Domestication of the emblematic white cheese-making fungus *Penicillium camemberti* and its diversification into two varieties

Jeanne Ropars[1], Estelle Didiot[1], Ricardo C. Rodríguez de la Vega[1], Bastien Bennetot[1], Monika Coton[2], Elisabeth Poirier[2], Emmanuel Coton[2], Alodie Snirc[1], Stéphanie Le Prieur[1] and Tatiana Giraud[1]

[1]Ecologie Systematique et Evolution, Universite Paris-Saclay, CNRS, AgroParisTech, 91400 Orsay, France

[2]Univ Brest, Laboratoire Universitaire de Biodiversité et Écologie Microbienne, F-29280 Plouzané, France

Correspondence: Jeanne Ropars, jeanne.ropars@u-psud.fr

Keywords: domestication, fungi, cheese, *Penicillium,* mycotoxin, horizontal gene transfer

Summary

Domestication involves recent adaptation under strong human selection and rapid diversification, and therefore constitutes a good model for studies of these processes. We studied the domestication of the emblematic white mold *Penicillium camemberti*, used for the maturation of soft cheeses, such as Camembert and Brie, about which surprisingly little was known, despite its economic and cultural importance. Whole genome-based analyses of genetic relationships and diversity revealed that an ancient domestication event led to the emergence of the gray-green *P. biforme* mold used in cheese-making, from the blue-green wild *P. fuscoglaucum* fungus. Another much more recent domestication event led to the generation of the *P. camemberti* clonal lineage from *P. biforme*. Both these domestication events were associated with strong bottlenecks. *Penicillium biforme* displayed signs of phenotypic adaptation to cheese-making relative to *P. fuscoglaucum*, in terms of its whiter color, faster growth on cheese medium under cave conditions, lower levels of toxin production and greater ability to prevent the growth of other fungi. The *P. camemberti* lineage displayed even stronger signs of domestication for all these phenotypic features. We also identified two differentiated *P. camemberti* lineages linked to different kinds of cheeses with contrasted phenotypic features in terms of color, growth, toxin production and competitive ability. We have, thus, identified footprints of domestication in these fungi, with genetic differentiation between cheese and wild strains and specific phenotypic traits beneficial for cheese-making. This study has not only fundamental implications for our understanding of domestication but can also have important impacts on cheese-making.

Introduction

Understanding how organisms adapt to their environment is a key issue in evolutionary biology, requiring investigations of population subdivision, and levels of genetic and phenotypic diversity or adaptive divergence. Domestication is a good model for studies of adaptive divergence, as it involves recent adaptation events affecting known traits under strong human selection and rapid diversification. Several studies on domesticated animals (e.g. horse, dog, pig[1–3]) and plants (e.g. maize, apricot[4,5]) have improved our understanding of adaptive divergence by revealing how adaptation and diversification into varieties occurred. Maize, for example, has undergone major changes in phenotype compared to its wild relative (teosinte), including a decrease in tillering, and the development of larger, non-dehiscent grains[6]. Several different maize varieties have been selected for different usages, with sugar-rich varieties grown for human consumption as kernels and field corn varieties grown for animal feed. Similarly, a number of different *Brassica* varieties have been selected for their leaves (cabbage and kale), stems (kohlrabi), flower shoots (broccoli and cauliflower) or buds (Brussels sprouts). Dog breeds have also diversified greatly under human selection, with different breeds displaying very different behaviors and phenotypes. Such notable adaptation is often associated with a decrease in fitness in natural environments, with, for example, a decrease in or loss of sexual reproduction ability in bulldogs[7] and bananas[8], smaller brains and less acute sense organs in most domesticated animals[9].

Fungi are excellent models for studying evolution and adaptation in eukaryotes, given their many experimental assets[10], including their small genomes and tractability for laboratory experiments. They can therefore be used to address questions in evolutionary biology through complementary approaches, including genomic and experimental methods. Humans have domesticated several fungi for the fermentation of foods (e.g. for beer, bread, wine, dried sausage and cheese), the production of secondary metabolites used in pharmaceutics (e.g. penicillin), and for their nutritional and gustatory qualities (e.g. button and shiitake mushrooms)[11]. Despite their economic and industrial importance, and their utility as biological models for studying adaptive divergence, fungi used by humans have been little studied, with the exception of the budding yeast *Saccharomyces cerevisiae* used for beer, wine and bread production[12–22], the filamentous fungus *Aspergillus oryzae* used to ferment soy and rice products in Asia[23–25] and the blue-cheese mold *Penicillium roqueforti*[26–28]. Whole-genome analyses have revealed that *P. roqueforti* has been domesticated twice, in two independent events, resulting in one population specific to the Roquefort protected designation of origin (PDO), the other population being used worldwide for all types of blue cheeses[26,29]. The Roquefort population displays some genetic diversity and had beneficial traits for cheese

71 production before the industrial era, growing more slowly in cheese and displaying greater spore
72 production on bread, the traditional multiplication medium. By contrast, the non-Roquefort cheese
73 population is a clonal lineage with traits beneficial for industrial cheese production, such as high
74 levels of lipolytic activity, efficient cheese cavity colonization and high salt tolerance. Some of these
75 beneficial traits have been conferred by two large horizontally transferred genomic regions, *Wallaby*
76 and *CheesyTer*[27,28], in the clonal non-Roquefort cheese population. The non-Roquefort cheese *P.*
77 *roqueforti* population grows more slowly in harsh conditions (minimal medium) than the Roquefort
78 cheese population[28]. Whole-genome analyses of *S. cerevisiae* have also revealed the existence of a
79 cheese-specific population of this yeast differentiated from the populations used for alcohol or
80 bread production. The cheese *S. cerevisiae* population assimilates galactose more rapidly than *S.*
81 *cerevisiae* populations thriving in other food environments (e.g. beer, bread, wine) or in natural
82 environments (oak)[30].

83 The white mold *Penicillium camemberti* is used for the maturation of soft cheeses, such as Camembert,
84 Brie and Neufchatel (Figure 1A-C). It is thought to be a white mutant selected from the gray-green
85 species *P. commune* for its color at the start of the 20th century[31], and cultured clonally ever since.
86 However, very little is known about its taxonomic status, origin and diversity, despite its great
87 economic and cultural importance. In particular, its relationships to the closely related species *P.*
88 *biforme, P. caseifulvum, P. commune* and P. *fuscoglaucum,* and even possible overlaps with these species,
89 remain unclear. *Penicillium camemberti* has only ever been found in the cheese/meat environment.
90 *Penicillium commune* is used for the maturation of other types of cheese (e.g. hard and blue cheeses)
91 and in the production of dried sausages, is commonly found as spoiler of dairy products[32] and is
92 also widespread in non-food environments. Genetic analyses however suggested that *P. commune*
93 was an invalid species and proposed to reinstate two ancient species, *P. biforme* (named by Thom in
94 1910) and *P. fuscoglaucum* (named by Biourge in 1923), based on a few genetic sequences separating
95 *P. commune* strains isolated from cheese (including the *P. biforme* type) from *P. commune* strains
96 isolated from other environments (including the *P. fuscoglaucum* type)[33]. However, the most recent
97 taxonomic reference study of the *Penicillium* genus did not recognize *P. fuscoglaucum* as a valid
98 species[34]. *Penicillium caseifulvum* has also sometimes been advocated to constitute a separate species
99 in this clade[34], including cheese strains isolated from Danish blue and other German and French
100 cheeses, on the basis of colony morphology and their lack of production of cyclopiazonic acid, a
101 mycotoxin produced by *P. camemberti.* However, a study based on a small number of genetic
102 markers was unable to differentiate these putative *P. caseifulvum* strains from *P. camemberti*[33]. Parts of
103 the horizontally transferred *CheesyTer* and *Wallaby* regions have also been found in the reference *P.*
104 *camemberti* and *P. biforme* genomes[27,28].

105 The taxonomy of this clade thus remains unclear, and population structure has not been studied
106 with powerful genetic markers, despite the paramount importance of genetic relationships to our
107 understanding of domestication. Differences in phenotypic traits have not been extensively studied
108 either, but are also essential for determining whether or not domestication has occurred. Humans
109 may, indeed, have selected strains for specific traits beneficial for cheese production, such as faster
110 growth in cheese, attractive color and a greater ability to compete against food-spoiling
111 microorganisms, but without the production of detrimental extrolites, called mycotoxins.
112 Mycotoxins are natural secondary metabolites produced by fungi under certain conditions that can
113 inhibit competitor growth and reproduction and which can have a toxicological impact on humans
114 depending on the exposure conditions[35].

115 We addressed these questions of genetic and phenotypic differentiation, by collecting and
116 sequencing the genomes (ca. 35 Mb) of 35 strains isolated from the crusts of different cheeses from
117 around the world (e.g. Camembert, tommes, blue cheeses) and 26 strains from other environments
118 (e.g. leaf litter, sausages, ice and wood). We performed a whole genome-based analysis of
119 population structure, which revealed that the three species *P. biforme, P. camemberti* and *P. fuscoglaucum*
120 formed separate and specific genetic clusters, thus confirming the division of *P. commune* into *P.*
121 *biforme* and *P. fuscoglaucum*. The species name *P. commune* should therefore be avoided. We found
122 that *P. camemberti* and *P. biforme* were sister species, both specific to the cheese environment, whereas
123 the more distantly related species *P. fuscoglaucum* was mostly found in natural environments. These
124 relationships suggest an ancient domestication event separating *P. biforme* from *P. fuscoglaucum* and
125 a much more recent domestication event generating the *P. camemberti* clonal lineage from *P. biforme.*
126 Consistent with this scenario, we found evidence of phenotypic adaptation to cheese-making in *P.*
127 *biforme* relative to *P. fuscoglaucum*, with a whiter color, faster growth on cheese medium under cave
128 conditions and lower levels of toxin production. We also identified two differentiated *P. camemberti*
129 lineages with contrasting phenotypic features. One of these *P. camemberti* lineages contained the
130 fluffy white strains isolated from Camembert or Brie including the *P. camemberti* type strain, which
131 grew more slowly in cave conditions than the other tested *Penicillium* strains, and produced similar
132 amounts of cyclopiazonic acid to *P. biforme*. This *P. camemberti* lineage carried the horizontally
133 transferred *CheesyTer* and *Wallaby* regions except three strains isolated from Camembert cheese
134 before 1905. The gray-green *P. camemberti* lineage was isolated from cheeses other than Camembert,
135 carries *CheesyTer* but not the entire *Wallaby* region, grew faster than the white *P. camemberti* lineage,
136 produced no cyclopiazonic acid, due to a frameshift mutation, and included the *P. caseifulvum* type
137 strain. Both *P. camemberti* lineages grew more slowly in harsh conditions than the other two clades,
138 suggesting that adaptation to cheese-making has led to a degeneration of functions useful in wild

139 environments. Both *P. camemberti* lineages excluded fungal competitors more effectively than *P.*
140 *biforme,* the fluffy white Camembert lineage being more effective at exclusion than the gray-green
141 *P. camemberti* lineage. This study thus reveals footprints of domestication in *P. camemberti* and *P.*
142 *biforme*, with genetic differentiation between cheese and wild strains, and the identification of
143 specific phenotypic traits beneficial for cheese-making. We also reveal the existence of
144 diversification into several varieties, also displaying genetic and phenotypic differentiation (in terms
145 of growth rate, color, competitive abilities and production of mycotoxin), used for the production
146 of different cheese types.

147

148 # Results and Discussion

149

150 ## *Penicillium camemberti, P. biforme* and *P. fuscoglaucum* each form separate
151 specific genetic clusters

152 We collected and sequenced 61 strains with Illumina® technology, including 35 strains isolated
153 from the rinds of various types of cheese (e.g., tommes, Camembert, blue cheeses), 9 from
154 sausages, and 17 from non-food environments (e.g., wood or leaf litter), which were attributed by
155 their collectors to the species *P. camemberti, P. biforme, P. commune* or *P. fuscoglaucum* (Table S1). We
156 resequenced the reference genome of *P. camemberti* (LCP06093, also known as FM013, and initially
157 sequenced with 454 technology[27]) with PacBio® long-read technology and used the sequence
158 obtained for mapping. We identified 392,072 SNPs across all strains (Supplemental Data 1). We
159 investigated population structure with two clustering methods based on genetic differences and
160 free from assumptions about mating system and mode of reproduction. Principal component
161 analysis (PCA; Figure 2A) and neighbor net (splitstree) analysis (Figure 2B) identified three genetic
162 clusters, corresponding to *P. camemberti* (n=19), *P. biforme* (n=28) and *P. fuscoglaucum* (n=14). A
163 population structure analysis also identified the same three genetic clusters at K=3 (Figure 2C), the
164 K value at which the structure was the strongest and clearest.
165 The strains present in public collections under the name "*P. commune*" did not form a single cluster
166 nor even a monophyletic clade (black arrows in Figures 2A-C); indeed, the *P. commune* strains
167 isolated from natural environments clustered with *P. fuscoglaucum* whereas the *P. commune* strains
168 isolated from cheeses grouped with *P. biforme*, being genetically closer to *P. camemberti* than to *P.*
169 *fuscoglaucum* (Figures 2A-C). The *P. caseifulvum* type strain (LCP05630) and the *P. camemberti* type

170 strain belonged to different clusters at *K* values ≥ 5 in the population structure analysis (red stars
171 on Figure 2C).

172 *Penicillium camemberti* was found only in food (cheese, sausage, food waste or animal feed; Figure 2).
173 Only three of the 28 strains in the *P. biforme* cluster were isolated from environments other than
174 food: the atmosphere, ice and leaf litter. These three strains were not genetically differentiated from
175 those isolated from food (bread, cheese or sausage) and may therefore be feral strains, i.e. escaped
176 from food. The *P. fuscoglaucum* cluster included three strains isolated from food environments (two
177 strains from cheese and one from sausage) and 13 strains from natural environments and was thus
178 likely to constitute a genuine wild population. The three food strains clustered with the wild strains.

179 *Penicillium fuscoglaucum* displayed the highest degree of nucleotide diversity level (π = 0.177), with
180 about 196,000 SNPs. The *P. biforme* cluster had a higher genetic diversity (π = 0.095; 142,771 SNPs;
181 wider point dispersion in the PCA on Figure 2A and in the neighbor net on Figure 2B) than *P.*
182 *camemberti* (π = 0.005). The very low genetic diversity detected within *P. camemberti,* with only 0.02%
183 sites identified as polymorphic (i.e., only 8,180 SNPs), suggests that *P. camemberti* is a clonal lineage.

184 The long branches and the lack of cross-linking observed in the neighbor-net analysis further
185 confirmed the clonality of *P. camemberti*, contrasting with the footprints of recombination detected
186 within both *P. biforme* and *P. fuscoglaucum* (Figure 2B). In *P. fuscoglaucum,* abundant cross-linking was
187 observed, right to the branch edges (Figure 2B), reinforcing the view that *P. fuscoglaucum*
188 corresponded to a sexual population thriving in natural environments, as strains used in the food
189 industry are replicated clonally. We studied the mating-type genes controlling fungal mating
190 compatibility to determine the likelihood of sexual reproduction occurrence within each species.
191 In heterothallic fungal species, sexual reproduction occurs only between two haploid individuals
192 carrying different alleles (MAT1-1 and MAT1-2) at the mating-type locus. A lack of sexual
193 reproduction leads to relaxed selection on the mating-type genes, as they are no longer used, and a
194 departure from a balanced ratio between mating types. Neither mating-type allele presented any
195 evidence of loss-of-function mutations or SNPs in any of the genomes. In *P. fuscoglaucum,* we found
196 no significant departure from the 1:1 ratio of the two mating-type alleles expected under conditions
197 of regular sexual reproduction (χ_2 = 1.14; df = 1; P = 0.29). By contrast, *P. biforme* displayed a
198 significant departure from the 1:1 ratio (χ_2 = 14.29; df = 1; P < 1e$_{-3}$), with MAT1-2 strains
199 predominating. All 19 *P. camemberti* strains carried the MAT1-2 allele, providing further evidence
200 of the clonality of *P. camemberti*[36,37]. The observed clonality in *P. camemberti* may be explained by the
201 recent selection of a white mutant and its use in industrial production[38], whereas *P. biforme* is more
202 widely used in the production of artisanal cheeses and farmers may make use of diverse local strains.

203

204 Two genetic clusters within *Penicillium camemberti*

205 The population structure analysis at $K = 5$ (Figure 2C), the PCA and the splitstree without *P.*
206 *fuscoglaucum* (Figure S1) supported the existence of two genetic clusters within *P. camemberti*,
207 separating strains isolated from soft cheeses, such as Camembert, and strains isolated from other
208 kinds, such as Rigotte de Condrieu and Saint Marcellin. The *P. caseifulvum* type strain (red star on
209 Figure 2C) clustered with *P. camemberti* strains isolated from non-soft cheeses. The genetic cluster
210 including the *P. caseifulvum* type strain will be referred to hereafter as *P. caseifulvum,* and the genetic
211 cluster including the *P. camemberti* type strain will be referred to as *P. camemberti sensu stricto.* A recent
212 study described rapid phenotypic change from green-gray to white in a "wild *P. commune"* strain[39],
213 interpreting these changes as evidence that domestication can occur within four weeks. However,
214 we actually found here this *P. commune* strain to belong to the cheese *P. biforme* clade and therefore
215 to an already domesticated clade (Figure S2).
216 The population structure analysis (Figure 2C) also suggested further genetic subdivision within
217 each of *P. fuscoglaucum* and *P. biforme*; these subdivisions did not correspond to the environment
218 from which the isolate was obtained, type of cheese, or any other variable for which we had data;
219 it did not correspond to either a strong subdivision in the PCA or the neighbor-net (Figures 2 and
220 S1).
221

222 Presence/absence polymorphism of the two horizontally transferred regions
223 likely to be advantageous in cheese

224 By contrast to the first 454-sequenced FM013 reference genome[27], in which the *Wallaby*
225 horizontally transferred region was found to be fragmented, the obtained PacBio genome
226 assembly of the same strain showed that the part of *Wallaby* present in the genome formed a
227 single block (scaffold 18, positions 102,711 to 538,265). Two genes in *Wallaby* have been
228 highlighted as putatively important for cheese-making. One, encoding the *Penicillium* antifungal
229 protein (PAF)[27], was absent from FM013, whereas the other, encoding Hce2, a protein with
230 known antimicrobial activities, was present. All *P. camemberti sensu stricto* strains carried the same
231 *Wallaby* region as FM013, whereas strains from the *P. caseifulvum* lineage carried fragmented and
232 duplicated segments of the *Wallaby* fragment present in FM013, in which the Hce2 gene was
233 systematically present. We also found *Wallaby* fragmented with some duplicated regions in 18 *P.*
234 *biforme* strains; it was completely absent from the 10 remaining strains (Figure 3 and Figure S3).
235 *Wallaby* appeared to be even more fragmented in *P. fuscoglaucum* strains, always lacking Hce2 but
236 in one strain, and it was completely absent in nine strains.

237 The other horizontally transferred region thought to be beneficial for growth on cheese[28], the 57
238 kb *CheesyTer* region, was not located at the very end of a scaffold in *P. camemberti* (scaffold 17,
239 positions 690,636 to 751,978 in FM013), like its position in *P. roqueforti*. We found no strict
240 association between *CheesyTer* and *Wallaby* in the clades studied here, whereas all strains of *P.*
241 *roqueforti* studied so far carried either both or neither of these regions[28]. In *P. camemberti,* 16 of 19
242 strains carried the whole *CheesyTer* region, with an identity of 100% (Figures 3 and S4). The other
243 three *P. camemberti* strains (LCP00584T, LCP05527 and UBOCC-A-108096), belonging to the *P.*
244 *camemberti s.s.* cluster, completely lacked *CheesyTer* and an 80 kb-downstream fragment. These three
245 strains were isolated from Camembert cheeses around 1905, suggesting that the horizontal transfer
246 of *CheesyTer* to *P. camemberti* occurred after 1905, in a single strain that was then clonally cultured.
247 In *P. biforme,* 25 of 28 strains carried *CheesyTer*, with an identity >99% between strains. The identity
248 with *P. camemberti s.l.* was also >99%, except in the 5 kb region at the start of *CheesyTer,* in which
249 the similarity between the two species dropped to 90%. The 26 kb between *CheesyTer* and the end
250 of scaffold 17 in *P. camemberti s.l.* also displayed lower similarity (around 90%) between *P. camemberti*
251 *s.l.* and *P. biforme,* whereas it was identical within species. The differences between *P. camemberti* and
252 *P. biforme* in the flanking regions and at the start of *CheesyTer* appeared to be mainly due to C:G to
253 T:A mutations in *P. camemberti,* corresponding to the repeat-induced point (RIP) mutations specific
254 to fungi occurring in repeated sequences during sexual reproduction[40] (Figure S5). As *P. camemberti*
255 *s.l.* was found to be a clonal lineage and no RIP footprints were detected elsewhere in the genome
256 of FM013, these findings suggest that *CheesyTer* was transferred to *P. camemberti s.l.* with these RIPed
257 sequences already present. *CheesyTer* was completely absent from six strains of *P. fuscoglaucum*
258 (Figure 3) and partially present in the other strains. One *P. fuscoglaucum* strain (LCP00218, isolated
259 from rubber) carried two copies of a 30 kb *CheesyTer* terminal region. The two *CheesyTer* genes
260 involved in lactose metabolism[28], one encoding a lactose permease and the other a β-galactosidase,
261 were present in all strains carrying at least one part of *CheesyTer* in *P. camemberti, P. biforme* and *P.*
262 *fuscoglaucum* (Figure 3). *Penicillium biforme, P. fuscoglaucum* and *P. roqueforti* had no RIP footprints in the
263 *CheesyTer* region for which RIP mutations were detected in *P. camemberti,* so it was not possible to
264 identify the species or strain from which *CheesyTer* was transferred to *P. camemberti.*

265

266 Better phenotypes for cheese production in *P. camemberti s.l.* and *P. biforme*
267 than in the closely related wild species *P. fuscoglaucum*

268 Strains selected by humans would be expected to display specific traits beneficial for cheese
269 production, such as faster growth in cheese, colonies of an attractive color, salt tolerance, an
270 absence of harmful toxin production and efficiency at excluding undesirable microorganisms. We

271 performed a set of laboratory experiments to determine whether cheese strains had evolved traits
272 beneficial for cheese production not present in the closely related wild *P. fuscoglaucum* lineage
273 occurring in other environments.
274
275 We first investigated whether *P. camemberti s.l.* and *P. biforme* had acquired traits enabling them to
276 grow more rapidly than *P. fuscoglaucum* in cheese-making conditions. Rapidly growing fungi can be
277 beneficial for cheese-making, as they are more capable of excluding spoiler bacteria, yeasts and
278 molds[11] and promote faster cheese maturation. The process of cheese maturation begins with the
279 transfer of salted milk curd to dark caves at a low temperature and high humidity. We therefore
280 grew 61 strains on various media (unsalted cheese, salted cheese, minimal and malt media), in the
281 dark and at different temperatures and humidity levels (25°C at ambient relative humidity, 10°C at
282 98% humidity or 8°C at 100% humidity), and we measured colony radial growth after 10 days. We
283 found no significant effect on growth of the substrate from which the isolate was originally
284 obtained (i.e. food versus non-food). We found a significant effect on growth of the culture media,
285 species, temperature/humidity conditions, lineage within *P. camemberti s.l.* and the interaction
286 between culture medium and species (Table S2, Figure 4A). The significance of the interaction
287 between culture medium and temperature/humidity conditions (Table S2) reflected the very slow
288 growth at low temperatures, making it difficult to detect significant differences between media.
289 The two *P. camemberti s.l.* genetic clusters were genetically very closely related, but the *P. caseifulvum*
290 lineage grew similarly to *P. biforme* under Camembert cave conditions (i.e. 10°C at 98% humidity,
291 post-hoc Tukey honestly significant difference - HSD test $P = 1$), with greater radial growth on
292 unsalted cheese and salted cheese, and slower growth on malt and minimal media than *P.*
293 *fuscoglaucum*. By contrast, the *P. camemberti s.s.* lineage displayed weaker radial growth than any other
294 lineage on all media, but was fluffier. Furthermore, *P. camemberti s.l.* and *P. biforme* grew less rapidly
295 than *P. fuscoglaucum* on minimal and malt media, indicating a disadvantage for growth in harsh
296 conditions, as expected for cheese strains, due to relaxed selection on functions useful only in wild
297 environments. These findings demonstrated that *P. caseifulvum* and *P. biforme* had acquired growth
298 traits that were beneficial under cheese maturation conditions. In addition, the two *P. camemberti s.l.*
299 lineages displayed opposite growth patterns on unsalted and salted cheese, with strains used for
300 making Camembert or Brie (*P. camemberti s.s.*) growing much less radially on cheese than any other
301 strains, but growing much more vertically (i.e. being fluffier), with the mycelium growing up to the
302 lid of the Petri dishes (Figure 1C).
303

304 High salt concentrations in cheeses prevent the growth of contaminants and cheese fungi may have
305 adapted to such conditions. We found an effect of salt on *P. biforme* growth, with strains growing
306 more rapidly on salted than unsalted cheese medium (post-hoc Tukey HSD test $P = 0.01$) and
307 more rapidly on salted cheese medium than *P. camemberti s.l.* (post-hoc Tukey HSD test $P = 0.007$).
308 Salt had no effect on the growth of *P. camemberti s.l.* (post-hoc test $P = 1$) or *P. fuscoglaucum* (post-
309 hoc test $P = 1$). There may have been stronger selection for salt tolerance in *P. biforme,* which is
310 used to inoculate blue and goat cheeses, both being more salty than soft cheeses, such as Brie or
311 Camembert[41], for which *P. camemberti s.s.* is used.
312
313 We investigated whether cheese lineages had become whiter, which can be more attractive to some
314 consumers than gray-green mold on cheese, by comparing the opacity of lineages on cheese
315 medium, as opacity increases with the brightness and fluffiness of a colony. We found significant
316 effects of species and lineage on color within *P. camemberti s.l.* (Figure 4C and Table S2), with *P.*
317 *camemberti s.l.* being significantly more opaque (i.e. brighter and fluffier) and *P. camemberti s.s.* even
318 more so, than *P. caseifulvum*. This is consistent with the white and fluffy aspect of the crust of
319 Camembert and Brie, made with *P. camemberti s.s.*, whereas *P. caseifulvum* is found in cheeses with a
320 grayer and less fluffy crust, such as Saint Marcellin or Rigotte de Condrieu (Figures 1 and 4B-C).
321 The *P. caseifulvum* and *P. biforme* lineages did not differ significantly from each other (post-hoc Tukey
322 HSD test $P = 1$) and both were brighter than the wild blue-green *P. fuscoglaucum* (Figure 4B).
323
324 We also investigated whether cheese lineages produced smaller amounts of cyclopiazonic acid
325 (CPA), which is cytotoxic to humans[35,42], than strains isolated from other environments. *Penicillium*
326 *camemberti* has been reported to produce CPA on yeast extract sucrose (YES) medium and at very
327 low, non-toxic concentrations in cheese[43]. None of the *P. caseifulvum* strains tested here produced
328 CPA on YES (Figure 4D), consistent with findings for the type strain[44]. By contrast, *P. camemberti*
329 *s.s., P. biforme* and *P. fuscoglaucum* produced CPA, the highest levels being obtained with *P. biforme*
330 and *P. fuscoglaucum.* The CPA biosynthesis cluster in the genome appeared to be functional, with six
331 genes present (*cpaA*, *cpaD*, *cpaO*, *cpaH*, *cpaM* and *cpaT*) but not the *cpa*R gene encoding a regulatory
332 protein (Figure S6). The only exceptions were the six *P. caseifulvum* strains, in which a 2 bp deletion
333 in the *cpaA* gene led to a frameshift. The *cpaA* gene encodes a polyketide synthase/non-ribosomal
334 peptide synthase responsible for the first step of the CPA biosynthetic pathway so that a non-
335 functional protein probably prevents CPA production on all substrates. Humans have often
336 selected fungal strains unable to produce harmful toxins for use in food, as exemplified by the
337 *Aspergillus oryzae* strains used to ferment Asian food products, which do not produce aflatoxins,

338 whereas the wild ancestor, *A. flavus,* does[45]. In the blue cheese fungus *P. roqueforti*, strains belonging
339 to the non-Roquefort population were also found unable to produce mycophenolic acid due to a
340 174 bp deletion in the *mpaC* gene[46].
341
342 Cheese is a nutrient-rich environment in which many microorganisms can thrive, including
343 undesirable food spoilage organisms. We therefore investigated whether the cheese lineages were
344 better able to outcompete challengers. We allowed strains of *P. camemberti s.s.* (*n*=3), *P. caseifulvum*
345 (*n*=2) or *P. biforme* (*n*=10) to grow as lawns on the surface of the cheese medium, which were
346 inoculated 24 h later with a single spot, in the middle of the Petri dish, using a competitor:
347 *Geotrichum candidum* (*n*=5, two strains isolated from cheese and three from other environments), *P.*
348 *biforme* (*n*=6, three strains isolated from cheese and three from other environments), *P. fuscoglaucum*
349 (*n*=5, two strains isolated from cheese and three from other environments), or *P roqueforti* (*n*=12,
350 three strains from each of the four known genetic clusters[26]). The species and the lineages of *P.*
351 *camemberti s.l.* used as the lawn had significant effects on the growth of the challenger (Table S2,
352 Figure 5). The two lineages of *P. camemberti s.l.* prevented the growth of all challengers more
353 effectively than *P. biforme*, and *P. camemberti s.s.* was even more effective. Furthermore, we observed
354 significant differences between the species used as challengers, with *Geotrichum candidum* having the
355 highest growth differential on the various lawns, growing less well on a *P. camemberti s.s.* lawn than
356 other lawns. *Geotrichum candidum* is a fungus that is also present on the surface of soft cheeses. The
357 exclusion effect may be mediated by the fluffy morphology, resulting in the fungus occupying more
358 space and using more resources, and/or by biochemical interactions.
359

360 # Conclusion

361 Population structure analyses based on whole-genome sequences revealed that the three species,
362 *P. biforme, P. camemberti s.l.* and *P. fuscoglaucum,* each formed a specific genetic cluster, confirming the
363 separation of *P. commune* into *P. biforme* and *P. fuscoglaucum; P. commune* should therefore be
364 considered as an invalid species name and should not be used anymore as considering all *"P.*
365 *commune"* strains to be wild leads to wrong inferences about trait evolution[39]. We found that *P.*
366 *camemberti s.l.* and *P. biforme* were sister clades specific to the cheese environment, whereas the more
367 distantly related species *P. fuscoglaucum* was found mostly in natural environments. These
368 relationships and the lower diversity of *P. biforme* than *P. fuscoglaucum*, with even lower levels of
369 diversity in *P. camemberti s.l.*, suggest that an ancient domestication event led to the generation of *P.*
370 *biforme* from *P. fuscoglaucum* and that a much more recent domestication event led to *P. camemberti*

371 *s.l.* emerging from *P. biforme*. Domestication often leads to bottlenecks[47–50], which can be very severe
372 in clonally multiplied fungi[26,51]. Consistent with this scenario, we found evidence of phenotypic
373 adaptation to cheese-making in *P. biforme* relative to the wild *P. fuscoglaucum* species, with a whiter
374 color, faster growth on cheese medium under cave conditions and lower levels of toxin production.
375 These signs of domestication were even more marked in *P. camemberti s.l.* than in *P. biforme*.

376 We observed a similar evolution of traits as in populations of the domesticated blue-cheese fungus
377 *P. roqueforti*, which grows more rapidly on cheese medium, is more competitive against contaminant
378 microorganisms, and grows less well under harsh conditions than wild strains[26,28]. Such convergent
379 evolution under similar selective pressures suggests that evolution may repeat itself, as already
380 reported in natural populations of *Anolis* lizards[52,53], three-spine sticklebacks[54], Mexican
381 cavefishes[55], and cichlid fishes in African rifts[56] or crater lakes in central America[57,58]. Phenotypic
382 convergent evolution has been also reported in domesticated organisms, particularly in crop plants,
383 for the loss of seed shattering, minimization of seed dormancy and increase in seed size and
384 number, for example, with most of these changes resulting from different genomic alterations in
385 different species[59,60].

386 We identified two genetically differentiated *P. camemberti* varieties, *P. camemberti s.s.* and *P. caseifulvum,*
387 with contrasting phenotypic features, used in the production of different kinds of cheese. The *P.*
388 *camemberti s.s.* strains were white and were isolated from Camembert or Brie. Their radial growth
389 was slower but their mycelia were much fluffier than the other *Penicillium* strains tested (Figure 1C)
390 and they produced similar amounts of CPA compared to *P. biforme* strains. They excluded fungal
391 competitors effectively, as previously suggested[61], probably due to their fluffy morphology, taking
392 up the available space and monopolizing resources. *Penicillium caseifulvum* strains were gray-green,
393 unable to produce CPA, and were isolated from cheeses other than Camembert, such as St
394 Marcellin or Rigotte de Condrieu (Figure 1A). They displayed more rapid radial growth than *P.*
395 *camemberti s.s.* and grew similarly to *P. biforme* strains on cheese medium in cave conditions. The
396 existence of two genetically and phenotypically differentiated lineages with different uses in cheese-
397 making suggested that these lineages emerged as a result of different human selection pressures, as
398 reported for the domesticated blue-cheese fungus *P. roquefort*[26], wine-making yeasts[12,15], maize, rice,
399 tomatoes, dogs, chickens and horses[1–6].

400 *Penicillium camemberti s.s.* is the emblematic species used to inoculate soft cheeses, such as
401 Camembert and Brie. According to the technical specifications for Brie de Meaux and Brie de
402 Melun PDOs and Camembert, the crust of the cheese must be white and fluffy, and inoculation

403 with *P. candidum* (a synonym of *P. camemberti s.l.*) is even specified for Brie de Meaux. These
404 specifications are recent (20th century) and seem to have had a negative impact on the diversity of
405 the fungi used for making these kinds of cheeses, as a single clonal lineage now predominates. The
406 genetic diversity of *P. roqueforti* has also been greatly reduced by recent industrialization[26,29], although
407 PDO specifications that local strains must be used have protected diversity to some extent in
408 Roquefort cheeses. Camembert and Brie cheeses were gray-green before the 20th century[38], as
409 illustrated by a 19th century painting by Marie Jules Justin entitled "Symphonie des fromages en
410 Brie majeur - Nature morte au fromage". These historical records are consistent with our findings,
411 which suggest a first domestication event leading to the emergence of the gray-green mold *P.*
412 *biforme,* subsequently followed, around 1900, by the domestication of *P. camemberti* from *P. biforme,*
413 with the selection of two different varieties displaying more marked signs of domestication than *P.*
414 *biforme*.

415 These findings have industrial implications, as they reveal the existence of three closely related but
416 different lineages that have evolved traits beneficial for cheese-making, with different phenotypic
417 traits selected according to usage. This study should foster further research, as it would be
418 interesting to assess other important traits, such as volatile compound production, and the
419 efficiencies of lipolysis and proteolysis. Our findings raise questions about the use of limited
420 numbers of clonal strains for cheese-making, which tends to lead to degeneration, limiting the
421 possibilities for further improvement, which is currently a major concern in the agrofood sector[62],
422 despite the great geneticist Vavilov known for having identified the centers of origin of cultivated
423 plants long ago highlighting the importance of genetic diversity in domesticated organisms for
424 variety improvement and diversification[63].

425

426 # Materials and Methods

427 ## Strain collection and DNA extraction

428 We analyzed strains isolated from 34 cheeses from five countries around the world (e.g.
429 Camembert, Saint Marcellin, tomme, Sein de Nounou). We also collected strains from dried
430 sausages and moldy food (e.g. bread). Spores were sampled from the food sources and spread on
431 Petri dishes containing malt-agar medium, which were then incubated for five days at 25°C. A total
432 of 42 strains were obtained from public strain collections (the Laboratoire de Cryptogamie (LCP
433 strains) at the National Museum of Natural History in Paris (France), the Laboratoire Universitaire
434 de Biodiversité et Ecologie Microbienne (LUBEM, UBOCC strains) in Plouzané (France) and the

Westerdijk Fungal Biodiversity Institute (CBS strains) in Utrecht (The Netherlands)). We obtained strains from natural environments (e.g. wood or natural cave walls) from nine different countries. Detailed information about the strains used in this study can be found in the Table S1. For each strain, single-spore cultures were generated by a dilution method, to ensure that only a single haploid genotype was cultured for each strain. We checked the species identification of all strains by Sanger sequencing of the β-tubulin gene (primers Bt2a - GGTAACCAAATCGGTGCTGCTTTC and bt2b - AACCTCAGTGTAGTGACCCTTGGC[64]) and the PC4 microsatellite flanking regions[33] (primers PC4F – CAAGCTGGCCGATAACCTG and PC4R – CCATCCGCTTGATTTCTCCT), to distinguish between *P. biforme* and *P. camemberti s. l.* All the strains were added to the ESE (Ecology Systematics Evolution Laboratory) collection, under an ESE accession number, and are available from the corresponding collections. For each strain, we used the Nucleospin Soil Kit (Machernary-Nagel, Düren, Germany) to extract DNA from fresh haploid mycelium grown for five days on malt agar.

Genome sequencing, assembly and mapping

Sequencing was performed with Illumina HiSeq 2500 paired-end technology (Illumina Inc.), with a mean insert size of 400 bp, at the INRA GenoToul platform, to obtain 10-50X coverage (Table S1). In addition, the genome of the *P. camemberti* LCP06093 (also known as FM013) reference strain, initially sequenced with 454 technology[27], was resequenced at 34x using the PacBio's long-read technology, assembled with Canu 1.8[65] and polished with pilon 1.23[66]. The PacBio assembly and reads for all samples have been deposited at the NCBI Sequence Read Archive under BioProject ID XXXX.

Reads were trimmed with Trimmomatic[67] (cut adapter, options PE, LEADING:3 TRAILING:3 SLIDINGWINDOW:4:25 MINLEN:36) and mapped onto the high-quality PacBio reference genome of LCP06093 with Bowtie2[68] (options very-sensitive-local, --phred33 PL:ILLUMINA -X 1000). Mapped reads were filtered for PCR duplication with picard tools (http://broadinstitute.github.io/picard) and realigned on the PacBio reference genome with GATK[69]. Single-nucleotide polymorphisms (SNPs) were called with GATK HaplotypeCaller, which provides one gVCF per strain (option -ERC GVCF). GVCFs were combined with GATK CombineGVCFs, genotyped with GATK genotype and hard filtered in accordance with GATK best practice workflow recommendations (https://software.broadinstitute.org/gatk/best-practices/).

Statistics of population genetics

469 Nucleotide diversity (π) was calculated with the R package pegas[70] for the three species *P. biforme,*
470 *P. fuscoglaucum* and *P. camemberti s.l.*. We also calculated π considering *P. camemberti s.s.* and *P.*
471 *caseifulvum* as separate genetic clusters.
472

473 ## Genetic structure

474 We used the dataset for 392,072 SNPs to infer a finer population structure. We inferred individual
475 ancestry from genotype likelihoods based on realigned reads, by assuming a known number of
476 admixing populations, ranging from $K = 2$ to $K = 6$, using NgsAdmix from the ANGSD package[71].
477 Neighbor-joining trees were generated with the R package ape[72]. We used the R package phangorn[73]
478 for neighbor-net analyses, and the prcomp function of R for principal component analysis (PCA).
479

480 ## Laboratory experiments

481 *Sampling and calibration*

482 All experiments were performed on the whole collection of *P. camemberti s.l.* (n = 20), *P. biforme* (n
483 = 28) and *P. fuscoglaucum* (n = 13) strains, including 34 strains isolated from cheeses, 8 from dried
484 sausages and 19 strains isolated from environments other than food. Experiments were initiated
485 with spore suspensions calibrated to 10^7 spores/mL with a hemocytometer, under the constraint
486 of the low rate of sporulation in *P. camemberti s.l.*.
487

488 *Growth in different conditions and on different media*

489 We investigated whether strains isolated from cheese displayed faster or slower radial growth than
490 strains isolated from other environments when cultured on cheese or other substrates, and under
491 conditions similar to those in maturation caves or other conditions. We prepared four different
492 culture media: a cheese medium without salt, a salted cheese medium (17g/L, corresponding to a
493 typical cheese), a malt medium and a minimal medium. The cheese media were produced from an
494 unsalted drained cow's milk cheese from Coubertin Farm in Saint Rémy-les-Chevreuse (France),
495 as previously described[61]; we added five drops of blue food coloring to these media, to make it
496 easier to distinguish white fungal colonies from the medium. The cheese media were rich in lipids
497 and proteins, whereas the malt medium (20 g/L) was rich in carbohydrates. The minimal medium[74]
498 contained only the trace elements necessary for fungal survival (i.e. iron sulfate, zinc sulfate, boric
499 acid, magnesium chloride, copper sulfate, ammonium heptamolybdate, cobalt chloride, EDTA,
500 sodium nitrate, potassium chloride, potassium phosphate and magnesium sulfate). All media were
501 sterilized in an autoclave (121°C for 20 min for the malt and minimal media, 110°C for 15 min for

502 the cheese media). Each 90 mm-diameter Petri dish was filled with 20 mL of the appropriate
503 medium.
504
505 We allowed the fungi to grow in the dark, under three different conditions for each strain and each
506 medium: 10°C with 98% humidity (Camembert cave conditions), 8°C with 85% humidity (cave
507 conditions for other cheeses) and 25°C with ambient relative humidity (ambient conditions). We
508 had 244 Petri dishes in total for each set of conditions, and we left the colonies to grow for 10
509 days. Images of the Petri dishes were obtained with a Scan 1200 from Interscience and analyzed
510 with IRIS[75] for growth and color-opacity measurements.
511
512 *Mycotoxin production*
513 For measurements of mycotoxin production, we used 1 µL of calibrated spore suspension (10^7
514 spores/mL) for each of the 61 strains to inoculate YES (yeast extract sucrose) agar medium
515 buffered at pH 4.5 with phosphate-citrate buffer and characterized by a high C/N ratio to favor
516 mycotoxin production as already described[46]. Each culture was performed in triplicate for myctoxin
517 analyses as well as for fungal dry weight measurements. The plates were incubated at 25°C in the
518 dark for 10 days and were then stored at -20°C until mycotoxin analysis.
519 For mycotoxin extractions[46], we homogenized the thawed samples with an Ultraturrax T25 (IKA,
520 Heidelberg, Germany) before recuperating 4g aliquots. Then, 25 mL of acetonitrile (ACN)
521 supplemented with 0.1% formic acid (v/v) was added, samples were vortexed for 30 sec followed
522 by 15 min sonication. Extracts were then centrifuged for 10 min at 5000g and the recuperated
523 supernatants were directly filtered through 0.2 µm PTFE membrane filters (GE Healthcare Life
524 Sciences, UK) into amber vials. All samples were stored at -20°C until analyses.
525 Mycotoxin detection and quantification were performed using an Agilent 6530 Accurate-Mass
526 Quadropole Time-of-Flight (Q-TOF) LC/MS system equipped with a Binary pump 1260 and
527 degasser, well plate autosampler set to 10°C and a thermostated column compartment. Filtered
528 samples (2µl) were injected into a ZORBAX Extend C-18 column (2.1x50mm and 1.8 µm, 600
529 bar) maintained at 35°C with a flow rate of 0.3 ml/min using mobile phase A (milli-Q water +
530 0.1% formic acid (v/v) and 0.1% ammonium formate (v/v) and mobile phase B (ACN + 0.1%
531 formic acid). Mobile phase B was maintained at 10% for 4 min followed by a gradient from 10 to
532 100% for 16 min. Then, mobile phase B was maintained at 100% for 2 min before a 5 min post-
533 time. Cyclopiazonic acid (CPA) was ionized in electrospray ionization mode ESI+ in the mass
534 spectrometer with the following parameters: capillary voltage 4 kV, source temperature 325°C,
535 nebulizer pressure 50 psig, drying gas 12 l/min, ion range 100-1000 m/z. CPA had a retention time

536 of 15.3 min and was quantified using the $[M+H]^+$ 337.1545 m/z ion and $[M+Na]^+$ 359.1366
537 qualifier ion (ESI+). Other extrolites included in our internal database that were targeted (in ESI+
538 and ESI- modes) were andrastin A, citreoviridin, citrinin, ermefortins A & B, (iso)-fumigaclavin A,
539 griseofulvin, meleagrin, mycophenolic acid, ochratoxin A, patulin, penitrem A, PR toxin,
540 roquefortin C, sterigmatocystin.
541 We used a matrix matched calibration curve (R^2 >0.99) for reliable mycotoxin quantification with
542 final concentrations ranging from 10 to 10000 ng/ml. Method performance and mycotoxin
543 determination was carried out as previously described[46]. Specific mycotoxin production was
544 expressed as ng per g fungal dry weight.
545
546 *Competition*
547 We inoculated salted cheese with 150 µL of a calibrated spore solution (10^7 spores/mL), which we
548 allowed to grow as a lawn. We used two *P. caseifulvum* strains (LCP05630 and ESE00019), three *P.*
549 *camemberti s.s.* strains (FM013, LCP04810 and LCP05527), five *P. biforme* strains isolated from cheese
550 (ESE00018, ESE00086, ESE00089, ESE00126, ESE00139) and five *P. biforme* strains isolated from
551 other environments (LCP05531, LCP06701, ESE00154, LCP06620 and LCP05496). After 24 h of
552 growth, we deposited a single 20 µL drop of calibrated spore solution from a challenger species
553 (10^7 spores/mL) in the middle of the Petri dish. The challenger species used were *G. candidum* (*n*=4,
554 ESE00163 and ESE00165 isolated from cheese, CBS11628 and VTTC4559 from other
555 environments), *P. biforme* (*n*=6, three isolated from cheese -ESE00126, ESE00089, ESE00018- and
556 three from other environments -ESE00154, LCP005531, LCP06701-), *P. fuscoglaucum* (*n*=5, two
557 isolated from cheese -ESE00090, LCP06621- and three from other environments -LCP000797,
558 LCP04799, LCP03239), and *P roqueforti* (*n*=12, three from each of the four genetic clusters
559 described in [26]: LCP06036, LCP06037, LCP06128, LCP06043, LCP06098, LCP06064, FM164,
560 LCP06137, LCP06131, LCP06173).
561

562 **Statistical tests**

563 All data were normalized with the bestNormalize R package[76]. Analysis of variance (ANOVA) was
564 performed, followed by post-ANOVA tests such as Tukey's HSD (honest significant difference)
565 test, performed with R.
566
567 We used the following normalizations for the various phenotypic traits tested. Ordered quantile
568 (ORQ) normalization, a one-to-one transformation for vectors with an arbitrary distribution that
569 generates normally distributed vectors, was applied to diameter measurements and competition

data. Yeo-Johnson transformation was used to obtain uniform distributions for color opacity and mycotoxin data.

Acknowledgments

We thank everyone who sent us cheese crusts, Steve Labrie for the LMA strain, and Joëlle Dupont for strains from the National Museum of Natural History in Paris.

This work was supported by the Fungadapt ANR-19-CE20-0002-02 ANR grant to TG, JR, EC, MC and a Fondation Louis D. grant from the Institut de France to TG.

Author contributions

JR and TG designed and supervised the study, and obtained funding. JR, ED, AS and SLP generated the data. JR, RRdlV and BB analyzed the genomes. JR, ED and AS performed the experiments. MC, EC and EP performed the mycotoxin and contributed to CPA biosynthetic gene cluster analyses. JR and ED analyzed the data from the laboratory experiments. JR and TG wrote the manuscript, with contributions from all the authors.

References

1. Warmuth, V. *et al.* European Domestic Horses Originated in Two Holocene Refugia. *PLoS ONE* **6**, e18194 (2011).
2. Axelsson, E. *et al.* The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360–364 (2013).
3. Frantz, L. A. F. *et al.* Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet* **47**, 1141–1148 (2015).
4. Hufford, M. B. *et al.* Comparative population genomics of maize domestication and improvement. *Nat Genet* **44**, 808–811 (2012).
5. Decroocq, S. *et al.* New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Mol. Ecol.* **25**, 4712–4729 (2016).
6. Doebley, J. The Genetics of Maize Evolution. *Annual Review of Genetics* **38**, 37–59 (2004).
7. Wydooghe, E., Berghmans, E., Rijsselaere, T. & Soom, A. V. International breeder inquiry into the reproduction of the English bulldog. *Vlaams Diergeneeskundig Tijdschrift* **82**, 38–43 (2013).
8. Perrier, X. *et al.* Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *PNAS* **108**, 11311–11318 (2011).
9. Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700–707 (2002).
10. Gladieux, P. *et al.* Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology* **23**, 753–773 (2014).
11. Dupont, J. *et al.* Fungi as a Source of Food. in *Microbiology Spectrum* vol. 5 (2017).
12. Almeida, P. *et al.* A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nature Communications* **5**, (2014).
13. Fay, J. C. & Benavides, J. A. Evidence for domesticated and wild populations of *Saccharomyces*

611 *cerevisiae*. *PLoS Genet* **1**, e5 (2005).
612 14. Gallone, B. *et al.* Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166**,
613 1397-1410.e16 (2016).
614 15. Legras, J.-L., Merdinoglu, D., Cornuet, J.-M. & Karst, F. Bread, beer and wine: *Saccharomyces*
615 *cerevisiae* diversity reflects human history. *Molecular Ecology* **16**, 2091–2102 (2007).
616 16. Libkind, D. *et al.* Microbe domestication and the identification of the wild genetic stock of
617 lager-brewing yeast. *PNAS* **108**, 14539–14544 (2011).
618 17. Marsit, S. *et al.* Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer
619 event in wine yeasts. *Molecular Biology and Evolution* **32**, 1695–1707 (2015).
620 18. Novo, M. *et al.* Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence
621 of the wine yeast *Saccharomyces cerevisiae* EC1118. *PNAS* **106**, 16333–16338 (2009).
622 19. Peter, J. *et al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–
623 344 (2018).
624 20. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
625 21. Spor, A. *et al.* Phenotypic and genotypic convergences are influenced by historical contingency
626 and environment in yeast. *Evolution* **68**, 772–790 (2014).
627 22. Dunn, B. *et al.* Recurrent rearrangement during adaptive evolution in an interspecific yeast
628 hybrid suggests a model for rapid introgression. *PLoS Genetics* **9**, e1003366 (2013).
629 23. Galagan, J. E. *et al.* Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus*
630 and *A. oryzae*. *Nature* **438**, 1105–1115 (2005).
631 24. Gibbons, J. G. *et al.* The evolutionary imprint of domestication on genome variation and
632 function of the filamentous fungus *Aspergillus oryzae*. *Current Biology* **22**, 1403–1409 (2012).
633 25. Machida, M. *et al.* Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* **438**, 1157–1161
634 (2005).
635 26. Dumas, E. *et al.* Independent domestication events in the blue-cheese fungus *Penicillium*
636 *roqueforti*. *Mol Ecol* mec.15359 (2020) doi:10.1111/mec.15359.
637 27. Cheeseman, K. *et al.* Multiple recent horizontal transfers of a large genomic region in cheese
638 making fungi. *Nat Commun* **5**, 2876 (2014).
639 28. Ropars, J. *et al.* Adaptive horizontal gene transfers between multiple cheese-associated Fungi.
640 *Current Biology* **25**, 2562–2569 (2015).
641 29. Gillot, G. *et al.* Insights into *Penicillium roqueforti* morphological and genetic diversity. *PLoS One*
642 **10**, e0129849 (2015).
643 30. Legras, J.-L. *et al.* Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable
644 genome plasticity and the footprints of domestication. *Mol Biol Evol* **35**, 1712–1727 (2018).
645 31. Pitt, J. I. *The genus Penicillium and its teleomorphic states Eupenicillium and Talaromyces.* (Academic
646 Press, 1979).
647 32. Garnier, L. *et al.* Diversity of spoilage fungi associated with various French dairy products.
648 *International Journal of Food Microbiology* **241**, 191–197 (2017).
649 33. Giraud, F. *et al.* Microsatellite loci to recognize species for the cheese starter and contaminating
650 strains associated with cheese manufacturing. *International Journal of Food Microbiology* **137**, 204–
651 213 (2010).
652 34. Visagie, C. M. *et al.* Identification and nomenclature of the genus *Penicillium*. *Studies in Mycology*
653 **78**, 343–371 (2014).
654 35. Hymery, N. *et al.* Filamentous Fungi and Mycotoxins in Cheese: A Review. *Comprehensive*
655 *Reviews in Food Science and Food Safety* **13**, 437–456 (2014).
656 36. Dyer, P. S. & O'Gorman, C. M. Sexual development and cryptic sexuality in fungi: insights
657 from Aspergillus species. *FEMS Microbiol. Rev.* **36**, 165–192 (2012).
658 37. Barve, M. P., Arie, T., Salimath, S. S., Muehlbauer, F. J. & Peever, T. L. Cloning and
659 characterization of the mating type (MAT) locus from *Ascochyta rabiei* (teleomorph: *Didymella*
660 *rabiei*) and a MAT phylogeny of legume-associated *Ascochyta* spp. *Fungal Genetics and Biology* **39**,
661 151–167 (2003).

662 38. Boisard, P. *Le camembert, mythe français.* (2007).
663 39. Bodinaku, I. *et al.* Rapid Phenotypic and Metabolomic Domestication of Wild *Penicillium* Molds
664 on Cheese. *mBio* **10**, e02445-19 (2019).
665 40. Selker, E. U., Cambareri, E. B., Jensen, B. C. & Haack, K. R. Rearrangement of duplicated
666 DNA in specialized cells of Neurospora. *Cell* **51**, 741–752 (1987).
667 41. Hashem, K. M., He, F. J., Jenner, K. H. & MacGregor, G. A. Cross-sectional survey of salt
668 content in cheese: a major contributor to salt intake in the UK. *BMJ Open* **4**, e005051–e005051
669 (2014).
670 42. Hymery, N., Masson, F., Barbier, G. & Coton, E. Cytotoxicity and immunotoxicity of
671 cyclopiazonic acid on human cells. *Toxicol In Vitro* **28**, 940–947 (2014).
672 43. Le Bars, J. Cyclopiazonic acid production by *Penicillium camemberti* Thom and natural
673 occurrence of this mycotoxin in cheese. *Appl Environ Microbiol* **38**, 1052–1055 (1979).
674 44. Lund, F., Filtenborg, O. & Frisvad, J. C. *Penicillium caseifulvum*, a new species found on
675 *Penicillium roqueforti* fermented cheeses. *Journal of Food Mycology* **1**, 95–101 (1998).
676 45. Barbesgaard, P., Heldt-Hansen, H. P. & Diderichsen, B. On the safety of Aspergillus oryzae:
677 a review. *Appl Microbiol Biotechnol* **36**, 569–572 (1992).
678 46. Gillot, G. *et al.* Functional diversity within the *Penicillium roqueforti* species. *Int. J. Food Microbiol.*
679 **241**, 141–150 (2017).
680 47. Zhu, Q., Zheng, X., Luo, J., Gaut, B. S. & Ge, S. Multilocus analysis of nucleotide variation
681 of Oryza sativa and its wild relatives: severe bottleneck during domestication of rice. *Mol. Biol.*
682 *Evol.* **24**, 875–888 (2007).
683 48. Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased
684 deleterious genetic variation in dogs. *PNAS* 201512501 (2015).
685 49. Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L. & Gaut, B. S. Investigation of the
686 bottleneck leading to the domestication of maize. *PNAS* **95**, 4441–4446 (1998).
687 50. Gaut, B. S., Díez, C. M. & Morrell, P. L. Genomics and the Contrasting Dynamics of Annual
688 and Perennial Domestication. *Trends in Genetics* **31**, 709–719 (2015).
689 51. Duan, S.-F. *et al.* The origin and adaptive evolution of domesticated populations of yeast from
690 Far East Asia. *Nat Commun* **9**, (2018).
691 52. Mahler, D. L., Ingram, T., Revell, L. J. & Losos, J. B. Exceptional convergence on the
692 macroevolutionary landscape in island lizard radiations. *Science* **341**, 292–295 (2013).
693 53. Thorpe, R. S., Barlow, A., Malhotra, A. & Surget-Groba, Y. Widespread parallel population
694 adaptation to climate variation across a radiation: implications for adaptation to climate
695 change. *Mol Ecol* **24**, 1019–1030 (2015).
696 54. Hohenlohe, P. A. *et al.* Population Genomics of Parallel Adaptation in Threespine Stickleback
697 using Sequenced RAD Tags. *PLoS Genetics* **6**, e1000862 (2010).
698 55. Duboué, E. R., Keene, A. C. & Borowsky, R. L. Evolutionary convergence on sleep loss in
699 cavefish populations. *Current Biology* **21**, 671–676 (2011).
700 56. Salzburger, W. The interaction of sexually and naturally selected traits in the adaptive
701 radiations of cichlid fishes. *Molecular Ecology* **18**, 169–185 (2009).
702 57. Elmer, K. R. *et al.* Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel
703 routes. *Nature Communications* **5**, 5168 (2014).
704 58. Elmer, K. R., Kusche, H., Lehtonen, T. K. & Meyer, A. Local variation and parallel evolution:
705 morphological and genetic diversity across a species complex of neotropical crater lake cichlid
706 fishes. *Philos Trans R Soc Lond B Biol Sci* **365**, 1763–1782 (2010).
707 59. Martínez-Ainsworth, N. E. & Tenaillon, M. I. Superheroes and masterminds of plant
708 domestication. *Comptes Rendus Biologies* **339**, 268–273 (2016).
709 60. Gross, B. L. & Olsen, K. M. Genetic perspectives on crop domestication. *Trends in Plant Science*
710 **15**, 529–537 (2010).
711 61. Decker, M. & Nielsen, P. V. The inhibitory effect of *Penicillium camemberti* and *Geotrichum*
712 *candidum* on the associated funga of white mould cheese. *International Journal of Food Microbiology*

104, 51–60 (2005).
62. Gouyon, P.-H., Leriche, H., Civard, A., Reeves, H. & Hulot, N. *Aux origines de l'environnement.* (Fayard, 2010).
63. Vavilov, N. I., Vavylov, M. I., Vavílov, N. Í. & Dorofeev, V. F. *Origin and Geography of Cultivated Plants.* (Cambridge University Press, 1992).
64. Glass, N. L. & Donaldson, G. C. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Appl Environ Microbiol* **61**, 1323–1330 (1995).
65. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
66. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9**, e112963 (2014).
67. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
69. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
70. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).
71. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
72. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
73. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
74. Hill, T. W. & Kafer, E. Improved protocols for *Aspergillus* minimal medium: trace element and minimal medium salt stock solutions. *Fungal Genetics Reports* **48**, 20–21 (2001).
75. Kritikos, G. *et al.* A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nature Microbiology* **2**, 17014 (2017).
76. Peterson, R. A. & Cavanaugh, J. E. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics* **0**, 1–16 (2019).

Figure Legends

Figure 1: Cheeses and fungi used for cheese-making.
A) A panel of soft cheeses with *Penicillium camemberti* or *P. biforme* on the crust. Left: Brie de Meaux protected designation of origin, top middle: Saint Marcellin, bottom middle: Camembert-like (buffalo milk), top right: Camembert protected designation of origin, bottom right: Selles fermier, Right picture: crottin de Chavignol (goat milk). B) Pictures of *P. biforme, P. camemberti* and *P. caseifulvum* on malt agar at 25°C after 10 days of growth. C) Petri dish with a white and fluffy *P. camemberti s.s.* strain (FM013) and a green rough *P. fuscoglaucum* strain (LCP06617).

Figure 2: Population structure of 61 strains from the *Penicillium camemberti* cheese-making fungal species complex, based on whole-genome data.
A) Principal component analysis (PCA), B) Neighbor-net analysis based on SNP data. Cross-linking indicates the likely occurrence of recombination. Branch lengths are shown and the scale bar represents 0.03 substitutions per site. C) Population subdivision inferred from $K = 2$ to $K = 6$. Each column represents a strain and colored bars represent their coefficients of membership for the various gene pools. For all panels, genetic clusters are represented by the same colors: dark purple for *P. fuscoglaucum,* light blue for *P. biforme,* dark green for *P. camemberti sensu stricto* and light green for *P. caseifulvum.* The strains identified as *P. commune* (invalid name) by collectors are shown in light purple. For PCA and neighbor-net analysis, symbols correspond to the environment of collection: circles for cheese, triangles for dried sausages and squares for other environments. On the structure analysis, letters below the plots indicate the origin of the strains: C for cheese, F for food strains other than cheese and E for other environments. Red asterisks indicated the type strains of *P. biforme, P. camemberti and P. caseifulvum.*

Figure 3: Presence/absence of the *Wallaby* and *CheesyTer* horizontally transferred regions in the 61 studied strains of the *Penicillium camemberti* species complex.
A) Neighbor-joining tree for the 61 strains. The strain name color indicates the environment of collection, black for food and blue for other environments. B) Presence/absence of the *Wallaby* region, as defined in the reference genome *P. camemberti s.s.* FM013. C) Presence/absence of the *CheeysyTer* region, as defined in the reference genome *P. camemberti s.s.* FM013. On panels B and C, the red bars represent the presence of the region in a given strain, white its absence, yellow the presence of a RIP region, light gray the presence of a highly fragmented region, dark gray the presence of a partially fragmented region and dark red a partially duplicated region. U before the strain ID indicates UBOCC strains.

Figure 4: Phenotypic traits distinguishing the four lineages *Penicillium biforme, P. caseifulvum, P. camemberti sensu stricto* and *P. fuscoglaucum.* A) Mean radial growth of the four lineages on unsalted cheese, salted cheese, malt and minimal media. B) Pictures of colonies on unsalted cheese, salted cheese, malt and minimal media; *P. fuscoglaucum*: LCP05471; *P. camemberti s.s.*: UBOCC-A-113052; *P. biforme*: ESE00211; *P. caseifulvum*: LCP06622. C) Difference in opacity between the four lineages on salted cheese at 25°C. D) Difference in production of cyclopiazonic acid between the four lineages on yeast extract sucrose. C-D) Horizontal lines of the boxplots represent the upper quartile, the median and the lower quartile. Each dot represents a strain and p-values are given for the tests of differences between the lineages (Wilcoxon test). The color indicates assignments to the four lineages as in other figures.

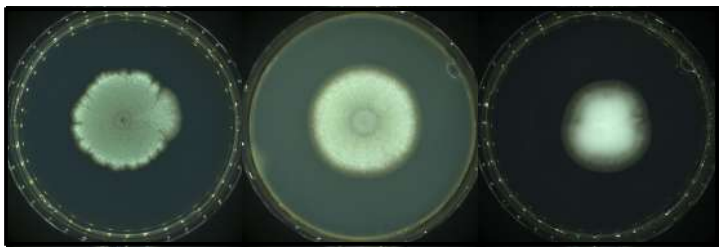
Figure 5: Competitive abilities of *Penicillium biforme, P. caseifulvum* and *P. camemberti sensu stricto* against *Geotrichum candidum, P. biforme, P. fuscoglaucum* and *P. roqueforti* competitors. Top: Pictures of *P. roqueforti* FM164, *Geotrichum candidum* ESE00163, *P. biforme* ESE00154 and *P. fuscoglaucum* ESE00221 on no lawn (first line), *P. biforme* ESE00086 lawn (second

797 line), *P. camemberti s.s.* LCP04810 (third line) and *P. caseifulvum* ESE00019 (fourth line) lawns, on
798 salted cheese. Bottom: Boxplots representing the differences of growth abilities of the competitors
799 on different species lawn. Horizontal lines of the boxplots represent the upper quartile, the median
800 and the lower quartile. Each dot represents a strain and p-values are given for the tests of
801 differences between the lineages (Wilcoxon test). The color indicates assignments to the three
802 lineages as in other figures.
803
804 **Figure S1: Population structure of 47 strains from the *Penicillium camemberti sensu lato***
805 **and *P. biforme* lineages, based on whole-genome data.**
806 A) Principal component analysis; B) Neighbor-net analysis. Cross-linking indicates the likely
807 occurrence of recombination Branch lengths are shown and the scale bar represents 0.01
808 substitutions per site. For both panels, the genetic clusters are represented by the same colors: light
809 blue for *P. biforme,* dark green for *P. camemberti sensu stricto* and light green for *P. caseifulvum.* The
810 strains identified as *P. commune* (invalid name) by collectors are shown in light purple. The symbols
811 correspond to the environment of collection: circles for cheese, triangles for dried sausages and
812 squares for other environments.
813
814 **Figure S2: Population structure of 61 strains from the *Penicillium camemberti* cheese-**
815 **making fungal species complex, based on whole-genome data with the addition of the**
816 **genome of the *P. commune* strain published in** [39]**.**
817 A) Principal component analysis; B) Neighbor-net analysis. Cross-linking indicates the likely
818 occurrence of recombination Branch lengths are shown and the scale bar represents 0.08
819 substitutions per site. For both panels, the genetic clusters are represented by the same colors: light
820 blue for *P. biforme,* dark green for *P. camemberti sensu stricto* and light green for *P. caseifulvum.* The
821 strain added in this analysis, 162 3FA WT8, identified as *P. commune* (invalid name) is shown in red.
822 The symbols correspond to the environment of collection: circles for cheese, triangles for dried
823 sausages and squares for other environments.
824
825 **Figure S3: Genomic scans of Illumina read coverage along the scaffold 18 where *Wallaby***
826 **is located.**
827 *Wallaby* is shown by red bars above plots. The vertical red line represents the position of the
828 Hce2 gene, involved in the competition against micro-organisms. When several strains showed
829 similar coverage patterns along the scaffold, only one representative plot is shown.
830
831 **Figure S4: Genomic scans of Illumina read coverage along the scaffold 17 from position**
832 **600,000 until the end of the scaffold where *CheesyTer* is located.**
833 *CheesyTer* is shown by gray rectangles on plots. Vertical red lines represent the position of the two
834 genes with relevant function for cheese-making previously identified in *CheesyTer*, i.e. a lactose
835 permease and a beta-galactosidase. When several strains showed similar coverage patterns along
836 the scaffold, only one representative plot is shown.
837
838 **Figure S5: Illustration of repeat-induced point mutation (RIP) footprints along the first**
839 **5kb of *CheesyTer* by comparing *Penicillium camemberti sensu lato, P. biforme, P.***
840 ***fuscoglaucum* and *P. roqueforti* genomes.** The nucleotide differences between species
841 corresponding to RIP, i.e., CpA to TpA or TpG to TpA are shown in red, while the other types
842 of variation are shown in blue and green.
843 A) Alignments of genomes from *P. camemberti sensu lato, P. biforme, P. fuscoglaucum and P. roqueforti*
844 genomes, showing in that mutations typical of RIP mostly occurred in *P. camemberti sensu lato* (red
845 traits). B) Percentages of nucleotide differences in *P. camemberti sensu lato* compared to the other
846 genomes showing that mutations typical of RIP are most abundant.
847

848 **Figure S6: Cyclopiazonic acid (CPA) biosynthesis cluster in the *Penicillium camemberti***
849 **species complex.** A) Schematic representation of the CPA biosynthesis cluster in the *P.*
850 *camemberti* reference genome FM013. B) Amino-acid alignment of the *CpaA* gene showing a 2bp
851 deletion in strains belonging to the *P. caseifulvum* genetic cluster, leading to a truncated protein.
852
853 **Table S1:** Description of the origin, donor and genome coverage of the 61 strains of the
854 *Penicillium camemberti* species complex used in this study.
855
856 **Table S2:** Results of statistical analyses performed for testing differences in traits between
857 *Penicillium camemberti s.s., P. caseifulvum, P. biforme* and *P. fuscoglaucum*. Significant P-values are indicated
858 with an asterisk and highlighted in gray.
859

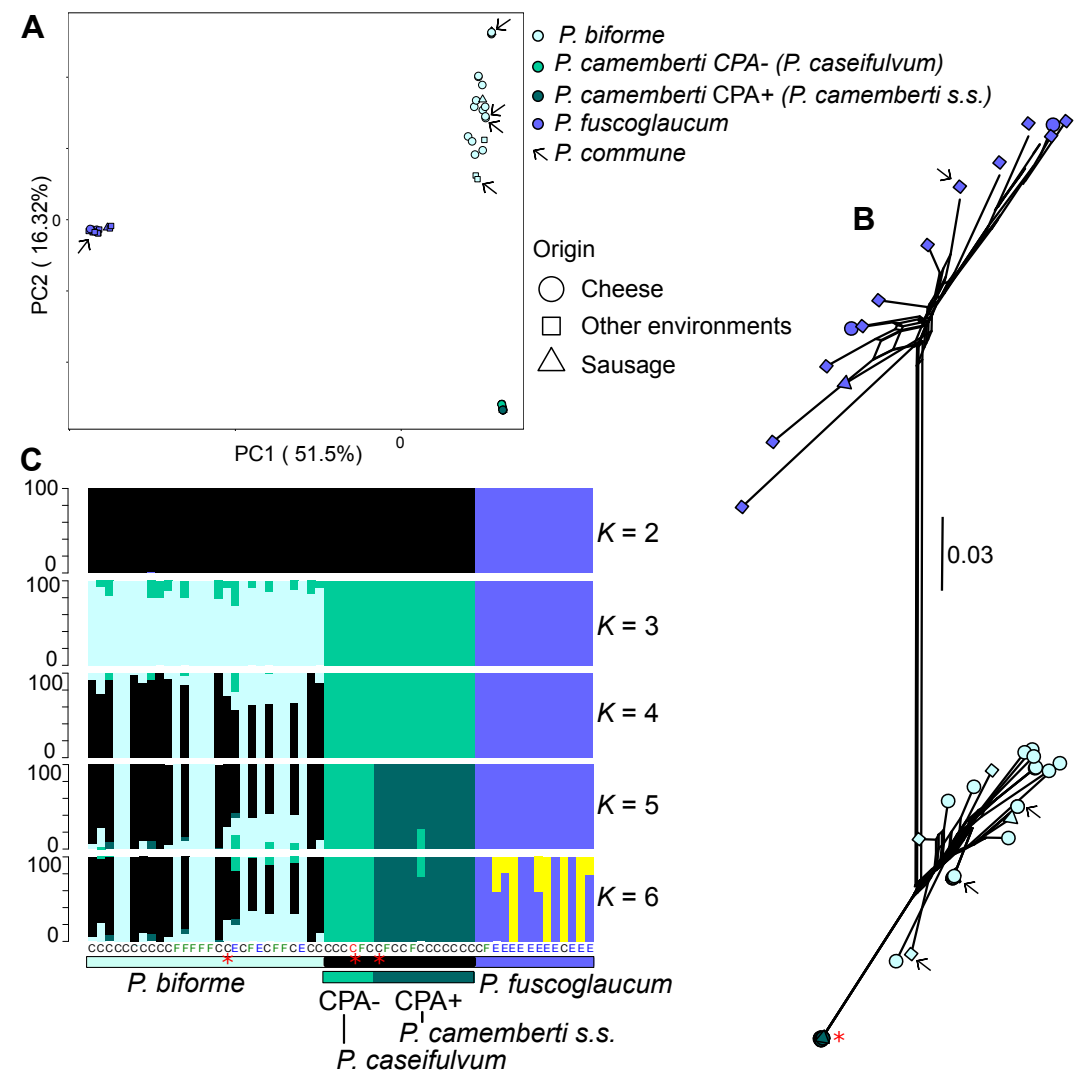
A

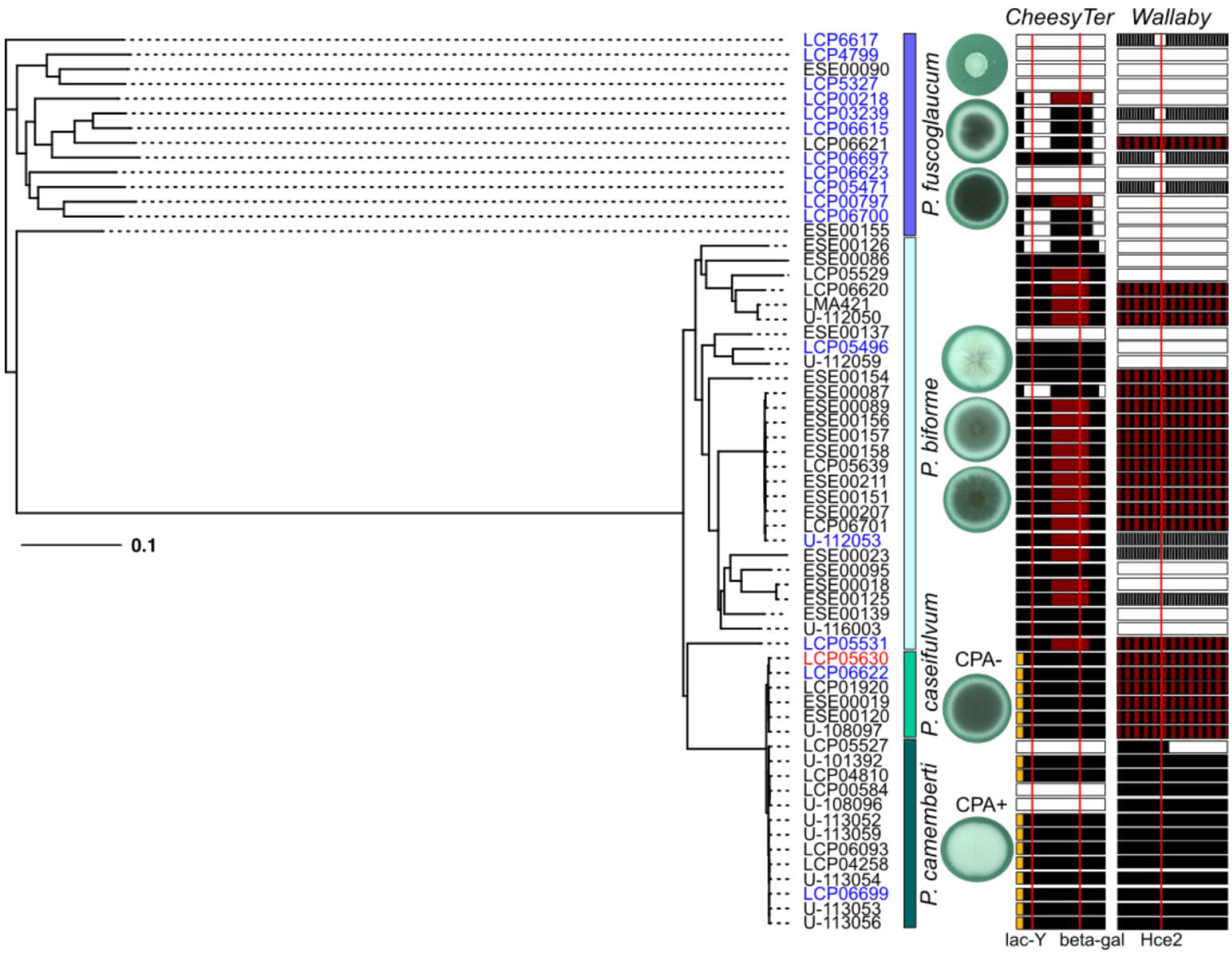
B

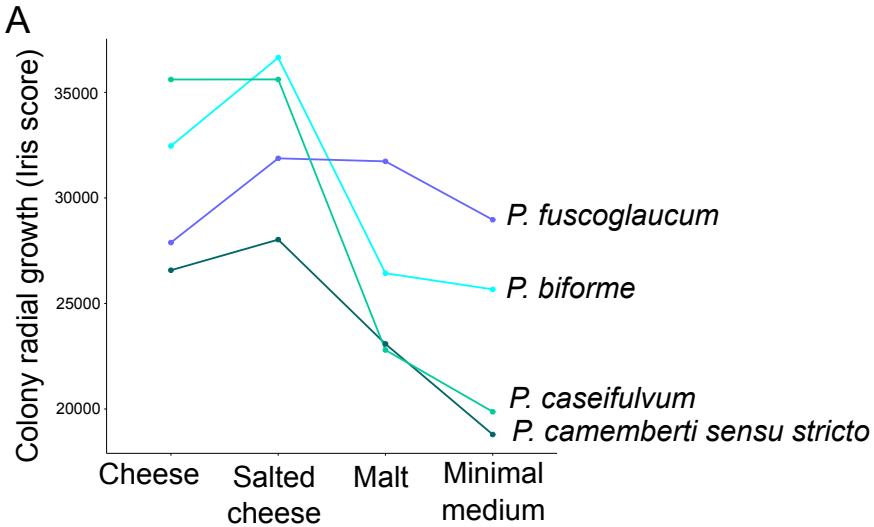


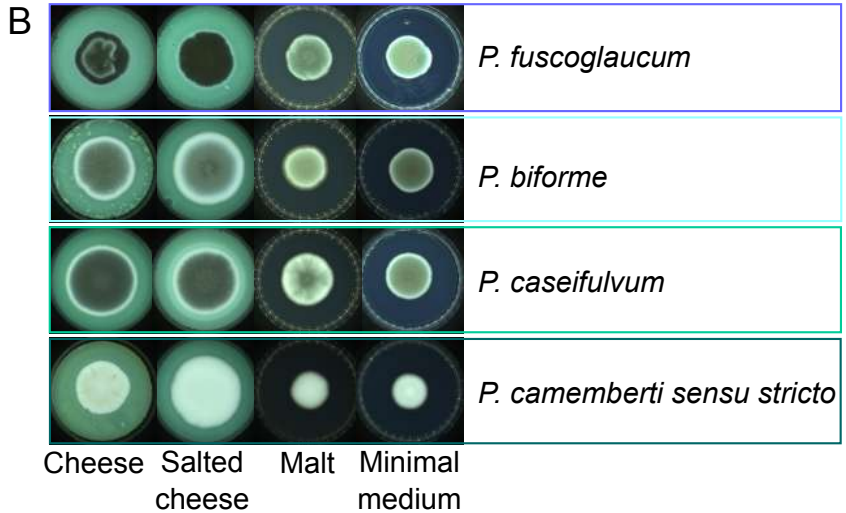
P. biforme *P. caseifulvum* *P. camemberti s.s.*

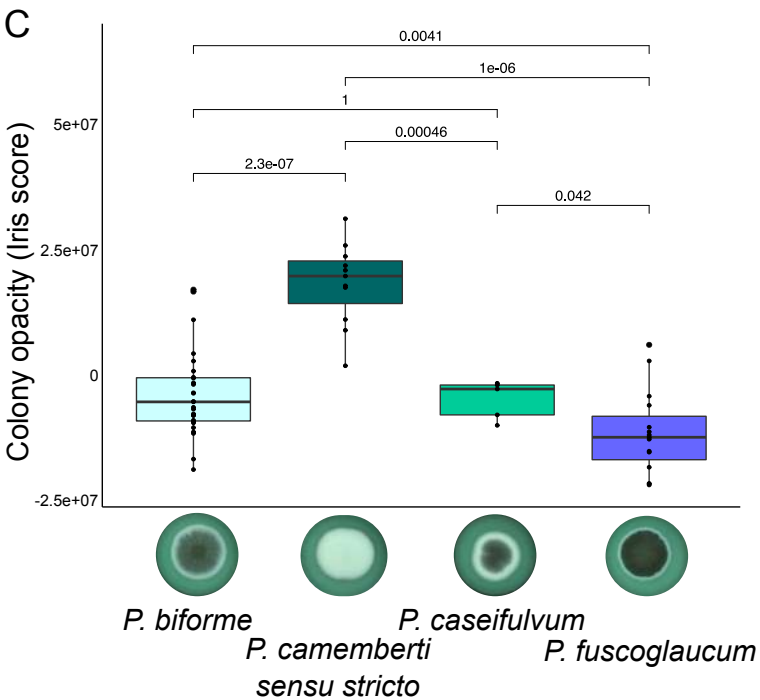
C

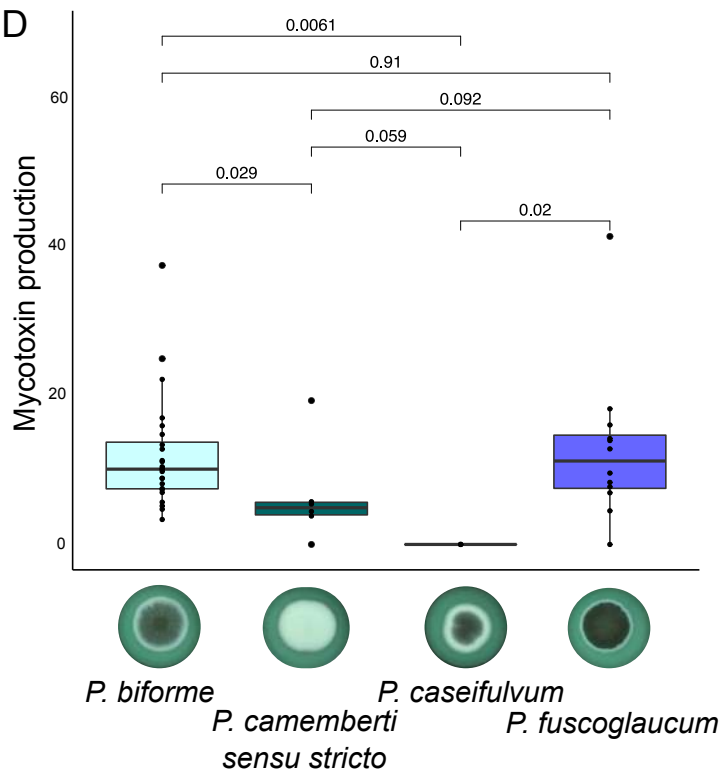

A
P. biforme
P. camemberti CPA- (P. caseifulvum)
P. camemberti CPA+ (P. camemberti s.s.)
P. fuscoglaucum
P. commune
Origin
Cheese
Other environments
Sausage
PC2 (16.32%)
PC1 (51.5%)
0
B
0.03
C
100
0
K = 2
K = 3
K = 4
K = 5
K = 6
P. biforme
P. fuscoglaucum
CPA-
CPA+
P. camemberti s.s.
P. caseifulvum

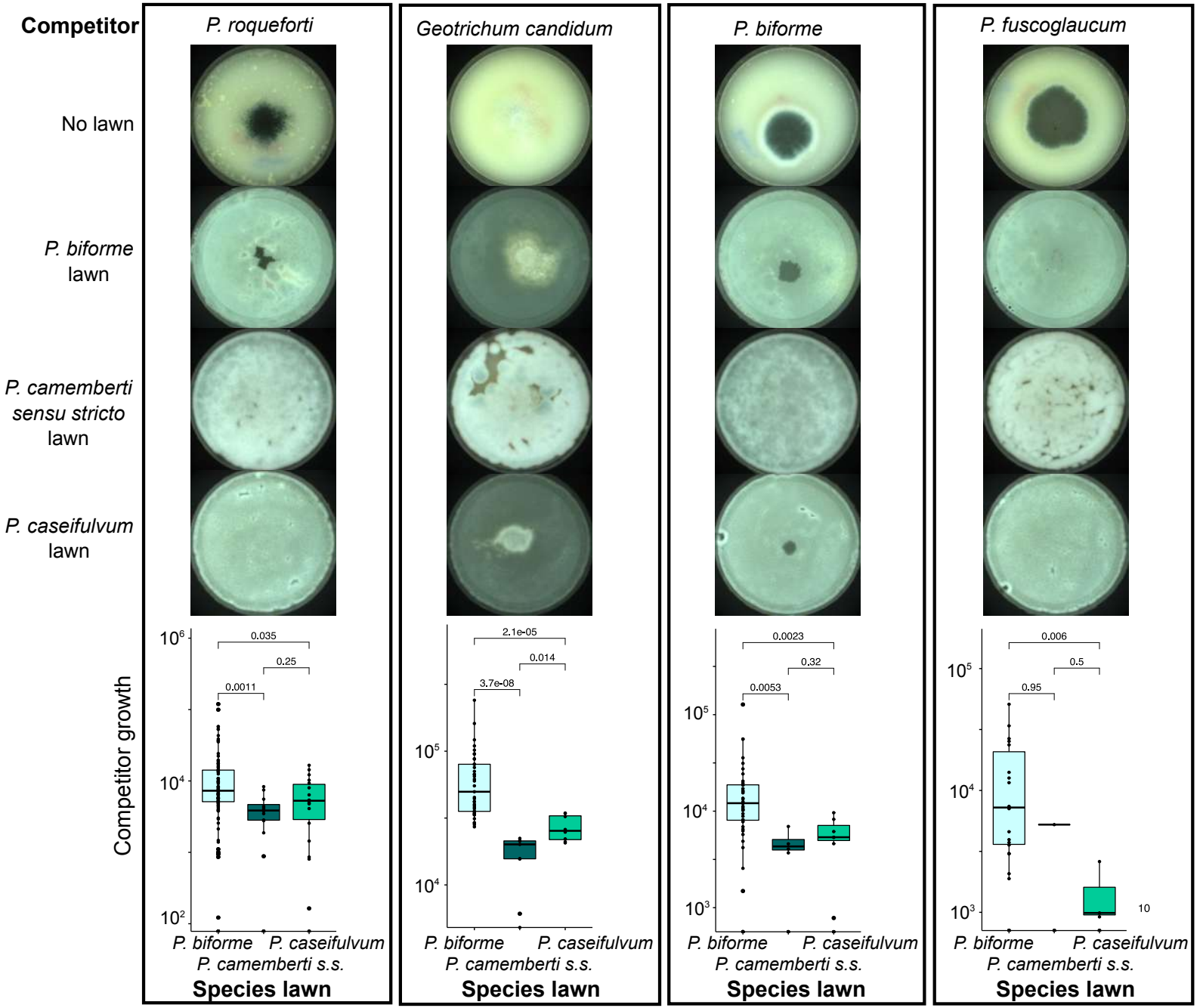

CheesyTer
Wallaby
LCP6617
LCP4799
ESE00090
LCP5327
LCP00218
LCP03239
LCP06615
LCP06621
LCP06697
LCP06623
LCP05471
LCP00797
LCP06700
ESE00155
ESE00126
ESE00086
LCP05529
LCP06620
LMA421
U-112050
ESE00137
LCP05496
U-112059
ESE00154
ESE00087
ESE00089
ESE00156
ESE00157
ESE00158
LCP05639
ESE00211
ESE00151
ESE00207
LCP06701
U-112053
ESE00023
ESE00095
ESE00018
ESE00125
ESE00139
U-116003
LCP05531
LCP05630
LCP06622
LCP01920
ESE00019
ESE00120
U-108097
LCP05527
U-101392
LCP04810
LCP00584
U-108096
U-113052
U-113059
LCP06093
LCP04258
U-113054
LCP06699
U-113053
U-113056
0.1
P. fuscoglaucum
P. biforme
P. caseifulvum
P. camemberti
CPA-
CPA+
lac-Y
beta-gal
Hce2


A
Colony radial growth (Iris score)
35000
30000
25000
20000
P. fuscoglaucum
P. biforme
P. caseifulvum
P. camemberti sensu stricto
Cheese
Salted cheese
Malt
Minimal medium


B
P. fuscoglaucum
P. biforme
P. caseifulvum
P. camemberti sensu stricto
Cheese
Salted cheese
Malt
Minimal medium


C
Colony opacity (Iris score)
5e+07
2.5e+07
0
-2.5e+07
0.0041
1e-06
1
0.00046
2.3e-07
0.042
P. biforme
P. caseifulvum
P. camemberti sensu stricto
P. fuscoglaucum


D
Mycotoxin production
60
40
20
0
0.0061
0.91
0.092
0.059
0.029
0.02
P. biforme
P. caseifulvum
P. camemberti sensu stricto
P. fuscoglaucum


Competitor
P. roqueforti
Geotrichum candidum
P. biforme
P. fuscoglaucum
No lawn
P. biforme lawn
P. camemberti sensu stricto lawn
P. caseifulvum lawn
Competitor growth
0.035
0.25
0.0011
2.1e-05
0.014
3.7e-08
0.0023
0.32
0.0053
0.006
0.5
0.95
10
P. biforme
P. caseifulvum
P. camemberti s.s.
Species lawn