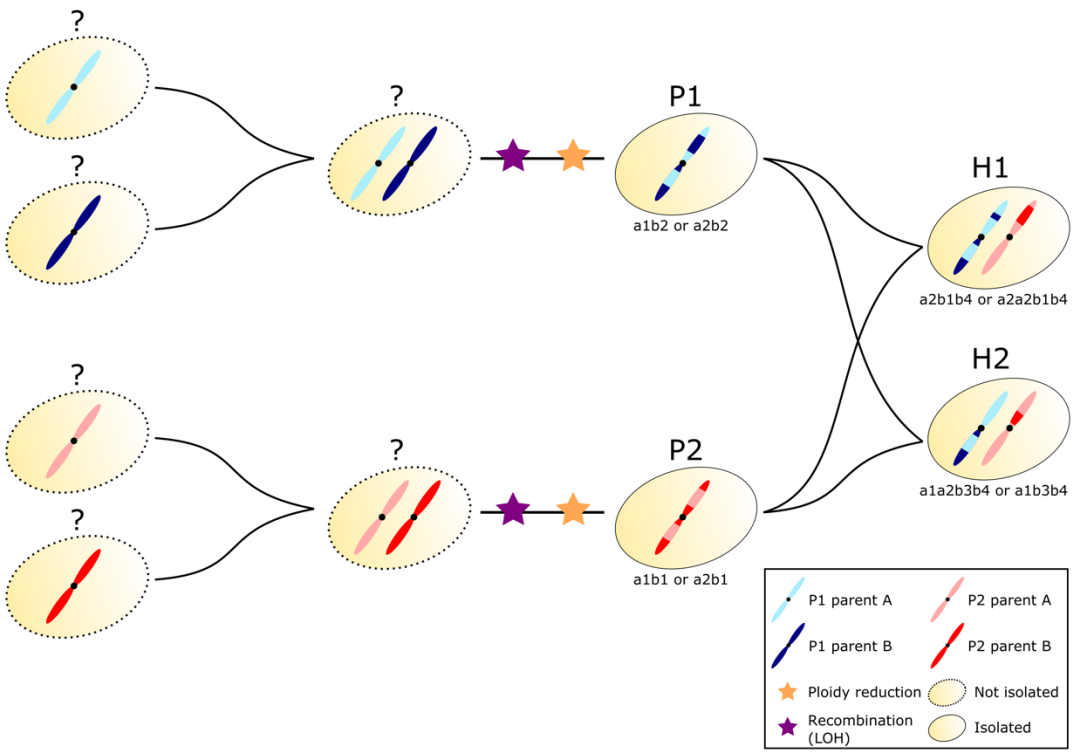

Figure 3. Comparative genomics analysis suggests multiple hybridization events in the evolutionary path of *M. furfur* lineages.

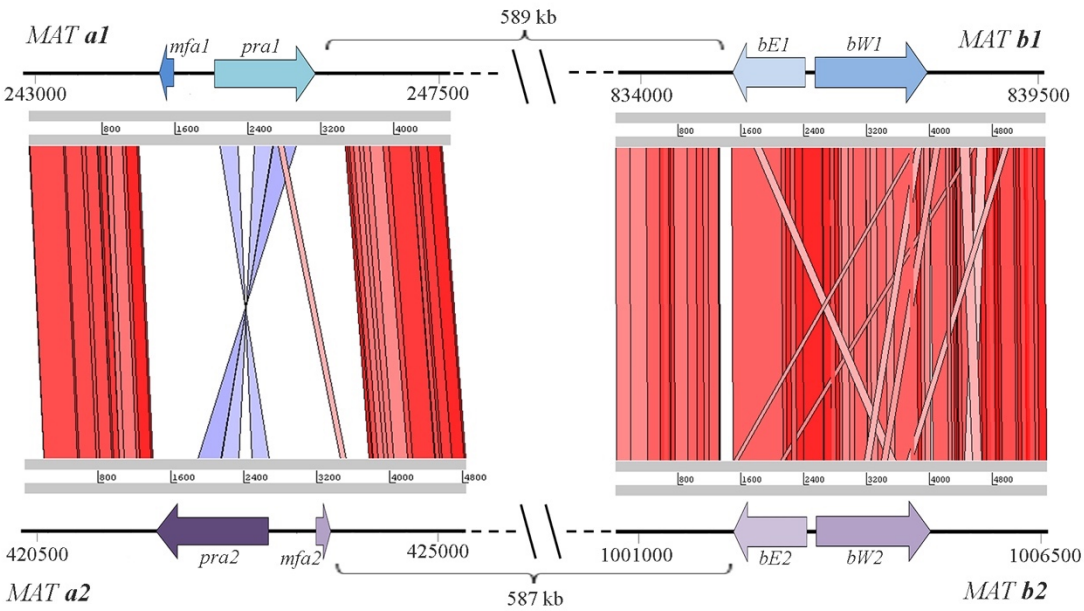

Figure 4. Schematic representation and comparison of mating loci for two mating compatible parental strains: *MAT* a1b1 loci of *M. furfur* strain CBS14139 (P2), and of the *MAT* a2b2 loci of *M. furfur* strain CBS7982 (P1). The coordinates of the genes in the genome scaffold are indicated. The two *MAT* regions were aligned with tBLASTx and visualized using ACT Artemis. The red and blue bars indicate regions of similarity, with red bars corresponding to regions of similar orientation and blue bars indicating regions oriented in opposite directions.

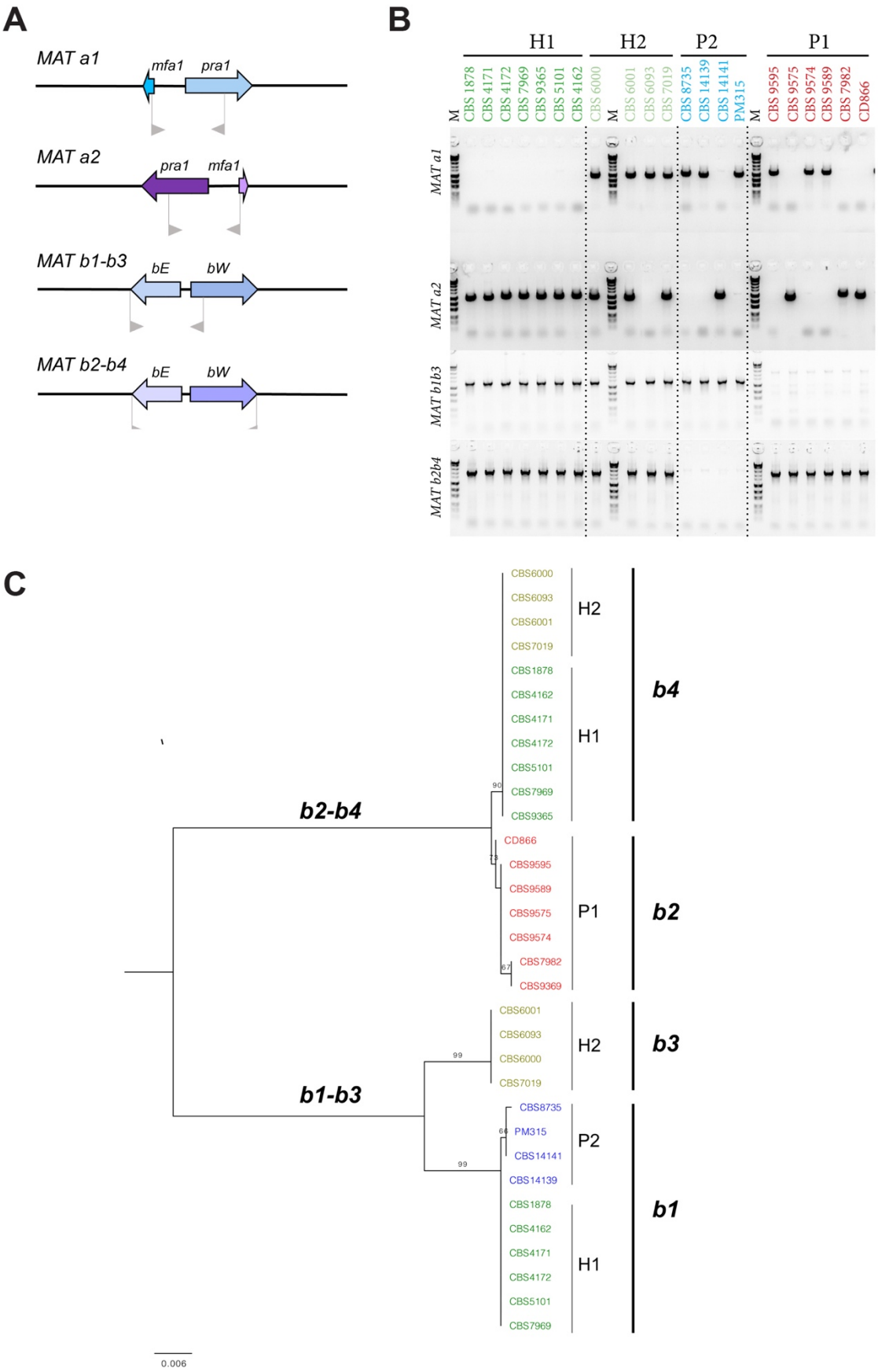
A
MAT a1
mfa1
pra1
MAT a2
pra1
mfa1
MAT b1-b3
bE
bW
MAT b2-b4
bE
bW
B
H1
H2
P2
P1
M
CBS 1878
CBS 4171
CBS 4172
CBS 7969
CBS 9365
CBS 5101
CBS 4162
CBS 6000
M
CBS 6001
CBS 6093
CBS 7019
CBS 8735
CBS 14139
CBS 14141
PM315
M
CBS 9595
CBS 9575
CBS 9574
CBS 9589
CBS 7982
CD866
MAT a1
MAT a2
MAT b1b3
MAT b2b4
C
CBS6000
CBS6093
CBS6001
CBS7019
H2
CBS1878
CBS4162
CBS4171
CBS4172
CBS5101
CBS7969
CBS9365
H1
b4
90
b2-b4
CD866
CBS9595
CBS9589
CBS9575
CBS9574
CBS7982
CBS9369
P1
b2
67
CBS6001
CBS6093
CBS6000
CBS7019
H2
b3
99
b1-b3
CBS8735
PM315
CBS14141
CBS14139
P2
66
99
CBS1878
CBS4162
CBS4171
CBS4172
CBS5101
CBS7969
H1
b1
0.006

Figure 5. Mating typing assay results. A) Schematic presentation of primer positions in the MAT genes. B) Agarose gel electrophoresis picture, showing PCR results for all assessed strains for Mat a1, MAT a2, combined MAT b1-b3, and MAT b2-b4. To distinguish between b1 and b3, and between b2 and b4, the MAT B PCR-positive products require sequencing. C) Phylogenetic tree based on the Maximum Likelihood method and Tamura-Nei model with 500 bootstrap replications, representing the MAT B loci, resulting in four main clusters.

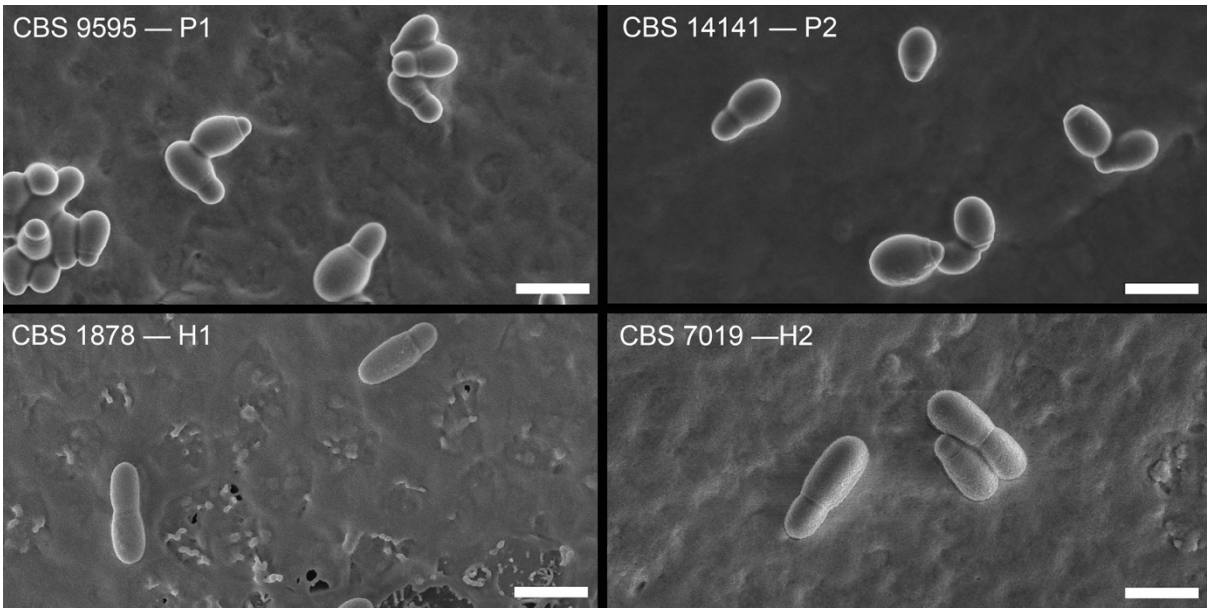

Figure 6. SEM photo plate for representative strains of each lineage, showcasing different morphologies between haploid parental strains and hybrid strains. Additional phenotypic data is available in Supplementary tables 8 and 9 for cell size measurements and traditional physiological data respectively. Bar = 5 µm