# Genome analysis reveals evolutionary mechanisms of adaptation in systemic dimorphic fungi

José F. Muñoz[1], Juan G. McEwen[2,3], Oliver K. Clay[3,4] , Christina A. Cuomo[1]*

[1]Broad Institute of MIT and Harvard, Cambridge, MA, United States.

[2] Cellular and Molecular Biology Unit, Corporación para Investigaciones Biológicas, Medellín, Colombia.

[3] School of Medicine, Universidad de Antioquia, Medellín, Colombia.

[4] School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia

* cuomo@broadinstitute.org

**Key Words:** Dimorphic fungi, comparative genomics, virulence evolution, Ajellomycetaceae

## ABSTRACT

Dimorphic fungal pathogens cause a significant human disease burden and unlike most fungal pathogens affect immunocompetent hosts. To examine the origin of virulence of these fungal pathogens, we compared genomes of classic systemic, opportunistic, and non-pathogenic species, including *Emmonsia* and two basal branching, non-pathogenic species in the Ajellomycetaceae, *Helicocarpus griseus* and *Polytolypa hystricis.* We found that gene families related to plant degradation, secondary metabolites synthesis, and amino acid and lipid metabolism are retained in *H. griseus* and *P. hystricis*. While genes involved in the virulence of dimorphic pathogenic fungi are conserved in saprophytes, changes in the copy number of proteases, kinases and transcription factors in systemic dimorphic relative to non-dimorphic species may have aided the evolution of specialized gene regulatory programs to rapidly adapt to higher temperatures and new nutritional environments. Notably, both of the basal branching, non-pathogenic species appear homothallic, with both mating type locus idiomorphs fused at a single locus, whereas all related pathogenic species are heterothallic. These differences revealed that independent changes in nutrient acquisition capacity have occurred in the Onygenaceae and Ajellomycetaceae, and underlie how the dimorphic pathogens have adapted to the human host and decreased their capacity for growth in environmental niches.

## INTRODUCTION

Among millions of ubiquitous fungal species that pose no threat to humans, a small number of species cause devastating diseases in immunocompetent individuals. Notably, many human pathogenic species are ascomycetes in the order Onygenales, which comprises dermatophytes (e.g. *Trichophyton* spp.), the spherule-forming dimorphic pathogen *Coccidioides*, and yeast-forming dimorphic pathogens *Histoplasma*, *Paracoccidioides*, and *Blastomyces*. Dimorphism is a specialized morphogenetic adaptation allowing both growth in the environment and colonization of a host and is critical for the lifecycle of dimorphic fungal pathogens [1]. In yeast-forming pathogens, dimorphism has been broadly defined as the ability to generate both yeast (e.g. blastoconidia) in the host or at 37 °C, and mycelia (e.g. conidia-producing or other hyphae) in the environment [1]. During fungal evolution, dimorphism has arisen independently multiple times in saprophytic fungi, as is manifested in the wide distribution of dimorphic species across the Ascomycota, Basidiomycota and Zygomycota phyla [2]. Notably, the dimorphic fungi in the order Onygenales are primary pathogens causing systemic mycosis in healthy humans, and collectively cause over 650,000 new infections a year in the United States alone [3]. Dimorphic fungi can persist as latent infections in tens of millions of people worldwide and may reactivate when the host becomes immune-deficient; symptoms of active infection include pneumonia, acute respiratory distress syndrome, and disseminated disease, which can affect multiple organ systems [1].

The largest cluster of thermally dimorphic human pathogenic fungi belongs to the onygenalean family Ajellomycetaceae, which includes *Histoplasma*, *Paracoccidioides*, and *Blastomyces*. In addition to these medically important genera, the Ajellomycetaceae family also includes more rarely observed pathogenic species including *Emmonsia parva* and *Ea. crescens*, which undergo a thermal dimorphic transition to produce adiaspores rather than yeast and cause adiospiromycosis sporadically in humans [4,5]. This group also includes *Lacazia loboi*, the yeast-like etiological agent of lobomycosis [6]. Of public health concern, recent reports have documented the worldwide emergence of yeast-forming Ajellomycetaceae that cause systemic mycoses predominantly in immunocompromised patients, often with high case-fatality rates [7,8]. Morphological and phylogenetic analyses were combined to demarcate species boundaries within the Ajellomycetaceae with clinical significance, which led to a proposed revision of the taxonomy; this included addition of the new genus *Emergomyces*, which includes the new species *Es. africanus* and *Es. orientalis*, the description of the new species *Blastomyces percursus*, and ongoing efforts to define additional *Emmonsia*-like and *Blastomyces*-like species [8–10].

63 Dimorphic species within the Ajellomycetaceae are typically restricted to specific ecological niches and
64 geographical regions. For example, while *H. capsulatum* and *Es. pasteurianus* are considered to have
65 worldwide distribution, *B. dermatitidis*, *P. brasiliensis*, and *Es. africanus* are restricted to North America,
66 Latin America and South Africa, respectively. Non-dimorphic species have also been documented as
67 early diverging Ajellomycetaceae, including *Emmonsiellopsis*, *Helicocarpus griseus* and *Polytolypa*
68 *hystricis*, neither of which has been associated with disease in mammals. These species are found in
69 environmental samples and do not appear to undergo a thermally regulated dimorphic transition [11–13].
70
71 We hypothesized that such phenotypic differences within the Ajellomycetaceae could be associated
72 with gene family expansions or contractions, as a consequence of gene duplication and gene loss
73 events, and that species comparison might reveal evolutionary mechanisms of adaptation in the
74 systemic dimorphic fungi. Previous comparative and population genomic studies within
75 Ajellomycetaceae had found evidence of gene family contractions and expansions associated with
76 virulence [5,14,15]. Other Onygenales genera outside the Ajellomycetaceae represented by published
77 genome sequences include *Coccidioides* and the non-pathogenic related species *Uncinocarpus reesii*
78 [16,17], as well as diverse dermatophytes that cause skin infections [18,19]. Recently, the genomes of four
79 non-pathogenic Onygenaceae species closely related to *Coccidioides* were described, providing
80 additional resolution into changes in gene content between pathogenic and non-pathogenic
81 Onygenales [20]. Together these studies helped establish how gene content is related to the life cycle of
82 different dimorphic pathogenic fungi and dermatophytes; for example, gene family contractions in
83 cellulases and other plant metabolism genes, and gene family expansions in proteases, keratinases
84 and other animal tissue metabolism genes, indicated that dimorphic fungi switched from a nutrition
85 system based on plants to a system based on animals, though mostly relative to outgroups to the
86 Onygenales such as *Aspergillus* [14,17–19]. In the Onygenales, this hypothesis was tested experimentally;
87 comparing growth on a wide variety of compounds revealed that in *Uncinocarpus reesii* hyphal growth
88 was restricted on carbohydrates and considerably improved on proteins [14].
89
90 To examine key transition points in evolution of virulence and host adaptation in the dimorphic fungi, we
91 increased the phylogenetic density within the Ajellomycetaceae by sequencing the genomes of non-
92 pathogenic and emerging species and then performed comparative genomic analyses of the systemic,
93 opportunistic, and non-pathogenic species. Although representative genomes for most of the described
94 pathogenic Onygenales genera have been sequenced, the only strictly non-pathogenic species that
95 have been sequenced are closely related to *Coccidioides* in the family Onygenaceae. Here, we have

96 sequenced four additional Onygenales species from the family Ajellomycetaceae; this includes *H.*
97 *griseus* and *P. hystricis*, neither of which has been associated with disease in mammals, and two
98 additional adiaspore-forming strains of *Ea. parva* and *Ea. crescens* with distinct phenotypic properties.
99 Leveraging our previous genomic studies and these additional genome sequences, we have
100 characterized the extent of gene family expansions and contractions to provide a better understanding
101 of the evolutionary mechanisms of adaptation in systemic dimorphic fungi in the Ajellomycetaceae.
102
103
104 **RESULTS**
105
106 **Genome comparisons of Ajellomycetaceae, the largest group of dimorphic human pathogenic**
107 **fungi**
108 Building on previous genomic analyses of the members from the Ajellomycetaceae [5,9,14,15,21], we
109 sequenced using Illumina technology the genomes of four species within Ajellomycetaceae with
110 demarcated phenotypic differences. In addition to *Emmonsia parva* UAMH139 and *Emmonsia crescens*
111 UAMH3008, we sequenced the genomes of one strain of *Emmonsia parva* UAMH130 (type strain) and
112 one additional strain of *E. crescens* UAMH4076. *E. parva* UAMH130 is from lungs of a rodent in USA,
113 and *E. crescens* UAMH4076 (type strain) is from a greenhouse source in Canada [22]. We also
114 sequenced two species that live in soil, where they are found in animal excrement but not associated
115 with disease; neither species appears to undergo a thermal dimorphic switch to a pathogenic phase, as
116 there is no observation of a yeast-like form [13,22]. This includes one strain of *Helicocarpus griseus*,
117 UAMH5409 (type strain; synonym *Spiromastix grisea*), isolated from gazelle dung in Algeria, and one
118 strain of *Polytolypa hystricis*, UAMH7299, isolated from porcupine dung in Canada.
119
120 The *de novo* genome assemblies of these four species have similar sizes to those of other genera
121 within the Ajellomycetaceae including *Histoplasma*, *Emergomyces* and *Paracoccidioides* [5,9,14,15], and
122 have high representation of conserved eukaryotic genes (**Figure S1, S2**). The assembly sizes were
123 27.8 Mb in *E. parva* UAMH130, 33.8 Mb in *E. crescens* UAMH4076, 32.0 Mb in *P. hystricis*, and 34.7
124 Mb in *H. griseus* (**Table 1**). In addition, using both the CEGMA [23] and BUSCO [24] conserved eukaryotic
125 gene sets, we found that the predicted genes for each sequenced species are nearly complete; all of
126 these genomes include 96-98% of the core genes (**Figure S2**). The total numbers of predicted genes in
127 *E. parva* UAMH130 ($n$ = 8,202), and *E. crescens* UAMH4076 ($n$ = 8,909) were similar to those found in
128 annotated genome assemblies of other Ajellomycetaceae ($n$ = 9,041 on average); however, relative to

129 other Ajellomycetaceae *P. hystricis*, and *H. griseus* have a modest increase in gene content (*n* = 9,935
130 and *n* = 10,225, respectively; **Table 1**; **Figure S1**).
131
132 Examining the guanine-cytosine (GC) content in the genomes *E. parva* UAMH130, *E. crescens*
133 UAMH4076, *H. griseus* UAMH5409 and *P. hystricis* UAMH7299 revealed that none of these genome
134 assemblies showed evidence of the pronounced bimodality of GC-content observed in *B. gilchristii* and
135 *B. dermatitidis* (**Figure S3**; **Table 1**; [5]). This highlights that the acquisition of repetitive elements
136 contributing to bimodal GC distribution were a unique phenomenon during the evolution of the *B.*
137 *gilchristii* and *B. dermatitidis* within the Onygenales. In addition, UAMH139 showed a small peak of low
138 GC that is largely absent in UAHM130 (44.4% and 47.5% GC, respectively; **Table 1**), which is
139 consistent with the phylogenetic divergence between these two strains, and highlights an independent
140 gain of repetitive sequences (10.4% and 2.6% repetitive bases in UAMH139 and UAMH130,
141 respectively; **Table 1**). Similarly, *E. crescens* UAMH3008 and UAHM4076 have demarcated differences
142 in GC and repetitive content; UAHM4076 has a higher peak of low GC sequences (42.9% GC, 15.0%
143 repetitive bases in UAMH4076; 45.2% GC, 5.1% repetitive bases in UAMH3008; **Table 1**). Additionally,
144 the more basal strain *P. hystricis* UAMH7299 has a representation of low GC sequences (45.6% GC,
145 12.8% repeat), but in *H. griseus* UAMH5409 they are almost absent (48.0% GC, 1.8% repeats; **Figure**
146 **S3**). These results highlight the independent expansion of repetitive elements during Ajellomycetaceae
147 evolution.
148
149 **Phylogenomic analysis revealed multiple transitions within dimorphic pathogens to enable**
150 **human infection**
151 We estimated a strongly supported phylogeny of the newly sequenced *E. parva* UAMH130*, E.*
152 *crescens* UAMH4075, *H. griseus* and *P. hystricis* strains relative to other dimorphic fungi using
153 maximum likelihood and 2,505 single copy core genes (**Figure 1**). Branch length and topology
154 supported the placement and relationships of *P. hystricis* and *H. griseus* within the Ajellomycetaceae as
155 basally branching. The two *E. parva* isolates (UAMH130 and UAMH139) do not form a single well-
156 defined clade, confirming that *E. parva* may not be a single species. In addition to longer branch
157 lengths suggesting higher divergence of these isolates, the average genome-wide identity between
158 UAMH130 and UAMH139 is much lower even comparing across groups that represent recently
159 subdivided species, for example, within *B. gilchristii* and *B. dermatitidis* or *P. lutzii* and *P. brasiliensis*,
160 with an average identity of 88.6% versus 95.8% and 91.1% respectively (**Figure S4**). *E. parva* (type
161 strain UAMH130) is the most basal member of the clade including *E. parva* UAMH139 and

162 *Blastomyces* species *B. percursus, B. gilchristii, and B. dermatitidis*. Previous molecular approaches
163 [12,25] and recent multi-locus approaches [8,9] also support the polyphyletic nature of *E. parva* isolates.
164 Although both isolates of *E. crescens* (UAMH3008 and UAMH4076) form a single clade, the genetic
165 variation between them is also greater than that observed within other species from the same genus in
166 the Ajellomycetaceae, with long branch lengths and slightly lower average genome identity (91.8%)
167 between these isolates than seen in other intraspecies comparisons within the family (**Figure S4**). The
168 *E. crescens* clade is closely related to the recently proposed genus *Emergomyces* including the *E.*
169 *pasteurianus/africanus* species [9]. The *E. crescens – Emergomyces* clade is a sister group of the clade
170 including *Histoplasma* and *Blastomyces*, and *Paracoccidioides* is in a basal position for the dimorphic
171 pathogens of the Ajellomycetaceae (**Figure 1**).
172
173 The intercalated clades of primary pathogens (e.g. *H. capsulatum*, *P. brasiliensis* and *B. dermatitidis*),
174 opportunistic pathogens (e.g. *E. pasteurianus*, *E. africanus*) producing a parasitic yeast phase, and
175 non-pathogens producing adiaspore-type cells rather than yeast cells (e.g. *E. parva* and *E. crescens*)
176 suggests that the dimorphic fungi in the Ajellomycetaceae have undergone numerous evolutionary
177 transitions allowing adaptation, infection and virulence to humans and other mammals, and that
178 interactions with other eukaryotes in the environment may help maintain the capacity for pathogenic
179 growth in a mammalian host (**Figure 1**; **Table 2**). The non-thermally-dimorphic species *H. griseus* and
180 *P. hystricis* are highly supported as early diverging lineages of Ajellomycetaceae as previously
181 suggested [13]; genomic analysis strongly supports *P. hystricis* as the earliest diverging lineage within the
182 sequenced Ajellomycetaceae, and large genetic variation between these early saprophytes. The basal
183 position of both *H. griseus* and *P. hystricis* in the family Ajellomycetaceae and the large divergence
184 between them suggests that the history of this family could involve evolution from saprophytic non-
185 thermotolerant and non-yeast-forming ancestral species to dimorphic human pathogenic species after
186 the divergence with *H. griseus*, including the adaptation to higher temperatures and the appearance of
187 a dimorphic switch and a yeast parasitic phase (**Figure 1**).
188
189 **Substantial gene family contractions underlie the shift from growth on plant to animals**
190 By comparing the gene conservation and functional annotation of primary and opportunistic
191 dimorphic pathogens with their saprophytic relatives *H. griseus* and *P. hystricis,* we identified significant
192 decreases in gene family size (PFAM domain and GO-term counts, corrected p-value $< 0.05$)
193 associated with a host/substrate shift from plants to animals in the pathogenic Ajellomycetaceae
194 (**Figure 2**). Plant cell wall degrading enzymes are reduced in number or totally absent in dimorphic

195 pathogens compared with *H. griseus* and *P. hystricis*, including several glycosyl hydrolases, and fungal
196 cellulose binding domain-containing proteins (**Figure 2; Table S2**). These gene family changes suggest
197 that dimorphic fungi from Ajellomycetaceae are not typical soil fungi in that they might maintain a close
198 association with living mammals or other organisms such as protozoa or insects. This transition
199 appears to have occurred within the family Ajellomycetaceae after divergence from *Helicocarpus* and
200 *Polytolypa* (**Figure 1**). Our analysis also recapitulates previous reports comparing smaller numbers of
201 genomes from the order Onygenales [5,14,17,20] that showed that genes coding for enzymes involved in the
202 deconstruction of plant cell walls were absent from all the human pathogens (e.g. *Paracoccidioides*,
203 *Blastomyces*, *Histoplasma*, *Coccidioides*) and non-pathogenic Onygenaceae (e.g. *Uncinocarpus reesii*,
204 *Byssoonygena ceratinophila*, *Amauroascus mutatus*, *Amauroascus niger*, *Chrysosporium*
205 *queenslandicum*) whereas these enzymes were commonly conserved in the non-pathogens outside the
206 order Onygenales (including *Aspergillus, Fusarium*, *Neurospora*; **Figure 2; Table S2**). Notably,
207 increasing the phylogenetic density within the order Onygenales in this study revealed that the loss of
208 plant cell wall degradative enzymes occurred independently in two families within the Onygenales, and
209 this highlights transitions between saprophytic and pathogenic species, where *Helicocarpus* and
210 *Polytolypa* have retained enzymes required to live on plant material in the soil, whereas non-pathogenic
211 species from the Onygenaceae, which are not adapted to survive in mammals, have also lost capacity
212 to degrade plant material.
213
214 Other notable changes in gene content included loss of protein families associated with secondary
215 metabolite biosynthesis such as polyketide synthase dehydratases, and beta-ketoacyl synthases
216 domains in dimorphic Ajellomycetaceae relative to *H. griseus* and *P. hystricis* (**Figure 2**; **Table S3**). We
217 identified predicted biosynthetic clusters using the program antiSMASH (antibiotics and Secondary
218 Metabolism Analysis Shell [26]). Overall, dimorphic pathogenic fungi from Ajellomycetaceae have fewer
219 polyketide synthase (PKs) gene clusters than the saprophytic *H. griseus* and *P. hystricis* (**Table S3;**
220 **Figure S5**). Dimorphic pathogens had lost essential genes or complete clusters of the type 1 and type
221 3 PKs, and terpene clusters (**Table S3**). Type 3 PKs are not common in fungi, and were not previously
222 reported in the Ajellomycetaceae. Furthermore, *H. griseus* and *P. hystricis* but not dimorphic fungi have
223 genes related to those involved in antibiotic biosynthetic pathways (ko01055; **Table S2**). Previous
224 studies had shown that *P. hystricis* secretes a pentacyclic triterpenoid exhibiting antifungal and
225 antibiotic activity, denominated Polytolypin [27], suggesting this terpene cluster is intact and may produce
226 this molecule. Loss of such secondary metabolic pathways may weaken responses to other microbes in
227 the environment. In addition to secondary metabolite biosynthesis, Ajellomycetaceae dimorphic fungi

228 showed significant decreases in pathways related with the production of sphingolipids, chloroalkane
229 and chloroalkene degradation, linoleic acid metabolism, and bisphenol degradation (KEGG-EC counts,
230 corrected p-value < 0.05; **Table S2**). Loss of degradative pathways for bisphenol and chloralkanes in
231 particular likely reflect encountering these molecules in the environment but not during pathogenic
232 growth. Altogether, these shifts in gene content highlight how the Ajellomycetaceae dimorphic
233 pathogenic fungi became less adapted to survive in the soil than the more basally diverging saprophytic
234 species *H. griseus* and *P. hystricis*.
235
236 **Transition among carbohydrate metabolism and protein catabolism**
237 We found that the evolution of carbohydrate metabolism and protein catabolism also occurred
238 independently within the Ajellomycetaceae. We annotated carbohydrate active enzymes (CAZy) and
239 peptidases (MEROPS; **Methods**), and using enrichment analysis found substantial shifts in the relative
240 proportion of these groups among dimorphic and saprophytic species. Notably, the pathogenic
241 Ajellomycetaceae species, but not the saprophytic species *H. griseus* and *P. hystricis*, showed a
242 dramatic reduction of carbohydrate active enzymes, including 44 categories that were totally absent
243 and 29 that were significantly depleted; these included glycoside hydrolases, glycosyltransferases,
244 carbohydrate esterases, and polysaccharide lyases (corrected p < 0.05; **Figure 3**). The fact that these
245 enzymes are reduced in dimorphic pathogenic Ajellomycetaceae but not in *H. griseus* and *P. hystricis*
246 confirms that reduced carbohydrate metabolism occurred independently Ajellomycetaceae from the
247 reduction in the Onygenaceae, originally noted in the spherule-dimorphic pathogen *Coccidioides* [17]. In
248 contrast to carbohydrate active enzymes, we found that the total number of peptidases is relatively
249 similar in dimorphic and saprophytic species within the Ajellomycetaceae. Significant protein family
250 expansions were found to be specific in *Blastomyces*, *Histoplasma* and Onygenaceae and
251 Arthrodermataceae species relative to *P. hystricis* and *H. griseus*, including the A11, S12, and S08A
252 peptidases (corrected p < 0.05; **Figure 3**). The peptidase family A11 contains endopeptidases involved
253 in the processing of polyproteins encoded by retrotransposons, and the increase in copy number is
254 correlated with the genome expansion due to the proliferation of repetitive elements in *Blastomyces* [5].
255 The peptidase family S12 contains serine-type D-Ala-D-Ala carboxypeptidases, and S8 contains the
256 serine endopeptidase subtilisin. Two families of fungal metalloendopeptidases (S8 and M35) were
257 previously reported to be expanded in *Coccidioides* relative to saprophytic species outside the order
258 Onygenales [17]. In addition, it has been reported that the Ajellomycetaceae includes fewer copies of
259 multiple classes of peptidases (M36, M43, S8) as well as an associated inhibitor (I9, inhibitor of S8
260 protease) relative to other Onygenales [5]. Comparing across all the Onygenales, we found that all

261 pathogenic Onygenales, as well as saprophytes related to *Coccidioides* such as *U. reesii*, had a higher
262 ratio of proteases to carbohydrate active enzymes than the saprophytic Ajellomycetaceae *H. griseus*
263 and *P. hystricis* (**Figure 3**). This strengthens the evidence that the adaptation and specialization of
264 dimorphic Ajellomycetaceae occurred independently and that non-pathogenic Onygenaceae may
265 represent an intermediate state in showing gene content changes associated with pathogens, which
266 allowed these fungi to use proteins as a main source of energy in nutrient-limited environments.
267
268
269 **Loss of transporters in Ajellomycetaceae dimorphic fungi**
270 In addition to major shifts in enzymes that support nutrient acquisition, the dimorphic
271 Ajellomycetaceae have undergone losses of families of transporters. These include the large classes of
272 major facilitator superfamily transporters (PF07690) and sugar transporters (PF00083) (**Figure 2)**. The
273 number of transporters correlates with the numbers of genes predicted to have transmembrane helices
274 (1,848 and 1,745 proteins have predicted transmembrane domains in *H. griseus* and *P. hystricis,*
275 respectively, relative to an average of 1,490 in dimorphic Ajellomycetaceae pathogens, **Figure 4**)**.** To
276 characterize gene gain and losses in transporters, we annotated transporter families (**Methods,** [28]), and
277 found significant expansion of 18 classes of transporters encompassing 9 families in *H. griseus* and *P.*
278 *hystricis* relative to dimorphic Ajellomycetaceae pathogens (corrected *p-value* < 0.05; **Figure 4**).
279 Transporter families that were totally absent or significantly reduced in pathogenic species included the
280 sugar porter, anion:cation symporter, AAA-ATPase, ammonium transporter and carnitine O-Acyl
281 transferase. The substrates of these transporters include galacturonate, Rip1, hexoses, dipeptides,
282 allantoate, ureidosuccinate, allantoin, glycerol, sulfate, sulfite, thiosulfate, sulfonates, tartrate, thiamine,
283 cellobiose, cellodextrin, ammonia, methylamine, phenylacetate, beta-lactams, trichothecene, and
284 carnitine O-octanoyl (**Figure 4**). Most of these compounds are only present in the plant cell wall,
285 including the cellulose, hemicellulose, or pectin types, supporting that dimorphic pathogens have
286 evolved to have more limited compounds from plant cell wall as source of energy.
287
288 **Pathogenesis-related genes are conserved in both saprophytic and pathogenic lifestyles**
289 Genes known to be important in the response to stress imposed by the host, including
290 virulence-associated or yeast-phase specific genes of central importance in dimorphic fungi in the
291 Onygenales, are conserved in *H. griseus* and *P. hystricis*. These include heat shock response proteins
292 (*HSF*, *HSP90*, *HSP70*), dimorphic switch related proteins (*RYP1, RYP2, RYP3*, *AGS1*, *FKS1*),
293 oxidative stress and hypoxia response proteins (*CATB*, *CATP*, *SOD3*, *SRB1*), as well as antigens

294 (*PbGP43*, *PbP27*, *BAD1*; **Table S4;**[21]). In dimorphic pathogenic fungi both morphological transitions
295 and growth temperature are linked, and loss of genes necessary for high-temperature growth and
296 morphological transition in these pathogens results in attenuated virulence, which highlights that they
297 are essential for pathogenesis [29–31]. Some of these genes such as the *RYP*1-3 transcriptional regulators
298 and the dimorphism-regulating histidine kinase *DRK1* have been connected with morphogenetic
299 adaptation in response to environmental stimuli, and with key determinants of the mycelia to yeast
300 transition in dimorphic human pathogenic fungi in the Ajellomycetaceae [29,30]. Although *H. griseus* and
301 *P. hystricis* are not dimorphic and cannot grow at 37 °C, the heat shock response has been shown to
302 be highly evolutionarily conserved among eukaryotes. The fact that genes of central importance for
303 high-temperature growth and dimorphic switch are conserved in *H. griseus* and *P. hystricis* suggests
304 that these saprophytes may sense different gradations of temperature, or that the mechanism that
305 triggers a rapid response to the adaptation may not be rapid enough to enable these fungi to grow at 37
306 °C.
307
308 Dimorphic Ajellomycetaceae may have evolved specialized mechanisms to regulate their transcriptional
309 responses; these include expansion of proteins related to the regulation of transcription, and protein
310 kinase activity (GO term enrichment analysis; $q$-value < 0.05; **Table S2**), which are all expanded in
311 dimorphic pathogens relative to *H. griseus* and *P. hystricis*. Thus, the gain of transcription factors,
312 transcriptional regulators and phosphotransferases suggests that rapid evolution of transcriptional
313 mechanisms may underlie the adaptation to both high temperature and the dimorphic transition. As one
314 of the largest categories enriched in dimorphic Ajellomycetaceae but not in saprophytic species was the
315 protein kinase activity (**Table S2**), we further classified protein kinases using Kinannote [32], including the
316 divergent fungal-specific protein kinase (*FunK1*; **Methods**). We found an expansion of the *FunK1*
317 family, in dimorphic Ajellomycetaceae as previously reported [5,14]. As we increased the phylogenetic
318 density in the Ajellomycetaceae, including species and strains with high genetic and phenotypic
319 variation this allowed resolution of the evolutionary history of this family of kinases (**Figure 5**). *FunK1*
320 family is enriched in the systemic pathogens *B. dermatitidis/gilchristii*, *H. capsulatum*, *Paracoccidioides*,
321 but limited in saprophytic species, *H. griseus* and *P.* hystricis. Notably, this family is also limited in the
322 adiaspore-forming *E. parva* (UAMH130, UAMH139), *E. crescens* (UAMH4076) but not in *E. crescens*
323 (UAMH3008). In addition, we found that the recently described primary pathogen *B. percursus* includes
324 a large expansion of FunK1 kinases, with almost twice the number as compared with its closest relative
325 *B. dermatitidis* (**Figure 5**). Phylogenetic analysis of *FunK1* genes revealed that this family undergone

326 independent lineage-specific expansions in primary dimorphic pathogens of both Ajellomycetaceae and
327 Onygenaceae, but not in non-pathogenic or opportunistic dimorphic species (**Figure 5**).
328
329 To identify specific genes that evolved within the Ajellomycetaceae to enable the dimorphic transition
330 and infection of mammals, we identified and examined ortholog clusters that were unique to these
331 pathogens and absent in both *H. griseus* and *P. hystricis*. We identified 75 ortholog clusters that were
332 present in dimorphic Ajellomycetaceae and dimorphic Onygenaceae species but absent in *H. griseus*
333 and *P. hystricis* (i.e. 'sporophyte gene loss events'), and 212 ortholog clusters that were present in
334 dimorphic Ajellomycetaceae, but absent in the dimorphic Onygenaceae species, *H. griseus* and *P.*
335 *hystricis* (i.e. 'dimorphic Ajellomycetaceae gene gain events'; **Table S5**). The gene loss events include
336 a mold specific protein (CIMG_00805), a phenol 2-monooxygenase (BDFG_05966), as well as several
337 protein kinases and transporters. The gene gain events include several protein kinases, transporters,
338 transcription factors, and peptidases (**Table S5**). In addition, we annotated genes that have been
339 broadly associated with host-pathogen interactions (Pathogen Host interaction database PHI), and
340 genes that have been reported as virulence factors in fungal species (Database of Fungal Virulence
341 Factors DFVF) [33,34]. We found that many genes that have been associated with host-pathogen
342 interactions are present in the saprophytic *P. hystricis* and *H. griseus,* and in fact some appear at
343 higher copy number counts (**Table S5**). This emphasizes that host-pathogen interaction and virulence
344 factors are not only important for pathogenicity in mammals, but also for saprophytic species to survive
345 and adapt to stress in the environment, including for example the ability to evade the engulfment by
346 amoebae, or the interaction with insects [35,36].
347
348 **Conservation of genes induced during *in vivo* infection**
349 We also hypothesized that genes induced during mammalian infection in dimorphic
350 Ajellomycetaceae pathogens might have different gene conservation patterns among *Blastomyces*,
351 *Histoplasma*, *Paracoccidioides*, and the novel species and saprophytes. Transcriptional responses
352 during an *in vivo* mouse model of infection had been studied in *B. dermatitidis* (strain ATCC26189; [5])
353 and *P. brasiliensis* (strain Pb18; [37]). In both *B. dermatitidis* and *P. brasiliensis* the high-affinity zinc
354 transporter, *ZRT1* (BDFG_09159; PADG_06417) and the flavodoxin-like protein *PST2* (BDFG_08006;
355 PADG_07749) were highly induced during infection; these genes are conserved in all Ajellomycetaceae
356 included in this analysis (**Table S6**). Of the 72 *in vivo*-induced genes in *B. dermatitidis* ATCC26188,
357 22% are absent from the *P. hystricis* genome and 11% are absent from the *H. griseus* genome. The
358 genes absent in *P. hystricis* include the cysteine dioxygenase (BDFG_08059), an acetyltransferase

359 (BDFG_07903), a thioesterase (BDFG_02348), a superoxide dismutase (BDFG_07895), a
360 sodium/hydrogen exchanger (BDFG_05427), and a response regulator protein (BDFG_01066). In
361 addition, we identified three secreted proteins that were found induced during the interaction of *B.*
362 *dermatitidis* with macrophages (BDFG_08689, BDFG_06057, BDFG_03876) [5] that are found
363 conserved only in the dimorphic Ajellomycetaceae pathogens but not in *H. griseus, P. hystricis,* or
364 outgroup species. These are strong candidates for effector proteins involved in host infection. Similar
365 numbers of genes found induced during *in vivo* infection in *P. brasiliensis* were absent in *P. hystricis*
366 and *H. griseus*, (23% and 17.5%, respectively; **Table S6**). Genes absent in both species include Gpr1
367 family protein (PADG_08695), isovaleryl-CoA dehydrogenase (PADG_05046), glutathione-dependent
368 formaldehyde-activating (PADG_05832), predicted transmembrane protein (PADG_05761), and ergot
369 alkaloid biosynthetic protein A (PADG_07739). Many of the genes that were absent in *P. hystricis* and
370 *H. griseus*, were also absent in at least one of the dimorphic pathogens in the Ajellomycetaceae,
371 highlighting species-specific gene gain events that may confer virulence specialization.
372
373 **Evolution of the mating type locus in the Onygenales family Ajellomycetaceae**
374 To elucidate the evolution of the mating system in the Ajellomycetaceae, we identified and
375 characterized the mating locus of *P. hystricis*, and *H. griseus*. Notably, these two species have both
376 mating type idiomorphs *HMG* box (*MAT 1-2*) and alpha box (*MAT1-1*); these are therefore the only
377 homothallic species identified so far within the Ajellomycetaceae (**Figure 6**). Comparative analysis
378 showed that the locus is not expanded relative to other Ajellomycetaceae species (21 kb between the
379 flanking genes *SLA2* to *APN2/COX13*), unlike the expansion that observed in the larger genomes of *B.*
380 *dermatitidis* or *B. gilchristii* (~60 kb; [38]). One difference is that a sequence inversion between *APN2* and
381 *COX13* is uniquely observed in *P. hystricis* relative to all other sequenced Ajellomycetaceae species,
382 including *H. griseus* and previously reported species [5,15,38,39]. This inversion observed in *P. hystricis*
383 (mating type idiomorph flanked by *SLA2* and *APN2/COX13*) is similar to the gene order in the
384 Onygenaceae species *Coccidioides* spp. but not in the Arthrodermataceae species *M. gypseum*
385 (*Nannizzia gypsea* [40]) or *T. rubrum* (**Figure 6; Table S7**) suggesting that independent inversion events
386 within the locus. In addition, we identified the mating type locus in non-pathogenic Onygenaceae, and
387 found that *Amauroascus mutatus* and *Byssoonygena ceratinophila* have both mating type idiomorphs
388 (**Table S7**). In *A. mutatus*, *HMG* box (*MAT 1-2*) and alpha box (*MAT1-1*) are linked in the same
389 scaffold, however they are separated by 184 kb, including 65 protein-coding genes (**Figure 6**). The
390 identification of homothallic mating loci in non-pathogenic species within both the Ajellomycetaceae and

391 the Onygenaceae suggests that there have been multiple transitions between homothallic and
392 heterothallic sexual states in dimorphic pathogens and non-pathogenic species in the Onygenales.
393

394 **DISCUSSION**

395

396 The evolution of the dimorphic Onygenales is characterized by multiple transitions between saprophytic
397 and human pathogenic growth. By sequencing the genomes of two additional dimorphic adiaspore-
398 forming species (*Emmonsia parva*, *E. crescens*) and two early diverging non-pathogenic, non-
399 dimorphic species (*Helicocarpus griseus* and *Polytolypa hystricis)*, we analyzed how changes in gene
400 content correlate with transitions to pathogenesis. Our results establish an outer bound on the timing of
401 two key events in the evolution of the Ajellomycetaceae; both the loss of genes involved in digesting
402 plant cell walls and the expansion of proteases, transcriptional regulators, and protein kinases,
403 occurred after divergence with *H. griseus*. The shift in these families was previously observed in both
404 *Paracoccidioides* within this group and in the related *Coccidioides–Uncinocarpus* clade [14,17]; more
405 recently the comparison with non-pathogenic Onygenaceae species supported previous reports of the
406 contraction of gene families involved in digesting plant cell walls throughout the Onygenales, and of the
407 expansion of gene families involved in digesting animal protein [20]. Whereas the studies on contraction
408 of gene families involved in digesting plant cell walls addressed other fungi outside the order
409 Onygenales, our analysis shows that *H. griseus* and *P. hystricis* did not experience such a contraction,
410 which suggests that contractions in gene families involved in digesting plant cell walls throughout the
411 Onygenales have occurred multiple times independently and more recently in Onygenaceae and
412 Ajellomycetaceae families.

413

414 Since many virulence factors and dimorphic switch related proteins were also conserved in the rarely
415 pathogenic (*Ea. parva* and *Ea. crescens*) and in the non-dimorphic Ajellomycetaceae species (*P.*
416 *hystricis* and *H. griseus*), this suggests that these proteins could be needed not only for survival in the
417 host, and that presence of virulence factors alone does not specify pathogenicity. In dimorphic
418 Ajellomycetaceae other mechanisms such as gene regulation or selection may play a role in how these
419 genes respond during distinct stimuli. In addition, it is possible that those virulence factors in mammals
420 were also needed for the survival in the soil, for example to survive interaction with amoebae or in
421 insects [41]. However, there are some virulence factors that are needed for virulence in mammals that are
422 not necessary for virulence in other eukaryotes. For example, the alpha mating factor locus in
423 *Cryptococcus neoformans* contributes to virulence in mice, but no difference was observed in the
424 interaction of congenic *MAT*α and *MAT***a** strains with amoebae [42]. While the human host is a very
425 different habitat in comparison with soil or excrements of animals that are the most common natural

426 niche for this group of fungi, interactions with other eukaryotes in the environment likely selected for
427 some properties that also predisposed species to become pathogenic to humans.
428
429 Selection pressures in the environment are responsible for the emergence and maintenance of traits
430 that confer upon some soil microbes the capacity for survival in animal hosts. This is important to
431 understand the evolution and emergence of novel pathogens. For example, *Emergomyces africanus*
432 has recently emerged and numerous cases of novel yeast-forming species have been reported .
433 However, *Emergomyces* species have not been found in other mammals like rodents or in the soil. In
434 addition, *E. africanus* has a very restricted geographic distribution. Across the Onygenales, nearly all
435 species have a high lower ratio of proteases to carbohydrate active enzymes; the exceptions are the
436 two saprophytes *P. hystricis* and *H. griseus*, but not the saprophytes related to *Coccidioides* such as
437 *Uncinocarpus reesii*. This genomic profile matches the results of growth assays for the non-pathogenic
438 species *U. reesii*; this species can grow on a wide range of proteinaceous substrates but only a very
439 limited range of carbohydrates, namely cellulose and its component glucose[14]. Together, these findings
440 suggest that fungi in the Onygenales transitioned from saprophytes that digest plant materials to
441 saprophytes that digest more limited plant materials and a wide range of animal proteins, and finally
442 those that are major animal pathogens.
443
444 We found many of the genes with roles in the dimorphic transition and virulence are conserved in *Ea.*
445 *parva* and *Ea. crescens*, and in saprophytic species *H. griseus* and *P. hystricis*, and we did not find
446 larger patterns of gain of functional classes in pathogenic species. However, we found smaller changes
447 in gene content related to the regulation of gene expression and signaling that may account for the
448 ability to grow at high temperature, species specific morphological transitions, and therefore their
449 differences in virulence, since rapid thermo-tolerance and morphological adaptation is essential for
450 pathogenesis [29,30,43]. Overall, we found significant expansions of the number of the protein kinase
451 *FunK1* family, transcription factors and other genes associated with the regulation of gene expression
452 in yeast-forming Ajellomycetaceae relative to adiaspore-forming and saprophytic species. Our
453 comparison with saprophytic Ajellomycetaceae highlights dramatic decreases and complete absences
454 of several classes of gene families that are important for survival and growth in a soil environment.
455
456 Intriguingly, we found that both more basal non-pathogenic Ajellomycetaceae species *H. griseus* and *P.*
457 *hystricis* appear homothallic, in striking contrast to all the previously described largely pathogenic
458 species in this group that are all heterothallic. *P. hystricis* has a fused mating type locus, with both

459 idiomorphs, *HMG* box (*MAT 1-2*) and alpha box (*MAT1-1*), fused at a single locus, and the same is
460 likely true of *H. griseus*, as genes are located at the ends of two scaffolds in this assembly. In addition,
461 we found that this is also true for the non-pathogenic Onygenaceae species *Amauroascus mutatus* and
462 *Byssoonygena ceratinophila*; these species have both mating type idiomorphs. *A. mutatus* has both
463 mating type idiomorphs located on the same scaffold; however, these are separated by a much larger
464 distance than is typically observed at the *MAT* locus of homothallic fungi. This suggests that this locus
465 has been subject to recombination events either to bring the two idiomorphs nearby on the same
466 scaffold or alternatively in expansion of a fused locus. While it has been previously suggested that the
467 ancestor of the Onygenales very likely was heterothallic [39], the placement of multiple lineages of
468 homothallic species in this group raises the possibility of a heterothallic ancestor, with more recent
469 transitions in pathogenic species to be homothallic and rearrangements of the mating locus in others
470 such as *P. hystricis* and *A. mutatus*. Both *P. hystricis* UAMH7299 and *H. griseus* UAMH5409 can
471 develop ascomata and ascospores, which have been shown to sporulate in *P. hystricis* [11,27,44]. Loss of
472 the capacity for homothallic self-mating could increase the frequency of out-crossing and this may be
473 beneficial to some species under pressure from their environment or host [45] such as the pathogenic
474 species in this group. Examining the mating type loci of additional non-pathogenic species in the
475 Onygenales would further reveal if there are novel configurations of homothallic or heterothallic loci,
476 which may help infer the mating ability of ancestral species and the timing of transitions in pathogens
477 and non-pathogens.
478
479 Our phylogenomic analysis revealed that dimorphic pathogens in the Onygenales have undergone
480 multiple evolutionary transitions that enabled infection of humans and other mammals. Some such as
481 the loss of plant cell wall degrading enzymes are convergent for pathogens in the two families of
482 Onygenales. Our comparisons also revealed major differences in non-pathogenic species between
483 these two families, which suggested that some species appear intermediate between saprophytes and
484 pathogens in terms of having a shift in metabolic capacity without associated pathogenic growth.
485 Targeting additional intermediate species within the Ajellomycetaceae with different phenotypes (e.g.
486 *Emmonsiellopsis*) for genome sequencing and comparison would help further refine the picture of the
487 overall evolution of this important group of fungi.
488

489 **MATERIALS AND METHODS**

490

491 **Genome sequencing**

492 Species were selected for sequencing based on previous estimates of high genetic variation
493 among the members of the genus *Emmonsia*, especially in the *E. parva* species, and based on
494 observations of the emergence of *Emmonsia*-like species in the Ajellomycetaceae causing systemic
495 human mycoses worldwide [8,9]. We selected one strain of *Emmonsia parva* (UAMH130; CBS139881;
496 type strain) and one additional strain of *Emmonsia crescens* (UAMH4076; CBS139868). *E. parva*
497 UAMH130 is from the lungs of a rodent in the USA, and *E. crescens* UAMH4076 is from a greenhouse
498 source in Canada, and mated with *E. crescens* UAMH129, 349 [22]. In addition, the first strictly non-
499 pathogenic non-dimorphic species from the Ajellomycetaceae, one strain of *Helicocarpus griseus*
500 (UAMH5409; a.k.a. *Spiromastix grisea*) and *Polytolypa hystricis* (UAMH7299) were selected for whole
501 genome sequencing. Genomic DNA was isolated from mycelial phase at 22 ºC. We constructed one
502 library with 180-base inserts and sequenced on the Illumina HiSeq 2000 platform to generate 101 bp
503 paired-end reads with high coverage (173X UAMH4076; 199X UAMH130; 205X UAMH5409; 163X
504 UAMH7299).

505

506 **Genome assembly, gene prediction and annotation**

507 The 101-bp Illumina reads of *E. parva* UAMH130, *E. crescens* UAMH4076, *P. hystricis* and *H.*
508 *griseus* were assembled using ALLPATHS-LG [46] with default parameters. All four *de novo* assemblies
509 were evaluated using the GAEMR package (http://software.broadinstitute.org/software/gaemr/), which
510 revealed no aberrant regions of coverage, GC content or contigs with sequence similarity suggestive of
511 contamination. Scaffolds representing the mitochondrial genome were separated out from the nuclear
512 assembly.

513

514 Genes were predicted and annotated by combining calls from multiple methods to obtain the best
515 consensus model for a given locus. These included *ab initio* predictions (GlimmerHMM, Augustus,
516 Snap, GeneMark-ES), homologous inference (Genewise, TBlastN), and gene model consolidation
517 programs (EvidenceModeler) [47,48]. For the protein coding-gene name assignment we combined
518 HMMER PFAM/TIGRFAM, Swissprot and Kegg products. Kinannote was used to annotate protein
519 kinases [32]. To evaluate the completeness of predicted gene sets, the representation of core eukaryotic
520 genes was analyzed using CEGMA genes [23] and BUSCO [24].

521

522 **Identification of orthologs and phylogenomic analysis**

523 To compare gene content and conservation, we identified orthologous gene clusters in these
524 sequenced genomes and in other Onygenales genomes using OrthoMCL (version 1.4) with a Markov
525 in a on index of 1.5 and a maximum e-value of 1e-5 [49]. We included the genomes of the classical
526 dimorphic pathogenic species (*Blastomyces dermatitidis, B. gilchristii*, *Histoplasma capsulatum,*
527 *Paracoccidioides brasiliensis, P. lutzii*, *Coccidioides immitis, and C. posadasii*), novel dimorphic human
528 pathogenic species (*Blastomyces percursus*, *Emergomyces africanus*, *E. pasteurianus*), and two
529 dimorphic non-human pathogenic species *Emmonsia parva* UAMH139 and *Emmonsia crescens*, along
530 with other species, including three *Aspergillus* (**Table S1; Figure S7**). In addition, we annotated and
531 included in the compare gene content and conservation analysis the available genomes of saprophytic
532 species from the family Onygenaceae *Byssoonygena ceratinophila* (UAMH5669), *Amauroascus*
533 *mutatus* (UAMH3576), *Amauroascus niger* (UAMH3544), and *Chrysosporium queenslandicum*
534 (CBS280.77), and the genome of the yeast-forming dimorphic pathogen *Emergomyces orientalis*
535 (5Z489, **Table S1**). Although, this is the largest genomic data set of Ajellomycetaceae strains so far, the
536 number of shared gene families after each addition of genome data did not clearly reach a plateau; this
537 suggests that the current estimate of a core of 4,200 genes is too high, and that more genomes are
538 needed to define the core-genome and pan-genome in the Ajellomycetaceae (**Figure S6**).
539
540 To examine the phylogenetic relationship of the newly sequenced *E. parva* and *E. crescens*, *P. hystricis*
541 and *H. griseus* relative to other dimorphic fungi we used single copy orthologs determined and
542 clustered using OrthoMCL. A total of 31 genomes from the Onygenales order and three *Aspergillus*
543 genomes were chosen to estimate the species phylogeny (**Table S1**). These include the four newly
544 strains described here: *E. parva* UAMH130, *E. crescens* UAMH4076, *P. hystricis* UAMH7299 and *H.*
545 *griseus* UAMH5409, as well as the following: three *Blastomyces dermatitidis* (ATCC26199,
546 ATCC18188, ER-3), one *Blastomyces gilchristii* (SLH14081), one *Emmonsia parva* (UAMH139), one
547 *Emmonsia crescens* (UAMH3008), two *Histoplasma* (WU24, G186AR), five *Paracoccidioides* (Pb01,
548 Pb03, Pb18, PbCnh, Pb300), two *Coccidioides* (RS, Silveira), *Uncinocarpus reesii* (UAMH1704),
549 *Microsporum gypseum* (CBS118893), *Trichophyton rubrum* (CBS118892), *Aspergillus nidulans* (FGSC
550 A4), *A. flavus* (NRRL3357) and *A. fumigatus* (Af293). Protein sequences were aligned using MUSCLE,
551 and a phylogeny was estimated from the concatenated alignments using RAxML v7.7.8 [50] with model
552 PROTCATWAG with a total of 1,000 bootstrap replicates.
553
554 **Gene family and protein domain analysis**

555 To identify gene content changes that could play a role in the evolution of the dimorphism and
556 pathogenesis within Ajellomycetaceae, we searched for expansions or contractions in functionally
557 classified genes compared to the other fungi from the order Onygenales. Genes were functionally
558 annotated by assigning PFAM domains, GO terms, and KEGG classification. HMMER3 [51] was used to
559 identify PFAM domains using release 27. GO terms were assigned using Blast2GO [52], with a minimum
560 e-value of $1x10^{-10}$. Protein kinases were identified using Kinannote [32] and divergent FunK1 kinases
561 were further identified using HMMER3. We further annotated carbohydrate active enzymes, peptidases,
562 and transporter families using the CAZY version 07-15-2016 [53], MEROPS version 9.12 [54], and TCDB
563 version 01-05-2017 [28] databases, respectively. Proteins were searched against their corresponding
564 databases using BLAST, with minimum e-values of $1\times10^{-80}$ for CAZY, $1\times10^{-20}$ for MEROPS, and
565 $1\times10^{-20}$ for TCDB.
566
567 To identify functional enrichments in the compared genomes, we used four gene classifications:
568 OrthoMCL similarity clusters, PFAM domains, KEGG pathways, and Gene Ontology (GO), including
569 different hierarchy levels, MEROPS and CAZY categories, and transporter families. Using a matrix of
570 gene class counts for each classification type, we identified enrichment comparing two subsets of
571 queried genomes using Fisher's exact test. Fisher's exact test was used to detect enrichment of PFAM,
572 KEGG, GO terms, CAZY, MEROPS, and transporter families between groups of interest, and p-values
573 were corrected for multiple comparisons [55]. Significant (corrected p-value $< 0.05$) gene class
574 expansions or depletions were examined for different comparisons.
575
576 **Data Availability Statement**
577 The raw sequences, genome assemblies and gene annotations were deposited in GenBank under the
578 following BioProject accession numbers: *Emmonsia parva* strain UAMH130 (PRJNA252752),
579 *Emmonsia crescens* strain UAMH4076 (PRJNA252751), *Polytolypa hystricis* strain UAMH7299
580 (PRJNA234736), and *Helicocarpus griseus* strain UAMH5409 (PRJNA234735).
581

**ACKNOWLEDGEMENTS**

We thank Lynne Sigler for her help in selecting strains for sequencing. We thank the Broad Genomics Platform for generating Illumina sequence for this study, Sarah Young and Margaret Priest for their assistance in assembly and annotation, Christopher Desjardins for helpful comments on the manuscript, and Leslie Gaffney for help with Figure 1. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under award N°: U19AI110818. This work was partly supported by Colciencias via the grants under Contract N°: 122256934875 and N°: 221365842971, and by the Universidad de Antioquia via a “Sostenibilidad 2016” grant.

**Author contributions**

Conceived and designed the experiments: JFM JGM OKC CAC. Performed the assembly and annotation: JFM CAC. Analyzed the data: JFM CAC. Contributed reagents/materials/analysis tools: JGM OKC. Wrote the paper: JFM CAC.

**Additional information**

Competing interests: The authors declare no competing financial interests.

601 **Tables and Figures**

602

603 **Table 1.** Statistics of the annotated genome assemblies of *Emmonsia*, *Helicocarpus* and *Polytolypa*

| Species | Strain | Assembly (Mb) | Scaffolds (count) | N50 (kb) | GC (%) | Repeat (%) | Genes (count) |
|---|---|---|---|---|---|---|---|
| *Emmonsia parva* | UAMH130 | 27.8 | 383 | 173.2 | 47.5 | 2.6 | 8,202 |
| *Emmonsia crescens* | UAMH4076 | 33.8 | 1,288 | 96.6 | 42.9 | 15.0 | 8,909 |
| *Helicocarpus griseus* | UAMH5409 | 34.7 | 641 | 135.7 | 48.0 | 1.8 | 10,225 |
| *Polytolypa hystricis* | UAMH7299 | 32.0 | 546 | 124.6 | 45.6 | 12.8 | 9,935 |

604

605 **Table 2.** Thermo-dimorphic pathogenic fungi, expansion and diversity within Ajellomycetaceae

| Species | Family | Dimorphism | Parasitic phase | Virulence in human |
|---|---|---|---|---|
| *Blastomyces gilchristii* | Ajellomycetaceae | Yes | Intermediate yeast | Primary pathogen |
| *Blastomyces dermatitidis* | Ajellomycetaceae | Yes | Intermediate yeast | Primary pathogen |
| *Blastomyces percursus* | Ajellomycetaceae | Yes | Large yeast | Primary pathogen |
| *Emmonsia parva* | Ajellomycetaceae | Yes | Small adiaspore | Occasional |
| *Histoplasma capsulatum* | Ajellomycetaceae | Yes | Small yeast | Primary pathogen |
| *Emmonsia crescens* | Ajellomycetaceae | Yes | Large adiaspore | Occasional |
| *Emergomyces pasteurianus* | Ajellomycetaceae | Yes | Small yeast | Opportunistic |
| *Emergomyces africanus* | Ajellomycetaceae | Yes | Small yeast | Opportunistic |
| *Paracoccidioides brasiliensis* | Ajellomycetaceae | Yes | Multibudding yeast | Primary pathogen |
| *Paracoccidioides lutzii* | Ajellomycetaceae | Yes | Multibudding yeast | Primary pathogen |
| *Lacazia loboi* | Ajellomycetaceae | ND | Multibudding yeast | Primary pathogen |
| *Helicocarpus griseus* | Ajellomycetaceae | No | No | Non-pathogen |
| *Polytolypa hystricis* | Ajellomycetaceae | No | No | Non-pathogen |
| *Coccidioides immitis* | Onygenaceae | Yes | Spherules | Primary pathogen |
| *Coccidioides posadasii* | Onygenaceae | Yes | Spherules | Primary pathogen |
| *Sporothrix schenckii* | Ophiostomataceae | Yes | Small yeast | Primary pathogen |
| *Talaromyces marneffei* | Trichocomaceae | Yes | Small septate yeast | Primary pathogen |

606

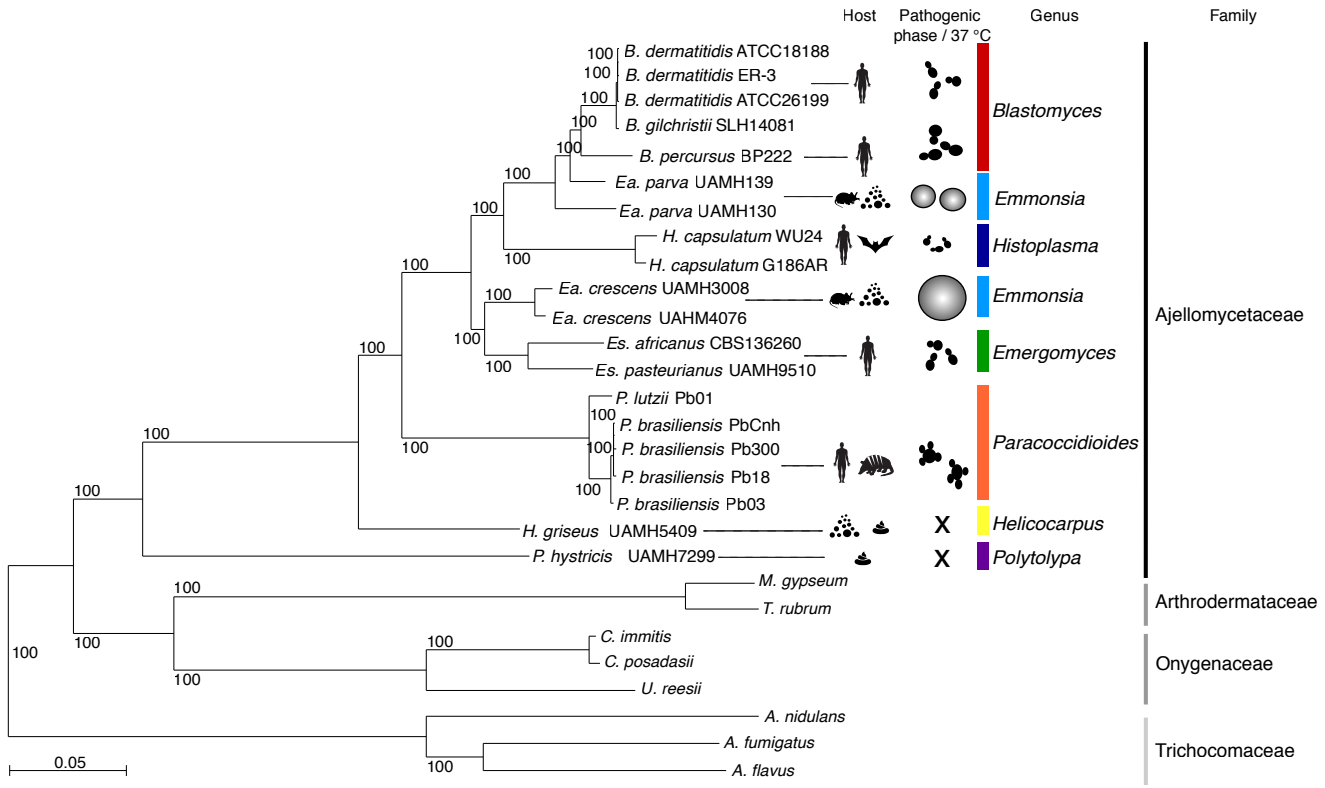


**Figure 1. Phylogenomic relationships and diversity in the family Ajellomycetaceae.** Maximum likelihood phylogeny (2,505 core genes based on 1,000 replicates) including 20 annotated genome assemblies of species within the family Ajellomycetaceae, five other Onygenales and three *Aspergilli*. All nodes were supported by 100% of bootstrap replicates. Color bars indicate the genus and icons represent most common source or host. Illustrations of the most common morphology at 37 °C or in the host are depicted for selected species highlighting the phenotypic variation within the Ajellomycetaceae and the transition to human hosts represented by pathogenicity traits.

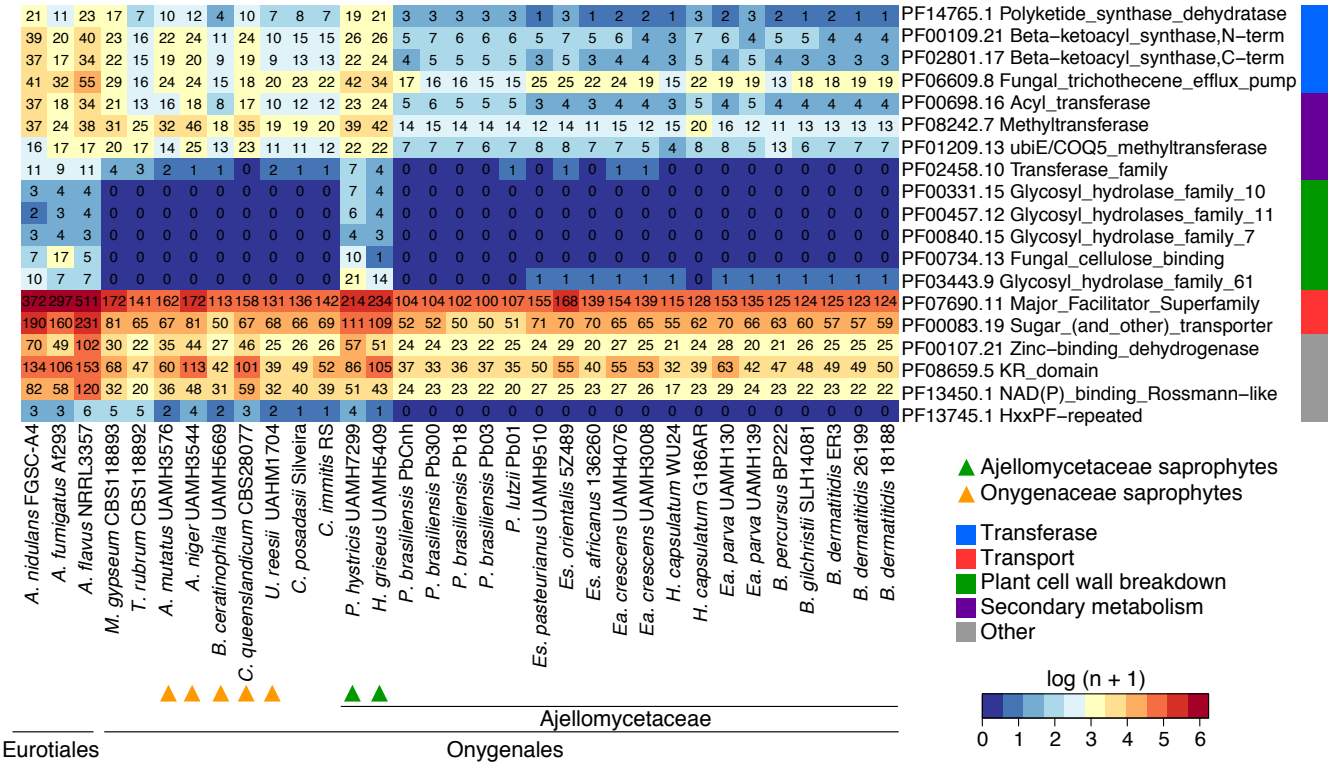


**Figure 2. Gene family changes across the Ajellomycetaceae.** Heatmap depicting results of a protein family enrichment analysis (PFAM domains; corrected *p-value* < 0.05) comparing the gene content of the non-dimorphic/non-pathogenic *Polytolypa hystricis* and *Helicocarpus griseus* with all other depicted members of the family Ajellomycetaceae including the dimorphic pathogens *Histoplasma*, *Emmonsia*, *Blastomyces*, *Emergomyces* and *Paracoccidioides*. Values are colored along a blue (low counts) to red (high counts) color scale, with color scaling relative to the low and high values of each row. Each protein family domain has a color code (right) indicating a global functional category. *Es*: *Emergomyces*; *Ea*: *Emmonsia*.

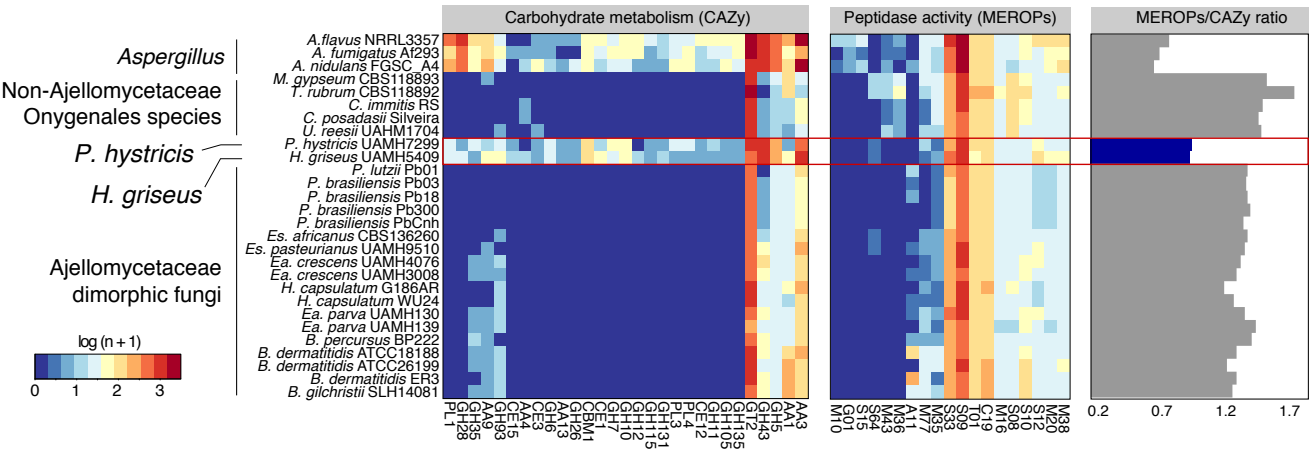


**Figure 3. Shifts in carbohydrate metabolism and peptidase activity in Ajellomycetaceae.** Heatmaps depicting for each taxon carbohydrate metabolism categories (CAZy) and peptidase families (MEROPs) across the Ajellomycetaceae species included in this study and other compared genomes. These categories were found significantly enriched (corrected *p-value* < 0.05) in the non-dimorphic systemic fungi *P. hystricis* and *H. griseus* relative to Ajellomycetaceae dimorphic fungi, or within Ajellomycetaceae species and other compared genomes from Onygenaceae and Arthrodermataceae families, or Eurotiales order. The ratio of MEROPS genes to CAZY genes for each genome across the Ajellomycetaceae species included in this study and other compared genomes is shown at right.

*Es*: *Emergomyces*; *Ea*: *Emmonsia*.

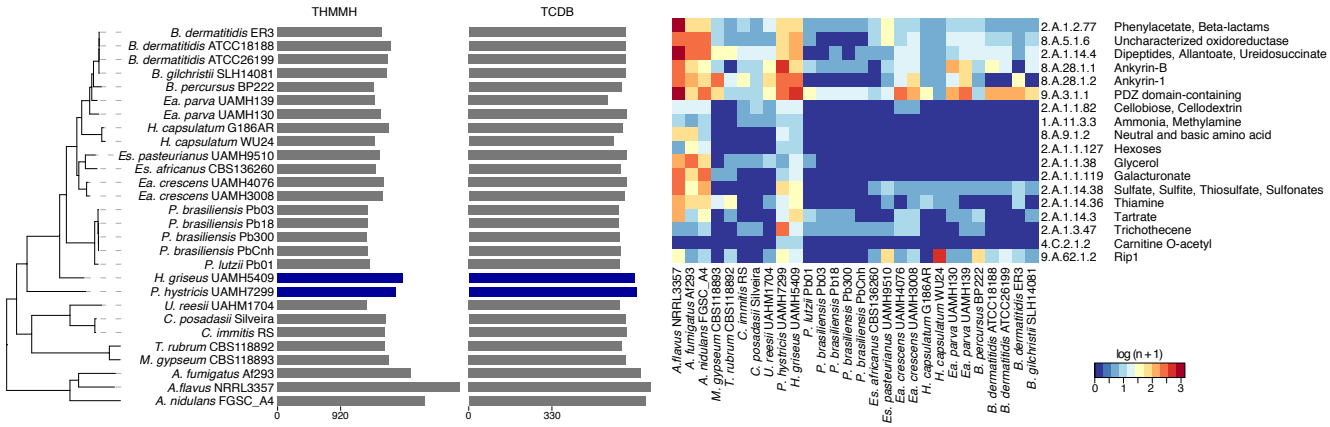


**Figure 4. Shifts in metabolism matched substrate preference within Ajellomycetaceae contraction of transporters within the Ajellomycetaceae.** (right) Heatmap depicting the number of family transporters (TCDB; color-code blue: low, red: high) for each taxon across the Ajellomycetaceae species and other compared genomes. These families were found significantly enriched (test, corrected *p-value* < 0.05) in the non-dimorphic *P. hystricis* and *H. griseus* relative to dimorphic Ajellomycetaceae fungi. (left) Total number of transporters identified in the genomes for each taxon included in this study. *Es*: *Emergomyces*; *Ea*: *Emmonsia*.

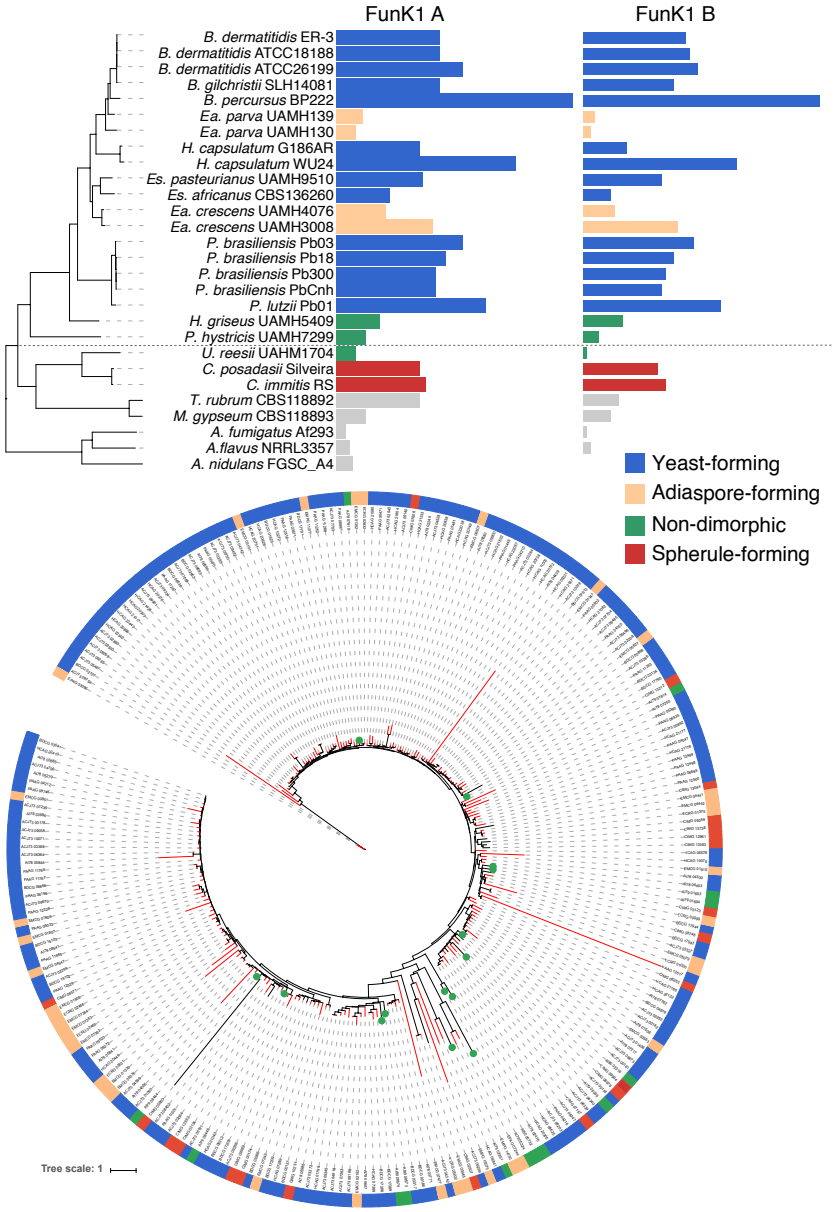


**Figure 5. Expansion and contraction of protein kinases class *FunK1* across the Ajellomycetaceae.** (top) *FunK1 A* and *FunK1 B* counts for each taxon across the Ajellomycetaceae species and other compared genomes. The tree shown corresponds to the maximum likelihood phylogenetic tree of the family Ajellomycetaceae from Fig. 1. (bottom) Maximum likelihood tree of *FunK1* family showing expansion in Ajellomycetaceae yeast-forming (blue) pathogenic species. Each species (top) or gene (bottom) has a color code indicating the parasitic phase at 37 °C (yeast, adiaspore, spherule, or non-pathogenic).

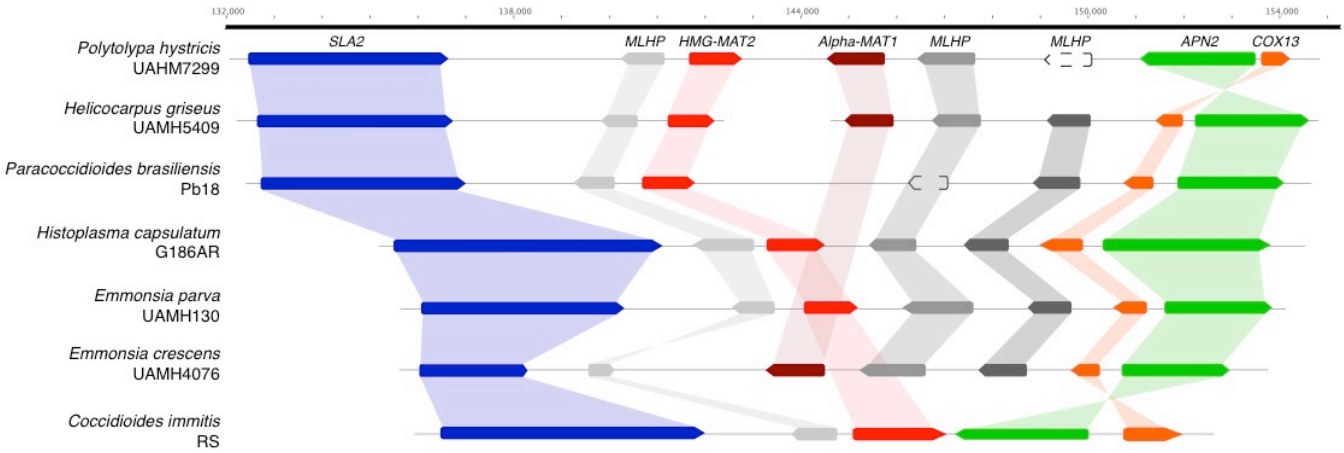


**Figure 6. Mating type evolution within the Onygenales family Ajellomycetaceae.** (top) Synteny schema depicting orientation and conservation of the genes adjacent to the mating type locus idiomorphs HMG box (*MAT 1-2*, red) and alpha box (*MAT1-1*, dark red) in the Ajellomycetaceae. The more basal species, *P. hystricis*, has both mating type idiomorphs intact, being the only homothallic species identified so far within the Ajellomycetaceae. The locus has similar size to other Ajellomycetaceae species with 21 kb average from *SLA2* (blue) to *APN2*/*COX13* (green/orange), but not *Blastomyces dermatitidis/gilchristii* (60 kb; Li W et al., 2013). A synteny inversion encompassing *APN2* and *COX13* is uniquely observed in this ancestor relative to the rest of Ajellomycetaceae species included in this study, including *H. griseus*. Grey genes are mating locus hypothetical proteins (MLHP).

666 **References**

667


668 1. Klein, B. S. & Tebbets, B. Dimorphism and virulence in fungi. *Curr Opin Microbiol* **10,** 314–9
669 (2007).
670 2. Sil, A. & Andrianopoulos, A. Thermally dimorphic human fungal pathogens--Polyphyletic
671 pathogens with a convergent pathogenicity trait. *Cold Spring Harb Perspect Med* **5,** a019794
672 (2015).
673 3. Gauthier, G. M. Dimorphism in fungal pathogens of mammals, plants, and insects. *PLoS Pathog*
674 **11,** e1004608 (2015).
675 4. England, D. M. & Hochholzer, L. Adiaspiromycosis: an unusual fungal infection of the lung.
676 Report of 11 cases. *Am J Surg Pathol* **17,** 876–86 (1993).
677 5. Munoz, J. F. *et al.* The dynamic genome and transcriptome of the human fungal pathogen
678 *Blastomyces* and close relative *Emmonsia*. *PLoS Genet* **11,** e1005493 (2015).
679 6. Herr, R. A. *et al.* Phylogenetic analysis of *Lacazia loboi* places this previously uncharacterized
680 pathogen within the dimorphic Onygenales. *J Clin Microbiol* **39,** 309–14 (2001).
681 7. Kenyon, C. *et al.* A dimorphic fungus causing disseminated infection in South Africa. *N Engl J*
682 *Med* **369,** 1416–1424 (2013).
683 8. Schwartz, I. S. *et al.* 50 Years of *Emmonsia* disease in humans: the dramatic emergence of a
684 cluster of novel fungal pathogens. *PLoS Pathog* **11,** e1005198 (2015).
685 9. Dukik, K. *et al.* Novel taxa of thermally dimorphic systemic pathogens in the Ajellomycetaceae
686 (Onygenales). *Mycoses* **60,** 296–309 (2017).
687 10. Wang, O. *et al.* A novel dimorphic pathogen, *Emergomyces orientalis* (Onygenales), agent of
688 disseminated infection. *Mycoses* **(in press),** (2017).
689 11. Marin-Felix, Y. *et al. Emmonsiellopsis*, a new genus related to the thermally dimorphic fungi of
690 the family Ajellomycetaceae. *Mycoses* **58,** 451–60 (2015).
691 12. Untereiner, W. A. *et al.* The Ajellomycetaceae, a new family of vertebrate-associated
692 Onygenales. *Mycologia* **96,** 812–821 (2004).
693 13. Untereiner, W. A., Scott, J. A., Naveau, F. A., Currah, R. S. & Bachewich, J. Phylogeny of
694 *Ajellomyces*, *Polytolypa* and *Spiromastix* (Onygenaceae) inferred from rDNA sequence and
695 non-molecular data. *Stud Mycol* **47,** 25–35 (2002).
696 14. Desjardins, C. A. *et al.* Comparative genomic analysis of human fungal pathogens causing
697 paracoccidioidomycosis. *PLoS Genet* **7,** e1002345 (2011).
698 15. Munoz, J. F. *et al.* Genome diversity, recombination, and virulence across the major lineages of
699 *Paracoccidioides*. *mSphere* **1,** (2016).
700 16. Neafsey, D. E. *et al.* Population genomic sequencing of *Coccidioides* fungi reveals recent
701 hybridization and transposon control. *Genome Res* **20,** 938–46 (2010).
702 17. Sharpton, T. J. *et al.* Comparative genomic analyses of the human fungal pathogens
703 *Coccidioides* and their relatives. *Genome Res* **19,** 1722–31 (2009).
704 18. Burmester, A. *et al.* Comparative and functional genomics provide insights into the pathogenicity
705 of dermatophytic fungi. *Genome Biol* **12,** R7 (2011).
706 19. Martinez, D. A. *et al.* Comparative genome analysis of *Trichophyton rubrum* and related
707 dermatophytes reveals candidate genes involved in infection. *MBio* **3,** e00259-12 (2012).
708 20. Whiston, E. & Taylor, J. W. Comparative phylogenomics of pathogenic and nonpathogenic
709 species. *G3 Bethesda* **6,** 235–44 (2015).
710 21. Munoz, J. F. *et al.* Genome update of the dimorphic human pathogenic fungi causing
711 paracoccidioidomycosis. *PLoS Negl Trop Dis* **8,** e3348 (2014).

712 22. Sigler, L. *Ajellomyces crescens* sp. nov., taxonomy of *Emmonsia* spp., and relatedness with
713 *Blastomyces dermatitidis* (teleomorph *Ajellomyces dermatitidis*). *J Med Vet Mycol* **34,** 303–14
714 (1996).
715 23. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in
716 eukaryotic genomes. *Bioinformatics* **23,** 1061–7 (2007).
717 24. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO:
718 assessing genome assembly and annotation completeness with single-copy orthologs.
719 *Bioinformatics* **31,** 3210–3212 (2015).
720 25. Peterson, S. W. & Sigler, L. Molecular genetic variation in *Emmonsia crescens* and *Emmonsia*
721 *parva*, etiologic agents of adiaspiromycosis, and their phylogenetic relationship to *Blastomyces*
722 *dermatitidis* (*Ajellomyces dermatitidis*) and other systemic fungal pathogens. *J Clin Microbiol* **36,**
723 2918–25 (1998).
724 26. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary
725 identification. *Nucleic Acids Res.* **45,** W36–W41 (2017).
726 27. Gamble, W. R., Gloer, J. B., Scott, J. A. & Malloch, D. Polytolypin, a new antifungal triterpenoid
727 from the coprophilous fungus *Polytolypa hystricis*. *J. Nat. Prod.* **58,** 1983–1986 (1995).
728 28. Saier, M. H., Jr., Tran, C. V. & Barabote, R. D. TCDB: the Transporter Classification Database
729 for membrane transport protein analyses and information. *Nucleic Acids Res* **34,** D181-6 (2006).
730 29. Beyhan, S., Gutierrez, M., Voorhies, M. & Sil, A. A temperature-responsive network links cell
731 shape and virulence traits in a primary fungal pathogen. *PLoS Biol* **11,** e1001614 (2013).
732 30. Nemecek, J. C., Wuthrich, M. & Klein, B. S. Global control of dimorphism and virulence in fungi.
733 *Science* **312,** 583–8 (2006).
734 31. Rappleye, C. A., Eissenberg, L. G. & Goldman, W. E. *Histoplasma capsulatum alpha*-(1,3)-
735 glucan blocks innate immune recognition by the beta-glucan receptor. *Proc Natl Acad Sci USA*
736 **104,** 1366–1370 (2007).
737 32. Goldberg, J. M. *et al.* Kinannote, a computer program to identify and classify members of the
738 eukaryotic protein kinase superfamily. *Bioinformatics* **29,** 2387–94 (2013).
739 33. Lu, T., Yao, B. & Zhang, C. DFVF: database of fungal virulence factors. *Database J. Biol.*
740 *Databases Curation* **2012,** (2012).
741 34. Winnenburg, R. *et al.* PHI-base: a new database for pathogen host interactions. *Nucleic Acids*
742 *Res.* **34,** D459–D464 (2006).
743 35. Casadevall, A., Steenbergen, J. N. & Nosanchuk, J. D. `Ready made' virulence and `dual use'
744 virulence factors in pathogenic environmental fungi-the *Cryptococcus neoformans* paradigm.
745 *Curr Opin Microbiol* **6,** 332–337 (2003).
746 36. Derengowski Lda, S. *et al.* The transcriptional response of *Cryptococcus neoformans* to
747 ingestion by *Acanthamoeba castellanii* and macrophages provides insights into the evolutionary
748 adaptation to the mammalian host. *Eukaryot Cell* **12,** 761–74 (2013).
749 37. Pigosso, L. L. *et al. Paracoccidioides brasiliensis* presents metabolic reprogramming and
750 secretes a serine proteinase during murine infection. *Virulence* **0,** 1–18 (2017).
751 38. Li, W. *et al.* Identification of the mating-type (MAT) locus that controls sexual reproduction of
752 *Blastomyces dermatitidis*. *Eukaryot Cell* **12,** 109–17 (2013).
753 39. Li, W., Metin, B., White, T. C. & Heitman, J. Organization and evolutionary trajectory of the
754 mating type (MAT) locus in dermatophyte and dimorphic fungal pathogens. *Eukaryot Cell* **9,** 46–
755 58 (2010).
756 40. Hoog, G. S. de *et al.* Toward a novel multilocus phylogenetic taxonomy for the dermatophytes.
757 *Mycopathologia* **182,** 5–31 (2017).
758 41. Steenbergen, J. N., Shuman, H. A. & Casadevall, A. *Cryptococcus neoformans* interactions with
759 amoebae suggest an explanation for its virulence and intracellular pathogenic strategy in
760 macrophages. *Proc. Natl. Acad. Sci.* **98,** 15245–15250 (2001).

761 42. Nielsen, K. *et al. Cryptococcus neoformans* α strains preferentially disseminate to the central
762 nervous system during coinfection. *Infect. Immun.* **73,** 4922–4933 (2005).
763 43. Rappleye, C. A. & Goldman, W. E. Defining virulence genes in the dimorphic fungi. *Annu Rev*
764 *Microbiol* **60,** 281–303 (2006).
765 44. Sugiyama, M., Ohara, A. & Mikawa, T. Molecular phylogeny of onygenalean fungi based on
766 small subunit ribosomal DNA (SSU rDNA) sequences. *Mycoscience* **40,** 251–258 (1999).
767 45. Heitman, J., Sun, S. & James, T. Y. Evolution of fungal sexual reproduction. *Mycologia* **105,** 1–
768 27 (2013).
769 46. Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel
770 sequence data. *Proc Natl Acad Sci U A* **108,** 1513–8 (2011).
771 47. Haas, B. J., Zeng, Q., Pearson, M. D., Cuomo, C. A. & Wortman, J. R. Approaches to Fungal
772 Genome Annotation. *Mycology* **2,** 118–141 (2011).
773 48. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and
774 the Program to Assemble Spliced Alignments. *Genome Biol* **9,** R7 (2008).
775 49. Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for
776 eukaryotic genomes. *Genome Res* **13,** 2178–89 (2003).
777 50. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
778 thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–90 (2006).
779 51. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7,** e1002195 (2011).
780 52. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in
781 functional genomics research. *Bioinformatics* **21,** 3674–6 (2005).
782 53. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource
783 for glycogenomics. *Nucleic Acids Res* **37,** D233-8 (2009).
784 54. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids*
785 *Res* **38,** D227-33 (2010).
786 55. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci*
787 *U A* **100,** 9440–5 (2003).
788
789