bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Independent domestication events in the blue-cheese fungus *Penicillium*
2 # *roqueforti*
3

4 Emilie Dumas[1], Alice Feurtey[1,2], Ricardo C Rodríguez de la Vega[1], Stéphanie Le Prieur[1], Alodie
5 Snirc[1], Monika Coton[3], Anne Thierry[4], Emmanuel Coton[3], Mélanie Le Piver[5], Daniel Roueyre[5],
6 Jeanne Ropars[1], Antoine Branca* [1] and Tatiana Giraud* [1]
7
8 [1]Ecologie Systématique et Evolution, Univ Paris-Sud, CNRS, AgroParisTech, Université Paris-
9 Saclay, 91400 Orsay, France
10 [2]Environmental Genomics, Max Planck Institute for Evolutionary Biology, August Thienemann-
11 Str. 2, 24306 Plön, Germany
12 [3]Université de Brest, EA 3882 Laboratoire Universitaire de Biodiversité et Ecologie Microbienne,
13 ESIAB, Technopôle Brest-Iroise, 29280 Plouzané, France
14 [4]Science et Technologie du Lait et de l'Œuf (STLO), UMR1253, INRA, Agrocampus Ouest, 65
15 rue de Saint Brieuc, 35000 Rennes, France
16 [5]Laboratoire Interprofessionnel de Production - SAS L.I.P, 34 Rue de Salers, 15000 Aurillac,
17 France
18
19 *These authors jointly supervised the study
20 Corresponding author: Antoine Branca antoine.branca@u-psud.fr
21

22 Short title: Independent domestications of the blue-cheese fungus
23 Keywords: domestication, convergence, parallel adaptation, fungi, cheese, *Penicillium*
24

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25 **Abstract**

26 Domestication provides an excellent framework for studying adaptive divergence. Using
27 population genomics and phenotypic assays, we reconstructed the domestication history of the
28 blue cheese mold *Penicillium roqueforti.* We showed than this fungus was domesticated twice
29 independently. The population used in Roquefort originated from an old domestication event
30 associated with weak bottlenecks and exhibited traits beneficial for pre-industrial cheese
31 production (slower growth in cheese and greater spore production on bread, the traditional
32 multiplication medium). The other cheese population originated more recently from the selection
33 of a single clonal lineage, was associated to all types of blue cheese worldwide but Roquefort, and
34 displayed phenotypes more suited for industrial cheese production (high lipolytic activity, efficient
35 cheese cavity colonization ability and salt tolerance). We detected genomic regions affected by
36 recent positive selection and putative horizontal gene transfers. This study sheds light on the
37 processes of rapid adaptation and raises questions about genetic resources conservation.

38

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39

40 **Introduction**

41

42 Mechanisms of adaptive divergence (population differentiation under selection) are key questions
43 in evolutionary biology for the understanding of how organisms adapt to their environment and
44 how biodiversity arises. Domestication studies can help us understand adaptive divergence as this
45 process involves strong and recent selection for traits that can be easily identified. Furthermore,
46 closely related non-domesticated populations are often available, making it possible to contrast
47 their traits and genomes with those of domesticated populations. This approach has already proved
48 to be powerful for reconstructing the origin of domesticated populations and the genetic
49 architecture of traits selected by humans. It has been applied to maize and teosinte, and to dog
50 breeds and wolves [1–6]. Independent domestication events from the same ancestral population
51 are particularly interesting because they provide replicates of the adaptation process and insights
52 into the evolution predictability and constraints [7–12]. Comparisons of domesticated varieties
53 selected for different phenotypes have also proved to be a powerful approach for elucidating the
54 mechanisms of adaptation, for example in dog breeds and pigeons [13,14]. Studies on genetic
55 diversity and subdivision in domesticated organisms provides also crucial information for the
56 conservation of genetic resources. Ancient domestication processes were slow and involved
57 contributions from large numbers of farmers. By contrast, recent breeding programs have been run
58 by a small number of companies; this and international trade demands have resulted in a massive
59 loss of genetic diversity in crops and breeds, potentially jeopardizing adaptive potential for
60 improvement [15–17].

61 Fungi are interesting eukaryotic models for adaptive divergence studies, with their small
62 genomes, easy access to the haploid phase and experimental tractability for *in vitro* experiments
63 [18,19]. Many fungi are used as food sources [20] and some have been domesticated for food

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

64 production. Propagation of the latter is controlled by humans, and this has resulted in genetic
65 differentiation from wild populations [21–24] and the evolution of specific phenotypes beneficial
66 for humans [24,25]. *Saccharomyces cerevisiae* yeasts domesticated for fermentation have
67 provided important insight into adaptive divergence mechanisms, with different yeast lineages
68 independently domesticated for different usages [23,26,27]. Studies about yeast adaptation for
69 alcohol and cheese production have highlighted the proximal genomic mechanisms involved,
70 including horizontal gene transfer, selective sweep, hybridization and introgression [25,27–30].

71 *Penicillium roqueforti,* a filamentous fungus used in the dairy industry to impart the typical
72 veins and flavor of blue cheeses, has recently emerged as an excellent model for studying adaptive
73 divergence [31,32]. Blue cheeses, including Roquefort, Gorgonzola and Stilton, are highly
74 emblematic foods that have been produced for centuries [33]. The strongest genetic subdivision
75 reported in *P. roqueforti* concerns the differentiation of a cheese-specific population that has
76 acquired faster growth in cheese than other populations and better excludes competitors, thanks to
77 very recent horizontal gene transfers, at the expense of slower growth on minimal medium
78 [31,34,35]. Such genetic differentiation and recent acquisition of traits beneficial to cheesemaking
79 in *P. roqueforti* suggests genuine domestication, i.e., adaptation under selection by humans for
80 traits beneficial for their food production. A second population identified in *P. roqueforti* and
81 lacking the horizontally-transferred regions includes strains isolated from cheese and other
82 environments, such as silage, lumber and spoiled food, suggesting that adaptive divergence may
83 have occurred [34–36]. The existence of further genetic subdivision separating populations
84 according to the original environment, or protected designation of origin (PDO) for cheese strains
85 has been suggested, but, because it was based only on a few microsatellite markers, the resolution
86 power was low [34–36]. Secondary metabolite production (aroma compounds and mycotoxins)
87 and proteolysis activity have been shown to differ between strains from different PDOs [37]. A
88 high-quality *P. roqueforti* genome reference is available [32], allowing more powerful analyses

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

89 based on population genomics.

90 Another asset of *P. roqueforti* as an evolutionary model is the availability of vast
91 collections of cheese strains and of historical records concerning cheesemaking [38–43]. While
92 the presence of *P. roqueforti* in cheeses was initially fortuitous, since the end of the 19th century,
93 milk or curd has been inoculated with the spores of this fungus for Roquefort cheese production.
94 Spores were initially multiplied on bread, before the advent of more controlled *in vitro* culture
95 techniques in the 20th century [38–43]. Bread was inoculated by recycling spores from the best
96 cheeses from the previous production [38–43]. This corresponds to yearly selection events since
97 the 19th century until ca. 20 years ago when strains were stored in freezers. After World War II,
98 strains were isolated in the laboratory for industrial use and selected based on their technological
99 and organoleptic impact in cheeses and compounds produced [44], which have likely accelerated
100 domestication. This history further suggests there may have been genuine domestication, i.e., an
101 adaptive evolution triggered by human selection for cheese quality. Unintentional selection may
102 also have been exerted on other traits, including growth and spore production on bread, the
103 traditional multiplication substrate.

104 By sequencing multiple *P. roqueforti* genomes from different environments and analyzing
105 large collections of cheese strains, we provide evidence for adaptive divergence. We identified
106 four genetically differentiated populations, two including only cheese strains and two other
107 populations including silage and food-spoiling strains. We inferred that the two cheese
108 populations corresponded to two independent domestication events. The first cheese population
109 correspond to strains used for Roquefort production and arose through a weaker and older
110 domestication event, with multiple strains probably originating from different cultures on local
111 farms in the PDO area, presumably initially selected for slow growth before the invention of
112 refrigeration systems. The second cheese population experienced an independent, more recent,
113 domestication event associated with a stronger genetic bottleneck. The non-Roquefort cheese

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

114 population showed a higher fitness for traits likely to be under selection for modern production of
115 cheese (e.g. growth in salted cheese and lipid degradation activities), while the Roquefort cheese
116 population showed greater spore production on bread, the traditional medium for spore
117 production. The two cheese populations also had different volatile compound profiles, with likely
118 effects on cheese flavor. Moreover, we detected genomic regions affected by recent positive
119 selection and genomic islands specific to a single cheese population. Some of these genomic
120 regions may have been acquired by horizontal gene transfers and have putative functions in the
121 biochemical pathways leading to the development of cheese flavor.
122
123 **Results**
124 ***Two out of four populations are used for cheesemaking: one specific to the Roquefort PDO and***
125 ***a worldwide clonal population***
126 We sequenced the genomes of 34 *P. roqueforti* strains from public collections [33], including 17
127 isolated from blue cheeses (e.g., Roquefort, Gorgonzola, Stilton), 17 isolated from non-cheese
128 environments (mainly spoiled food, silage, and lumber), and 11 outgroup genomes from three
129 *Penicillium* species closely related to *P. roqueforti* (Supplementary Table 1). After data filtering,
130 we identified a total of 115,544 SNPs from the reads mapped against the reference *P. roqueforti*
131 FM164 genome (48 scaffolds).

132 Three clustering methods free from assumptions about mating system and mode of reproduction
133 separated the *P. roqueforti* strains into four genetic clusters (Figs 1, 2 and 3A), two of which
134 almost exclusively contained cheese strains (the exceptions being two strains isolated from a
135 brewery and brioche, Figs. 1 and 2, probably corresponding to feral strains). One cluster contained
136 both silage strains (N=4) and food-spoiling strains (N=4), and the last cluster contained mostly
137 food-spoiling strains (N=5) plus strains from lumber (N=2) (Figs. 1, 2, and Supplementary Table
138 1). Noteworthy, these two clusters corresponding to strains from other environments did not

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

139 include a single cheese strain. The two cheese clusters were not the most closely related,
140 suggesting independent domestication events (Figs 1 and 2). Moreover, cheese clusters were much
141 less diverse than non-cheese clusters, as shown by their shorter branch lengths in the tree, their
142 low genetic diversity represented by small □ values and more homogeneous colors in distance-
143 based clustering (Figs. 1, 2 and 4A). One of the two cheese clusters displayed a particularly low
144 level of genetic diversity (Figs. 1, 2 and 4A, with only 0.03% polymorphic sites, and a lack of
145 recombination footprints (i.e., a higher level of linkage disequilibrium, as shown by r^2 values, Fig.
146 4B, and by the large single-color blocks along the genomes, Fig. 2). These findings suggest that
147 the second cheese population is a single clonal lineage.

148 We used genome sequences to design genetic markers (Supplementary Table 2) for assigning a
149 collection of 65 strains provided by the main French supplier of *P. roqueforti* spores for artisanal
150 and industrial cheesemakers, 18 additional strains from the National History Museum collection in
151 Paris (LCP) and 31 strains from the adjunct collection of the Université de Bretagne Occidentale
152 (UBOCC, Supplementary Table 1) to the four genetic clusters. Out of these 148 strains, 55 were
153 assigned to the more genetically diverse of the two cheese clusters. The majority of these strains
154 included strains used for Roquefort PDO cheese production (N=30); three strains originated from
155 Bleu des Causses cheeses (Supplementary Figure 1, Supplementary Table 1), produced in the
156 same area as Roquefort and using similarly long storage in caves. The remaining strains of this
157 cluster included samples from other blue cheeses (N=13), unknown blue cheeses (N=5) or other
158 environments (N=4), the latter likely associated with feral strains. Because of the strong bias of
159 usage toward Roquefort production, we refer to this cluster hereafter as the "Roquefort
160 population". Of the remaining 95 strains, 60 belonged to the second cheese cluster, which was less
161 genetically diverse and contained mainly commercial strains used to produce a wide range of blue
162 cheeses, but only one from the Roquefort PDO (Supplementary Figure 1, Supplementary Table 1).
163 This cluster was therefore named the "non-Roquefort population". The Roquefort population also

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

164 included 13 strains used to inoculate other types of blue cheese (e.g. Gorgonzola or Bleu
165 d'Auvergne), but strains from these types of cheeses were more common in the non-Roquefort
166 population. The non-Roquefort cluster contained strains carrying *Wallaby* and *CheesyTer*, two
167 large genomic regions recently shown to have been transferred horizontally between different
168 *Penicillium* species from the cheese environment and conferring faster growth on cheese [31,32],
169 whereas all the strains in the Roquefort cluster lacked those regions.

170

171 ***Two independent domestication events in* Penicillium roqueforti *for cheesemaking***

172 By comparing 11 demographic scenarios in approximate Bayesian computation (ABC), we
173 showed that the two *P. roqueforti* cheese populations (Roquefort and non-Roquefort) resulted
174 from two independent domestication events (Fig. 5, Supplementary Figure 2). The highest
175 posterior probabilities were obtained for the S4 scenario, in which the two cheese populations
176 formed two lineages independently derived from the common ancestral population of all *P.*
177 *roqueforti* strains (Fig. 5, model choice and parameter estimates in Supplementary Figure 2). We
178 inferred much stronger bottlenecks in the two cheese populations than in the non-cheese
179 populations, with the most severe bottleneck found in the non-Roquefort cheese population. Some
180 gene flow (m=0.1) was inferred between the two non-cheese populations but none with cheese
181 populations. The bottleneck date estimates in ABC had too large credibility intervals to allow
182 inferring domestication dates (Supplementary Figure 2E). We therefore used the multiple
183 sequentially Markovian coalescent (MSMC) method [45] to estimate times since domestication,
184 considering that they corresponded to the last time there was gene flow between genotypes within
185 populations, as this also corresponds to bottleneck date estimates in coalescence. The
186 domestication for the Roquefort cheese population was inferred seven times longer ago than for
187 the non-Roquefort cheese population, both domestication events being recent (ca. 760 versus 140

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

188 generations ago, Figure 5B-C). Unfortunately, generation time, and even generation definition, are
189 too uncertain in the clonal *P. roqueforti* populations to infer domestication dates in years. In
190 addition, the MSMC analysis detected two bottlenecks in the history of the Roquefort cheese
191 population (Figure 5B).

192

193 ***Contrasting fitness traits between cheese populations***

194 We tested whether different phenotypes relevant for cheesemaking had evolved in the two cheese
195 clusters, relative to other populations (Figs. 3B and 6, Supplementary Table 3). We first produced
196 experimental cheeses inoculated with strains from the different *P. roqueforti* populations to assess
197 their ability to colonize cheese cavities, a trait that may have been subject to human selection to
198 choose inocula producing the most visually attractive blue cheeses. The fungus requires oxygen
199 and can therefore sporulate only in the cheese cavities, its spores being responsible for the
200 characteristic color of blue cheeses. Strains from the non-Roquefort cheese population were the
201 most efficient colonizers of cheese cavities (Supplementary Table 4); no difference was detected
202 between strains from the Roquefort and non-cheese populations (Fig. 6E).

203 As *P. roqueforti* strains were traditionally multiplied on bread loaves for cheese inoculation, they
204 may have been subject to unintentional selection for faster growth on bread. However, growth rate
205 on bread did not significantly differ between populations (Fig. 6A, Supplementary Table 4).

206 We then assessed lipolytic and proteolytic activities in the *P. roqueforti* populations. These
207 activities are important for energy and nutrient uptake, as well as for cheese texture and the
208 production of volatile compounds responsible for cheese flavors [37,46]. Lipolysis was faster in
209 the non-Roquefort cheese population than in the Roquefort and silage/food spoiling populations
210 (Fig. 6B, Supplementary Table 4). A strong population effect was found for proteolytic activity

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

211 (Supplementary Table 4) although without significant differences in the post-hoc pairwise analysis
212 (Fig. 6C). However, variances showed significant differences between populations (Levene test F-
213 ratio=5.97, d.f.=3, P<0.0017), with the two cheese populations showing the highest variances, and
214 with extreme values above and below those in non-cheese populations (Fig. 6C). Noteworthy,
215 proteolysis is a choice criterion for making different kinds of blue cheeses (https://www.lip-
216 sas.fr/index.php/nos-produits/penicillium-roquefortii/18-penicillium-roquefortii). This suggests
217 that some cheese strains may have been selected for higher and others for lower proteolytic
218 activity. Alternatively, selection could have been relaxed on this trait in the cheese populations,
219 leading to some mutations decreasing and other increasing proteolysis in different strains, thus
220 increasing variance in the populations.

221 The ability of *P. roqueforti* strains to produce spores may also have been selected by humans, both
222 unwittingly, due to the collection of spores from moldy bread, and deliberately, through the choice
223 of inocula producing bluer cheeses. We detected no difference in spore production between the *P.*
224 *roqueforti* populations grown on cheese medium or malt. However, we observed significant
225 differences in spore production on bread medium. The Roquefort population produced the highest
226 number of spores and significantly more than the non-Roquefort population (Fig. 6D,
227 Supplementary Table 4).

228 High salt concentrations have long been used in cheesemaking to prevent the growth of spoiler
229 and pathogenic microorganisms. We found that the ability to grow on salted malt and cheese
230 media decreased in all *P. roqueforti* populations (Fig. 6F, Supplementary Table 4). We found a
231 significant interaction between salt and population factors, and post-hoc tests indicated that the
232 Roquefort population was more affected by salt than the other populations (Fig. 6F,
233 Supplementary Table 4).

234 Volatile compound production was also investigated in the two cheese populations, as these

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

235 compounds are important for cheese flavor [46]. We identified 52 volatile compounds, including
236 several involved in cheese aroma properties, such as ketones, free fatty acids, sulfur compounds,
237 alcohols, aldehydes, pyrazines, esters, lactones and phenols [47] (Supplementary Figure 3). The
238 two cheese populations presented significantly different volatile compound profiles, differing by
239 three ketones, one alcohol and two pyrazines (Supplementary Figure 3). The Roquefort population
240 produced the highest diversity of volatile compounds (Supplementary Figure 3A).

241

242 ***Detection of genomic regions affected by recent positive selection and population-specific***
243 ***genomic islands***

244 We identified five regions present in the genomes of strains from the non-Roquefort cheese
245 population and absent from the other populations. We also detected five other genomic islands
246 present in several *P. roqueforti* strains but absent from the non-Roquefort cheese strains
247 (Supplementary Figure 4). Nine of these ten genomic regions were not found in the genomes of
248 the outgroup *Penicillium* species analyzed here and they displayed no genetic diversity in *P.*
249 *roqueforti*. No SNPs were detected even at synonymous sites or in non-coding regions, suggesting
250 recent acquisitions, by horizontal gene transfer. Only FM164-C, one of the genomic islands
251 specific to the non-Roquefort population, was present in the outgroup genomes, in which it
252 displayed variability, indicating a loss in the other lineages rather than a gain in the non-Roquefort
253 population and the outgroup species (Supplementary Figure 4A). The closest hits in the NCBI
254 database for genes in the ten genomic islands were in *Penicillium* genomes. Most of the putative
255 functions proposed for the genes within these genomic regions were related to lipolysis,
256 carbohydrate or amino-acid catabolism and metabolite transport. Other putative functions
257 concerned fungal development, including spore production and hyphal growth (Supplementary
258 Figure 4). In the genomic regions specific to the non-Roquefort cheese population, we also

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

259 identified putative functions potentially relevant for competition against other microorganisms,
260 such as phospholipases, proteins carrying peptidoglycan- or chitin-binding domains and chitinases
261 (Supplementary Figure 4) [48]. Enrichment tests were non-significant, probably due to the small
262 number of genes in these regions.

263 Footprints of positive selection in *P. roqueforti* genomes were detected using an extension of the
264 McDonald-Kreitman test [49] which identifies genes with more frequent amino-acid changes than
265 expected under neutrality, neutral substitution rates being assessed by comparing the rates of
266 synonymous and non-synonymous substitutions within and between species or populations to
267 account for gene-specific mutation rates. We ran the test with three levels of population
268 subdivision. First, no significant footprint of positive selection was detected for any gene by
269 comparing the whole *P. roqueforti* species with *P. paneum*. In a second test, a set of 15 genes was
270 identified as evolving under positive selection in the Roquefort cheese population but not in the
271 other pooled *P. roqueforti* populations (Supplementary Figure 5B). Interestingly, eight of these 15
272 genes clustered at the end of the largest scaffold (Supplementary Figure 5A). In a third test, four
273 genes were identified as evolving under positive selection in the non-Roquefort cheese population
274 but not in the pooled non-cheese *P. roqueforti* populations (Supplementary Figure 5B). Two of
275 these genes corresponded to a putative aromatic ring hydroxylase and a putative cyclin evolving
276 under purifying selection in Roquefort and non-cheese *P. roqueforti* populations (Supplementary
277 Figure 5B). Aromatic ring hydroxylases are known to be involved in the catabolism of aromatic
278 amino acids, which are precursors of flavor compounds [50,51].

279

280 **Discussion**

281 We report here the genetic subdivision of *P. roqueforti*, the fungus used worldwide for blue
282 cheese production, with unprecedented resolution, providing insights into its domestication

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

283 history. Population genomics studies on strains from various substrates and from a large collection
284 of cheese identified four genetically differentiated populations, two of which being cheese
285 populations originating from independent and recent domestication events. One *P. roqueforti*
286 cheese population included all the genotyped strains but one used for PDO Roquefort cheeses,
287 produced in the French town Roquefort-sur-Soulzon, where blue cheeses have been made since at
288 least the 15th century, and probably for much longer [33,38–43]. The strains from this Roquefort
289 cheese population lacks the horizontally-transferred *Wallaby* and *CheesyTer* genomic islands
290 contrary to the other non-Roquefort cheese population.

291 We observed that the two *P. roqueforti* cheese populations differed on several traits important for
292 cheese production, probably corresponding to historical discrepancies. Indeed, the Roquefort
293 population has retained moderate genetic diversity, consistent with soft selection during pre-
294 industrial times on multiple farms near Roquefort-sur-Soulzon, where specific strains were kept
295 for several centuries. The Roquefort cheese population grew slower in cheese [31] and had weaker
296 lipolytic activity. Slow maturation is particularly crucial for the storage of Roquefort cheeses for
297 long periods in the absence of refrigeration [41] because they are made of ewe's milk, a product
298 available only between February and July. During storage, cheeses could become over degraded
299 by too high rates of lipolysis, thus likely explaining the low lipolysis activity in Roquefort strains.
300 By contrast, most other blue cheeses are produced from cow's milk, which is available all year.
301 The Roquefort population showed greater sporulation on bread, which is consistent with
302 unconscious selection for this trait when strains were cultured on bread in Roquefort-sur-Soulzon
303 farms before cheese inoculation during the end of the 19th and beginning of the 20th centuries.

304

305 Lipolytic activity is known to impact texture and the production of volatile compounds affecting
306 cheese pungency [54–59]. The Roquefort and non-Roquefort populations showed different volatile

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

307 compound profiles, suggesting also different flavor profiles. The discovery of different
308 phenotypes in the two cheese populations, together with the availability of a protocol for inducing
309 sexual reproduction in *P. roqueforti* [36], pave the way for crosses to counteract degeneration after
310 clonal multiplication and bottlenecks, for variety improvement and the generation of diversity.

311 Both cheese populations were found to have gone through bottlenecks. A previous study showed
312 that these bottlenecks, together with clonal multiplication, decreased fertility, with different stages
313 in sexual reproduction affected in the two populations identified here as the Roquefort and non-
314 Roquefort lineages [52]. The non-Roquefort cheese population, despite suffering from a more
315 severe and more recent bottleneck, was found to be used in the production of all types of blue
316 cheese worldwide, including Gorgonzola, Bleu d'Auvergne, Stilton, Cabrales and Fourme
317 d'Ambert. We showed that it grows more rapidly on cheese [31], exhibits greater ability to
318 colonize cheese cavities, higher salt tolerance and faster lipolysis than the Roquefort population.
319 These characteristics are consistent with the non-Roquefort cheese population resulting from a
320 very recent strong selection of traits beneficial for modern, accelerated, production of blue cheese
321 using refrigeration techniques, followed by a worldwide dissemination for the production of all
322 types of blue. Such drastic losses of genetic diversity in domesticated organisms are typical of
323 strong selection for industrial use by a few international firms and raise concerns about the
324 conservation of genetic resources, the loss of which may hinder future innovation. More generally,
325 in crops, the impoverishment in genetic diversity decreases the ability of cultivated populations to
326 adapt to environmental and biotic changes to meet future needs [15–17]. The PDO label, which
327 imposes the use of local strains, has probably contributed to the conservation of genetic diversity
328 in the Roquefort population (see "Cahier des charges de l'appellation d'origine protégée
329 Roquefort", i.e., the technical specifications for Roquefort PDO). We inferred two bottlenecks in
330 the Roquefort population, more ancient than in the non-Roquefort population, likely
331 corresponding to a pre-industrial domestication event when multiple local farms multiplied their

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

332 strains, followed by a second bottleneck when fewer strains were kept by the first industrial
333 societies. For other blue cheeses, even if their production was also ancient, the performant non-
334 Roquefort clonal lineage could have been recently chosen to fit modern industrial production
335 demands due to the lack of PDO rules imposing the use of local strains. However, despite a much
336 lower genome-wide diversity in domesticated populations, proteolysis and volatile compounds
337 diversity was found higher in cheese than in non-cheese populations. In fact, different strains with
338 more or less rapid proteolysis and lipolysis are sold for specific blue cheese types (e.g., milder or
339 stronger), in particular by the French LIP company (https://www.lip-sas.fr/index.php/nos-
340 produits/penicillium-roquefortii/18-penicillium-roquefortii). Such a high phenotypic diversity
341 within the cheese populations is consistent with diversification of usage under domestication, and
342 in particular when different characteristics are desired according to cheese type. This has already
343 been observed in relation to the diversification of crop varieties or breeds in domesticated animals
344 [13,14].

345 When studying adaptation in domesticated organisms, it is often useful to contrast traits and
346 genomic variants between domesticated and closely related wild populations to determine the
347 nature of the adaptive changes occurring under artificial selection [53,54]. The only known non-
348 cheese populations of *P. roqueforti* occur essentially in human-made environments (silage, food
349 and lumber), consistent with the specific adaptation of these populations to these environments.
350 The two non-cheese populations were inferred to have diverged very recently and they displayed
351 footprints of recombination and marked differentiation from the cheese populations. Domesticated
352 populations are expected to be nested within their source populations, suggesting that we have not
353 sampled the wild population that is the most closely related from cheese strains yet. The high level
354 of diversity and inferred demographic history of *P. roqueforti* indicate that most food-spoiling
355 strains belong to differentiated populations and are not feral cheese strains. In addition, not a
356 single cheese strain was found in the food spoiling and silage populations. This was shown by

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

357 both genome sequences and by the genotyping of a larger number of strains using a few selected
358 markers, in the present study and based on microsatellite markers in a previous work [35].
359 Consequently, *P. roqueforti* spores from blue cheeses may, rarely, spoil food and food-spoiling
360 and silage strains are not used for cheesemaking nor recombine with cheese strains. Such a lack of
361 incoming gene flow into cheese populations allowed trait differentiation in cheese strains as
362 expected under domestication.

363 It came as a surprise that the two non-cheese populations split more recently from each other than
364 from the cheese lineages. In particular, the non-Roquefort population diverged the earliest from
365 the unidentified ancestral population, and this has likely occurred in another environment than
366 cheese. Much more recently, selection in industrial times has likely only kept the most performant
367 clonal lineage of this population for cheesemaking, losing most of the initial diversity, as indicated
368 by the very strong and recent bottleneck inferred in this lineage. Possible scenarios to explain the
369 existence of two separated clusters thriving in food and silage differentiated from cheese strains
370 include the very recent adaptive differentiation of a population from silage on human food or vice
371 versa. The finding that silage strains are only found in one cluster (the orange in our
372 representation) suggests an adaptation to this ecological niche, although experiments will be
373 required to test this hypothesis. Food spoiling strains are in contrast found in three clusters and
374 may thus not constitute a specific population adapted to this environment and may instead
375 represent migrants from several populations belonging to other ecological niches. Green and
376 orange clusters may alternatively represent populations thriving in yet unidentified environments,
377 dispersing to silage and food.

378 The history of blue cheese production may provide circumstantial clues as to the origin of *P.*
379 *roqueforti* cheese populations. Indeed, the first blue cheeses likely resulted from the sporadic
380 accidental contamination of cheeses with spores from the environment, such as moldy food.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

381 However, this would not be consistent with the demographic history inferred here for cheese and
382 food-spoiling strains, as the cheese strains were not found to be nested within the food-spoiling
383 strains, some of which originated from moldy bread. Furthermore, old French texts suggest that
384 the blue mold colonized the cheese from within [33,39,40], which would indicate that the milk or
385 curd was contaminated. French cheese producers began to inoculate cheeses with *P. roqueforti*
386 spores from moldy rye bread at the end of the 19th century [33,39,40]. Breads were specifically
387 made with a 2:1 mixture of wheat and rye flour and were baked rapidly at high temperature
388 (500°C), to yield a protective crust, around a moist, undercooked interior [38,41]; the mold
389 developed from the inside of the bread after one to five months in the Roquefort caves [33,39,40].
390 Surveys of the microorganisms present in their caves [41,55,56] and our unsuccessful attempts to
391 obtain samples from a maturing cellar suggest that *P. roqueforti* spores did not originate from the
392 caves, which were nevertheless crucial due to the ideal conditions provided for *P. roqueforti*
393 development [41]. Bread may have been colonized from the environment or from rye flour if the
394 source *P. roqueforti* population was a rye endophyte or pathogen. This last hypothesis would be
395 consistent with the lifestyle of many *Penicillium* species, which live in close association with
396 plants, often acting as plant pathogens or necrotrophs [57], and with the occurrence of a *P.*
397 *roqueforti* population in lumber and silage. If this hypothesis is correct, then cheeses may
398 historically have become contaminated with *P. roqueforti* from fodder during milking.

399 Comparison between non-cheese and cheese populations allowed us to identify specific traits and
400 genes that have been under selection in cheese as opposed to other environments. Furthermore, the
401 two independently domesticated *P. roqueforti* cheese populations, exhibiting different traits,
402 represent a good model for studying the genomic processes involved in adaptation. We were able
403 to identify candidate genes and evolutionary mechanisms potentially involved in adaptation to
404 cheese in *P. roqueforti*. The horizontally-transferred *CheesyTer* genomic island probably
405 contributes to the faster growth of the strains identified here as constituting the non-Roquefort

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

406 population [31]. Indeed, *CheesyTer* includes genes with putative functions involved in
407 carbohydrate utilization (e.g. β-galactosidase and lactose permease genes) that are specifically
408 expressed at the beginning of cheese maturation, when lactose and galactose are available. This
409 horizontal gene transfer may thus have been involved in adaptation to recently developed
410 industrial cheese production processes in the non-Roquefort cheese population, conferring faster
411 growth. We also identified additional genomic islands specific to the non-Roquefort cheese
412 population, probably acquired recently and including genes putatively involved in fungal growth
413 and spore production. In the genomic islands specific to the cheese populations, several genes
414 appeared to be involved in lipolysis, carbohydrate or amino-acid catabolism and metabolite
415 transport, all of which are important biochemical processes in the development of cheese flavor. In
416 the Roquefort cheese population, a genomic region harboring genes with footprints for positive
417 selection included several genes encoding proteins potentially involved in aromatic amino-acid
418 catabolism corresponding to precursors of volatile compounds. Further studies are required to
419 determine the role of these genes in cheese flavor development.

420 In conclusion, we show that *P. roqueforti* cheese populations represent genuine domestication. Of
421 course, the domestication process in cheese fungi has been more recent and different from the
422 ones in emblematic crops or animals, and may not even fit exactly some definition of
423 domestication. Nevertheless, we did observe strong genetic differentiation from non-cheese
424 populations, strong bottlenecks and trait differentiation with likely benefits for cheese production.
425 Furthermore, a previous study has shown that the non-Roquefort cheese strains have acquired
426 genes conferring better growth in cheese [31]. The two independent domestication events
427 identified here represent parallel adaptations to the same new environment, a particularly powerful
428 situation for studies of adaptation [7,8,11]. Our findings concerning the history of *P. roqueforti*
429 domestication shed light on the processes of adaptation to rapid environmental change, but they
430 also have industrial implications and raise questions about the conservation of genetic resources in

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

431 the agri-food context.

432
433
434 **Methods**

435 ***Isolation attempts of* Penicillium roqueforti *in ripening cellar and dairy environments***

436 In order to investigate whether a wild *P. roqueforti* population occurred in ripening cellars or dairy
437 environments that could be at the origin of the observed cheese populations, we sampled spores
438 from the air in an artisanal cheese dairy company (GAEC Le Lévejac, Saint Georges de Lévejac,
439 France, ca 60 km from Roquefort-sur-Soulzon, producing no blue cheese to avoid feral strains, i.e.
440 dispersal from inoculated cheeses), sampling was performed in the sheepfold, milking parlour,
441 cheese dairy and ripening cellar. We also sampled spores from the air in an abandoned ripening
442 cellar in the town of Meyrueis (ca 70 km from Roquefort-sous-Soulzon) where Roquefort cheeses
443 used to be produced and stored in the early 19th century. In total, 55 Petri dishes containing malt
444 (2% cristomalt, Difal) and 3% ampicillin were left open for six days as traps for airborne spores
445 (35 Petri dishes in the abandoned ripening cellar and 20 Petri dishes in the artisanal cheese dairy
446 company). Numerous fungal colonies were obtained on the Petri dishes. One monospore was
447 isolated from each of the 22 *Penicillium*-like colonies. DNA was extracted using the Nucleospin
448 Soil Kit (Macherey-Nagel, Düren, Germany) and a fragment of the β-tubulin gene was amplified
449 using the primer set Bt2a/Bt2b [59], and then sequenced. Sequences were blasted against the
450 NCBI database to assign monospores to species. Based on β-tubulin sequences, ten strains were
451 assigned to *P. solitum*, six to *P. brevicompactum*, two to *P. bialowienzense*, one to *P. echinulatum*
452 and two to the *Cladosporium* genus. No *P. roqueforti* strain could thus be isolated from this
453 sampling procedure.

454

455 ***Genome sequencing and analysis***

456 The genomic DNAs of cheesemaking strains obtained from public collections belonging to *P.*
457 *roqueforti*, seven strains of *P. paneum,* one strain of *P. carneum* and one strain of *P.*
458 *psychrosexualis* (Supplementary Table 1) were extracted from fresh haploid mycelium after
459 monospore isolation and growth for five days on malt agar using the Nucleospin Soil Kit
460 (Macherey-Nagel, Düren, Germany). Sequencing was performed using the Illumina HiSeq 2500

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

461 paired-end technology (Illumina Inc.) with an average insert size of 400 bp at the GenoToul INRA
462 platform and resulted in a 50x-100x coverage. In addition, the genomes of four strains
463 (LCP05885, LCP06096, LCP06097 and LCP06098) were used that had previously been
464 sequenced using the ABI SOLID technology [32]. GenBank accession numbers are HG792015-
465 HG792062.

466 Identification of presence/absence polymorphism of blocks larger than 10 kbp in genomes was
467 performed based on coverage using mapping against the FM164 *P. roqueforti* reference genome.
468 In order to identify genomic regions that would be lacking in the FM164 genome but present in
469 other strains, we used a second assembled genome, that of the UASWS *P. roqueforti* strain
470 collected from bread, sequenced using Illumina HiSeq shotgun and displaying 428 contigs
471 (Genbank accession numbers: JNNS01000420-JNNS01000428). Blocks larger than 10 kbp
472 present in the UASWS genome and absent in the FM164 genome were identified using the
473 *nucmer* program v3.1 [68]. Gene models for the UASWS genome were predicted with EuGene
474 following the same pipeline as for the FM164 genome [32,69]. The presence/absence of these
475 regions in the *P. roqueforti* genomes was then determined using the coverage obtained by
476 mapping reads against the UASWS genome with the start/end positions identified by *nucmer*. The
477 absence of regions was inferred when less than five reads were mapped. In order to determine
478 their presence/absence in other *Penicillium* species, the sequences of these regions were blasted
479 against nine *Penicillium* reference genomes (Supplementary Table 1). PCR primer pairs were
480 designed using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/)
481 in the flanking sequences of these genomic regions in order to check their presence/absence in a
482 broader collection of *P. roqueforti* strains based on PCR tests (Supplementary Table 2). For each
483 genomic island, two primer pairs were designed when possible (i.e. when sufficiently far from the
484 ends of the scaffolds and not in repeated regions): one yielding a PCR product when the region
485 was present and another one giving a band when the region was absent, in order to avoid relying
486 only on lack of amplification for inferring the absence of a genomic region. PCRs were performed
487 in a volume of 25 µL, containing 12,5 µL template DNA (ten folds diluted), 0.625 U Taq DNA
488 Polymerase (MP Biomedicals), 2.5 µL 10x PCR buffer, 1 µL of 2.5 mM dNTPs, 1 µL of each of
489 10 µM primer. Amplification was performed using the following program: 5 min at 94°C and 30
490 cycles of 30 s at 94°C, 30 s at 60°C and 1 min at 72°C, followed by a final extension of 5 min at
491 72°C. PCR products were visualized using stained agarose gel electrophoresis.Data were
492 deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena/) under the accession
493 number: PRJEB20132 for whole genome sequencing and PRJEB20413 for Sanger sequencing.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

494 For each strain, reads were mapped using stampy v1.0.21 [60] against the high-quality reference
495 genome of the FM164 *P. roqueforti* strain [32]. In order to minimize the number of mismatches,
496 reads were locally realigned using the genome analysis toolkit (GATK) IndelRealigner v3.2-2
497 [61]. SNP detection was performed using the GATK Unified Genotyper [61], based on the
498 reference genome in which repeated sequences were detected using RepeatMasker [62] and
499 masked, so that SNPs were not called in these regions. In total 483,831 bp were masked,
500 corresponding to 1.67% of the FM164 genome sequence. The 1% and 99% quantiles of the
501 distribution of coverage depth were assessed across each sequenced genome and SNPs called at
502 positions where depth values fell in these extreme quantiles were removed from the dataset. Only
503 SNPs with less than 10% of missing data were kept. After filtering, a total of 115,544 SNPs were
504 kept.

505 The strain tree was inferred by maximum likelihood using the RAxML program v7.0.3 [63] under
506 the GTRCAT model using 6905 concatenated genes. To take into account possible differences in
507 nucleotide substitution rates, the dataset was divided into two partitions, one including the 1st and
508 2nd codon positions and one including the 3rd codon positions. To assess node confidence, 1000
509 bootstraps were computed.

510 Population structure was assessed using a discriminant analysis of principal components (DAPC)
511 with the Adegenet R package [64]. The genetic structure was also inferred along the genome by
512 clustering the strains according to similarities of their genotypes, in windows of 50 SNPs, using
513 the Mclust function of the mclust R package [65,66] with Gower's distance and a Gaussian
514 mixture clustering with K=7 (as the above analyses indicated the existence of four *P. roqueforti*
515 populations and there were three outgroup species).

516 The nucleotidic diversity, genetic diversity and linkage disequilibrium were estimated using the
517 θ_{π}, θ_{w} and r^2 statistics, respectively, with the *compute* and *rsq* programs associated to libsequence
518 v1.8.9 [67] on 1145 sliding windows of 50 kb with 25 kb of overlap distributed along the longest
519 eleven scaffolds of the FM164 assembly (> 200 kb).

520 To identify genes evolving under positive selection in *P. roqueforti* genomes, we used the method
521 implemented in SnIPRE [49], a Bayesian generalization of the log-linear model underlying the
522 McDonald-Kreitman test. This method detects genes in which amino-acid changes are more
523 frequent than expected under neutrality, by contrasting synonymous and non-synonymous SNPs,
524 polymorphic or fixed in two groups, to account for gene-specific mutation rates.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

525

526 ***Strain genotyping***

527 We identified two genomic regions with multiple diagnostic SNPs allowing discriminating the two
528 cheese clusters. Two PCR primer pairs were designed (Supplementary Table 2) to sequence these
529 regions in order to assign the 65 strains (Supplementary Table 1) that can be purchased at the
530 Laboratoire Interprofessionnel de Production d'Aurillac (LIP) (the main French supplier of *P.*
531 *roqueforti* spores for artisanal and industrial cheese-makers; *https://www.lip-sas.fr/*) to the
532 identified clusters. PCR products were then purified and sequenced at Eurofins (France). Because
533 one of the cheese clusters included strains carrying the *Wallaby* and *CheesyTer* genomic islands
534 while the second cluster strains lacked these genomic regions [31], we used previously developed
535 primer pairs to check for the presence/absence of *CheesyTer* and *Wallaby* [31].

536 Sequences were first aligned together with those extracted from sequenced genomes, allowing
537 assignation of LIP strains to one of the two cheese populations using MAFFT software[70] and
538 then the alignments were visually checked. Then a tree reconstruction was made using RAxML
539 following GTRCAT substitution model, using 2 partitions corresponding to the two fragments and
540 1000 bootstraps tree were generated [63].

541

542 ***Strain phenotyping***

543 Experimental cheeses were produced in an artisanal dairy company (GAEC Le Lévejac, Saint
544 Georges de Lévejac, France). The same ewe curd was used for all produced cheeses. Seven *P.*
545 *roqueforti* strains were used for inoculation (two from each of the Roquefort, non-Roquefort and
546 silage/food spoiler clusters, and one from the lumber/food spoiler cluster; their identity is given in
547 Supplementary Table 1) using 17.8 mg of lyophilized spores. Three cheeses were produced for
548 each strain in cheese strainers (in oval pots with opposite diameters of 8 and 9 cm, respectively),
549 as well as a control cheese without inoculation. After 48 h of draining, cheeses were salted (by
550 surface scrubbing with coarse salt), pierced and placed in a maturing cellar for four weeks at 11°C.
551 Cheeses were then sliced into six equal pieces and a picture of each slice was taken using a Nikon
552 D7000 (zoom lens: Nikon 18-105mm f:3.5-5.6G). Pictures were analyzed using the geospatial
553 image processing software ENVI (Harris Geospatial Solution) (Fig. 6F). This software enables
554 pixel classification according to their level of blue, red, green, and grey into two to four classes
555 depending on the analyzed image. This classification allowed assigning pixels to two classes

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

556 corresponding to the inner white part and the cavities of the cheese, respectively (Fig. 6F). For
557 each picture, the percentage of pixels corresponding to the cavities was then quantified. Because
558 the software could not reliably assign pixels to the presence versus absence of the fungus in
559 cavities, we visually determined the cavity areas that were colonized by *P. roqueforti* using
560 images. This allowed calculating a cheese cavity colonization rate. Because *Penicillium* spores
561 have a high dispersal ability which could cause contaminations, we confirmed strain identity
562 present in cheeses by performing Sanger sequencing of four diagnostic markers designed based on
563 SNPs and specific to each strain (Supplementary Table 2). For each cheese, three random
564 monospore isolates were genotyped, and no contamination was detected (i.e. all the sequences
565 obtained corresponded to the inoculated strains).

566 To compare the growth rates of the different *P. roqueforti* clusters on bread (i.e. the traditional
567 multiplication medium), 24 strains were used (eight from each of the Roquefort and non-
568 Roquefort cheese clusters, five from the silage/food spoiler cluster, and three from the
569 lumber/food spoiler cluster; the identities of the strains are shown in Supplementary Table 1).
570 Each strain was inoculated in a central point in three Petri dishes by depositing 10 µL of a
571 standardized spore suspension (0.7×10^9 spores/mL). Petri dishes contained agar (2%) and crushed
572 organic cereal bread including rye (200 g/L). After three days at 25°C in the dark, two
573 perpendicular diameters were measured for each colony to assess colony size.

574 The lipolytic and proteolytic activities of *P. roqueforti* strains were measured as follows:
575 standardized spore suspensions (2500 spores/inoculation) for each strain (n=47: 15 from the
576 Roquefort cluster, 15 from the non-Roquefort cheese cluster, 10 from the silage/food spoiler
577 cluster and seven from the lumber/food spoiler cluster, identity in Supplementary Table 1) were
578 inoculated on the top of a test tube containing agar and tributyrin for lypolytic activity measure
579 (10 mL/L, ACROS Organics, Belgium) or semi-skimmed milk for the proteolytic activity measure
580 (40 g/L, from large retailers). The lipolytic and proteolytic activities were estimated by the
581 degradation degree of the compounds, which changes the media from opaque to translucent. For
582 each media, three independent experiments have been conducted. For each strain, duplicates were
583 performed in each experiment and the degree of enzymatic activity level in the medium was
584 marked. Measures were highly repeatable between the two replicates (Pearson's product-moment
585 correlation coefficient of 0.93 in pairwise comparison between replicates, $P<0.0001$). We
586 measured the distance between the initial mark and the hydrolysis, translucent front, after 7, 14, 21
587 and 28 days of growth at 20°C in the dark.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

588 A total of 47 strains were used to compare spore production between the four *P. roqueforti*
589 clusters (Supplementary Table 1), 15 belonging to the non-Roquefort cluster, 15 to the Roquefort
590 cluster, 10 to the silage/food spoiler cluster and seven to the lumber/food spoiler cluster. After
591 seven days of growth on malt agar in Petri dishes of 60 mm diameter at room temperature, we
592 scraped all the fungal material by adding 5 mL of tween water 0.005%. We counted the number of
593 spores per mL in the solution with a Malassez hemocytometer (mean of four squares per strain)
594 for calibrating spore solution. We spread 50 µL of the calibrated spore solution (i.e. 7.10^6
595 spores.mL^{-1}) for each strain on Petri dishes of 60 mm diameter containing three different media,
596 malt, cheese and bread agar (organic "La Vie Claire" bread mixed with agar), in duplicates (two
597 plates per medium and per strain). After eight days of growth at room temperature, we took off a
598 circular plug of medium with spores and mycelium at the top, using Falcon 15 mL canonical
599 centrifuge tubes (diameter of 15 mm). We inserted the plugs into 5 mL Eppendorf tubes
600 containing 2 mL of tween water 0.005% and vortexed for 15 seconds to detach spores from the
601 medium. Using a plate spectrophotometer, we measured the optical density (OD) at 600 nm for
602 each culture in the supernatant after a four-fold dilution (Supplementary Table 3).
603 To compare salt tolerance between *P. roqueforti* clusters, 26 strains were used (eight from the
604 Roquefort cluster, ten from the non-Roquefort cluster, three from the silage/food spoiler cluster,
605 and five from the lumber/food spoiler cluster; strain identities are shown in Supplementary Table
606 1). For each strain and each medium, three Petri dishes were inoculated by depositing 10 µL of
607 standardized spore suspension (0.7×10^9 spores/mL) on Petri dishes containing either only malt (20
608 g/L), malt and salt (NaCl 8%, which corresponds to the salt concentration used before fridge use
609 to avoid contaminants in blue cheeses), only goat cheese, or goat cheese and salt (NaCl 8%). The
610 goat cheese medium was prepared as described in a previous study [31]. Strains were grown at
611 25°C and colony size measured daily for 24 days.

612 Volatile production assays were performed on 16 Roquefort strains and 19 non-Roquefort cheese
613 strains grown on model cheeses as previously described [37]. Briefly, model cheeses were
614 prepared in Petri dishes and incubated for 14 days at 25 °C before removing three 10 mm-
615 diameter plugs (equivalent to approximately 1 g). The plugs were then placed into 22 mL Perkin
616 Elmer vials that were tightly closed with polytetrafluorethylene (PTFE)/silicone septa and stored
617 at -80°C prior to analyses [37]. Analyses and data processing were carried out by headspace trap-
618 gas chromatography-mass spectrometry (HS-trap-GC-MS) using a Perkin Elmer turbomatrix HS-
619 40 trap sampler, a Clarus 680 gas chromatograph coupled to a Clarus 600T quadrupole MS
620 (Perkin Elmer, Courtaboeuf, France), and the open source XCMS package of the R software

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

621 (http://www.r-project.org/), respectively, as previously described [71].

622 All phenotypic measures are reported in Supplementary Table 3. Statistical analyses for testing
623 differences in phenotypes between populations and/or media (Supplementary Table 4) were
624 performed with R software (http://ww.r-project.org).

625 Differences in volatile profiles among the two *P. roqueforti* cheese populations were analyzed
626 using a supervised multivariate analysis method, orthogonal partial least squares discriminant
627 analysis (OPLS-DA). OPLS is an extension of principal components analysis (PCA), that is more
628 powerful when the number of explained variables (Y) is much higher than the number of
629 explanatory variables (X). PCA is an unsupervised method maximizing the variance explained in
630 Y, while partial least squares (PLS) maximizes the covariance between X and Y(s). OPLS is a
631 supervised method that aims at discriminating samples. It is a variant of PLS which uses
632 orthogonal (uncorrelated) signal correction to maximize the explained covariance between X and
633 Y on the first latent variable, and components >1 capture variance in X which is orthogonal
634 (uncorrelated) to Y. The optimal number of latent variables was evaluated by cross-validation
635 [72]. Finally, to identify the volatile compounds that were produced in significantly different
636 quantities between the two populations, a t-test was performed using the R software
637 (http://www.r-project.org/).

638

639 ***Demographic modeling using approximate Bayesian computation (ABC)***

640 The likelihoods of 11 demographic scenarios for the *P. roqueforti* populations were compared
641 using approximate Bayesian computation (ABC) [73,74]. The scenarios differed in the order of
642 demographic events, and included 21 parameters to be estimated (Supplementary Figure 2). A
643 total of 262 fragments, ranging from 5 kb to 15 kb, were generated from observed SNPs by
644 compiling in a fragment all adjacent SNPs in complete linkage disequilibrium. The population
645 mutation rate θ (the product of the mutation rate and the effective population size) used for
646 coalescent simulations was obtained from data using $\theta_{w,}$ the Watterson's estimator. Simulated data
647 were generated using the same fragment number and sizes as the SNP dataset generated from the
648 genomes. Priors were sampled in a log-uniform distribution (Supplementary Figure 2C). For each
649 scenario, one million coalescent simulations were run and the following summary statistics were
650 calculated on observed and simulated data using msABC [75] : the number of segregating sites,
651 the estimators π [76] and θ_w [77] of nucleotide diversity, Tajima's D [78], the intragenic linkage

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

652 disequilibrium coefficient ZnS [79], F_{ST} [80], the percentage of shared polymorphisms between
653 populations, the percentage of private SNPs for each population, the percentage of fixed SNPs in
654 each population, Fay and Wu's H [81], the number of haplotypes [82] and the haplotype diversity
655 [82]. For each summary statistic, both average and variance values across simulated fragments
656 were calculated.
657
658 The choice of summary statistics to estimate posterior parameters is a crucial step in ABC [83].
659 We chose the summary statistics based on their capacity to discriminate scenarios, by testing if
660 their values were significantly different among scenarios across simulations running a Kruskal-
661 Wallis test. We finally kept 34 summary statistics for model choice: average and variance of
662 shared polymorphism percentages between pairs of population, variance of private SNP
663 percentages in lumber/food spoiler and non-Roquefort populations, average and variance of fixed
664 SNPs percentage between pairs of populations, average of F_{ST} between pairs of population,
665 average of Fay and Wu's H and number of haplotypes in the Roquefort population.

666 The posterior probability distributions of the parameters, the goodness of fit for each model and
667 model selection (Supplementary Figure 2E) were calculated using a rejection-regression procedure
668 [73]. Acceptance values of 0.005 were used for all analyses. Regression analyses was performed
669 using the "abc" R package (http://cran.rproject.org/web/packages/abc/index.html).
670

671 ***Estimate of time since domestication***

672 The multiple sequentially Markovian coalescent (MSMC) software was used to estimate the
673 domestication times of cheese populations. The estimate of the last time gene flow occurred
674 within each cheese population was taken as a proxy of time since domestication as it also
675 corresponds in such methods to bottleneck date estimates and is more precisely estimated.
676 Recombination rate was set at zero because sexual reproduction has likely not occurred since
677 domestication in cheese populations (see results). Segments were set to 21*1+1*2+1*3 for the
678 Roquefort population which contains three haplotypes (Figure 1) and to 10*1+15*2 for the non-
679 Roquefort population, which contains two closely related haplotypes (Figure 1). In both cases,
680 MSMC was run for 15 iterations and otherwise default parameters. The mutation rate was set to
681 10^{-8}.
682
683

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

684 **References**

685 1. Albert, F.W., Somel, M., Carneiro, M., Aximu-Petri, A., Halbwax, M., Thalmann, O., Blanco-
686 Aguiar, J.A., Plyusnina, I.Z., Trut, L., and Villafuerte, R. (2012). A comparison of brain gene
687 expression levels in domesticated and wild animals. PLoS Genet. *8*, e1002962.

688 2. Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M.T., Perloski, M.,
689 Liberg, O., Arnemo, J.M., Hedhammar, \AAke, and Lindblad-Toh, K. (2013). The genomic
690 signature of dog domestication reveals adaptation to a starch-rich diet. Nature *495*, 360–364.

691 3. Freedman, A.H., Lohmueller, K.E., and Wayne, R.K. (2016). Evolutionary history, selective
692 sweeps, and deleterious variation in the dog. Annu. Rev. Ecol. Evol. Syst. *47*, 73–96.

693 4. Hake, S., and Ross-Ibarra, J. (2015). The natural history of model organisms: genetic,
694 evolutionary and plant breeding insights from the domestication of maize. Elife *4*, e05861.

695 5. Li, D., Wang, X., Zhang, X., Chen, Q., Xu, G., Xu, D., Wang, C., Liang, Y., Wu, L., and
696 Huang, C. (2016). The genetic architecture of leaf number and its genetic relationship to
697 flowering time in maize. New Phytol. *210*, 256–268.

698 6. Wang, H., Studer, A.J., Zhao, Q., Meeley, R., and Doebley, J.F. (2015). Evidence that the
699 origin of naked kernels during maize domestication was caused by a single amino acid
700 substitution in tga1. Genetics *200*, 965–974.

701 7. Elmer, K.R., and Meyer, A. (2011). Adaptation in the age of ecological genomics: insights
702 from parallelism and convergence. Trends Ecol. Evol. *26*, 298–306.

703 8. Elmer, K.R., Fan, S., Kusche, H., Spreitzer, M.L., Kautt, A.F., Franchini, P., and Meyer, A.
704 (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat.
705 Commun. *5*, 5168.

706 9. Bao, Y., Hu, G., Flagel, L.E., Salmon, A., Bezanilla, M., Paterson, A.H., Wang, Z., and
707 Wendel, J.F. (2011). Parallel up-regulation of the profilin gene family following independent
708 domestication of diploid and allopolyploid cotton (Gossypium). Proc. Natl. Acad. Sci. *108*,
709 21152–21157.

710 10. McTavish, E.J., Decker, J.E., Schnabel, R.D., Taylor, J.F., and Hillis, D.M. (2013). New
711 World cattle show ancestry from multiple independent domestication events. Proc. Natl. Acad.
712 Sci. *110*, E1398–E1406.

713 11. Wake, D.B., Wake, M.H., and Specht, C.D. (2011). Homoplasy: from detecting pattern to
714 determining process and mechanism of evolution. Science *331*, 1032–1035.

715 12. Wang, M., Yu, Y., Haberer, G., Marri, P.R., Fan, C., Goicoechea, J.L., Zuccolo, A., Song, X.,
716 Kudrna, D., and Ammiraju, J.S. (2014). The genome sequence of African rice (Oryza
717 glaberrima) and evidence for independent domestication. Nat. Genet. *46*, 982.

718 13. Parker, H.G., Harris, A., Dreger, D.L., Davis, B.W., and Ostrander, E.A. (2017). The bald and
719 the beautiful: hairlessness in domestic dog breeds. Phil Trans R Soc B *372*, 20150488.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

720 14. Shapiro, M.D., Kronenberg, Z., Li, C., Domyan, E.T., Pan, H., Campbell, M., Tan, H., Huff,
721 C.D., Hu, H., and Vickrey, A.I. (2013). Genomic diversity and evolution of the head crest in
722 the rock pigeon. Science *339*, 1063–1067.

723 15. Gouyon, P.-H., Leriche, H., Civard, A., Reeves, H., and Hulot, N. (2010). Aux origines de
724 l'environnement, Fayard Ed. (France : Paris).

725 16. Harlan, J.R. (1992). Crops and man. American Society of Agronomy Ed. (USA: Madison,WI).

726 17. Vavilov, N. (1992) Origin and geography of cultivated plants, Cambridge University Press
727 (Great Britain: Cambridge).

728 18. Giraud, T., Koskella, B., and Laine, A.-L. (2017). Introduction: microbial local adaptation:
729 insights from natural populations, genomics and experimental evolution. Mol. Ecol. *26*, 1703–
730 1710.

731 19. Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguileta, G., Vienne, D.M., Rodríguez de la
732 Vega, R.C., Branco, S., and Giraud, T. (2014). Fungal evolutionary genomics provides insight
733 into the mechanisms of adaptive divergence in eukaryotes. Mol. Ecol. *23*, 753–773.

734 20. Dupont, J., Dequin, S., Giraud, T., Le Tacon, F., Marsit, S., Ropars, J., Richard, F., and
735 Selosse, M.-A. (2016). Fungi as a Source of Food. Spectrum *5*.

736 21. Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P., and Sampaio, J.P. (2017). Adaptive
737 divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions
738 and rapid evolution at noncoding sites. Mol. Ecol. *26*, 2167–2182.

739 22. Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I.,
740 Albertin, W., Durrens, P., Sherman, D.J., and Marullo, P. (2014). A Gondwanan imprint on
741 global diversity and domestication of wine and cider yeast Saccharomyces uvarum. Nat.
742 Commun. *5*, 4044.

743 23. Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M.,
744 Gonçalves, P., and Sampaio, J.P. (2016). Distinct domestication trajectories in top-fermenting
745 beer yeasts and wine yeasts. Curr. Biol. *26*, 2750–2761.

746 24. Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede,
747 A., Roncoroni, M., Voordeckers, K., and Miraglia, L. (2016). Domestication and divergence
748 of Saccharomyces cerevisiae beer yeasts. Cell *166*, 1397–1410.

749 25. Marsit, S., Mena, A., Bigey, F., Sauvage, F.-X., Couloux, A., Guy, J., Legras, J.-L., Barrio, E.,
750 Dequin, S., and Galeote, V. (2015). Evolutionary advantage conferred by an eukaryote-to-
751 eukaryote gene transfer event in wine yeasts. Mol. Biol. Evol. *32*, 1695–1707.

752 26. Borneman, A.R., Desany, B.A., Riches, D., Affourtit, J.P., Forgan, A.H., Pretorius, I.S.,
753 Egholm, M., and Chambers, P.J. (2011). Whole-genome comparison reveals novel genetic
754 elements that characterize the genome of industrial strains of Saccharomyces cerevisiae. PLoS
755 Genet. *7*, e1001287.

756 27. Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A.,
757 Barre, B., Freel, K., and Llored, A. (2018). Genome evolution across 1,011 Saccharomyces

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

758 cerevisiae isolates. Nature *556*, 339.

759 28. Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.-L.,
760 Wincker, P., and Casaregola, S. (2009). Eukaryote-to-eukaryote gene transfer events revealed
761 by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proc. Natl.
762 Acad. Sci. *106*, 16333–16338.

763 29. Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I.,
764 Couloux, A., Guy, J., and Franco-Duarte, R. (2018). Adaptation of S. cerevisiae to fermented
765 food environments reveals remarkable genome plasticity and the footprints of domestication.
766 Mol. Biol. Evol. *35*, 1712–1727.

767 30. Morales, L., and Dujon, B. (2012). Evolutionary role of interspecies hybridization and genetic
768 exchanges in yeasts. Microbiol. Mol. Biol. Rev. *76*, 721–739.

769 31. Ropars, J., de la Vega, R.C.R., López-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, É.,
770 Lacoste, S., Debuchy, R., Dupont, J., Branca, A., *et al.* (2015). Adaptive Horizontal Gene
771 Transfers between Multiple Cheese-Associated Fungi. Curr. Biol., *25*, 1-8 .

772 32. Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A.-L.,
773 Ceppi, M., Conseiller, E., and Debuchy, R. (2014). Multiple recent horizontal transfers of a
774 large genomic region in cheese making fungi. Nat. Commun. *5*, 1–9.

775 33. Vabre, S. (2015). Le sacre du Roquefort. Press. Univ. Fr. Ed. (France: Paris)

776 34. Gillot, G., Jany, J.-L., Coton, M., Le Floch, G., Debaets, S., Ropars, J., López-Villavicencio,
777 M., Dupont, J., Branca, A., and Giraud, T. (2015). Insights into Penicillium roqueforti
778 morphological and genetic diversity. PloS One *10*, e0129849.

779 35. Ropars, J., López-Villavicencio, M., Snirc, A., Lacoste, S., and Giraud, T. (2017). Blue
780 cheese-making has shaped the population genetic structure of the mould Penicillium
781 roqueforti. PloS One *12*, e0171387.

782 36. Ropars, J., López-Villavicencio, M., Dupont, J., Snirc, A., Gillot, G., Coton, M., Jany, J.-L.,
783 Coton, E., and Giraud, T. (2014). Induction of sexual reproduction and genetic diversity in the
784 cheese fungus Penicillium roqueforti. Evol. Appl. *7*, 433–441.

785 37. Gillot, G., Jany, J.-L., Poirier, E., Maillard, M.-B., Debaets, S., Thierry, A., Coton, E., and
786 Coton, M. (2017). Functional diversity within the Penicillium roqueforti species. Int. J. Food
787 Microbiol. *241*, 141–150.

788 38. Aussibal, R. (1983). Les caves de roquefort, Musée du Rouergue Ed. (France: Cahors).

789 39. Labbe, M., and Serres, J.P. (2004). Chroniques du Roquefort—De la préhistoire à l'aube
790 industrielle. Maurice Labbé Ed.(France: Saint-Rome-de-Cernon).

791 40. Labbe, M., and Serres, J.P. (2009). Chroniques du Roquefort-Des hommes, des entreprises,
792 des marques, période moderne. Maurice Labbé Ed.(France: Saint-Rome-de-Cernon).

793 41. Marre, E. (1906). Le roquefort, E. Carrère Ed. (France : Rodez).

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

794 42. Marres, P. (1935). Les Grands Causses, Arrault et Cie Ed. (France : Tours).

795 43. Vabre, S. (2010). Roquefort Société: une industrie agro-alimentaire en Aveyron (vers 1840-
796 1914).

797 44. Besana, C., D'Errico, R., and Ghezzi, R. (2017). Cheese Manufacturing in the Twentieth
798 Century. The Italian Experience in an International Context, Peter Lang Ed (Belgium:
799 Bruxelles).

800 45. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history
801 from multiple genome sequences. Nat. Genet. *46*, 919.

802 46. McSweeney, P.L. (2004). Biochemistry of cheese ripening. Int. J. Dairy Technol. *57*, 127–
803 144.

804 47. Curioni, P.M.G., and Bosset, J.O. (2002). Key odorants in various cheese types as determined
805 by gas chromatography-olfactometry. Int. Dairy J. *12*, 959–984.

806 48. Gooday, G.W., Zhu, W.-Y., and O'Donnell, R.W. (1992). What are the roles of chitinases in
807 the growing fungus? FEMS Microbiol. Lett. *100*, 387–391.

808 49. Eilertson, K.E., Booth, J.G., and Bustamante, C.D. (2012). SnIPRE: selection inference using
809 a Poisson random effects model. PLoS Comput. Biol. *8*, e1002806.

810 50. Yvon, M., and Rijnen, L. (2001). Cheese flavour formation by amino acid catabolism. Int.
811 Dairy J. *11*, 185–201.

812 51. Ardö, Y. (2006). Flavour formation by amino acid catabolism. Biotechnol. Adv. *24*, 238–242.

813 52. Ropars, J., Lo, Y.-C., Dumas, E., Snirc, A., Begerow, D., Rollnik, T., Lacoste, S., Dupont, J.,
814 Giraud, T., and López-Villavicencio, M. (2016). Fertility depression among cheese-making
815 Penicillium roqueforti strains suggests degeneration during domestication. Evolution *70* (9)
816 2099-2109.

817 53. Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M.B., Ross-Ibarra, J., Myers, C.L.,
818 Tiffin, P., and Springer, N.M. (2012). Reshaping of the maize transcriptome by domestication.
819 Proc. Natl. Acad. Sci. *109*, 11878–11883.

820 54. Xue, S., Bradbury, P.J., Casstevens, T., and Holland, J.B. (2016). Genetic architecture of
821 domestication-related traits in maize. Genetics *204*, 99–113.

822 55. Chaptal, J.-A. (1789). Observations sur les caves et le fromage de Roquefort. Ann. Chimie *4*,
823 31–61.

824 56. Marcorelle, J.-F., and Chaptal, J.-A. (1833). Observations sur les caves et le fromage de
825 Roquefort. In Art de faire le beurre et les meilleurs fromages (Chez Madame Huzard).

826 57. Ropars, J., de La Vega, R.R., López-Villavicencio, M., Gouzy, J., Dupont, J., Swennen, D.,
827 Dumas, E., Giraud, T., and Branca, A. (2016). Diversity and mechanisms of genomic
828 adaptation in Penicillium. In: Aspergillus Penicillium Post-Genomic Era DeVries RG
829 Andersen MR Eds, 27–42.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

830 58. Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). pcaMethods—a
831 bioconductor package providing PCA methods for incomplete data. Bioinformatics *23*, 1164–
832 1167.

833 59. Glass, N.L., and Donaldson, G.C. (1995). Development of primer sets designed for use with
834 the PCR to amplify conserved genes from filamentous ascomycetes. Appl. Environ.
835 Microbiol. *61*, 1323–1330.

836 60. Lunter, G., and Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast
837 mapping of Illumina sequence reads. Genome Res. *21*, 936–939.

838 61. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella,
839 K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis Toolkit: a
840 MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res.
841 *20*, 1297–1303.

842 62. Smit, A., Hubley, R., and Green, P. (2013). RepeatMasker Open-4.0.

843 63. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
844 with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

845 64. Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.
846 Bioinformatics *24*, 1403–1405.

847 65. Fraley, C., Raftery, A.E., and Scrucca, L. (2012). Normal mixture modeling for model-based
848 clustering, classification, and density estimation. Dep. Stat. Univ. Wash. *23*, 2012.

849 66. Fraley, C., and Raftery, A.E. (2002). Model-based clustering, discriminant analysis, and
850 density estimation. J. Am. Stat. Assoc. *97*, 611–631.

851 67. Thornton, K. (2003). Libsequence: a C++ class library for evolutionary genetic analysis.
852 Bioinformatics *19*, 2325–2327.

853 68. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and
854 Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome
855 Biol. *5*, R12.

856 69. Foissac, S., Gouzy, J., Rombauts, S., Mathé, C., Amselem, J., Sterck, L., de Peer, Y.V.,
857 Rouzé, P., and Schiex, T. (2008). Genome annotation in plants and fungi: EuGene as a model
858 platform. Curr. Bioinforma. *3*, 87–97.

859 70. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version
860 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

861 71. Pogačić, T., Maillard, M.-B., Leclerc, A., Hervé, C., Chuat, V., Yee, A.L., Valence, F., and
862 Thierry, A. (2015). A methodological approach to screen diverse cheese-related bacteria for
863 their ability to produce aroma compounds. Food Microbiol. *46*, 145–153.

864 72. Pierre, P.S., Jansen, J.J., Hordijk, C.A., Van Dam, N.M., Cortesero, A.-M., and Dugravot, S.
865 (2011). Differences in volatile profiles of turnip plants subjected to single and dual herbivory
866 above-and belowground. J. Chem. Ecol. *37*, 368.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

867 73. Beaumont, M.A. (2010). Approximate bayesian computation in evolution and ecology. Annu.
868 Rev. Ecol. Evol. Syst., null.

869 74. Lopes, J.S., and Beaumont, M.A. (2010). ABC: a useful Bayesian tool for the analysis of
870 population data. Infect. Genet. Evol. *10*, 825–832.

871 75. Pavlidis, P., Laurent, S., and Stephan, W. (2010). msABC: a modification of Hudson's ms to
872 facilitate multi-locus ABC analysis. Mol. Ecol. Resour. *10*, 723–727.

873 76. Nei, M. (1987). Molecular evolutionary genetics, Columbia university press (USA: Columbia,
874 NY).

875 77. Watterson, G.A. (1975). On the number of segregating sites in genetical models without
876 recombination. Theor. Popul. Biol. *7*, 256–276.

877 78. Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA
878 polymorphism. Genetics *123*, 585–595.

879 79. Kelly, J.K. (1997). A test of neutrality based on interlocus associations. Genetics *146*, 1197–
880 1206.

881 80. Hudson, R.R., Slatkin, M., and Maddison, W.P. (1992). Estimation of levels of gene flow
882 from DNA sequence data. Genetics *132*, 583–589.

883 81. Fay, J.C., and Wu, C.-I. (2000). Hitchhiking under positive Darwinian selection. Genetics *155*,
884 1405–1413.

885 82. Depaulis, F., and Veuille, M. (1998). Neutrality tests based on the distribution of haplotypes
886 under an infinite-site model. Mol. Biol. Evol. *15*, 1788–1790.

887 83. Csilléry, K., Blum, M.G., Gaggiotti, O.E., and François, O. (2010). Approximate Bayesian
888 computation (ABC) in practice. Trends Ecol. Evol. *25*, 410–418.

889 84. Marcet-Houben, M., Ballester, A.-R., de la Fuente, B., Harries, E., Marcos, J.F., González-
890 Candelas, L., and Gabaldón, T. (2012). Genome sequence of the necrotrophic fungus
891 Penicillium digitatum, the main postharvest pathogen of citrus. BMC Genomics *13*, 646.

892 85. van den Berg, M.A., Albang, R., Albermann, K., Badger, J.H., Daran, J.-M., Driessen, A.J.M.,
893 Garcia-Estrada, C., Fedorova, N.D., Harris, D.M., Heijne, W.H.M., *et al.* (2008). Genome
894 sequencing and analysis of the filamentous fungus Penicillium chrysogenum. Nat. Biotechnol.
895 *26*, 1161–1168.

896 86. Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design
897 program Primer3. Bioinformatics *23*, 1289.

898 87. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS:
899 processing mass spectrometry data for metabolite profiling using nonlinear peak alignment,
900 matching, and identification. Anal. Chem. *78*, 779–787.

901 88. Csilléry, K., François, O., and Blum, M.G. (2012). abc: an R package for approximate
902 Bayesian computation (ABC). Methods Ecol. Evol. *3*, 475–479.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

903

904 **Acknowledgments**

905 This work was supported by the ERC starting grant GenomeFun 309403 awarded to TG, the ANR
906 FROMA-GEN grant (ANR-12-PDOC-0030) to AB, and an "*Attractivité*" grant from Paris-Sud
907 University to AB. We thank Kamel Soudani for help with image analysis and Aurélien Tellier for
908 advice concerning ABC analyses. We are grateful to Coralie Benel and Francis Roujon of GAEC
909 Le Lévejac for assistance with cheesemaking and Paul Villain for experimental help. Sequencing
910 was performed at GenoToul INRA platform. We thank INRA and MNHN for granting access to
911 four genomes sequenced with the help of Joëlle Dupont, Sandrine Lacoste, Yves Brygoo and
912 Jeanne Ropars in the framework of the ANR 'Food Microbiomes' project (ANR-08-ALIA-007-
913 02) coordinated by Pierre Renault.

914

915 **Author contributions**

916 TG and AB acquired the funding, designed and supervised the study. SL and AS produced the
917 genomes. ED, AB and RdlV analyzed the genomes. ED, SL, JR, AS, MC, AT, EC, MLP and DR
918 performed the experiments. ED, AB and TG analyzed the data from the experiments. ED, AB and
919 AF performed ABC analyses. ED and TG wrote the manuscript with contributions from the other
920 authors.

921 **Figure legends**

922 **Figure 1: Diversity and population subdivision in *Penicillium roqueforti*.** Unrooted maximum
923 likelihood tree of *P. roqueforti* strains generated with RAxML from concatenated sequences of
924 6905 single-copy genes. The *P. roqueforti* ingroup is rooted on the closely related *P.*
925 *psychrosexualis*, *P. carneum* and *P. paneum* species. Node support values are based on 1,000

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

926 bootstrap replicates. The scale bar indicates the number of substitutions per site. The letters
927 indicate the origin of the strains, C = Cheese, F = Food, S = Silage and L = Lumber. The color
928 indicates assignment to one of the four *P. roqueforti* populations identified, as in the other figures.
929 Blue, non-Roquefort; purple, Roquefort; green, lumber/food spoilage, and; orange, silage/food
930 spoilage.

931 **Figure 2: Clustering of *Penicillium roqueforti* along the FM164 reference genome using non-**
932 **overlapping 50 SNP sliding windows.** Clustering was done in each window using the mclust
933 function with Gaussian mixture modelling and using the Gower's distance between haplotypes.
934 The maximum number of clusters was fixed to seven, corresponding to the three outgroup species
935 plus the four populations of *P. roqueforti*. Each color corresponds to a cluster. Windows
936 containing fewer than 50 SNPs at the edge of scaffolds are not represented. The dendrogram on
937 the left side was reconstructed using hierarchical clustering based on the Gower's distance
938 between clusters for the entire genome. The histogram on the top left represents the distribution of
939 the number of clusters inferred for the whole genome. The letters indicate the origin of the strains,
940 C = Cheese, F = Food, S = Silage and L = Lumber.

941

942 **Figure 3**: **Genetic and phenotypic differentiation among *Penicillium roqueforti* populations.**
943 **Colors correspond to the genetic clusters as in other figures**. A: genetic differentiation assessed
944 by a discriminant analysis of principal components (DAPC) based on genome-wide single-
945 nucleotide polymorphisms (SNPs). The dots represent the strains and the colors the four
946 populations identified based on the genealogical tree in Fig. 1 as well as the similarity clustering
947 in Fig. 2. The insets show the distribution of eigenvalues for the principal component analysis
948 (PCA) and for the discriminant analysis (DA). B: phenotypic differentiation among *P. roqueforti*
949 genetic clusters illustrated by a PCA based on all tested phenotypes. Colors correspond to the

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

950 genetic clusters as in other figures. Missing data correction has been done using Bayesian
951 correction in the pcaMethods package [58].

952 **Figure 4**: **Genetic diversity and linkage disequilibrium levels in the four *Penicillium*
953 *roqueforti* populations**. The colors denote the four populations as in other figures. Horizontal
954 lines on the boxplots represent the upper quartile, the median and the lower quartile. Dots
955 represent the outlier values. Different letters indicate significant differences (Supplementary Table
956 4). A: Nucleotidic diversity estimated using the average heterozygosity, Θ_{Π} per site; B: Extent of
957 the r^2 index (i.e., a measure of linkage disequilibrium). C: Genetic diversity estimated using the
958 Watterson estimator, Θ_W per site. All estimates were calculated on 1145 sliding windows of 50 kb
959 with 25 kb overlap.

960 **Figure 5: Demographic history of *Penicillium roqueforti* populations.** A. Demographic
961 scenario (S4) with the highest posterior probability for the history of *Penicillium roqueforti*
962 populations. Estimates of time since divergence are indicated in units of $2N_e$ generations
963 (Supplementary Figure 2 E); effective population sizes and their variation (bottlenecks) are
964 represented by the widths of the genealogy branches, with relative sizes being represented to scale.
965 The color indicates assignment to the *P. roqueforti* populations as in the other figures. B.
966 Estimated past migration rate (gene flow) within each of the two cheese populations backward in
967 time (t=0 represents the present time). The dashed red lines represent the inferred times of
968 domestication, estimated as the last time gene flow occurred within cheese populations. C.
969 Estimated demographic history for the Roquefort population using the multiple sequentially
970 Markovian coalescent (MSMC) method. The inferred population effective size is plotted along
971 generations backward in time (t=0 represents the present time). The dashed red line represents the
972 inferred domestication time, estimated as the last time gene flow occurred within the Roquefort
973 population (Fig. 5B). The scheme above the figure represents a schematic view of the effective

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

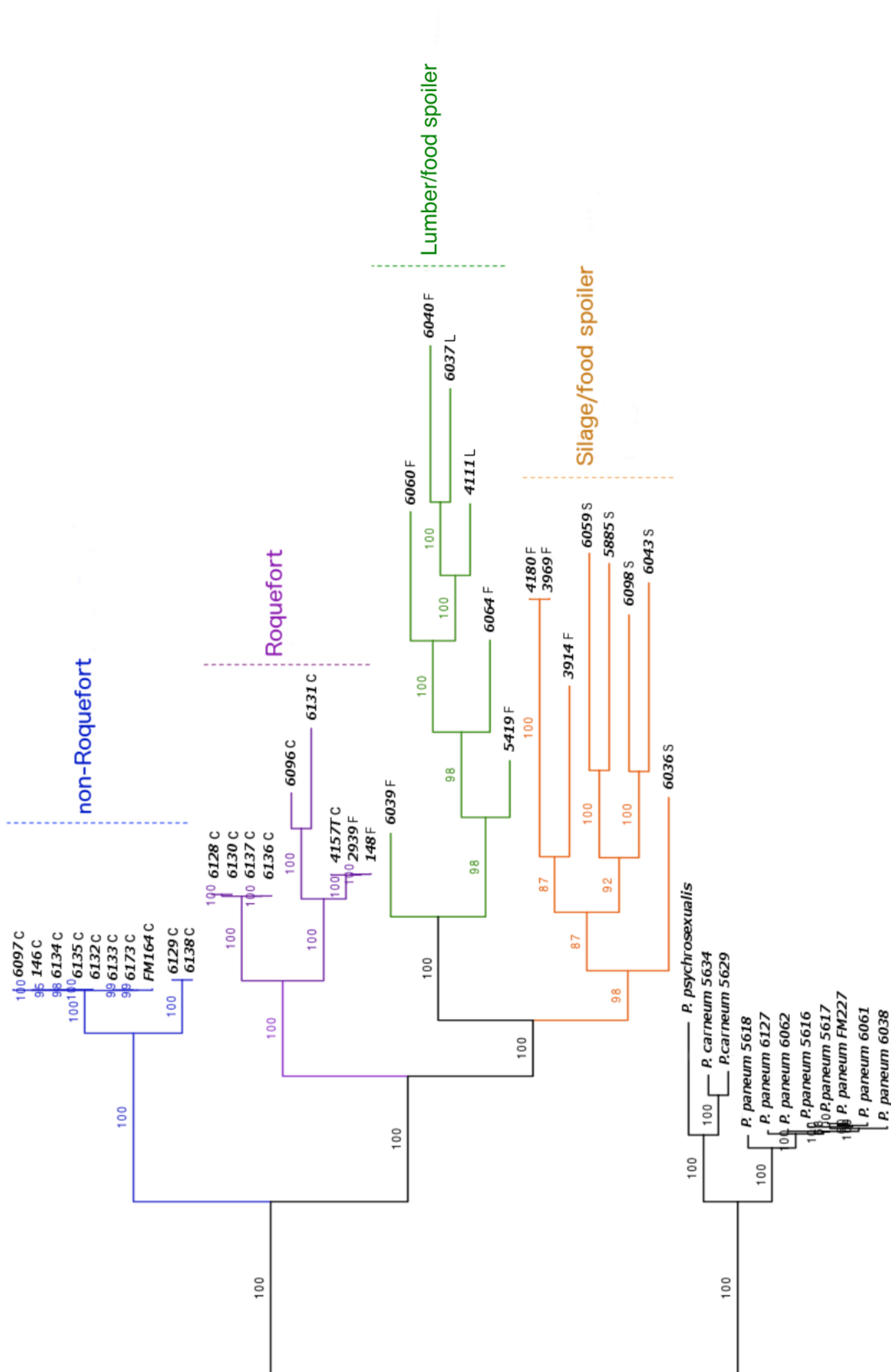
974 population size along generations, representing the two bottlenecks.

975

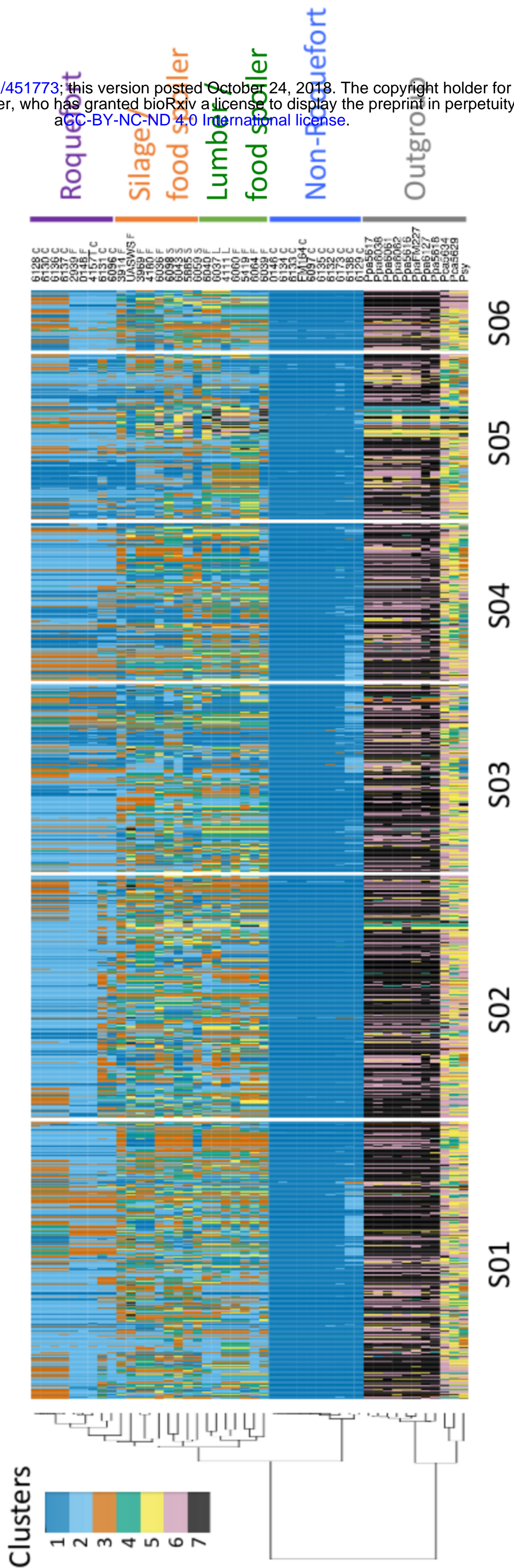
976 **Figure 6: Differences in phenotype between *Penicillium roqueforti* populations for various**
977 **traits relevant for cheesemaking.** The color indicates assignment to the *P. roqueforti* populations
978 identified, as in the other figures. Horizontal lines on the boxplots represent the upper quartile, the
979 median and the lower quartile. Dots represent the outlier values. Different letters indicate
980 significant differences (Supplementary Table 4). **A**: Growth on bread medium (colony size in mm
981 after three days of growth); **B**: Lipolytic activity measured at four different dates; **C**: Proteolytic
982 activity measured at four different dates; **D**: Spore production on cheese, bread and malt media
983 measured as optical density by spectrophotometer; **E**: Salt tolerance, i.e., growth rates on Petri
984 dishes containing cheese and malt media with (8%) and without (0%) NaCl. **F**: Cheese cavity
985 occupation (i.e., percentage of total cheese cavity space colonized by the fungus, as measured on
986 images) estimated in experimental cheeses by image analysis. The two clusters of non-cheese
987 strains were pooled, as there were too few strains per cluster to test differences between the
988 lumber/food spoiler and silage/food spoiler clusters. (a) Picture of a cheese slice. (b)
989 Corresponding image analysis using the geospatial image processing software ENVI (Harris
990 Geospatial Solution). Colors correspond to pixel classification based on their color on the picture.
991 In yellow and blue: the inner white part of the cheese; in green and red: cavities.

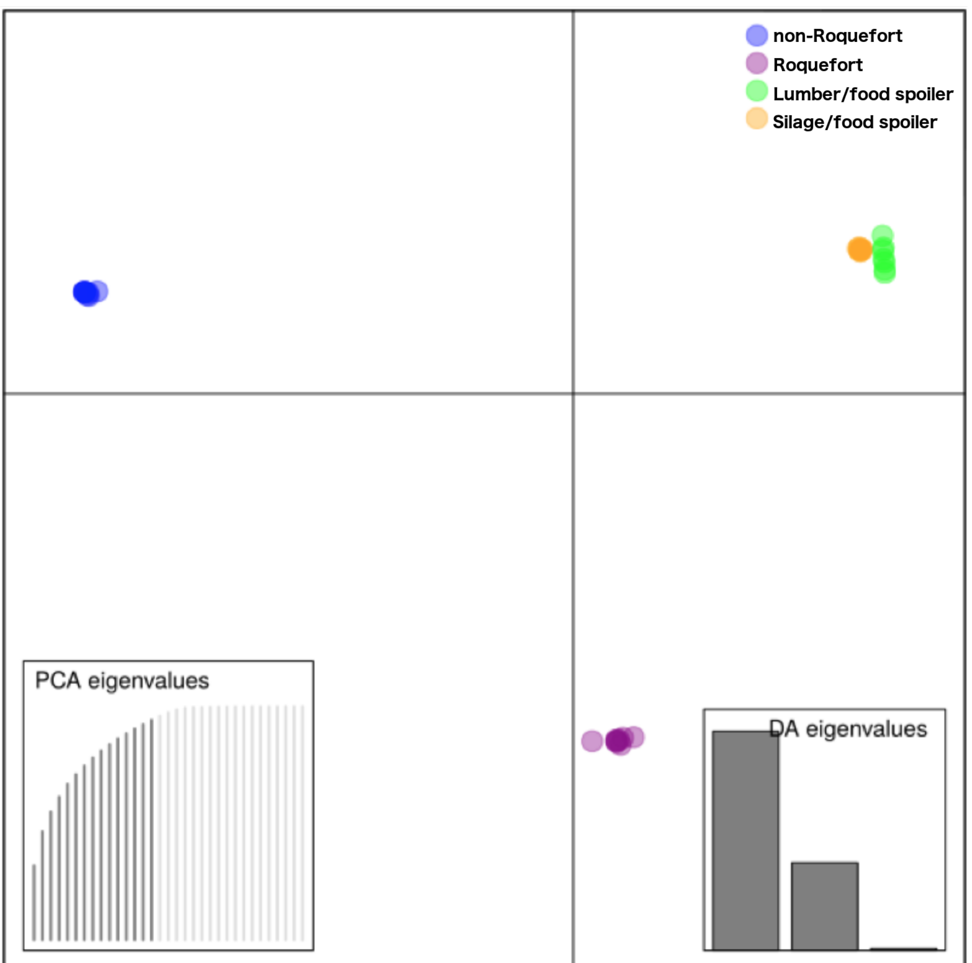
992

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



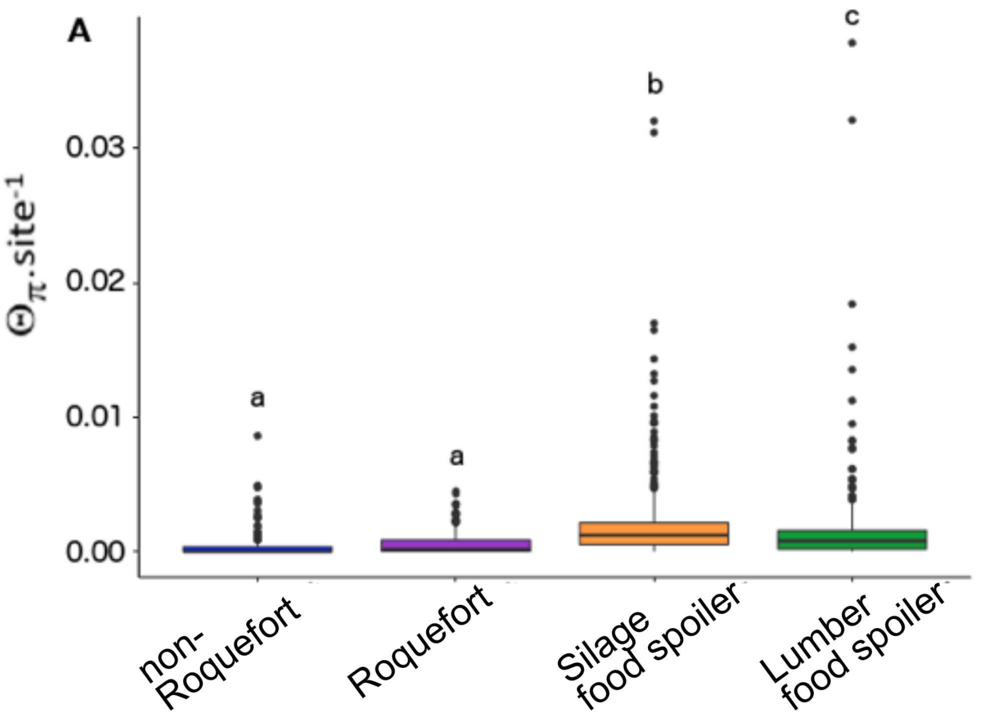
bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

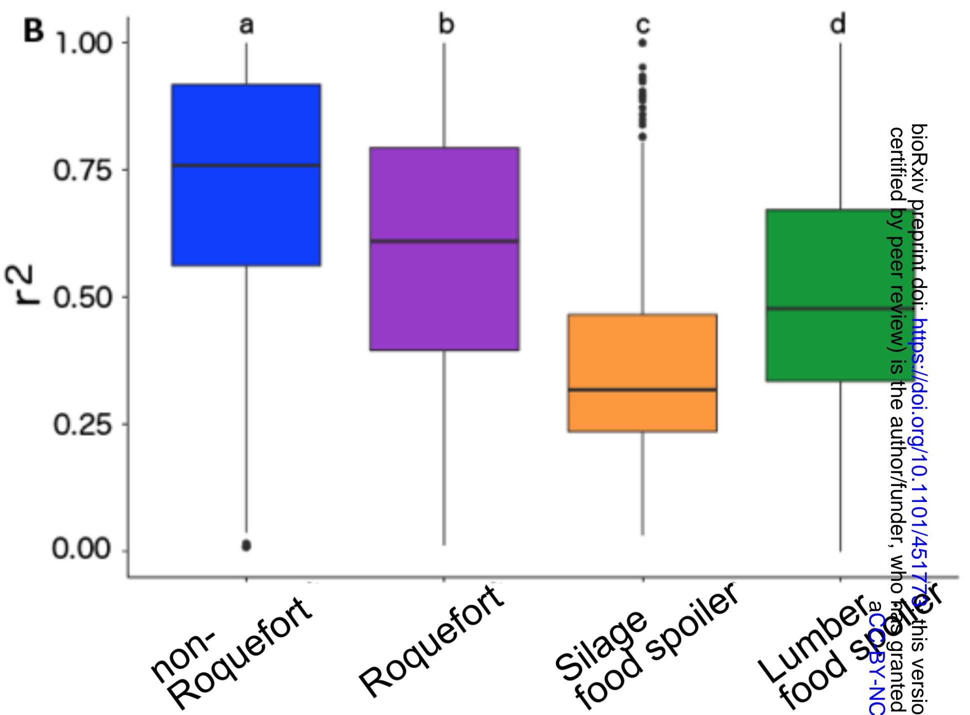


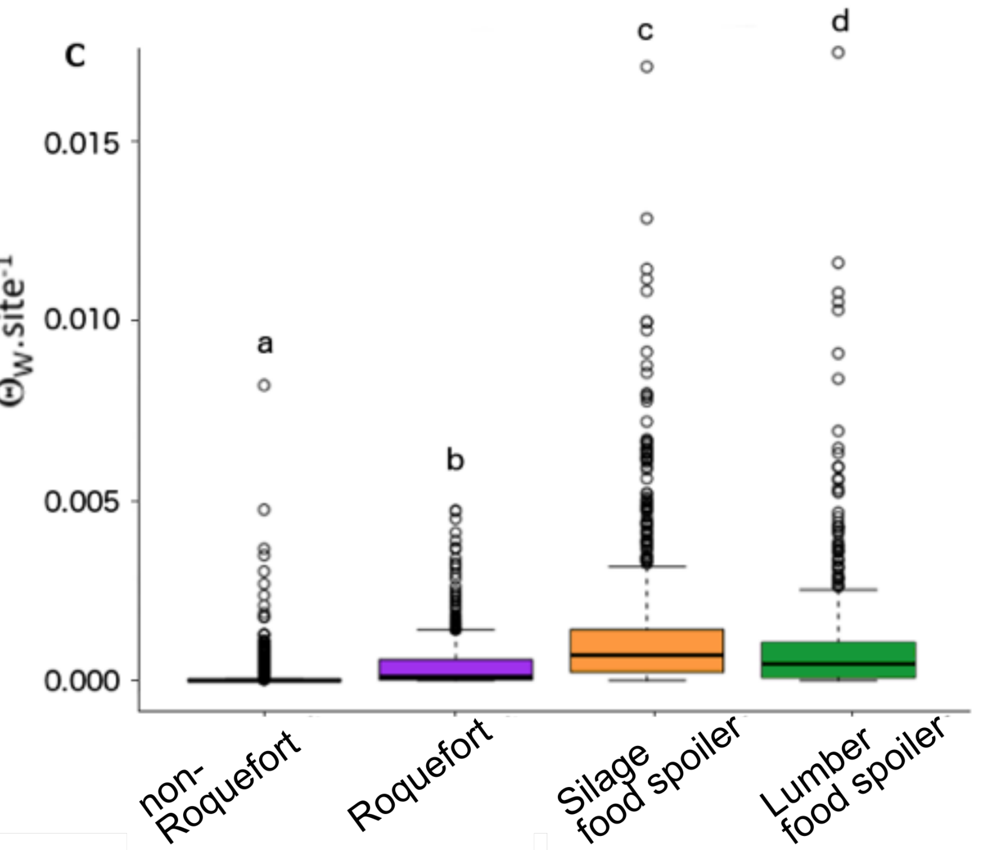


bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.





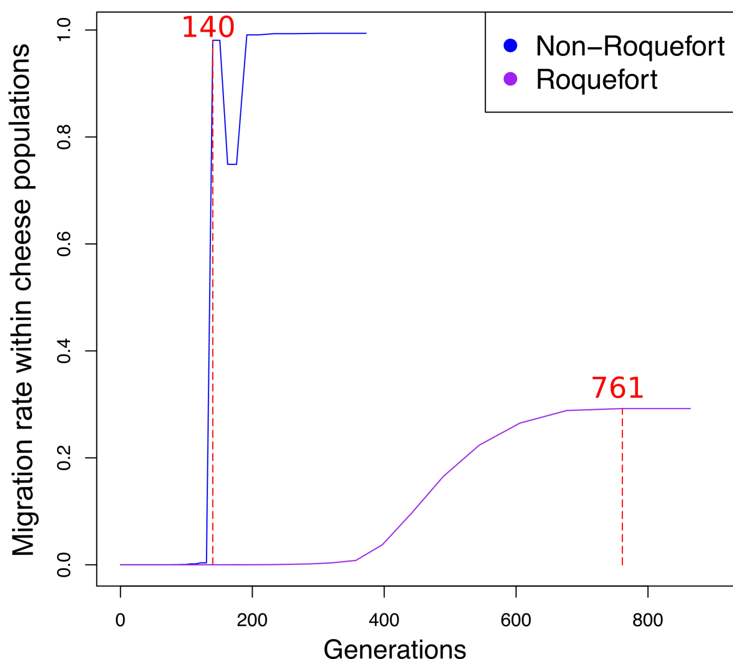


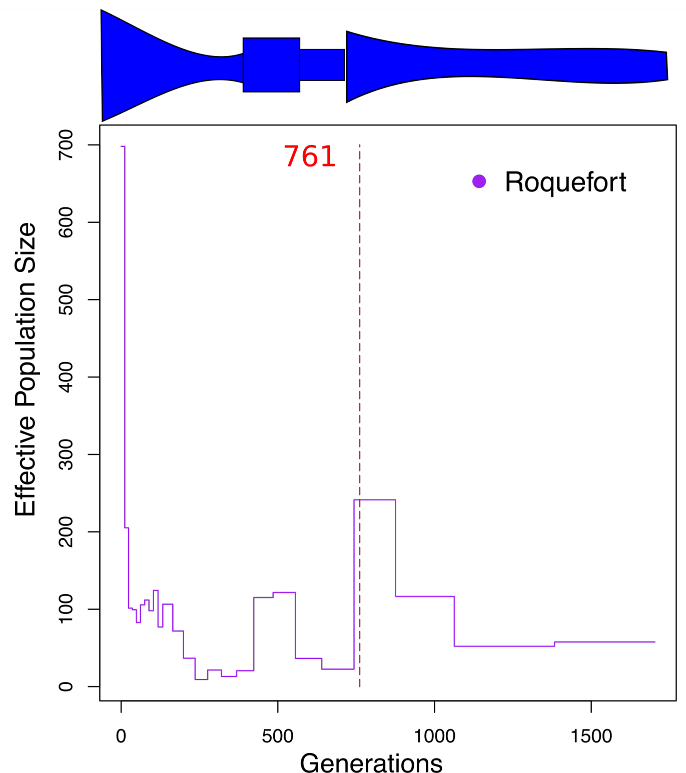
bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

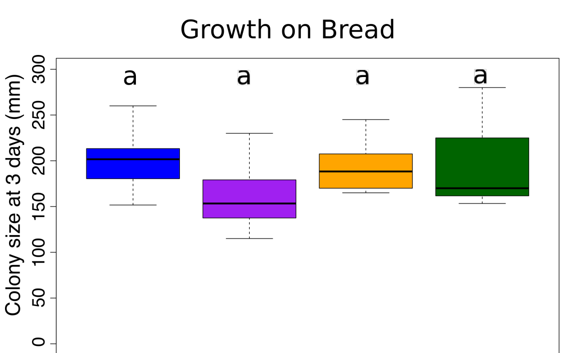
A

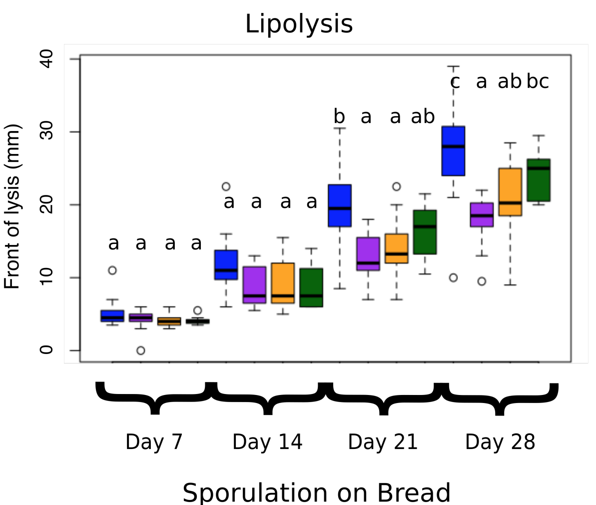
t11: 0.27
t8: 0.08
t10: 0.05
t3: 0.03
t7: 0.02
t9: 0.017
t12: 0.01 / t5: 0.01
t6: 0.008
t2: 0.004
t1: 0.001 / t4: 0.001

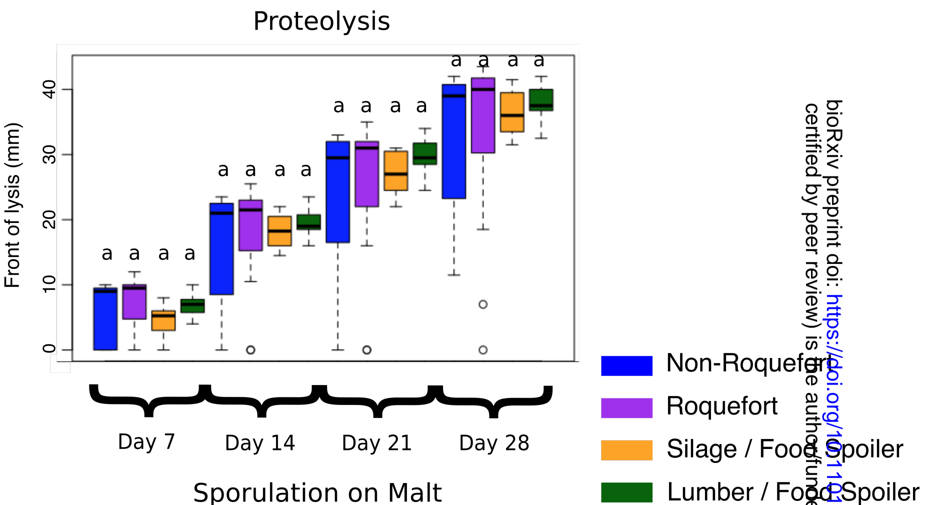
Lumber / food spoiler
Silage / food spoiler
Roquefort
Non-Roquefort

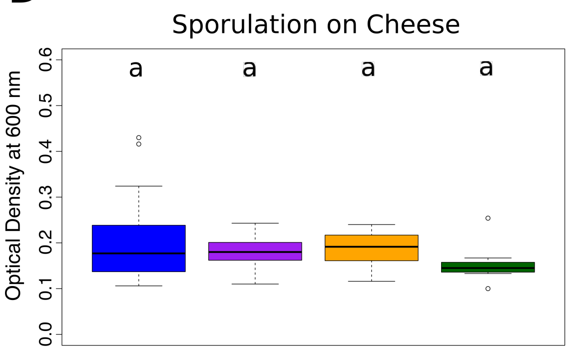






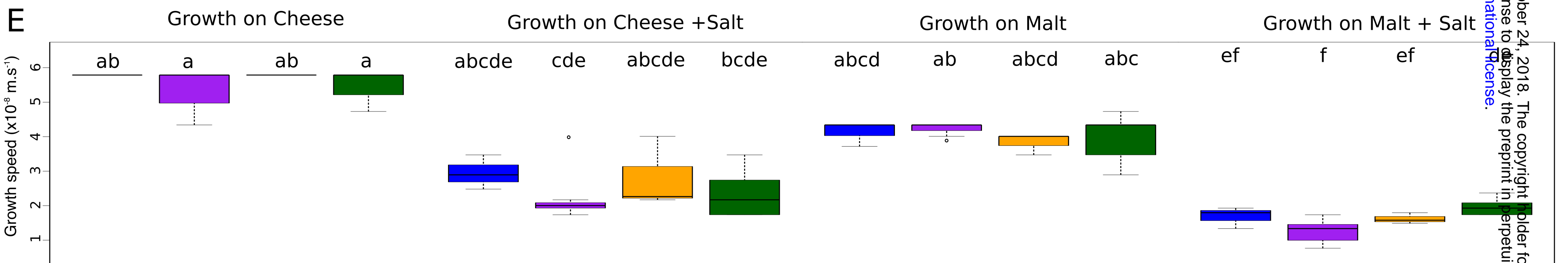






Sporulation on Bread

Sporulation on Malt



bioRxiv preprint doi: https://doi.org/10.1101/451773; this version posted October 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.