bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Independent domestication events in the blue-cheese fungus *Penicillium roqueforti*

Emilie Dumas[1], Alice Feurtey[1,2], Ricardo C Rodríguez de la Vega[1], Stéphanie Le Prieur[1], Alodie Snirc[1], Monika Coton[3], Anne Thierry[4], Emmanuel Coton[3], Mélanie Le Piver[5], Daniel Roueyre[5], Jeanne Ropars[1], Antoine Branca[1]* and Tatiana Giraud[1]*

[1]Ecologie Systématique et Evolution, Univ Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91400 Orsay, France

[2]Environmental Genomics, Max Planck Institute for Evolutionary Biology, August Thienemann-Str. 2, 24306 Plön, Germany

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

[4]Science et Technologie du Lait et de l'Œuf (STLO), UMR1253, INRA, Agrocampus Ouest, 65 rue de Saint Brieuc, 35000 Rennes, France

[5]Laboratoire Interprofessionnel de Production - SAS L.I.P, 34 Rue de Salers, 15000 Aurillac, France

*These authors jointly supervised the study

Corresponding author: Antoine Branca antoine.branca@u-psud.fr

Short title: Independent domestications of the blue-cheese fungus

Keywords: domestication, convergence, parallel adaptation, fungi, cheese, *Penicillium*

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Abstract

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

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Introduction

What are the mechanisms of adaptive divergence (population differentiation under selection) is a key question in evolutionary biology for understanding how organisms adapt to their environment and how biodiversity arises. Domestication is a special case of adaptive divergence, involving strong and recent selection for traits that can be easily identified. Furthermore, closely related non-domesticated populations are often available, making it possible to contrast their traits and genomes with those of domesticated populations. Studying domestication can therefore provide a deeper understanding of the mechanisms of adaptive divergence. This approach has proved to be powerful for reconstructing the history of divergence and the genetic architecture of traits selected by humans when applied to maize and teosinte or to dog breeds and wolves (Albert et al., 2012; Axelsson et al., 2013; Freedman et al., 2016; Hake and Ross-Ibarra, 2015; Li et al., 2016; Wang et al., 2015) Comparisons of domesticated varieties selected for different phenotypes have also proved to be a powerful approach for elucidating the mechanisms of adaptation, for example in dog breeds and pigeons (Parker et al., 2017; Shapiro et al., 2013)]. Studies on genetic diversity and subdivision in domesticated organisms provides also crucial information for the conservation of genetic resources. Indeed, recent breeding programs have resulted in a massive loss of genetic diversity in crops and breeds, potentially jeopardizing adaptive potential for improvement (Gouyon et al., 2010; Harlan, 1992; Vavilov, 1992).

Fungi are interesting eukaryotic models for adaptive divergence studies, with their small genomes, easy access to the haploid phase and experimental tractability for *in vitro* experiments (Giraud et al., 2017; Gladieux et al., 2014). Many fungi are used as food sources (Dupont et al., 2016) and some have been domesticated for food production. Propagation of the latter is controlled by humans, and this has resulted in genetic differentiation from wild populations (Almeida et al.,

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

2017, 2014; Gallone et al., 2016; Gibbons et al., 2012; Gonçalves et al., 2016) and the evolution of specific phenotypes beneficial for humans (Dupont et al., 2016; Gallone et al., 2016; Gibbons et al., 2012; Gibbons and Rinker, 2015; Marsit et al., 2015). *Saccharomyces cerevisiae* yeasts domesticated for fermentation have provided important insight into adaptive divergence mechanisms, with different yeast lineages independently domesticated for different usages (Borneman et al., 2011; Gonçalves et al., 2016; Peter et al., 2018). Studies about yeast adaptation for alcohol and cheese production have highlighted the proximal genomic mechanisms involved, including horizontal gene transfer, selective sweep, hybridization and introgression (Legras et al., 2018; Marsit et al., 2015; Morales and Dujon, 2012; Novo et al., 2009; Peter et al., 2018).

Penicillium roqueforti, a filamentous fungus used in the dairy industry to impart the typical veins and flavor of blue cheeses, has recently emerged as an excellent model for studying adaptive divergence (Cheeseman et al., 2014; Ropars et al., 2015). Blue cheeses, including Roquefort, Gorgonzola and Stilton, are highly emblematic foods that have been produced for centuries (Vabre, 2015). The strongest genetic subdivision reported in *P. roqueforti* concerns the differentiation of a cheese-specific population that has acquired faster growth in cheese than other populations and better excludes competitors, thanks to very recent horizontal gene transfers, at the expense of slower growth on minimal medium (Gillot et al., 2015; Ropars et al., 2015, 2017). Such genetic differentiation and recent acquisition of traits beneficial to cheesemaking in *P. roqueforti* suggests genuine domestication, i.e., adaptation under selection by humans for traits beneficial for food production. A second population identified in *P. roqueforti* and lacking the horizontally-transferred regions includes strains isolated from cheese and other environments, such as silage, lumber and spoiled food (Gillot et al., 2015; Ropars et al., 2014, 2017). *Penicillium roqueforti* is the main contaminant of silage, spoilage typically occurring following breaks in plastic or after opening the stack for cattle feeding. In this context, it can produce harmful mycotoxins causing health disorder in cattle (Malekinejad et al., 2015). In addition, *P. roqueforti* is one of the most common *Penicillium*

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

88 species in spoiled food, where it is also responsible for mycotoxin production (Rundberget et al.,
89 2004). The existence of further genetic subdivision separating populations according to the original
90 environment, or protected designation of origin (PDO) for cheese strains has been suggested, but,
91 because it was based only on a few microsatellite markers, the resolution power was low (Gillot et
92 al., 2015; Ropars et al., 2014, 2017). Secondary metabolite production (aroma compounds and
93 mycotoxins) and proteolysis activity have been shown to differ between strains from different PDOs
94 (Gillot et al., 2017). A high-quality *P. roqueforti* genome reference is available (Cheeseman et al.,
95 2014), allowing more powerful analyses based on population genomics.

96 Another asset of *P. roqueforti* as an evolutionary model is the availability of vast collections
97 of cheese strains and of historical records concerning cheesemaking (Aussibal, 1983; Labbe and
98 Serres, 2009, 2004; Marre, 1906; Marres, 1935; Vabre, 2010). While the presence of *P. roqueforti*
99 in cheeses was initially fortuitous, since the end of the 19th century, milk or curd has been inoculated
100 with the spores of this fungus for Roquefort cheese production. Spores were initially multiplied on
101 bread, before the advent of more controlled *in vitro* culture techniques in the 20th century (Aussibal,
102 1983; Labbe and Serres, 2009, 2004; Marre, 1906; Marres, 1935; Vabre, 2010). Bread was
103 inoculated by recycling spores from the best cheeses from the previous production (i.e., back-
104 slopping) (Aussibal, 1983; Labbe and Serres, 2009, 2004; Marre, 1906; Marres, 1935; Vabre, 2010).
105 This corresponds to yearly selection events since the 19th century until ca. 20 years ago when strains
106 were stored in freezers. After World War II, strains were isolated in the laboratory for industrial use
107 and selected based on their technological and organoleptic impact in cheeses and compounds
108 produced (Besana et al., 2017), which have likely accelerated domestication. This history further
109 suggests that there may have been genuine domestication, i.e., an adaptive evolution triggered by
110 human selection for cheese quality. Unintentional selection may also have been exerted on other
111 traits, including growth and spore production on bread, the traditional multiplication substrate.

112 By sequencing multiple *P. roqueforti* genomes from different environments and analyzing

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

large collections of cheese strains, we provide evidence for adaptive divergence. We identified four genetically differentiated populations, two including only cheese strains and two other populations including silage- and food-spoiling strains. We inferred that the two cheese populations corresponded to two independent domestication events. The first cheese population corresponded to strains used for Roquefort production and arose through a weaker and older domestication event, with multiple strains probably originating from different cultures on local farms in the PDO area, presumably initially selected for slow growth before the invention of refrigeration systems. The second cheese population experienced an independent and more recent domestication event associated with a stronger genetic bottleneck. The non-Roquefort cheese population showed beneficial traits for modern industrial production of cheese (e.g. faster growth in salted cheese, more efficient cheese cavity colonization and faster lipid degradation activities), while the Roquefort cheese population showed greater spore production on bread, the traditional medium for spore production. The four populations further showed differences in proteolysis activities, with a higher variance in the cheese populations. The two cheese populations also had different volatile compound profiles, with likely effects on cheese flavor. These phenotypic differences might be associated with genomic regions affected by recent positive selection and genomic islands specific to a single cheese population. Some of these genomic regions may have been acquired by horizontal gene transfers and have putative functions in the biochemical pathways leading to the development of cheese flavor.

Results

Two out of four populations are used for cheesemaking: one specific to the Roquefort PDO and a worldwide clonal population

We sequenced the genomes of 34 *P. roqueforti* strains from public collections (Ropars et al., 2017), including 17 isolated from blue cheeses (e.g., Roquefort, Gorgonzola, Stilton), 17 isolated from

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

non-cheese environments (mainly spoiled food, silage, and lumber), and 11 outgroup genomes from three *Penicillium* species closely related to *P. roqueforti* (Supplementary Table 1). After data filtering, we identified a total of 115,544 SNPs from the reads mapped against the reference *P. roqueforti* FM164 genome ($29x10^6$ bp, 48 scaffolds).

We used three clustering methods free from assumptions about mating system and mode of reproduction, based on genetic differences (principal component analyses, SplitsTree and clustering based on similarities between genotypes along the genomes in 50 SNP-windows). The three methods separated the *P. roqueforti* strains into four genetic clusters (Figs. 1, 2 and 3), two of which almost exclusively contained cheese strains (the exceptions being two strains isolated from a brewery and brioche, Figs. 1 and 2, probably corresponding to feral strains). One cluster contained both silage strains (N=4) and food-spoiling strains (N=4), and the last cluster contained mostly food-spoiling strains (N=5) plus strains from lumber (N=2) (Figs. 1 and 2, and Supplementary Table 1). Noteworthy, these two clusters corresponding to strains from other environments did not include a single cheese strain. The two cheese clusters were not the most closely related one to each other, suggesting independent domestication events (Figs. 1 and 2). Moreover, cheese clusters displayed much lower genetic diversity than non-cheese clusters, as shown by their small Θ values (corresponding to $4N_e\mu$, i.e., the product of the effective population size and the mutation rate) and more homogeneous colors in distance-based clustering (Table 1 and Fig. 2). One of the two cheese clusters displayed a particularly low level of genetic diversity (Table 1 and Fig. 2) with only 0.03% polymorphic sites, and a lack of recombination footprints (i.e., a higher level of linkage disequilibrium, as shown by the more gradual decay of r^2 values (Supplementary Fig. 1), and by the large single-color blocks along the genomes, Fig. 2). These findings suggest that the second cheese population is a single clonal lineage. The first cheese population also appears to lack recombination footprints, while including several clonal lineages (Fig. 2). Given such a lack of recombination footprints, clustering methods free of assumptions on modes of recombination were better suited to

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

analyse the dataset. The Structure software, that assumes random mating, nevertheless yielded similar results (Supplementary Fig. 2).

We used genome sequences to design genetic markers (Supplementary Table 2) for assigning a collection of 65 strains provided by the main French supplier of *P. roqueforti* spores for artisanal and industrial cheesemakers, 18 additional strains from the National History Museum collection in Paris (LCP) and 31 strains from the collection of the Université de Bretagne Occidentale (UBOCC, Supplementary Table 1) to the four genetic clusters. Out of these 148 strains, 55 were assigned to the more genetically diverse of the two cheese clusters. The majority of these strains included strains used for Roquefort PDO cheese production (N=30); three strains originated from Bleu des Causses cheeses (Supplementary Fig. 3, Supplementary Table 1), produced in the same area as Roquefort and using similarly long storage in caves. The remaining strains of this cluster included samples from other blue cheeses (N=13), unknown blue cheeses (N=5) or other environments (N=4), the latter likely associated with feral strains. Because of the strong bias of usage toward Roquefort production, we refer to this cluster hereafter as the "Roquefort population". Of the remaining 95 strains, 60 belonged to the second cheese cluster, which was less genetically diverse and contained mainly commercial strains used to produce a wide range of blue cheeses (Supplementary Fig. 3, Supplementary Table 1). This cluster was therefore named the "non-Roquefort population". A single strain (LCP00146) in this non-Roquefort population had been likely sampled from a Roquefort cheese, but it did not appear phenotypically different from other strains in its genetic group; the "Roquefort" origin may however be dubious as no brand was recorded for this strain from an old collection. The Roquefort population also included 13 strains used to inoculate other types of blue cheese (e.g. Gorgonzola or Bleu d'Auvergne), but strains from these types of cheeses were more common in the non-Roquefort population. The non-Roquefort cluster contained strains harbouring *Wallaby* and *CheesyTer*, two large genomic regions recently shown to have been transferred horizontally between different *Penicillium* species from the cheese environment and

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

conferring faster growth on cheese (Cheeseman et al., 2014; Ropars et al., 2015), whereas all the strains in the Roquefort cluster lacked those regions.

Two independent domestication events in Penicillium roqueforti *for cheesemaking*

We compared 11 demographic scenarios with approximate Bayesian computation (ABC), simulating either a single domestication event (the most recent divergence event then separating the two cheese populations) or two independent domestication events, with different population tree topologies and with or without gene flow (Supplementary Fig. 4). Parameters in the scenarios modeled corresponded to the divergence dates, the strength and dates of bottlenecks and population growth, and rates of gene flow. ABC simulates sequence evolution under the various scenarios using the coalescent theory framework and compares various population statistics under a Bayesian framework between the simulation outputs and the observed data to identify the most likely scenario (Beaumont et al., 2002). The ABC results showed that the two *P. roqueforti* cheese populations (Roquefort and non-Roquefort) resulted from two independent domestication events (Fig. 4). The highest posterior probabilities were obtained for the S4 scenario, in which the two cheese populations formed two lineages independently derived from the common ancestral population of all *P. roqueforti* strains (Fig. 4, model choice and parameter estimates in Supplementary Fig. 4). We inferred much stronger bottlenecks in the two cheese populations than in the non-cheese populations, with the most severe bottleneck found in the non-Roquefort cheese population. Some gene flow (m=0.1) was inferred between the two non-cheese populations but none with cheese populations. The bottleneck date estimates in ABC had too large credibility intervals to allow inferring domestication dates (Supplementary Fig. 4E). We therefore used the multiple sequentially Markovian coalescent (MSMC) method to estimate times since domestication, considering that they corresponded to the last time there was gene flow between genotypes within populations, given the

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

lack of recombination footprints in cheese population and the mode of conservation and clonal growth of cheese strains by humans, and given that this also corresponds to bottleneck date estimates in coalescence. The domestication for the Roquefort cheese population was inferred seven times longer ago than for the non-Roquefort cheese population, both domestication events being recent (ca. 760 versus 140 generations ago, Fig. 5B-C). Unfortunately, generation time, and even generation definition, are too uncertain in the clonal *P. roqueforti* populations to infer domestication dates in years. In addition, the MSMC analysis detected two bottlenecks in the history of the Roquefort cheese population (Fig. 4B).

Contrasting fitness traits between cheese populations

We tested whether different phenotypes relevant for cheesemaking had evolved in the two cheese clusters, relative to other populations (Fig. 5, Supplementary Table 3). We first produced experimental cheeses inoculated with strains from the different *P. roqueforti* populations to assess their ability to colonize cheese cavities, a trait that may have been subject to human selection to choose inocula producing the most visually attractive blue cheeses. The fungus requires oxygen and can therefore sporulate only in the cheese cavities, its spores being responsible for the typical color of blue-veined cheeses; the application of highly salted solutions followed by tin foil wrapping prevents sporulation on the surface of cheeses. Strains from the non-Roquefort cheese population were the most efficient colonizers of cheese cavities (Supplementary Table 4); no difference was detected between strains from the Roquefort and non-cheese populations (Fig. 5).

As *P. roqueforti* strains were traditionally multiplied on bread loaves for cheese inoculation, they may have been subject to unintentional selection for faster growth on bread. However, growth rate on bread did not significantly differ between populations (Fig. 5, Supplementary Table 4).

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

We then assessed lipolytic and proteolytic activities in the *P. roqueforti* populations. These activities are important for energy and nutrient uptake, as well as for cheese texture and the production of volatile compounds responsible for cheese flavors (Gillot et al., 2017; McSweeney, 2004). Lipolysis was faster in the non-Roquefort cheese population than in the Roquefort and silage/food spoiling populations (Fig. 5, Supplementary Table 4). A strong population effect was found for proteolytic activity (Supplementary Table 4), with faster proteolysis activities in cheese populations (Fig. 5), although posthoc pairwise tests were not significant. Variances showed significant differences between populations (Levene test F-ratio=5.97, d.f.=3, P<0.0017), with the two cheese populations showing the highest variances, and with extreme values above and below those in non-cheese populations (Fig. 5). Noteworthy, proteolysis is a choice criterion for making different kinds of blue cheeses that is often showcased by culture producers (e.g. https://www.lip-sas.fr/index.php/nos-produits/penicillium-roquefortii/18-penicillium-roquefortii). This suggests that some cheese strains may have been selected for higher and others for lower proteolytic activity. Alternatively, selection could have been relaxed on this trait in the cheese populations, leading to some mutations decreasing and other increasing proteolysis in different strains, thus increasing variance in the populations.

The ability of *P. roqueforti* strains to produce spores may also have been selected by humans, both unwittingly, due to the collection of spores from moldy bread, and deliberately, through the choice of inocula producing bluer cheeses. We detected no difference in spore production between the *P. roqueforti* populations grown on cheese medium or malt. However, we observed significant differences in spore production on bread medium. The Roquefort population produced the highest number of spores and significantly more than the non-Roquefort population (Fig. 5, Supplementary Table 4).

High salt concentrations have long been used in cheesemaking to prevent the growth of spoiler and pathogenic microorganisms. We found that the ability to grow on salted malt and cheese media

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

decreased in all *P. roqueforti* populations (Supplementary Table 4). We found a significant interaction between salt and population factors, and post hoc tests indicated that the Roquefort population was more affected by salt than the other populations (Supplementary Fig. S5, Supplementary Table 4).

Volatile compound production was also investigated in the two cheese populations, as these compounds are important for cheese flavor (McSweeney, 2004). We identified 52 volatile compounds, including several involved in cheese aroma properties, such as ketones, free fatty acids, sulfur compounds, alcohols, aldehydes, pyrazines, esters, lactones and phenols (Curioni and Bosset, 2002) (Fig. 6). The two cheese populations presented significantly different volatile compound profiles, differing by three ketones, one alcohol and two pyrazines (Fig. 6). The Roquefort population produced the highest diversity of volatile compounds (Fig. 6A).

Detection of genomic regions population specific or affected by recent positive selection

We identified five regions present in the genomes of strains from the non-Roquefort cheese population and absent from the other populations. We also detected five other genomic islands present in several *P. roqueforti* strains but absent from the non-Roquefort cheese strains (Fig. 7). Nine of these ten genomic regions were not found in the genomes of the outgroup *Penicillium* species analyzed here and they displayed no genetic diversity in *P. roqueforti*. No SNPs were detected even at synonymous sites or in non-coding regions, suggesting recent acquisitions, by horizontal gene transfer. The absence of the genomic islands in some populations and outgroups prevented running gene topology analyses designed for horizontal gene transfer analyses but were even stronger evidence for the existence of horizontal gene transfer. Only FM164-C, one of the genomic islands specific to the non-Roquefort population, was present in the outgroup genomes, in which it displayed variability, indicating a loss in the other lineages rather than a gain in the non-

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

283 Roquefort population and the outgroup species (Fig. 7A). The closest hits in the NCBI database for
284 genes in the ten genomic islands were in *Penicillium* genomes. Most of the putative functions
285 proposed for the genes within these genomic regions were related to lipolysis, carbohydrate or
286 amino-acid catabolism and metabolite transport. Other putative functions concerned fungal
287 development, including spore production and hyphal growth (Fig. 7). In the genomic regions
288 specific to the non-Roquefort cheese population, we also identified putative functions potentially
289 relevant for competition against other microorganisms, such as phospholipases, proteins carrying
290 peptidoglycan- or chitin-binding domains and chitinases (Fig. 7) (Gooday et al., 1992). Enrichment
291 tests were non-significant, probably due to the small number of genes in these regions.

292 Footprints of positive selection in *P. roqueforti* genomes were first detected using an extension of
293 the McDonald-Kreitman test which identifies genes with more frequent amino-acid changes than
294 expected under neutrality, neutral substitution rates being assessed by comparing the rates of
295 synonymous and non-synonymous substitutions within and between species or populations to
296 account for gene-specific mutation rates. We ran the test with three levels of population subdivision.
297 First, no significant footprint of positive selection was detected for any gene by comparing the whole
298 *P. roqueforti* species with *P. paneum*. In a second test, a set of 15 genes was identified as evolving
299 under positive selection in the Roquefort cheese population but not in the other pooled *P. roqueforti*
300 populations (Fig. 8A). Interestingly, eight of these 15 genes clustered at the end of the largest
301 scaffold (Fig. 8B). In a third test, four genes were identified as evolving under positive selection in
302 the non-Roquefort cheese population but not in the pooled non-cheese *P. roqueforti* populations
303 (Fig. 8A). Two of these genes corresponded to a putative aromatic ring hydroxylase and a putative
304 cyclin evolving under purifying selection in Roquefort and non-cheese *P. roqueforti* populations
305 (Fig. 8A). Aromatic ring hydroxylases are known to be involved in the catabolism of aromatic amino
306 acids, which are precursors of flavor compounds (Ardö, 2006; Yvon and Rijnen, 2001).

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Secondly, we looked for regions of low diversity and high divergence between the two cheese populations as these are footprints of recent divergent selection, i.e. positive selection in one or both of the two cheese populations but for differentiated alleles. The identified regions showed a good overlap with those detected in the Snipre analysis (Fig. 9); in particular, the same genomic island at the end of scaffold 1 stood out (Fig. 9). In the regions of high divergence and low diversity, we found a significant enrichment in transcription related genes (GO:0000981 RNA polymerase II transcription factor activity, sequence-specific DNA binding; Fisher's exact test p-value<0.01; Supplementary Fig. 6). We found a particularly high divergence on the gene coding for RPB2 subunit of RNA polymerase II with a high number of fixed differences that were specific to the Roquefort population; fixed differences were synonymous, suggesting that important changes concern rather the regulation level than the protein itself.

Discussion

We report here the genetic subdivision of *P. roqueforti*, the fungus used worldwide for blue cheese production, with unprecedented resolution, providing insights into its domestication history. Population genomics studies on strains from various substrates and from a large collection of cheeses identified four genetically differentiated populations, two of which being cheese populations likely originating from independent and recent domestication events. One *P. roqueforti* cheese population included all the genotyped strains but one used for PDO Roquefort cheeses, produced in the French town of Roquefort-sur-Soulzon, where blue cheeses have been made since at least the 15th century, and probably long before (Aussibal, 1983; Labbe and Serres, 2009, 2004; Marre, 1906; Marres, 1935; Vabre, 2015, 2010). The strains from this Roquefort cheese population lack the horizontally-transferred *Wallaby* and *CheesyTer* genomic islands contrary to the other non-Roquefort cheese population.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

We observed that the two *P. roqueforti* cheese populations differed on several traits important for cheese production, probably corresponding to historical differences. Indeed, the Roquefort population has retained moderate genetic diversity, consistent with soft selection during pre-industrial times on multiple farms near Roquefort-sur-Soulzon, where specific strains were kept for several centuries. The Roquefort cheese population grew slower in cheese (Ropars et al., 2015) and had weaker lipolytic activity. Slow maturation is particularly crucial for the storage of Roquefort cheeses for long periods in the absence of refrigeration (Marre, 1906) because they are made of ewe's milk, a product available only between February and July. During storage, cheeses could become over degraded by too high rates of lipolysis, thus likely explaining the low lipolysis activity in Roquefort strains. By contrast, most other blue cheeses are produced from cow's milk, which is available all year. The Roquefort population showed greater sporulation on bread, which is consistent with unconscious selection for this trait when strains were cultured on bread in Roquefort-sur-Soulzon farms before cheese inoculation during the end of the 19th and beginning of the 20th centuries.

Lipolytic activity is known to impact texture and the production of volatile compounds affecting cheese pungency (Alonso et al., 1987; De Llano et al., 1992, 1990; Martín and Coton, 2016; Thierry et al., 2017; Woo and Lindsay, 1984). The Roquefort and non-Roquefort populations showed different volatile compound profiles, suggesting also different flavor profiles. The discovery of different phenotypes in the two cheese populations, together with the availability of a protocol for inducing sexual reproduction in *P. roqueforti* (Ropars et al., 2014), pave the way for crosses to counteract degeneration after clonal multiplication and bottlenecks, for variety improvement and the generation of diversity.

Both cheese populations were found to have gone through bottlenecks. The cheese populations were

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

355 the easiest to sample compared to other environments, where *P. roqueforti* is relatively rarely found.
356 It seems therefore highly unlikely that the lower genetic diversity in the cheese populations would
357 reflect sampling biases. In particular, the least diverse cheese population was the one including the
358 highest numbers of countries and sampled cheese types, indicating genuine strong bottleneck. There
359 was no particular sampling bias regarding geography either (Table S1). A previous study showed
360 that these bottlenecks, together with clonal multiplication, decreased fertility, with different stages
361 in sexual reproduction affected in the two populations identified here as the Roquefort and non-
362 Roquefort lineages (Ropars et al., 2016b). The non-Roquefort cheese population, despite suffering
363 from a more severe and more recent bottleneck, was found to be used in the production of all types
364 of blue cheese worldwide, including Gorgonzola, Bleu d'Auvergne, Stilton, Cabrales and Fourme
365 d'Ambert. The non-Roquefort cheese population grows more rapidly on cheese (Ropars et al.,
366 2015), exhibits greater ability to colonize cheese cavities, higher salt tolerance and faster lipolysis
367 than the Roquefort population. These characteristics are consistent with the non-Roquefort cheese
368 population resulting from a very recent strong selection of traits beneficial for modern and
369 accelerated production of blue cheese using refrigeration techniques, followed by a worldwide
370 dissemination for the production of all types of blue cheeses. Such drastic losses of genetic diversity
371 in domesticated organisms are typical of strong selection for industrial use by a few international
372 firms and raise concerns about the conservation of genetic resources, the loss of which may hinder
373 future innovation. More generally in crops, the impoverishment in genetic diversity decreases the
374 ability of cultivated populations to adapt to environmental and biotic changes to meet future needs
375 (Gouyon et al., 2010; Harlan, 1992; Vavilov, 1992). The PDO label, which imposes the use of local
376 strains, has probably contributed to the conservation of genetic diversity in the Roquefort population
377 (see "Cahier des charges de l'appellation d'origine protégée Roquefort", i.e., the technical
378 specifications for Roquefort PDO). We inferred two bottlenecks in the Roquefort population, more
379 ancient than in the non-Roquefort population, likely corresponding to a pre-industrial domestication

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

event when multiple local farms multiplied their strains, followed by a second bottleneck when fewer strains were kept by the first industrial societies. For other blue cheeses, even if their production was also ancient, the performant non-Roquefort clonal lineage could have been recently chosen to fit modern industrial production demands due to the lack of PDO rules imposing the use of local strains. However, despite a much lower genome-wide diversity in domesticated populations, proteolysis and volatile compounds diversity was found higher in cheese than in non-cheese populations. In fact, different strains with more or less rapid proteolysis and lipolysis are sold for specific blue cheese types (e.g., milder or stronger), in particular by the French LIP company (https://www.lip-sas.fr/index.php/nos-produits/penicillium-roquefortii/18-penicillium-roquefortii). Such a high phenotypic diversity within the cheese populations is consistent with diversification of usage under domestication, and in particular when different characteristics are desired according to cheese type. This has already been observed in relation to the diversification of crop varieties or breeds in domesticated animals (Parker et al., 2017; Shapiro et al., 2013).

When studying adaptation in domesticated organisms, it is often useful to contrast traits and genomic variants between domesticated and closely related wild populations to determine the nature of the adaptive changes occurring under artificial selection (Swanson-Wagner et al., 2012; Xue et al., 2016). The only known non-cheese populations of *P. roqueforti* occur essentially in human-made environments (silage, food and lumber), consistent with the specific adaptation of these populations to these environments. The two non-cheese populations were inferred to have diverged very recently, and displayed footprints of recombination and marked differentiation from the cheese populations. Domesticated populations are expected to be nested within their source populations, suggesting that we have not sampled the wild population that is the most closely related from cheese strains yet. The high level of diversity and inferred demographic history of *P. roqueforti* indicate that most food-spoiling strains belong to differentiated populations and are not feral cheese strains. In addition, not a single cheese strain was found in the food spoiling and silage populations. This

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

was shown by both genome sequences and by the genotyping of a larger number of strains using a few selected markers, in the present study and based on microsatellite markers in a previous work (Ropars et al., 2017). Consequently, *P. roqueforti* spores from blue cheeses may, rarely, spoil food and food-spoiling and silage strains are not used for cheesemaking nor recombine with cheese strains. Such a lack of incoming gene flow into cheese populations allowed trait differentiation in cheese strains as expected under domestication.

It came as a surprise that the two non-cheese populations split more recently from each other than from the cheese lineages. In particular, the non-Roquefort population diverged the earliest from the unidentified ancestral population, and this has likely occurred in another environment than cheese. Much more recently, selection in industrial times has likely only kept the most performant clonal lineage of this population for cheesemaking, losing most of the initial diversity as indicated by the very strong and recent bottleneck inferred in this lineage. Possible scenarios to explain the existence of two separated clusters thriving in food and silage differentiated from cheese strains include the very recent adaptive differentiation of a population from silage on human food or vice versa. The finding that silage strains are only found in one cluster (the orange one in Fig.1 to 5) suggests an adaptation to this ecological niche, although experiments will be required to test this hypothesis. Food spoiling strains are in contrast found in three clusters and may thus not constitute a specific population adapted to this environment and may instead represent migrants from several populations belonging to other ecological niches. Green and orange clusters may alternatively represent populations thriving in yet unidentified environments, dispersing to silage and food. Another hypothesis would be a single domestication event for cheesemaking before the divergence of the four lineages, followed by an escape and subsequent differentiation of the orange and green lineages in other human related habitats. This hypothesis however would not predict such high genetic diversity in the green and orange populations, and in particular the similar nucleotidic diversity levels in the two non-cheese populations as in the *P. carneum* and *P. paneum* outgroups. Given the

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

very low genetic diversity in the cheese populations, coalescence events occurred recently in the past, preventing tests of the occurrence of bottlenecks in the common ancestor of the four *P. roqueforti* populations.

The history of blue cheese production may provide circumstantial clues as to the origin of *P. roqueforti* cheese populations. Indeed, the first blue cheeses likely resulted from the sporadic accidental contamination of cheese with spores from the environment, such as moldy food. However, this would not be consistent with the demographic history inferred here for cheese and food-spoiling strains, as the cheese strains were not found to be nested within the food-spoiling strains, some of which originated from moldy bread. Furthermore, old French texts suggest that the blue mold colonized the cheese from within (Labbe and Serres, 2009, 2004; Vabre, 2015), which would indicate that the milk or curd was contaminated. French cheese producers began to inoculate cheeses with *P. roqueforti* spores from moldy rye bread at the end of the 19th century (Labbe and Serres, 2009, 2004; Vabre, 2015). Breads were specifically made with a 2:1 mixture of wheat and rye flour and were baked rapidly at high temperature (500°C), to yield a protective crust, around a moist, undercooked interior (Aussibal, 1983; Marre, 1906); the mold developed from the inside of the bread after one to five months in the Roquefort caves (Labbe and Serres, 2009, 2004; Vabre, 2015). Surveys of the microorganisms present in their caves (Chaptal, 1789; Marcorelle and Chaptal, 1833; Marre, 1906) and our unsuccessful attempts to obtain samples from a maturing cellar suggest that *P. roqueforti* spores did not originate from the caves, which were nevertheless crucial due to the ideal conditions provided for *P. roqueforti* development (Marre, 1906). Bread may have been colonized from the environment or from rye flour if the source *P. roqueforti* population was a rye endophyte or pathogen. This last hypothesis would be consistent with the lifestyle of many *Penicillium* species, which live in close association with plants, often acting as plant pathogens or necrotrophs (Ropars et al., 2016a), and with the occurrence of a *P. roqueforti* population in lumber and silage. Actually, a recent study reports the finding of *P. roqueforti* as an endophyte and could

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

be inoculated on wheat (Ikram et al., 2018), although species identification should be checked with more powerful markers. If this hypothesis is correct, then cheeses may historically have become contaminated with *P. roqueforti* from fodder during milking.

Comparison between non-cheese and cheese populations allowed us to identify specific traits and genes that have been under selection in cheese as opposed to other environments. Furthermore, the two independently domesticated *P. roqueforti* cheese populations, exhibiting different traits, represent a good model for studying the genomic processes involved in adaptation. We could not run analyses of selective sweep detection based on local decrease in genetic diversity in the genomes; indeed, because of the clonality of cheese populations, the whole genome will have hitchhiked with any selected locus. This effect has likely contributed to the strong bottlenecks. We were nevertheless able to identify candidate genes and evolutionary mechanisms potentially involved in adaptation to cheese in *P. roqueforti*. The horizontally-transferred *CheesyTer* genomic island probably contributes to the faster growth of the strains identified here as constituting the non-Roquefort population (Ropars et al., 2015). Indeed, *CheesyTer* includes genes with putative functions involved in carbohydrate utilization (e.g. β-galactosidase and lactose permease genes) that are specifically expressed at the beginning of cheese maturation, when lactose and galactose are available. This horizontal gene transfer may thus have been involved in adaptation to recently developed industrial cheese production processes in the non-Roquefort cheese population, conferring faster growth. We also identified additional genomic islands specific to the non-Roquefort cheese population, probably acquired recently and including genes putatively involved in fungal growth and spore production. In the genomic islands specific to the cheese populations, several genes appeared to be involved in lipolysis, carbohydrate or amino-acid catabolism and metabolite transport, all of which are important biochemical processes in the development of cheese flavor. In the Roquefort cheese population, a genomic region harboring genes with footprints for positive selection included several genes encoding proteins potentially involved in aromatic amino-

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

acid catabolism corresponding to precursors of volatile compounds. Further studies are required to determine the role of these genes in cheese flavor development.

In conclusion, we show that *P. roqueforti* cheese populations represent genuine domestication. Of course, the domestication process in cheese fungi has been more recent and different from the ones in emblematic crops or animals. Nevertheless, we did observe strong genetic differentiation from non-cheese populations, strong bottlenecks and trait differentiation with likely benefits for cheese production. This suggests genuine domestication, as has been reported previously in other fungi (Almeida et al., 2014; Baker et al., 2015; Gallone et al., 2016; Gibbons et al., 2012; Gonçalves et al., 2016; Libkind et al., 2011; Sicard and Legras, 2011), and defined as "the genetic modification of a species by breeding it in isolation from its ancestral population in an effort to enhance its utility to humans" (Gibbons and Rinker, 2015). Furthermore, a previous study has shown that the non-Roquefort cheese strains have acquired genes conferring better growth in cheese (Ropars et al., 2015). Our study revealed genetic divergence of cheese population from non-cheese populations, as well as the evolution of specific traits, with beneficial characteristics for cheese production. These findings therefore indicate the occurrence of domestication, a special case of adaptive divergence. We found that gene flow was prevented by clonality of cheese lineages and lack of migration between cheese and non-cheese populations, and that adaptation occurred on several traits beneficial for cheese production (lipolysis, proteolysis, spore production, volatile compound production, growth in salted cheese, cheese cavity colonization ability). Genomic footprints of adaptation were found in terms of rapid amino-acid changes and horizontal gene transfers. The two independent domestication events identified here interestingly represent adaptations to different production modes. Our findings concerning the history of *P. roqueforti* domestication thus shed light on the processes of adaptation to rapid environmental change, but they also have industrial implications and raise questions about the conservation of genetic resources in the agri-food context.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Methods

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

In order to investigate whether a wild *P. roqueforti* population occurred in ripening cellars or dairy environments that could be at the origin of the observed cheese populations, we sampled spores from the air in an artisanal cheese dairy company (GAEC Le Lévejac, Saint Georges de Lévejac, France, ca 60 km from Roquefort-sur-Soulzon, producing no blue cheese to avoid feral strains, i.e. dispersal from inoculated cheeses), sampling was performed in the sheepfold, milking parlour, cheese dairy and ripening cellar. We also sampled spores from the air in an abandoned ripening cellar in the town of Meyrueis (ca 70 km from Roquefort-sous-Soulzon) where Roquefort cheeses used to be produced and stored in the early 19th century. In total, 55 Petri dishes containing malt (2% cristomalt, Difal) and 3% ampicillin were left open for six days as traps for airborne spores (35 Petri dishes in the abandoned ripening cellar and 20 Petri dishes in the artisanal cheese dairy company). Numerous fungal colonies were obtained on the Petri dishes. One monospore was isolated from each of the 22 *Penicillium*-like colonies. DNA was extracted using the Nucleospin Soil Kit (Macherey-Nagel, Düren, Germany) and a fragment of the β-tubulin gene was amplified using the primer set Bt2a/Bt2b (Glass and Donaldson, 1995), and then sequenced. Sequences were blasted against the NCBI database to assign monospores to species. Based on β-tubulin sequences, ten strains were assigned to *P. solitum*, six to *P. brevicompactum*, two to *P. bialowienzense*, one to *P. echinulatum* and two to the *Cladosporium* genus. No *P. roqueforti* strain could thus be isolated from this sampling procedure.

Genome sequencing and analysis

The genomic DNAs of cheesemaking strains obtained from public collections belonging to *P. roqueforti*, seven strains of *P. paneum,* one strain of *P. carneum* and one strain of *P. psychrosexualis* (Supplementary Table 1) were extracted from fresh haploid mycelium after monospore isolation and growth for five days on malt agar using the Nucleospin Soil Kit (Macherey-Nagel, Düren, Germany). Sequencing was performed using the Illumina HiSeq 2500 paired-end technology (Illumina Inc.) with an average insert size of 400 bp at the GenoToul INRA platform and resulted in a 50x-100x coverage. In addition, the genomes of four strains (LCP05885, LCP06096, LCP06097

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

535 and LCP06098) were used that had previously been sequenced using the ABI SOLID technology
536 (Cheeseman et al., 2014). GenBank accession numbers are HG792015-HG792062.

537 Identification of presence/absence polymorphism of blocks larger than 10 kbp in genomes was
538 performed based on coverage using mapping against the FM164 *P. roqueforti* reference genome. In
539 order to identify genomic regions that would be lacking in the FM164 genome but present in other
540 strains, we used a second assembled genome, that of the UASWS *P. roqueforti* strain collected from
541 bread, sequenced using Illumina HiSeq shotgun and displaying 428 contigs (Genbank accession
542 numbers: JNNS01000420-JNNS01000428). Blocks larger than 10 kbp present in the UASWS
543 genome and absent in the FM164 genome were identified using the *nucmer* program v3.1 (Kurtz et
544 al., 2004). Gene models for the UASWS genome were predicted with EuGene following the same
545 pipeline as for the FM164 genome (Cheeseman et al., 2014; Foissac et al., 2008). The
546 presence/absence of these regions in the *P. roqueforti* genomes was then determined using the
547 coverage obtained by mapping reads against the UASWS genome with the start/end positions
548 identified by *nucmer*. The absence of regions was inferred when less than five reads were mapped.
549 In order to determine their presence/absence in other *Penicillium* species, the sequences of these
550 regions were blasted against nine *Penicillium* reference genomes (Supplementary Table 1). PCR
551 primer pairs were designed using Primer3Plus (http://www.bioinformatics.nl/cgi-
552 bin/primer3plus/primer3plus.cgi/) in the flanking sequences of these genomic regions in order to
553 check their presence/absence in a broader collection of *P. roqueforti* strains based on PCR tests
554 (Supplementary Table 2). For each genomic island, two primer pairs were designed when possible
555 (i.e. when sufficiently far from the ends of the scaffolds and not in repeated regions): one yielding
556 a PCR product when the region was present and another one giving a band when the region was
557 absent, in order to avoid relying only on lack of amplification for inferring the absence of a genomic
558 region. PCRs were performed in a volume of 25 µL, containing 12,5 µL template DNA (ten folds
559 diluted), 0.625 U Taq DNA Polymerase (MP Biomedicals), 2.5 µL 10x PCR buffer, 1 µL of 2.5
560 mM dNTPs, 1 µL of each of 10 µM primer. Amplification was performed using the following
561 program: 5 min at 94°C and 30 cycles of 30 s at 94°C, 30 s at 60°C and 1 min at 72°C, followed by
562 a final extension of 5 min at 72°C. PCR products were visualized using stained agarose gel
563 electrophoresis. Data were deposited at the European Nucleotide Archive
564 (http://www.ebi.ac.uk/ena/) under the accession number: PRJEB20132 for whole genome
565 sequencing and PRJEB20413 for Sanger sequencing.

566 For each strain, reads were mapped using stampy v1.0.21 (Lunter and Goodson, 2011) against the

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

high-quality reference genome of the FM164 *P. roqueforti* strain (Cheeseman et al., 2014). In order to minimize the number of mismatches, reads were locally realigned using the genome analysis toolkit (GATK) IndelRealigner v3.2-2 (McKenna et al., 2010). SNP detection was performed using the GATK Unified Genotyper (McKenna et al., 2010) , based on the reference genome in which repeated sequences were detected using RepeatMasker (Smit et al., 2013) and masked, so that SNPs were not called in these regions. In total 483,831 bp were masked, corresponding to 1.67% of the FM164 genome sequence. The 1% and 99% quantiles of the distribution of coverage depth were assessed across each sequenced genome and SNPs called at positions where depth values fell in these extreme quantiles were removed from the dataset. Only SNPs with less than 10% of missing data were kept. After filtering, a total of 115,544 SNPs were kept.

Population structure was assessed using a discriminant analysis of principal components (DAPC) with the Adegenet R package (Jombart, 2008). The genetic structure was also inferred along the genome by clustering the strains according to similarities of their genotypes, in windows of 50 SNPs, using the Mclust function of the mclust R package (Fraley et al., 2012; Fraley and Raftery, 2002) with Gower's distance and a Gaussian mixture clustering with K=7 (as the above analyses indicated the existence of four *P. roqueforti* populations and there were three outgroup species).

We performed a neighbor-net analysis using the network approach to visualize possible recombination events within and between populations with the phangorn R package (Schliep, 2010). The substitution model used for building the distance matrix was JC69 (Jukes and Cantor, 1969)

The genetic diversity were estimated using the $\theta\pi$ and, θw with the compute programs associated to libsequence v1.8.9 (Thornton, 2003) on 1145 sliding windows of 50 kb with 25 kb of overlap distributed along the longest eleven scaffolds of the FM164 assembly (> 200 kb). Linkage disequilibrium per genetic cluster (i.e. non-Roquefort, Roquefort, Lumber/food spoiler and silage/food spoiler) was estimated using the r2 statistics, with VCFtools v 0.1.15 (Danecek et al., 2011) and the following parameters: --geno-r2 --ld-window-bp 15000. Plots were generated using R.

To identify genes evolving under positive selection in *P. roqueforti* genomes, first, we used the method implemented in SnIPRE (Eilertson et al., 2012), a Bayesian generalization of the log-linear model underlying the McDonald-Kreitman test. This method detects genes in which amino-acid changes are more frequent than expected under neutrality, by contrasting synonymous and non-synonymous SNPs, polymorphic or fixed in two groups, to account for gene-specific mutation rates.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

598 Secondly, we performed a scan of the divergence statistics dxy between the two cheese populations,
599 calculated using a custom R script in 50kbp windows overlapping over 25 kbp along the genome. .
600 We considered genes belonging to the 1% most divergent regions and the 5% least genetically
601 diverse (π values) as under positive selection in one of the populations. We did not consider the
602 other pairwise comparisons, i.e. using orange and green populations (Figs. 1 to 5), because most
603 SNPs in those populations were shared by several strains, as shown by high diversity, positive Dt
604 and low FST values (Table 1). Consequently, islands of high divergence and low diversity were
605 restricted to cheese populations that were already found using pairwise comparison between cheese
606 populations. We performed GO annotation enrichment tests using separate Fisher's exact tests on
607 the three ontologies (BP: biological process; CC: cellular component; MF: metabolic function).

608

609

610 ***Strain genotyping***

611 We identified two genomic regions with multiple diagnostic SNPs allowing discriminating the two
612 cheese clusters. Two PCR primer pairs were designed (Supplementary Table 2) to sequence these
613 regions in order to assign the 65 strains (Supplementary Table 1) that can be purchased at the
614 Laboratoire Interprofessionnel de Production d'Aurillac (LIP) (the main French supplier of *P.*
615 *roqueforti* spores for artisanal and industrial cheese-makers; *https://www.lip-sas.fr/*) to the identified
616 clusters. PCR products were then purified and sequenced at Eurofins (France). Because one of the
617 cheese clusters included strains carrying the *Wallaby* and *CheesyTer* genomic islands while the
618 second cluster strains lacked these genomic regions (Ropars et al., 2015), we used previously
619 developed primer pairs to check for the presence/absence of *CheesyTer* and *Wallaby* (Ropars et al.,
620 2015).

621 Sequences were first aligned together with those extracted from sequenced genomes, allowing
622 assignation of LIP strains to one of the two cheese populations using MAFFT software (Katoh and
623 Standley, 2013) and then the alignments were visually checked. Then a tree reconstruction was
624 made using RAxML following GTRCAT substitution model, using 2 partitions corresponding to
625 the two fragments and 1000 bootstraps tree were generated (Stamatakis, 2006).

626

627 ***Strain phenotyping***

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

For all experiments, strains were picked up at random in each group. Experimental cheeses were produced in an artisanal dairy company (GAEC Le Lévejac, Saint Georges de Lévejac, France). The same ewe curd was used for all produced cheeses. Seven *P. roqueforti* strains were used for inoculation (two from each of the Roquefort, non-Roquefort and silage/food spoiler clusters, and one from the lumber/food spoiler cluster; their identity is given in Supplementary Table 1) using 17.8 mg of lyophilized spores. Three cheeses were produced for each strain in cheese strainers (in oval pots with opposite diameters of 8 and 9 cm, respectively), as well as a control cheese without inoculation. After 48 h of draining, cheeses were salted (by surface scrubbing with coarse salt), pierced and placed in a maturing cellar for four weeks at 11°C. Cheeses were then sliced into six equal pieces and a picture of each slice was taken using a Nikon D7000 (zoom lens: Nikon 18-105mm f:3.5-5.6G). Pictures were analyzed using the geospatial image processing software ENVI (Harris Geospatial Solution) (Fig. 6). This software enables pixel classification according to their level of blue, red, green, and grey into two to four classes depending on the analyzed image. This classification allowed assigning pixels to two classes corresponding to the inner white part and the cavities of the cheese, respectively (Fig. 6). For each picture, the percentage of pixels corresponding to the cavities was then quantified. Because the software could not reliably assign pixels to the presence versus absence of the fungus in cavities, we visually determined the cavity areas that were colonized by *P. roqueforti* using images. This allowed calculating a cheese cavity colonization rate. Because *Penicillium* spores have a high dispersal ability which could cause contaminations, we confirmed strain identity present in cheeses by performing Sanger sequencing of four diagnostic markers designed based on SNPs and specific to each strain (Supplementary Table 2). For each cheese, three random monospore isolates were genotyped, and no contamination was detected (i.e. all the sequences obtained corresponded to the inoculated strains).

To compare the growth rates of the different *P. roqueforti* clusters on bread (i.e. the traditional multiplication medium), 24 strains were used (eight from each of the Roquefort and non-Roquefort cheese clusters, five from the silage/food spoiler cluster, and three from the lumber/food spoiler cluster; the identities of the strains are shown in Supplementary Table 1). Each strain was inoculated in a central point in three Petri dishes by depositing 10 µL of a standardized spore suspension ($0.7x10^9$ spores/mL). Petri dishes contained agar (2%) and crushed organic cereal bread including rye (200 g/L). After three days at 25°C in the dark, two perpendicular diameters were measured for each colony to assess colony size.

The lipolytic and proteolytic activities of *P. roqueforti* strains were measured as follows:

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

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

A total of 47 strains were used to compare spore production between the four *P. roqueforti* clusters (Supplementary Table 1), 15 belonging to the non-Roquefort cluster, 15 to the Roquefort cluster, 10 to the silage/food spoiler cluster and seven to the lumber/food spoiler cluster. After seven days of growth on malt agar in Petri dishes of 60 mm diameter at room temperature, we scraped all the fungal material by adding 5 mL of tween water 0.005%. We counted the number of spores per mL in the solution with a Malassez hemocytometer (mean of four squares per strain) for calibrating spore solution. We spread 50 µL of the calibrated spore solution (i.e. 7.10^6 spores.mL^{-1}) for each strain on Petri dishes of 60 mm diameter containing three different media, malt, cheese and bread agar (organic “La Vie Claire” bread mixed with agar), in duplicates (two plates per medium and per strain). After eight days of growth at room temperature, we took off a circular plug of medium with spores and mycelium at the top, using Falcon 15 mL canonical centrifuge tubes (diameter of 15 mm). We inserted the plugs into 5 mL Eppendorf tubes containing 2 mL of tween water 0.005% and vortexed for 15 seconds to detach spores from the medium. Using a plate spectrophotometer, we measured the optical density (OD) at 600 nm for each culture in the supernatant after a four-fold dilution (Supplementary Table 3).

To compare salt tolerance between *P. roqueforti* clusters, 26 strains were used (eight from the Roquefort cluster, ten from the non-Roquefort cluster, three from the silage/food spoiler cluster, and five from the lumber/food spoiler cluster; strain identities are shown in Supplementary Table 1). For each strain and each medium, three Petri dishes were inoculated by depositing 10 µL of standardized spore suspension ($0.7x10^9$ spores/mL) on Petri dishes containing either only malt (20

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

g/L), malt and salt (NaCl 8%, which corresponds to the salt concentration used before fridge use to avoid contaminants in blue cheeses), only goat cheese, or goat cheese and salt (NaCl 8%). The goat cheese medium was prepared as described in a previous study (Ropars et al., 2015). Strains were grown at 25°C and colony size measured daily for 24 days.

Volatile production assays were performed on 16 Roquefort strains and 19 non-Roquefort cheese strains grown on model cheeses as previously described (Gillot et al., 2017). Briefly, model cheeses were prepared in Petri dishes and incubated for 14 days at 25 °C before removing three 10 mm-diameter plugs (equivalent to approximately 1 g). The plugs were then placed into 22 mL Perkin Elmer vials that were tightly closed with polytetrafluorethylene (PTFE)/silicone septa and stored at -80°C prior to analyses (Gillot et al., 2017). Analyses and data processing were carried out by headspace trap-gas chromatography-mass spectrometry (HS-trap-GC-MS) using a Perkin Elmer turbomatrix HS-40 trap sampler, a Clarus 680 gas chromatograph coupled to a Clarus 600T quadrupole MS (Perkin Elmer, Courtaboeuf, France), and the open source XCMS package of the R software (http://www.r-project.org/), respectively, as previously described (Pogačić et al., 2015).

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

Differences in volatile profiles among the two *P. roqueforti* cheese populations were analyzed using a supervised multivariate analysis method, orthogonal partial least squares discriminant analysis (OPLS-DA). OPLS is an extension of principal components analysis (PCA), that is more powerful when the number of explained variables (Y) is much higher than the number of explanatory variables (X). PCA is an unsupervised method maximizing the variance explained in Y, while partial least squares (PLS) maximizes the covariance between X and Y(s). OPLS is a supervised method that aims at discriminating samples. It is a variant of PLS which uses orthogonal (uncorrelated) signal correction to maximize the explained covariance between X and Y on the first latent variable, and components >1 capture variance in X which is orthogonal (uncorrelated) to Y. The optimal number of latent variables was evaluated by cross-validation (Pierre et al., 2011). Finally, to identify the volatile compounds that were produced in significantly different quantities between the two populations, a t-test was performed using the R software (http://www.r-project.org/).

Demographic modeling using approximate Bayesian computation (ABC)

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

The likelihoods of 11 demographic scenarios for the *P. roqueforti* populations were compared using approximate Bayesian computation (ABC) (Beaumont, 2010; Lopes and Beaumont, 2010). The scenarios differed in the order of demographic events, and included 21 parameters to be estimated (Supplementary Fig. 4). A total of 262 fragments, ranging from 5 kb to 15 kb, were generated from observed SNPs by compiling in a fragment all adjacent SNPs in complete linkage disequilibrium. The population mutation rate θ (the product of the mutation rate and the effective population size) used for coalescent simulations was obtained from data using θ_w, the Watterson's estimator. Simulated data were generated using the same fragment number and sizes as the SNP dataset generated from the genomes. Priors were sampled in a log-uniform distribution (Supplementary Fig. 4C). For each scenario, one million coalescent simulations were run and the following summary statistics were calculated on observed and simulated data using msABC (Pavlidis et al., 2010) : the number of segregating sites, the estimators π (Nei, 1987) and θ_w (Watterson, 1975) of nucleotide diversity, Tajima's D (Tajima, 1989), the intragenic linkage disequilibrium coefficient ZnS (Kelly, 1997), F_{ST} (Hudson et al., 1992), the percentage of shared polymorphisms between populations, the percentage of private SNPs for each population, the percentage of fixed SNPs in each population, Fay and Wu's H (Fay and Wu, 2000), the number of haplotypes (Depaulis and Veuille, 1998) and the haplotype diversity (Depaulis and Veuille, 1998). For each summary statistic, both average and variance values across simulated fragments were calculated. The choice of summary statistics to estimate posterior parameters is a crucial step in ABC (Csilléry et al., 2010). Summary statistics were selected using the AS.select() function with the neuralnet method in the "abctools" R package (Nunes and Prangle, 2015). In total, 101 summary statistics were kept for subsequent analyses. Cross validation was run with the neuralnet method using 100 samples and a tolerance of 0.01 (Supplementary Fig. 4D). Model selection was performed using four tolerance rates ranging from 0.005 to 0.1 and rejection, logistic regression and neural network methods. Because there was still an uncertainty on the choice between scenarios 4 and 5 after model selection (i.e. whether it was the non-Roquefort or Roquefort population that diverged first from the ancestral population) (Supplementary Fig. 4 F), an extra one million simulations were run for each of those two scenarios and model selection was performed again. All tolerance rates and methods favored scenario 4 over scenario 5 with absolute confidence of 1.000.

The posterior probability distributions of the parameters, the goodness of fit for each model and model selection (Supplementary Fig. 4E) were calculated using a rejection-regression procedure (Beaumont, 2010). Acceptance values of 0.005 were used for all analyses. Regression analyses was performed using the "abc" R package (Csilléry et al., 2012)

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(http://cran.rproject.org/web/packages/abc/index.html).

Estimate of time since domestication

The multiple sequentially Markovian coalescent (MSMC) software was used to estimate the domestication times of cheese populations (Schiffels and Durbin, 2014). The estimate of the last time gene flow occurred within each cheese population was taken as a proxy of time since domestication as it also corresponds in such methods to bottleneck date estimates and is more precisely estimated. Recombination rate was set at zero because sexual reproduction has likely not occurred since domestication in cheese populations (see results). Segments were set to 21*1+1*2+1*3 for the Roquefort population which contains three haplotypes (Fig. 2) and to 10*1+15*2 for the non-Roquefort population, which contains two closely related haplotypes (Fig. 2). In both cases, MSMC was run for 15 iterations and otherwise default parameters. The mutation rate was set to 10^{-8}.

References

Albert FW, Somel M, Carneiro M, Aximu-Petri A, Halbwax M, Thalmann O, Blanco-Aguiar JA, Plyusnina IZ, Trut L, Villafuerte R. 2012. A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet* **8**:e1002962.

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

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

Alonso L, Juarez M, Ramose M, Martin-Alvarez PJ. 1987. Overall composition, nitrogen fractions and fat characteristics of Cabrales cheese during ripening. *Z Für Lebensm-Unters Forsch* **185**:481–486. doi:10.1007/BF01042813

Ardö Y. 2006. Flavour formation by amino acid catabolism. *Biotechnol Adv* **24**:238–242.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

786 Aussibal R. 1983. Les caves de roquefort. France: Cahors.

787 Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM,
788 Hedhammar \AAke, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals
789 adaptation to a starch-rich diet. *Nature* **495**:360–364.

790 Baker E, Wang B, Bellora N, Peris D, Hulfachor AB, Koshalek JA, Adams M, Libkind D, Hittinger CT.
791 2015. The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing
792 yeasts. *Mol Biol Evol* **32**:2818–2831.

793 Beaumont MA. 2010. Approximate bayesian computation in evolution and ecology. *Annu Rev Ecol Evol Syst*
794 null. doi:doi: 10.1146/annurev-ecolsys-102209-144621

795 Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics.
796 *Genetics* **162**:2025.

797 Besana C, D'Errico R, Ghezzi R. 2017. Cheese Manufacturing in the Twentieth Century. The Italian
798 Experience in an International Context.

799 Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ. 2011.
800 Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial
801 strains of *Saccharomyces cerevisiae*. *PLoS Genet* **7**:e1001287.

802 Chaptal J-A. 1789. Observations sur les caves et le fromage de Roquefort. *Ann Chime* **4**:31–61.

803 Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M, Conseiller E,
804 Debuchy R. 2014. Multiple recent horizontal transfers of a large genomic region in cheese making
805 fungi. *Nat Commun* **5**:1–9.

806 Csilléry K, Blum MG, Gaggiotti OE, François O. 2010. Approximate Bayesian computation (ABC) in
807 practice. *Trends Ecol Evol* **25**:410–418.

808 Csilléry K, François O, Blum MG. 2012. abc: an R package for approximate Bayesian computation (ABC).
809 *Methods Ecol Evol* **3**:475–479.

810 Curioni PMG, Bosset JO. 2002. Key odorants in various cheese types as determined by gas chromatography-
811 olfactometry. *Int Dairy J* **12**:959–984.

812 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT,

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Sherry ST. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.

De Llano DG, Ramos M, Polo C, Sanz J, Martinez-Castro I. 1990. Evolution of the volatile components of an artisanal blue cheese during ripening. *J Dairy Sci* **73**:1676–1683.

De Llano DG, Ramos M, Rodriguez A, Montilla A, Juárez M. 1992. Microbiological and physicochemical characteristics of Gamonedo blue cheese during ripening. *Int Dairy J* **2**:121–135.

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

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

Eilertson KE, Booth JG, Bustamante CD. 2012. SnIPRE: selection inference using a Poisson random effects model. *PLoS Comput Biol* **8**:e1002806.

Fay JC, Wu C-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**:1405–1413.

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

Fraley C, Raftery AE. 2002. Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* **97**:611–631.

Fraley C, Raftery AE, Scrucca L. 2012. Normal mixture modeling for model-based clustering, classification, and density estimation. *Dep Stat Univ Wash* **23**:2012.

Freedman AH, Lohmueller KE, Wayne RK. 2016. Evolutionary history, selective sweeps, and deleterious variation in the dog. *Annu Rev Ecol Evol Syst* **47**:73–96.

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

Gibbons JG, Rinker DC. 2015. The genomics of microbial domestication in the fermented food environment. *Curr Opin Genet Dev* **35**:1–8.

Gibbons JG, Salichos L, Slot JC, Rinker DC, McGary KL, King JG, Klich MA, Tabb DL, McDonald WH, Rokas A. 2012. The evolutionary imprint of domestication on genome variation and function of the

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

filamentous fungus *Aspergillus oryzae*. *Curr Biol* **22**:1403–1409.

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

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

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

Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, Vienne DM, Rodríguez de la Vega RC, Branco S, Giraud T. 2014. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Mol Ecol* **23**:753–773.

Glass NL, Donaldson GC. 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Appl Environ Microbiol* **61**:1323–1330.

Gonçalves M, Pontes A, Almeida P, Barbosa R, Serra M, Libkind D, Hutzler M, Gonçalves P, Sampaio JP. 2016. Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr Biol* **26**:2750–2761.

Gooday GW, Zhu W-Y, O'Donnell RW. 1992. What are the roles of chitinases in the growing fungus? *FEMS Microbiol Lett* **100**:387–391.

Gouyon P-H, Leriche H, Civard A, Reeves H, Hulot N. 2010. Aux origines de l'environnement. France: Paris: Fayard.

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

Harlan JR. 1992. Crops and man. USA: Madison,WI: American Society of Agronomy.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**:583–589.

Ikram M, Ali N, Jan G, Jan FG, Rahman IU, Iqbal A, Hamayun M. 2018. IAA producing fungal endophyte *Penicillium roqueforti* Thom., enhances stress tolerance and nutrients uptake in wheat plants grown

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

on heavy metal contaminated soils. *PloS One* **13**:e0208150.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403–1405.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. *Mamm Protein Metab* **3**:132.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.

Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics* **146**:1197–1206.

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

Labbe M, Serres JP. 2009. Chroniques du Roquefort-Des hommes, des entreprises, des marques, période moderne. *Graphi Impr Primaube Fr*.

Labbe M, Serres JP. 2004. Chroniques du Roquefort—De la préhistoire à l'aube industrielle. *Graphi Impr Primaube Fr*.

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

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

Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, Johnston M, Gonçalves P, Sampaio JP. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci* **108**:14539–14544. doi:10.1073/pnas.1105430108

Lopes JS, Beaumont MA. 2010. ABC: a useful Bayesian tool for the analysis of population data. *Infect Genet Evol* **10**:825–832.

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

Malekinejad H, Aghazadeh-Attari J, Rezabakhsh A, Sattari M, Ghasemsoltani-Momtaz B. 2015.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Neurotoxicity of mycotoxins produced *in vitro* by *Penicillium roqueforti* isolated from maize and grass silage. *Hum Exp Toxicol* **34**:997–1005.

Marcorelle J-F, Chaptal J-A. 1833. Observations sur les caves et le fromage de Roquefort. Art de faire le beurre et les meilleurs fromages. Chez Madame Huzard.

Marre E(. 1906. Le roquefort. Rodez: E. Carrère.

Marres P. 1935. Les Grands Causses (PhD Thesis). Arrault.

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

Martín JF, Coton M. 2016. Blue cheese: microbiota and fungal metabolites. Fermented Foods in Health and Disease Prevention. Elsevier. pp. 275–303.

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

McSweeney PL. 2004. Biochemistry of cheese ripening. *Int J Dairy Technol* **57**:127–144.

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

Nei M. 1987. Molecular evolutionary genetics. Columbia university press.

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

Nunes MA, Prangle D. 2015. abctools: an R package for tuning approximate Bayesian computation analyses. *R J* **7**:189–205.

Parker HG, Harris A, Dreger DL, Davis BW, Ostrander EA. 2017. The bald and the beautiful: hairlessness in domestic dog breeds. *Phil Trans R Soc B* **372**:20150488.

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

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

921 Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored
922 A. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**:339.
923 Pierre PS, Jansen JJ, Hordijk CA, Van Dam NM, Cortesero A-M, Dugravot S. 2011. Differences in volatile
924 profiles of turnip plants subjected to single and dual herbivory above-and belowground. *J Chem Ecol*
925 **37**:368.
926 Pogačić T, Maillard M-B, Leclerc A, Hervé C, Chuat V, Yee AL, Valence F, Thierry A. 2015. A
927 methodological approach to screen diverse cheese-related bacteria for their ability to produce aroma
928 compounds. *Food Microbiol* **46**:145–153.
929 Ropars J, de la Vega RCR, López-Villavicencio M, Gouzy J, Sallet E, Dumas É, Lacoste S, Debuchy R,
930 Dupont J, Branca A, others. 2015. Adaptive Horizontal Gene Transfers between Multiple Cheese-
931 Associated Fungi. *Curr Biol*.
932 Ropars J, de La Vega RR, López-Villavicencio M, Gouzy J, Dupont J, Swennen D, Dumas E, Giraud T,
933 Branca A. 2016a. Diversity and mechanisms of genomic adaptation in *Penicillium*. *Aspergillus*
934 *Penicillium Post-Genomic Era DeVries RG Andersen MR Eds* 27–42.
935 Ropars J, Lo Y-C, Dumas E, Snirc A, Begerow D, Rollnik T, Lacoste S, Dupont J, Giraud T, López-
936 Villavicencio M. 2016b. Fertility depression among cheese-making *Penicillium roqueforti* strains
937 suggests degeneration during domestication. *Evolution*.
938 Ropars J, López-Villavicencio M, Dupont J, Snirc A, Gillot G, Coton M, Jany J-L, Coton E, Giraud T. 2014.
939 Induction of sexual reproduction and genetic diversity in the cheese fungus *Penicillium roqueforti*.
940 *Evol Appl* **7**:433–441. doi:10.1111/eva.12140
941 Ropars J, López-Villavicencio M, Snirc A, Lacoste S, Giraud T. 2017. Blue cheese-making has shaped the
942 population genetic structure of the mould *Penicillium roqueforti*. *PloS One* **12**:e0171387.
943 Rundberget T, Skaar I, Fl\a aøyen A. 2004. The presence of *Penicillium* and *Penicillium* mycotoxins in food
944 wastes. *Int J Food Microbiol* **90**:181–188.
945 Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome
946 sequences. *Nat Genet* **46**:919.
947 Schliep KP. 2010. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**:592–593.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

948 Shapiro MD, Kronenberg Z, Li C, Domyan ET, Pan H, Campbell M, Tan H, Huff CD, Hu H, Vickrey AI.
949 2013. Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**:1063–1067.
950 Sicard D, Legras J-L. 2011. Bread, beer and wine: Yeast domestication in the< i> Saccharomyces sensu
951 stricto</i> complex. *C R Biol* **334**:229–236.
952 Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0.
953 Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. 2007. pcaMethods - a bioconductor package
954 providing PCA methods for incomplete data. *Bioinformatics* **23**:1164–1167.
955 Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of
956 taxa and mixed models. *Bioinformatics* **22**:2688–2690.
957 Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM.
958 2012. Reshaping of the maize transcriptome by domestication. *Proc Natl Acad Sci* **109**:11878–
959 11883.
960 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
961 *Genetics* **123**:585–595.
962 Thierry A, Collins YF, Mukdsi MA, McSweeney PL, Wilkinson MG, Spinnler HE. 2017. Lipolysis and
963 metabolism of fatty acids in cheese. Cheese (Fourth Edition). Elsevier. pp. 423–444.
964 Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics*
965 **19**:2325–2327.
966 Vabre S. 2015. Le sacre du Roquefort. *Press Univ Fr*.
967 Vabre S. 2010. Roquefort Société: une industrie agro-alimentaire en Aveyron (vers 1840-1914) (PhD
968 Thesis). Toulouse 2.
969 Vavilov N. 1992. Origin and geography of cultivated plants. Great Britain: Cambridge: Cambridge
970 University Press.
971 Wang H, Studer AJ, Zhao Q, Meeley R, Doebley JF. 2015. Evidence that the origin of naked kernels during
972 maize domestication was caused by a single amino acid substitution in tga1. *Genetics* **200**:965–974.
973 Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor*
974 *Popul Biol* **7**:256–276.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

975 Woo AH, Lindsay RC. 1984. Concentrations of major free fatty acids and flavor development in Italian
976 cheese varieties. *J Dairy Sci*.
977 Xue S, Bradbury PJ, Casstevens T, Holland JB. 2016. Genetic architecture of domestication-related traits in
978 maize. *Genetics* **204**:99–113.
979 Yvon M, Rijnen L. 2001. Cheese flavour formation by amino acid catabolism. *Int Dairy J* **11**:185–201.
980

981 **Acknowledgments**

982 This work was supported by the ERC starting grant GenomeFun 309403 awarded to TG, the ANR
983 FROMA-GEN grant (ANR-12-PDOC-0030) to AB, and an "*Attractivité*" grant from Paris-Sud
984 University to AB. We thank Kamel Soudani for help with image analysis and Aurélien Tellier for
985 advice concerning ABC analyses. We are grateful to Coralie Benel and Francis Roujon of GAEC
986 Le Lévejac for assistance with cheesemaking and Paul Villain for experimental help. Sequencing
987 was performed at GenoToul INRA platform. We thank INRA and MNHN for granting access to
988 four genomes sequenced with the help of Joëlle Dupont, Sandrine Lacoste, Yves Brygoo and Jeanne
989 Ropars in the framework of the ANR 'Food Microbiomes' project (ANR-08-ALIA-007-02)
990 coordinated by Pierre Renault.
991

992 **Author contributions**

993 TG and AB acquired the funding, designed and supervised the study. SL and AS produced the
994 genomes. ED, AB and RdlV analyzed the genomes. ED, SL, JR, AS, MC, AT, EC, MLP and DR
995 performed the experiments. ED, AB and TG analyzed the data from the experiments. ED, AB and
996 AF performed ABC analyses. ED and TG wrote the manuscript with contributions from the other
997 authors.

998

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

999 # Figure legends

000 **Figure 1: Diversity and population subdivision in *Penicillium roqueforti*.** Unrooted phylogenetic
001 network of P. roqueforti strains generated with SplitsTree4 from SNP variation. The scale bar
002 indicates the number of substitutions per site. The letters indicate the origin of the strains, C =
003 cheese, F = spoiled food, S = silage and L = lumber. The color indicates assignment to one of the
004 four *P. roqueforti* populations identified, as in the other figures. Blue, non-Roquefort; purple,
005 Roquefort; green, lumber/food spoilage, and; orange, silage/food spoilage.

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

016

017 **Figure 3**: **Genetic and phenotypic differentiation among *Penicillium roqueforti* populations.**
018 **Colors correspond to the genetic clusters as in other figures**. A: genetic differentiation assessed
019 by a discriminant analysis of principal components (DAPC) based on genome-wide single-
020 nucleotide polymorphisms (SNPs). The dots represent the strains and the colors the four populations
021 identified based on the genealogical tree in Fig. 1 as well as the similarity clustering in Fig. 2. The

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

insets show the distribution of eigenvalues for the principal component analysis (PCA) and for the discriminant analysis (DA). B: phenotypic differentiation among *P. roqueforti* genetic clusters illustrated by a PCA based on all tested phenotypes. Colors correspond to the genetic clusters as in other figures. Missing data correction has been done using Bayesian correction in the pcaMethods package (Stacklies et al., 2007).

Figure 4: Demographic history of *Penicillium roqueforti* populations. A. Demographic scenario (S4) with the highest posterior probability for the history of *Penicillium roqueforti* populations. Estimates of time since divergence are indicated in units of $2N_e$ generations (Supplementary Figure 4 E); effective population sizes and their variation (bottlenecks) are represented by the widths of the genealogy branches, with relative sizes being represented to scale. The color indicates assignment to the *P. roqueforti* populations as in the other figures. B. Estimated past migration rate (gene flow) within each of the two cheese populations backward in time (t=0 represents the present time). The dashed red lines represent the inferred times of domestication, estimated as the last time gene flow occurred within cheese populations. C. Estimated demographic history for the Roquefort population using the multiple sequentially Markovian coalescent (MSMC) method. The inferred population effective size is plotted along generations backward in time (t=0 represents the present time). The dashed red line represents the inferred domestication time, estimated as the last time gene flow occurred within the Roquefort population (Fig. 4B). The scheme above the figure represents a schematic view of the effective population size along generations, representing the two bottlenecks.

Figure 5: Differences in phenotype between *Penicillium roqueforti* populations for various traits relevant for cheesemaking. The color indicates assignment to the *P. roqueforti* populations identified, as in the other figures. Horizontal lines on the boxplots represent the upper quartile, the median and the lower quartile. Dots represent the outlier values. Different letters indicate significant

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

046 differences (Supplementary Table 4). **A**: Lipolytic activity measured at four different dates; **B**:
047 Proteolytic activity measured at four different dates; **C**: Spore production on bread medium
048 measured as optical density by spectrophotometer; **D**: Cheese cavity occupation (i.e., percentage of
049 total cheese cavity space colonized by the fungus, as measured on images) estimated in experimental
050 cheeses by image analysis. The two clusters of non-cheese strains were pooled, as there were too
051 few strains per cluster to test differences between the lumber/food spoiler and silage/food spoiler
052 clusters. (a) Picture of a cheese slice. (b) Corresponding image analysis using the geospatial image
053 processing software ENVI (Harris Geospatial Solution). Colors correspond to pixel classification
054 based on their color on the picture. In yellow and blue: the inner white part of the cheese; in green
055 and red: cavities.

056 **Figure 6: A: Differences in volatile compound profiles of the two *Penicillium roqueforti***
057 **cheese populations.** Orthogonal projection of the latent structure discriminant analysis (OPLS-
058 DA), with each dot representing the score of the averaged volatile profile of a strain from the non-
059 Roquefort cheese population (in red) or the Roquefort population (in blue) in the two principal
060 components. **B: Identified volatile compounds emitted by the non-Roquefort and the**
061 **Roquefort populations**, chemical class, quantification ion: mass (m) to charge (z) ratio, and results
062 of t-test statistical comparisons between the two populations: quantification estimate, standard error,
063 degrees of freedom (Df), t values and P values (Pr(>|t|)). In bold are the volatile compounds whose
064 quantity was found significantly different between the two populations.
065
066 **Figure 7: A: Presence/absence of the ten genomic islands identified in this study in the 35**
067 ***Penicillium roqueforti* and nine *Penicillium* outgroup species, in addition to the *CheesyTer* and**
068 ***Wallaby* horizontally-transferred regions identified in a previous study.** The ten genomic
069 islands were detected as absent from one of the two *P. roqueforti* genomes with high-quality

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

assemblies, while present in the second reference genome; the two reference *P. roqueforti* genomes are those of the FM164 strain (isolated from Gorgonzola cheese) and of the UASWS strain (isolated from bread Supplementary Table1 for information on outgroup reference genomes. For each genomic island, its name is indicated, together with its scaffold or contig and its start/end positions. Each strain is represented as a line, the presence of a genomic island is indicated by a colored box and its absence by a white box. The grey intensity indicates the percentage of sequence identity in these genomic islands, either within *P. roqueforti* or compared to outgroups. Strain assignment to the identified genetic clusters is indicated, with the same colors as in other figures. **B: Fisher exact test for function enrichment of the genes identified in the presence/absence regions** based on the InterPro annotation. For each annotation, the Table gives the InterPro number, the number of occurrences in the presence/absence regions and in the FM164 reference genome, the p-value before and after FDR correction and the functional annotation. Annotations are shown only for genes with significant enrichment before multiple testing correction. Annotations followed by a star refer to putative functions related to fungal growth and sporulation. Annotations followed by two stars refer to putative functions related lipolysis, carbohydrate or amino-acid catabolism and metabolite transporter.

Figure 8: A: Genes detected as evolving under positive selection using the SnIPRE software (i.e. genes with higher numbers of non-synonymous substitutions than expected under neutrality, controlling for gene-specific mutation rates). Values represent the estimates of the γ selection coefficient. In red, genes under positive selection ($\gamma > 0$), in blue genes under purifying selection ($\gamma < 0$), as detected based on analyses in the Roquefort cluster, the non-Roquefort cluster and in the pooled *Penicillium roqueforti* strains from the four clusters. The asterisks after the gene names highlight the eight genes clustered in the ProqFM164S01 scaffold in B. **B:** Selection effect (γ) estimated per gene along the ProqFM164S01 scaffold in the Roquefort population. The selection

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

coefficient γ was calculated with SnIPRE. The red dots correspond to genes evolving under positive selection (γ significantly greater than 0), the blue dots to genes evolving under purifying selection (γ significantly **lower** than 0), and the gray dots to genes evolving under neutrality (γ not significantly different from 0).

Figure 9: Scans of genetic differentiation (d_{xy}) between non-Roquefort and Roquefort *Penicillium roqueforti* populations, and of genetic diversities (π) within non-Roquefort and Roquefort populations. Values were calculated in 50 kb sliding windows, overlapping over 25 kb. Red dots correspond to windows located in the 1% highest d_{xy} (small dashed line) and 5 % lowest π values (long dashed line). Outliers detected in Snipre (Fig 8) are shown as green dots.

Table 1: **Population genetics statistics in the four *Penicillium roqueforti* populations.** A: Statistics calculated by averaging values on 1144 sliding windows of 50 kb with 25 kb overlap. B: F_{ST} values calculated on pairwise comparisons.

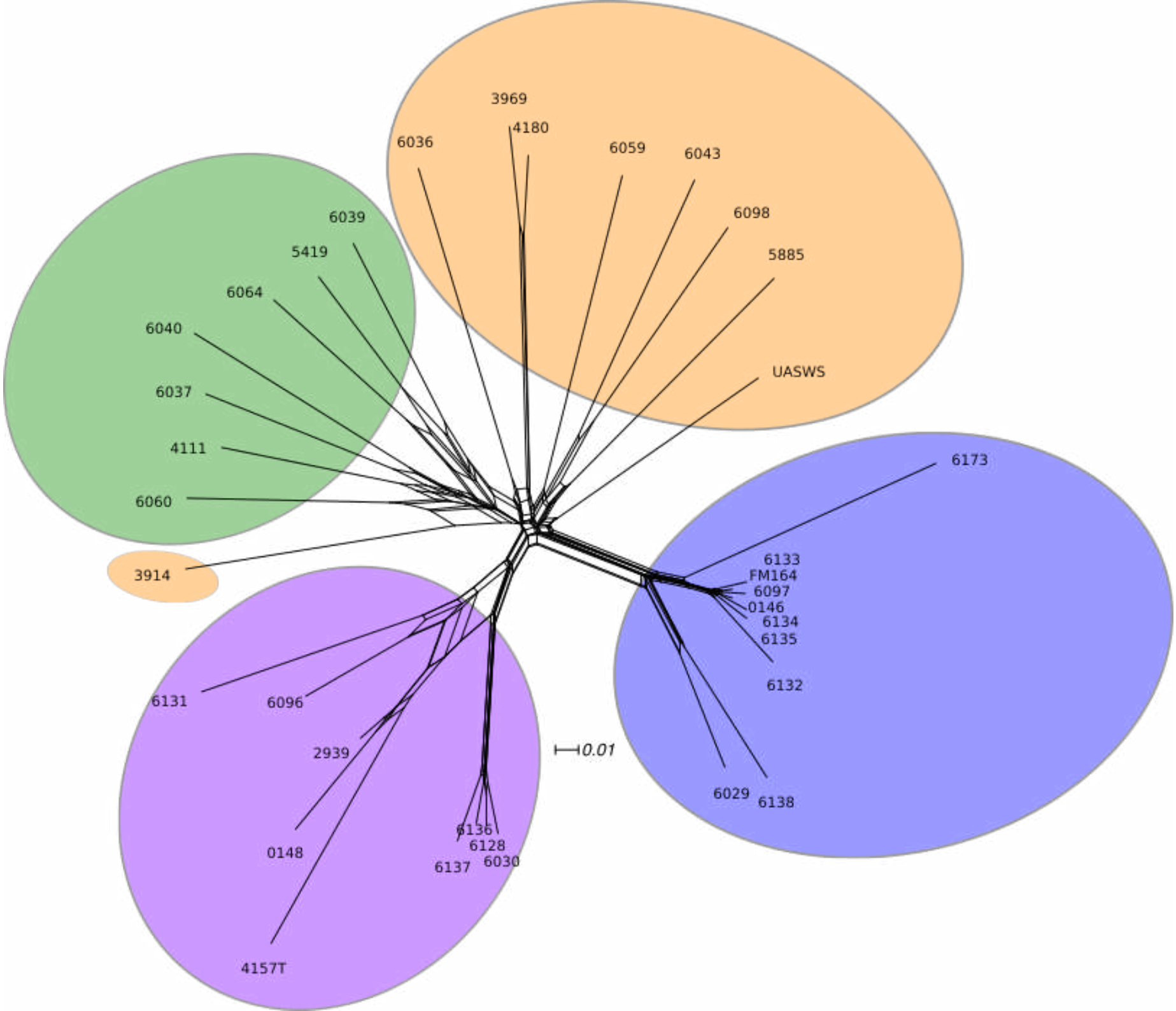
bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

A)

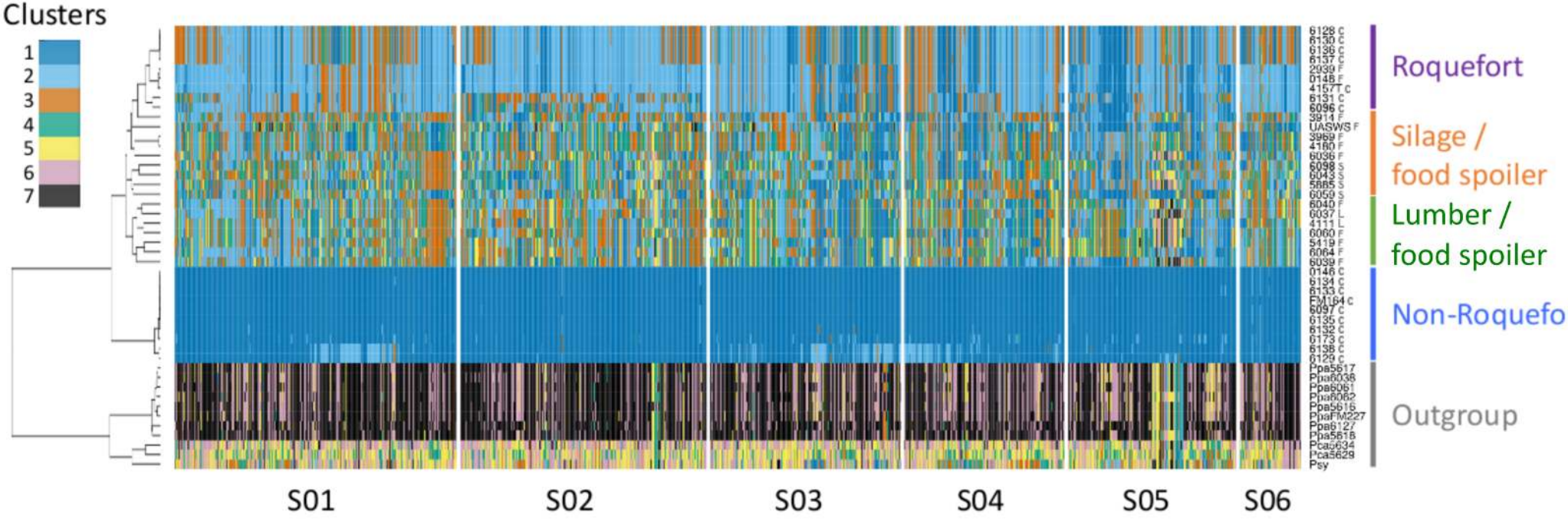
	Number of segregating sites per kilobase	π per site	Watterson's Θ per site	D_t	H_f
Silage/Food spoiler	2.28	0.00098	0.00084	0.75689	0.00001
Lumber/ Food spoiler	1.59	0.00078	0.00070	0.77300	-0.00004
Non-Roquefort	0.25	0.00008	0.00011	-1.27191	-0.00021
Roquefort	1.03	0.00043	0.00040	0.56090	-0.00007
Penicillium roqueforti	2.75	0.00107	0.00070	1.80833	0.48170

B)

	Silage/Food spoiler	Lumber/ Food spoiler	Non-Roquefort
Roquefort	0.21	0.27	0.62
Non-Roquefort	0.38	0.49	
Lumber/ Food spoiler	0.08		



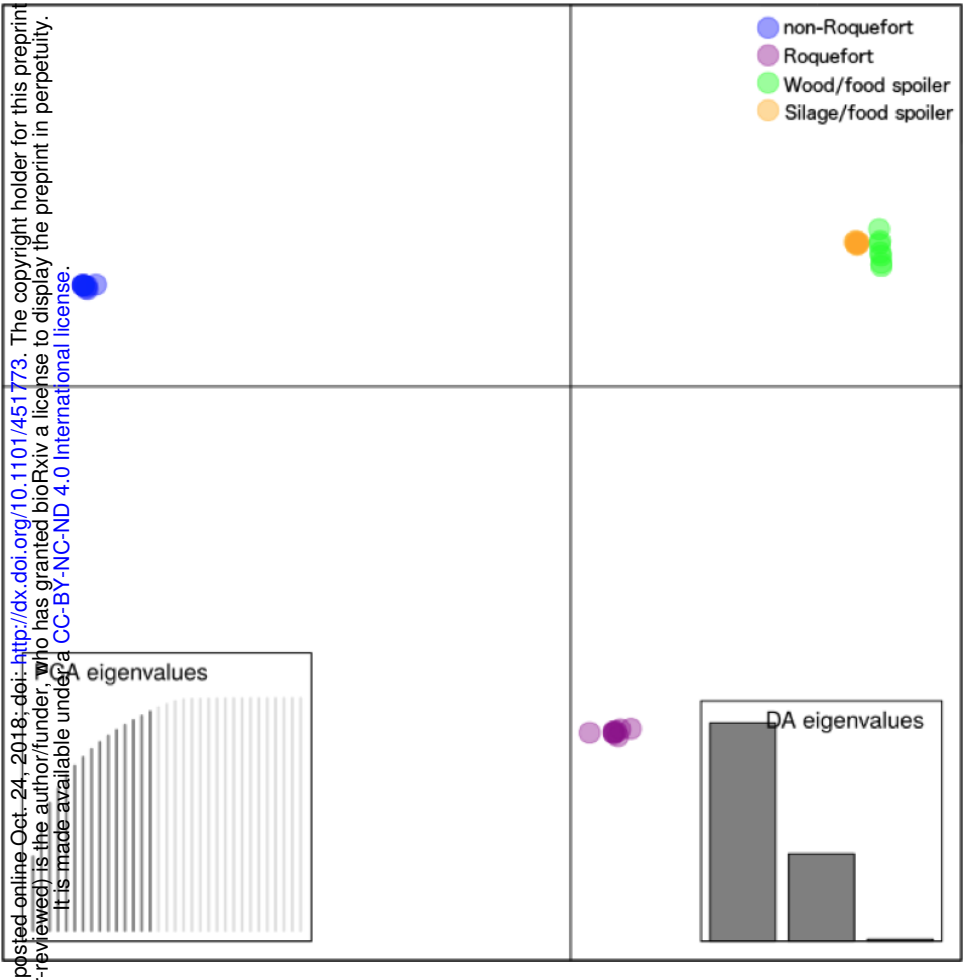
bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



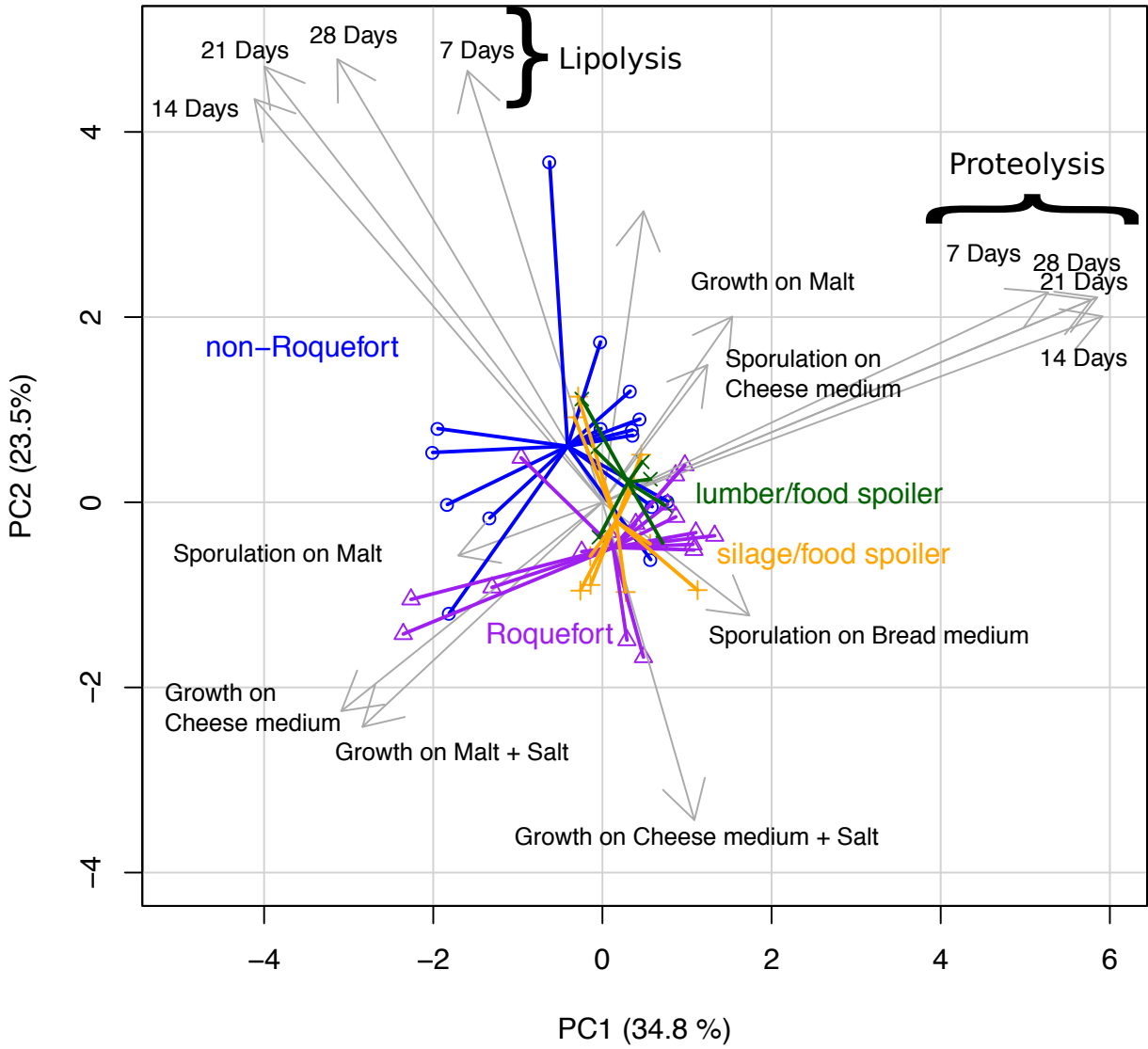
bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

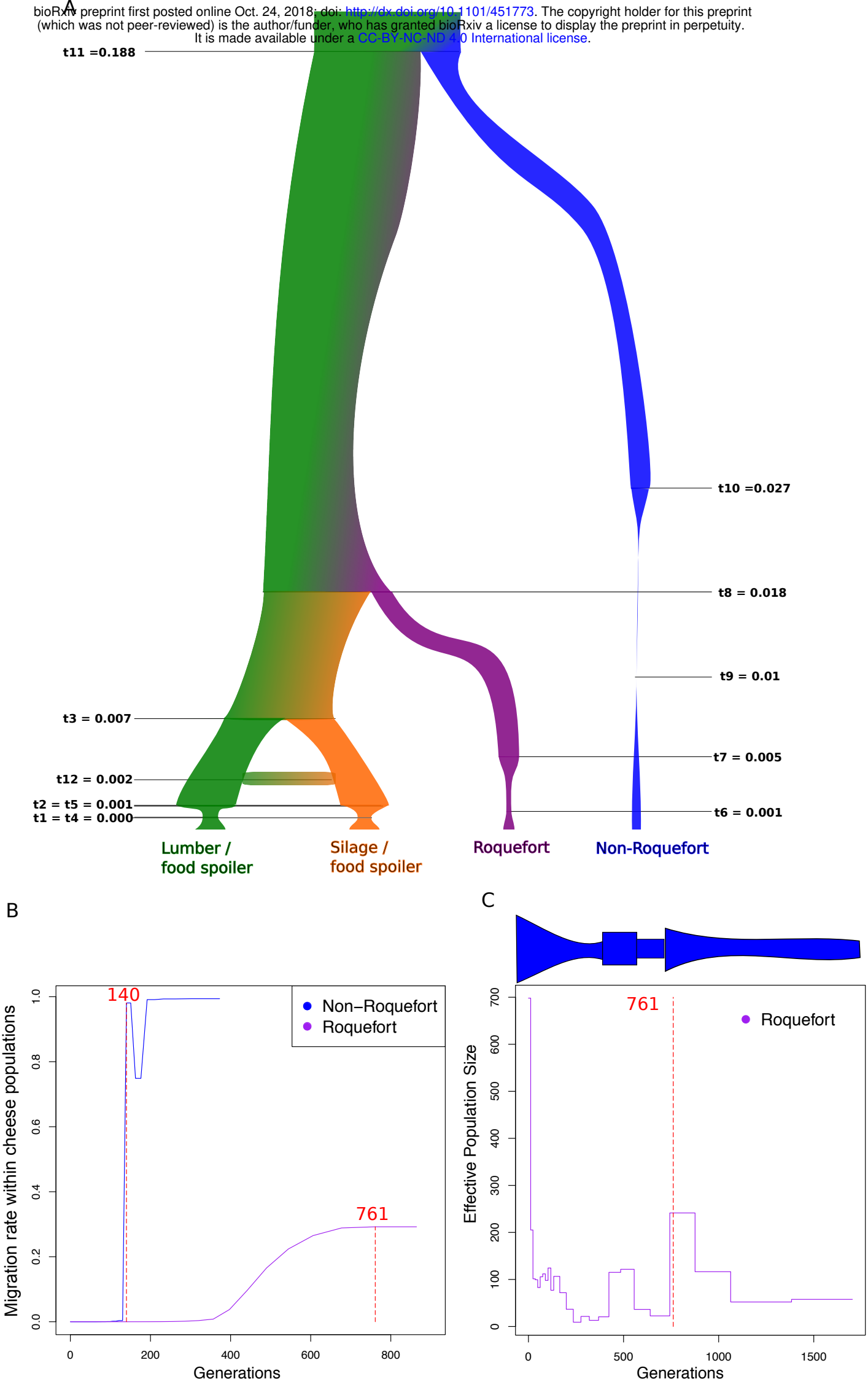
A



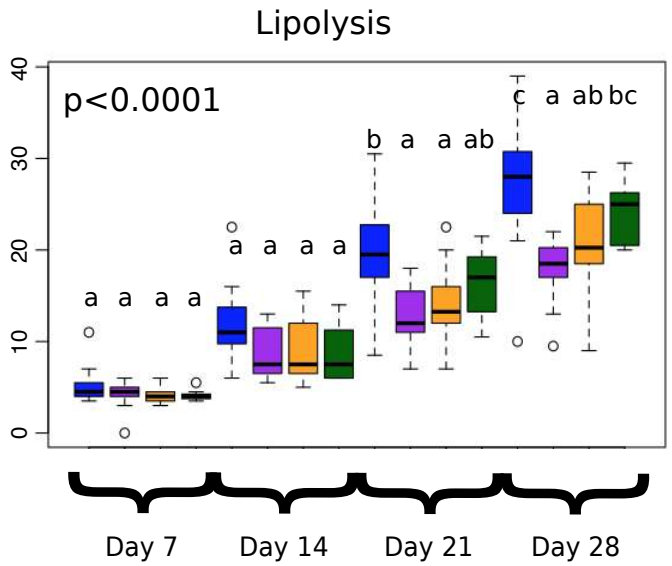
B

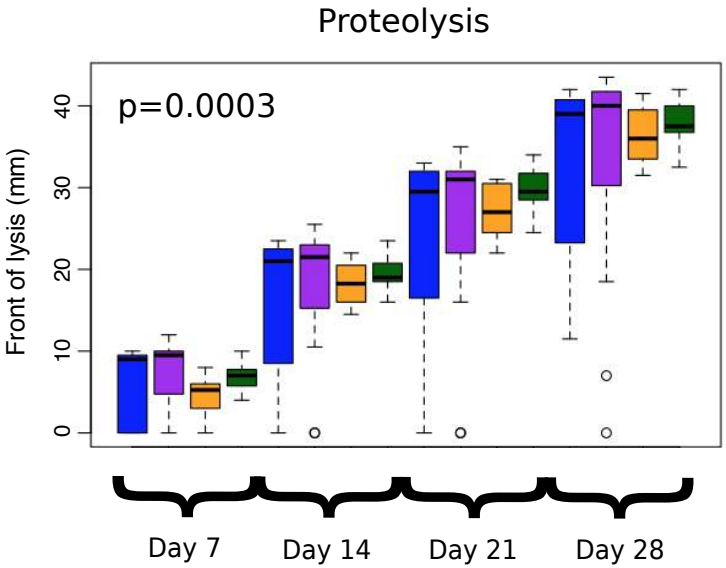


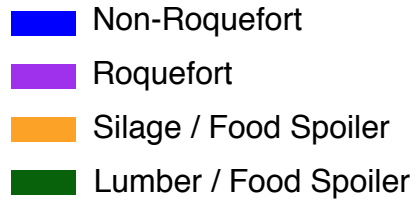
bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

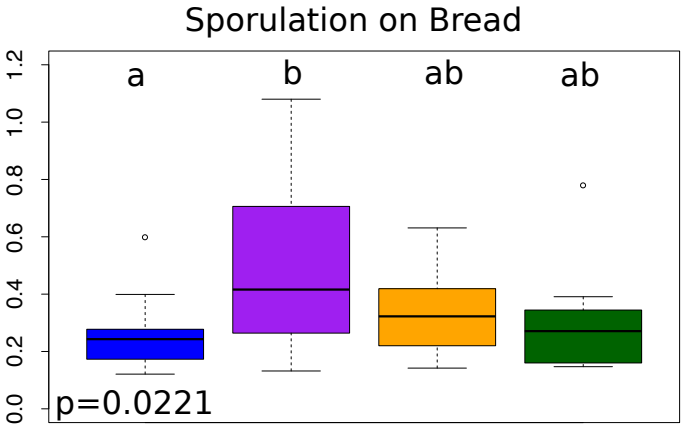


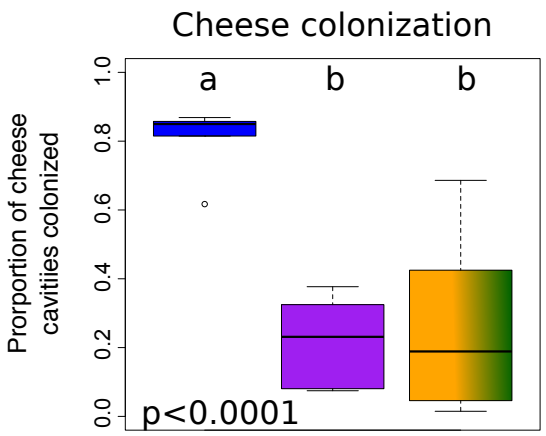
bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

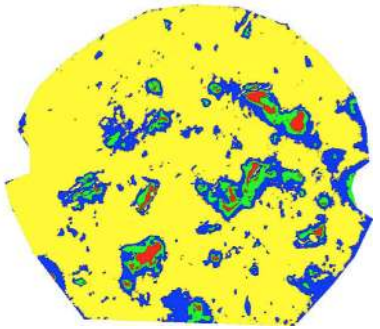




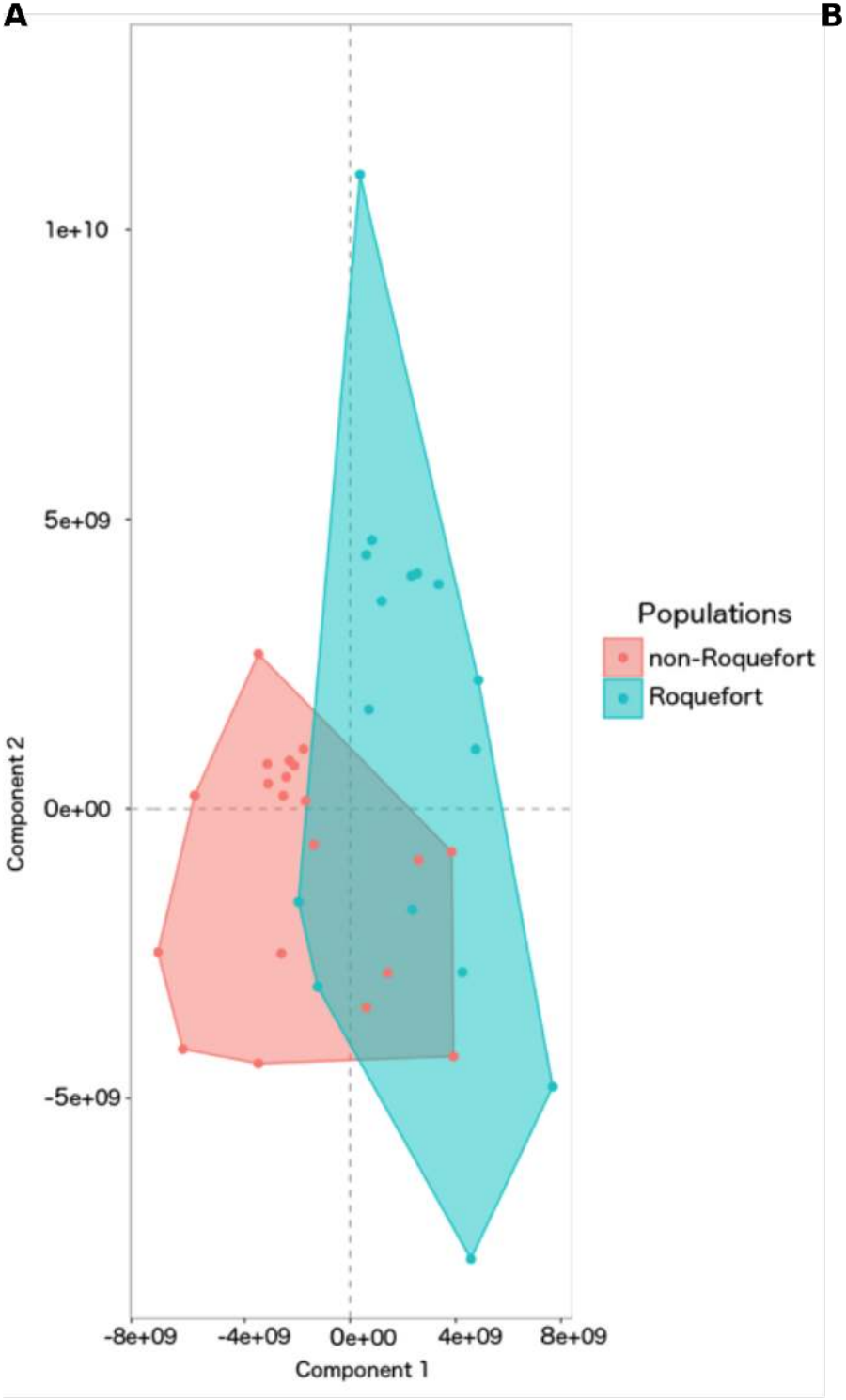








A



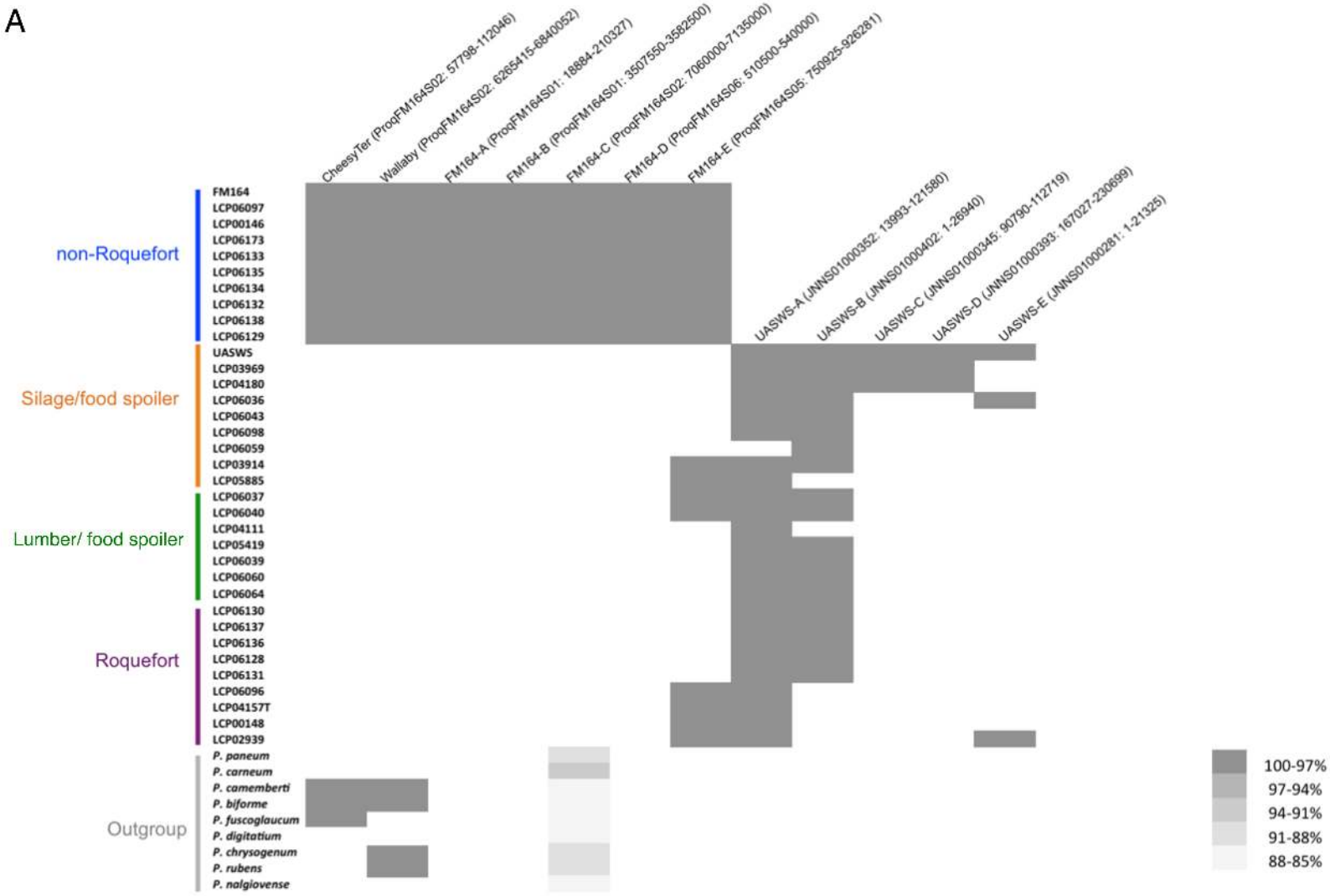
B

Chemical class	Volatile compounds (other name)	m/z	Estimate	Standard error	Df	t value	Pr(>|t|)
Ketones	propan-2-one (acetone)	58	5.1105e-12	8.7161e-12	35	0.5863	0.561413
	butan-2-one	43	9.6893e-13	2.1519e-11	35	0.0450	0.964341
	pentan-2-one	58	3.5722e-12	4.5743e-12	35	0.7809	0.440102
	butane-2,3-dione (diacetyl)	**86**	**1.7389e-11**	**8.2570e-12**	**35**	**2.1059**	**0.042452**
	2-methylpentan-3-one	100	-3.6707e-14	2.2278e-13	35	-0.1648	0.870075
	4-methylpentan-2-one	100	-2.8174e-13	1.4679e-13	35	-1.9194	0.063118
	heptan-2-one	58	2.1430e-12	2.8510e-11	35	0.0752	0.940512
	6-methylheptan-2-one	110	-4.2094e-15	6.7433e-15	35	-0.6242	0.536517
	octan-3-one	**99**	**3.3524e-12**	**1.3443e-12**	**35**	**2.4937**	**0.017515**
	3-hydroxybutan-2-one (acetoin)	**45**	**1.1266e-10**	**3.2646e-11**	**35**	**3.4511**	**0.001475**
	2-hydroxypentan-3-one	102	7.9416e-14	5.1472e-14	35	1.5429	0.131848
	nonan-2-one	57	1.6631e-11	1.7272e-11	35	0.9628	0.342228
	non-8-en-2-one	82	1.3217e-12	1.5882e-12	35	0.8322	0.410927
	decan-2-one	59	4.2918e-14	2.5445e-14	35	1.6867	0.100549
	undecan-2-one	58	1.3548e-11	1.1917e-11	35	1.1368	0.263334
	1-(2-aminophenyl)ethanone	120	-2.1713e-14	7.5074e-14	35	-0.2892	0.774122
Acids	acetic acid	60	8.1443e-12	1.5289e-11	35	0.5327	0.597618
	propanoic acid	74	-1.3298e-12	9.8625e-12	35	-0.1348	0.893512
	2-methylpropanoic acid	73	3.0534e-12	1.6745e-11	35	0.1823	0.856368
	butanoic acid	60	-6.0001e-12	2.1437e-11	35	-0.2799	0.781209
	3-methylbutanoic acid	43	-5.1382e-12	1.8056e-11	35	-0.2846	0.777645
	4-methylpentanoic acid	74	-7.1382e-12	2.1132e-11	35	-0.3378	0.737534
	hexanoic acid	60	-3.8153e-12	9.5320e-12	35	-0.4003	0.691398
	octanoic acid	84	-3.6948e-14	8.4859e-14	35	-0.4354	0.665945
Sulfur compounds	methanethiol	48	-1.7684e-13	4.0058e-13	35	-0.4415	0.661586
	methylsulfanylmethane	35	2.4421e-15	1.7160e-14	35	0.1423	0.887649
	(methyldisulfanyl)methane	94	1.0665e-11	1.8999e-11	35	0.5613	0.578155
	S-methyl butanethioate	118	-1.2664e-13	1.4019e-13	35	-0.9033	0.372535
	(methyltrisulfanyl)methane	126	-9.0796e-12	1.1893e-11	35	-0.7634	0.450326
	(methyldisulfanyl)-methylsulfanylmethane	140	-1.1100e-13	1.6581e-13	35	-0.6695	0.507589
	methylsulfonylmethane	94	1.8433e-13	2.4150e-13	35	0.7633	0.450405
Alcohols	butan-2-ol	59	-9.2360e-13	1.1768e-12	35	-0.7848	0.437828
	2-methylpropan-1-ol	43	2.8563e-12	1.4753e-11	35	0.1936	0.847607
	3-methylbutan-1-ol	**55**	**7.0951e-11**	**2.0846e-11**	**35**	**3.4035**	**0.001682**
	oct-1-en-3-ol	57	4.6461e-11	2.6848e-11	35	1.7305	0.092344
	nonan-2-ol	83	1.1241e-13	1.2719e-13	35	0.8839	0.382805
	2-phenylethanol	92	-7.3661e-14	1.2387e-13	35	-0.5947	0.555889
Aldehydes	butanal	55	-1.9131e-12	5.6494e-12	35	-0.3386	0.736903
	3-methylbutanal	58	-3.6169e-12	4.4886e-11	35	-0.0806	0.936236
	3-methylbut-2-enal	84	-3.5077e-15	1.6735e-14	35	-0.2096	0.835193
	benzaldehyde	106	-1.4806e-14	1.0074e-13	35	-0.1470	0.883997
	2-phenylacetaldehyde	92	2.6495e-14	8.7604e-14	35	0.3024	0.764109
Pyrazines	**2-methylpyrazine**	**94**	**-6.2491e-13**	**2.6627e-13**	**35**	**-2.3469**	**0.024719**
	2,5-dimethylpyrazine	**42**	**-3.7734e-11**	**1.5290e-11**	**35**	**-2.4679**	**0.018622**
	2,3,5-trimethylpyrazine	122	2.2775e-13	6.7771e-13	35	0.3361	0.738838
	2,3-dimethyl-5-[(E)-prop-1-enyl]pyrazine	148	3.3785e-14	3.5925e-14	35	0.9404	0.353448
Alkanes	octane	85	-1.8613e-12	1.1278e-12	35	-1.6504	0.107809
	dodecane	98	-1.0454e-13	7.2413e-14	35	-1.4436	0.157739
Esters	methyl 3-methylbut-2-enoate	83	3.4011e-12	3.3609e-12	35	1.0120	0.318496
Lactones	4-methyl-2,3-dihydropyran-6-one	54	-2.2734e-13	2.1772e-13	35	-1.0442	0.303561
Phenols	4-methylphenol (p-cresol)	107	9.2738e-14	2.7672e-13	35	0.3351	0.739528
Terpenes	7-methyl-3-methylideneocta-1,6-diene (β-myrcene)	93	2.4345e-13	5.9613e-13	35	0.4084	0.685477

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

A



B

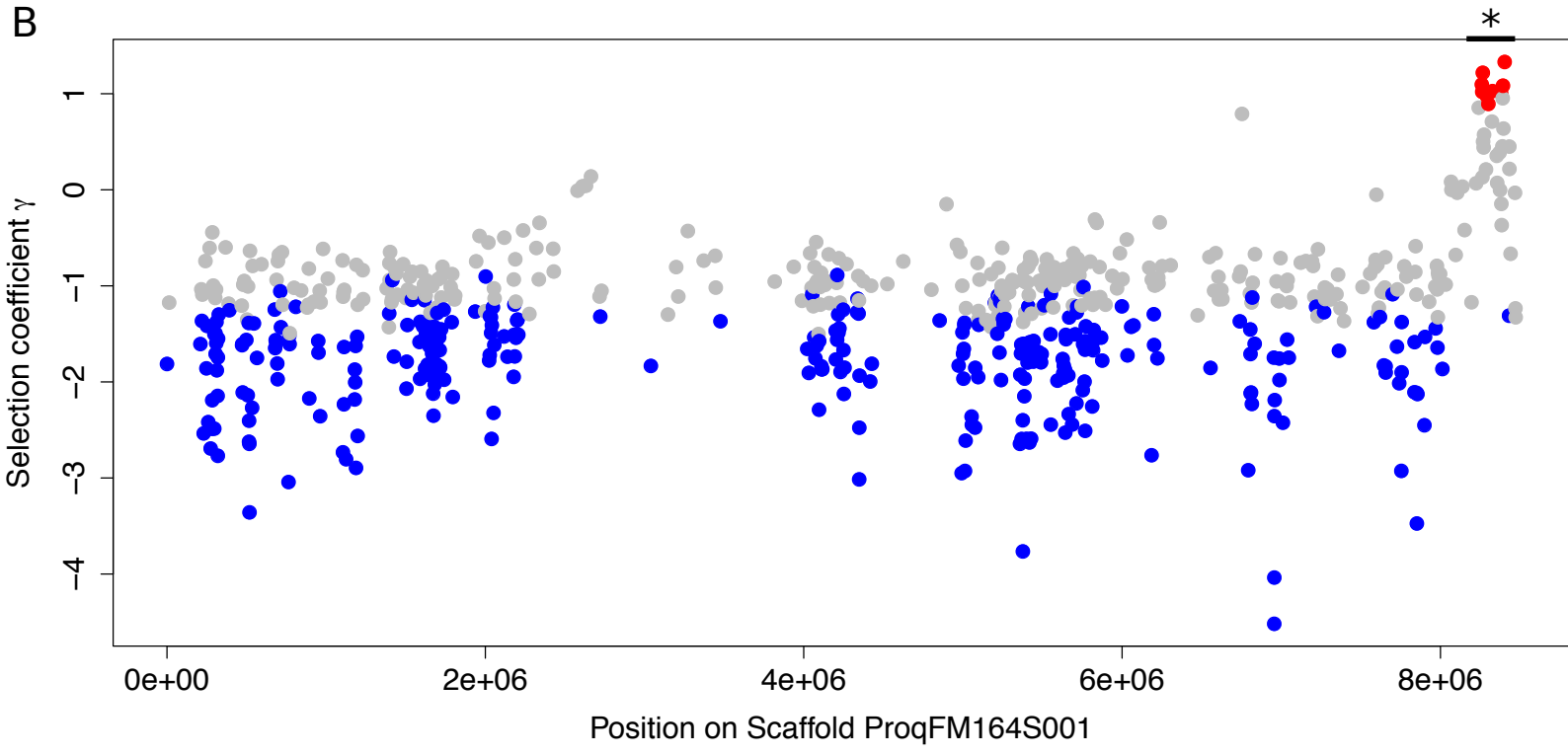
InterPro Number	Count in the region	Count in FM164	Pvalue Fisher test	Pvalue Fisher test after correction	InterPro domain annotation
IPR000641	2	3	0.00484	0.15383	CbxX/CfqX* (Foulger, D. & Errington, J. 1991)
IPR001138	2	350	0.02629	0.18840	Zn(2)-C6 fungal-type DNA-binding domain
IPR001223	3	20	0.01446	0.15548	Glycoside hydrolase family 18, catalytic domain* (Tzelepis, GD. et al. 2012)
IPR002100	2	13	0.04378	0.20115	Transcription factor, MADS-box
IPR002641	1	1	0.04444	0.20115	Patatin-like phospholipase domain *; ** (La Camera, S. et al. 2005; Zimmermann, R. et al. 2004)
IPR003286	2	5	0.00987	0.15548	Reverse transcriptase
IPR005197	2	10	0.02877	0.19033	Glycoside hydrolase family 71 * (Hasegawa, S. & Nordin JH. 1969)
IPR006600	2	6	0.01296	0.15548	HTH CenpB-type DNA-binding domain
IPR007087	4	42	0.01990	0.17570	Zinc finger, C2H2
IPR008160	1	1	0.04444	0.20115	Collagen triple helix repeat
IPR011547	1	1	0.04444	0.20115	SLC26A/SulP transporter domain ** (Bradfield, G. et al. 1970)
IPR013069	2	14	0.04929	0.21197	BTB/POZ domain
IPR013103	2	4	0.00715	0.15383	Reverse transcriptase
IPR018122	1	1	0.04444	0.20115	Fork head domain conserved site1
IPR018834	2	4	0.00715	0.15383	DNA/RNA-binding domain, Est1-type
IPR020683	8	101	0.00348	0.15383	Ankyrin repeat-containing domain
IPR020829	1	1	0.04444	0.20115	Glyceraldehyde 3-phosphate dehydrogenase, catalytic domain ** (Rogers, S. et al. 1986)
IPR022198	2	6	0.01296	0.15548	Protein of unknown function DUF3723
IPR024088	1	1	0.04444	0.20115	Tyrosine-tRNA ligase, bacterial-type
IPR028343	1	0	0.02247	0.17570	Fructose-1,6-bisphosphatase ** (Rogers, DT. et al. 1988)

bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

A

Gene	Roquefort	Non Roquefort	*Penicillium roqueforti*	Protein Length	Annotation
ProqFM164S01g002533	Neutral	0.659	-2.164	504	Aromatic-ring-hydroxylase-like
ProqFM164S01g002740	Neutral	0.965	-1.912	461	F-box-domain C-cyclin-like
ProqFM164S01g003510 *	1.097	Neutral	Neutral	253	Unknown function
ProqFM164S01g003511 *	1.021	Neutral	Neutral	166	Unknown function
ProqFM164S01g003514 *	1.219	Neutral	-2.228	332	Unknown function
ProqFM164S01g003523 *	0.980	Neutral	-2.869	940	UDP-glucuronosyl/UDP-glucosyltransferase
ProqFM164S01g003529 *	0.895	Neutral	Neutral	534	Putative glycosyl transferase
ProqFM164S01g003542 *	1.029	Neutral	-1.119	3848	Transcription associated protein
ProqFM164S01g003561 *	1.010	Neutral	-1.614	503	sap61, CWF-complex-protein
ProqFM164S01g003566 *	1.084	Neutral	-1.809	601	Beta-lactamase/transpeptidase-like
ProqFM164S01g003570 *	1.332	Neutral	Neutral	65	Unknown function
ProqFM164S03g000676	1.244	Neutral	Neutral	232	Unknown function
ProqFM164S03g001307	0.776	Neutral	Neutral	1635	Regulator-of-chromosome-condensation/beta-lactamase-inhibitor-protein-II
ProqFM164S04g000246	1.001	Neutral	Neutral	528	Major-facilitator-superfamily
ProqFM164S04g000250	1.442	0.912	Neutral	335	RPB3, DNA-directed-RNA-polymerase-II-subunit
ProqFM164S04g000252	1.241	Neutral	Neutral	435	Acyl-CoA-N-acyltransferase
ProqFM164S04g000579	1.183	Neutral	Neutral	78	Unknown function
ProqFM164S04g000895	Neutral	0.912	Neutral	124	Unknown function
ProqFM164S06g000156	Neutral	0.956	Neutral	400	Unknown function

B



bioRxiv preprint first posted online Oct. 24, 2018; doi: http://dx.doi.org/10.1101/451773. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.