Domestication of different varieties in the cheese-making fungus *Geotrichum candidum*

Bastien Bennetot[1], Jean-Philippe Vernadet[1], Vincent Perkins[2], Sophie Hautefeuille[1], Ricardo C. Rodríguez de la Vega[1], Samuel O'Donnell[1], Alodie Snirc[1], Cécile Grondin[2], Marie-Hélène Lessard[3], Anne-Claire Peron[4], Steve Labrie[3], Sophie Landaud[4], Tatiana Giraud[1], Jeanne Ropars[1]

[1]Laboratoire Ecologie Systématique et Evolution, Université Paris-Saclay, CNRS, AgroParisTech, site IDEEV, Gif-sur-Yvette, France
[2]SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France
[3]Département de sciences des aliments et de nutrition, Université Laval, Québec, Canada
[4]UMR SayFood, Université Paris-Saclay, INRAe, AgroParisTech, Thiverval Grignon, France

Keywords: domestication, adaptive divergence, lipolysis, proteolysis, genomics, volatiles, laboratory experiments, *Galactomyces candidus*

Abstract

Domestication is an excellent model for studying adaptation processes, involving recent adaptation and diversification, convergence following adaptation to similar conditions, as well as degeneration of unused functions. *Geotrichum candidum* is a fungus used for cheese-making and is also found in other environments such as soil and plants. By analyzing whole-genome data from 98 strains, we found that all strains isolated from cheese formed a monophyletic clade. Within the cheese clade, we identified three differentiated populations and we detected footprints of recombination and admixture. The genetic diversity in the cheese clade was high, indicating a lack of strong bottleneck. Commercial starter strains were scattered across the cheese clade, thus not constituting a single clonal lineage. The cheese populations were phenotypically differentiated from other populations, with a slower growth on all media, even cheese, a prominent production of attractive cheese flavors and a lower proteolytic activity. Furthermore, one of the cheese populations displayed footprints of a more advanced state of domestication, with much lower genetic diversity, denser and fluffier colones and excluding more efficiently cheese spoiler fungi. Cheese populations lost two beta lactamase-like genes, involved in xenobiotic clearance, and displayed transposable element expansion, likely due to relaxed selection. Our findings suggest the existence of genuine domestication in *G. candidum*, which led to diversification into different varieties with contrasted phenotypes. Some of the traits acquired by cheese strains indicate convergence with other, distantly related fungi used for cheese maturation.

Introduction

Understanding how populations adapt to their environment is a key question in evolutionary biology. Domestication, the change in the genetic and phenotypic make-up of populations under human artificial selection, is an excellent model for studying adaptation processes, as it involves recent adaptation events under strong selection on known traits and rapid diversification. Numerous studies have documented the specific traits acquired in domesticated animals (dog, horse, pig, cow) (Diamond, 2002; Frantz et al., 2015; Petersen et al., 2013; Warmuth et al., 2011) and plants (cabbage, corn, wheat) (Hufford et al., 2012; Mabry et al., 2021; Peng et al., 2011), as well as their genetic differentiation from wild populations and their adaptive genomic changes. For example, domesticated animals (dog, horse and cattle) have been selected for coat color, size, rapidity and docility (Liu et al., 2022; Plassais et al., 2022; Qanbari et al., 2014). In plants too, similar traits have been selected in different lineages, such as bigger grains with more nutrients and lack of dormancy (Cornille et al., 2014; Hufford et al., 2012; Peng et al., 2011; Purugganan, 2019). On the other hand, functions essential in wild environments but unused in anthropic environments have often degenerated due to relaxed selection, for example, reductions in defense mechanisms in plants (Cornille et al., 2014; Hufford et al., 2012). Domestication also often leads to strong reduction in genetic diversity due to bottlenecks in animals (e.g. dog) (Marsden et al., 2016) and annual plants (e.g. rice) (Zhu et al., 2007).

Humans have domesticated several fungi for the fermentation of foods (e.g. beer, bread, wine, dry-cured meat and cheese), to produce secondary metabolites used in pharmaceutics (e.g. penicillin), or for their nutritional and gustatory values (e.g. button and shiitake mushrooms) (Steensels et al., 2021). Fungi are excellent models for studying evolution and adaptation in eukaryotes, given their many experimental assets (Gladieux et al., 2014): fungi have relatively small genomes, many are easy to culture in laboratory conditions, and spores can survive long periods in the freezer. However, despite their economic and industrial importance, and their utility as biological models for studying adaptive divergence, the fungi used by humans have yet been little studied. An exception is the budding yeast *Saccharomyces cerevisiae* used in the production of beer, wine and bread (Bai et al., 2022; Duan et al., 2018; Fay and Benavides, 2005; Gallone et al., 2016; Lahue et al., 2020; Legras et al., 2018; Libkind et al., 2011; Liti et al., 2009; Peter et al., 2018), and to a lesser extent the filamentous fungus *Aspergillus oryzae* used to ferment soy and rice products in Asia (Galagan et al., 2005; Gibbons et al., 2012; Machida et al., 2005) and the *Penicillium* species used for cheese ripening, e.g. *P. camemberti* for soft cheeses (Ropars et al., 2020), and *P. roqueforti* for blue cheeses (Cheeseman et al., 2014; Dumas et al., 2020; Ropars et al., 2015). Phenotypic traits beneficial for food production have been acquired in fungal domesticated populations, being different from wild populations. The domestication having led to *P. camemberti* occurred in several steps, with the successive differentiation of several lineages displaying decreasing diversity and increasing beneficial traits for cheese maturation, from the wild *P. fuscoglaucum*, to *P. biforme* and then the two clonal *P. camemberti* varieties, *caseifulvum* and *camemberti* (Ropars et al., 2020b). Domesticated populations of fermented food microorganisms can for example better assimilate the carbon sources present in the anthropic environment, *e.g.* lactose for *Penicillium* cheese fungi (Ropars et al., 2015) and maltose for *S. cerevisiae* sourdough strains (Bigey et al., 2021). Furthermore, volatile organic compounds crucial for cheese flavor are more appealing in cheese populations compared to wild populations in *P. roqueforti* (Caron et al., 2021).

The genomic processes involved in adaptation to human-made environments in domesticated fungi include gene family expansion for specific metabolism pathways, gene gain, inter-specific hybridization, introgression and horizontal gene transfer (Almeida et al., 2014; Barros Lopes et al., 2002; Borneman et al., 2016; Cheeseman et al., 2014; Gallone et al., 2016; Libkind et al., 2011; Machida et al., 2005; Naumova et al., 2005; Novo et al., 2009; Ropars et al., 2015). Domesticated fungi also have lost parts of genes no longer useful in the food environment; for example a cheese *P. roqueforti* population and *P. camemberti* var. *caseifulvum* are no longer able to produce some of their toxins due to deletions in the corresponding genes (Gillot et al., 2017; Ropars et al., 2020b). Such losses are likely due to relaxed selection in terms of competition ability in cheese, in which desired fungi are often inoculated in large quantities compared to possible competitors. Bottlenecks and degeneration have also been documented in domesticated fungi, with reduced fertility and genetic diversity in the cheese fungi *P. roqueforti* and *P. camemberti* (Dumas et al., 2020; Ropars et al., 2020b).

While several cheese-making fungi have been studied recently, it is important to add study cases in additional lineages, as it allows addressing the question of whether adaptation to a similar medium leads to convergent traits. In the case of cheese-making fungi, one can expect convergence for example for more or less rapid growth on cheese, higher proteolysis and lipolysis abilities, higher competitive abilities and greater production of positive volatile compounds (Ropars et al., 2020a, 2020b). *Geotrichum candidum* (teleomorph *Galactomyces candidus*) is a dimorphic fungus (i.e., able to grow as a yeast or a mycelial form), commonly used for cheese-making, but also thriving in other environments such as soil, plants and fruits. *Geotrichum candidum* is naturally present in raw milk and is also often added as a starter culture for the production of semi-hard, mold-ripened, smeared soft cheeses, fresh goat and ewe cheeses. Analyses based on genetic markers have revealed genetic differentiation between cheese and wild strains (Alper et al., 2013; Jacques et al., 2017; Perkins et al., 2020; Tinsley et al., 2022). Phenotypic diversity within *G. candidum* has been reported in terms of carbon and nitrogen assimilation, lipolysis and proteolysis (Boutrou and Gueguen, 2005; Perkins et al., 2020). However, it has not been tested whether *G. candidum* cheese populations have evolved specific traits that could be beneficial for cheese-making.

By analyzing the genomes of 98 strains isolated from different kinds of cheeses and other environments, we confirmed the genetic differentiation between cheese and wild strains. Within the cheese clade, we identified three differentiated cheese populations, including one with all goat cheese strains, as well as footprints of recombination and admixture. The genetic diversity in the cheese clade remained high indicating a lack of strong bottlenecks. Commercial strains were scattered within the cheese clade, some corresponding to hybrid strains. We found phenotypic differentiation between cheese and wild populations, and between cheese populations, in terms of growth, proteolysis and volatile compounds. We revealed the loss, in the cheese clade, of two tandem beta lactamase-like genes involved in xenobiotic clearance. Altogether our findings suggest the existence of genuine domestication in *G. candidum,* resulting in both the genetic and phenotypic differentiation of cheese strains from their wild counterparts, and their convergence with other domesticated cheese fungi. We also found diversification within the cheese clade, with three genetic clusters with contrasted traits and levels of diversity.

Results

Genetic differentiation between wild and cheese strains in *Geotrichum candidum*

We collected and sequenced the genomes of 88 *G. candidum* strains with Illumina technology and included in our analyses ten available genomes (Illumina and PacBio) (Perkins et al., 2020). Our dataset included 61 strains isolated from different kinds of cheeses (semi-hard, mold-ripened, smeared soft and fresh goat cheeses), 16 industrial strains used for cheese-making, seven strains from dairy products, four strains from other food substrates (e.g., sausage or vegetables) and all the 10 wild strains available in public collections worldwide (isolated for exemple from soil or plant) (Table S1). We identified 699,755 SNPs across the 98 strains by mapping against the CLIB 918 reference genome (cheese strain, NCBI accession: PRJEB5752).

The maximum likelihood tree, principal component analysis (PCA) and neighbor-net (SplitsTree) analyses all identified the same three clades (Figure 1A), with one containing mostly wild strains (corresponding to the GeoC group identified previously based on genetic markers) (Perkins et al., 2020), one composed of strains of varying origins (i.e. dairy products and other environments, corresponding to the group previously named GeoB) and one containing mostly cheese and dairy strains (previously named GeoA). The larger sampling and the genome sequencing of the present study further revealed genetic subdivision in the cheese clade, with three clearly differentiated populations and several admixed strains (Figure 1B).

We performed an admixture analysis by inferring *K* populations, *K* ranging from two to ten. At *K*=3, the cheese, mixed-origin and wild clades were separated (Figure S1). At *K*=5, the cheese clade was divided into three genetic clusters, corresponding to monophyletic groups in the maximum likelihood tree and well-separated genetic clusters in the PCA and the neighbor-net (Figure 1 B, D). At higher K, new populations inferred were either too small (two individuals) or not monophyletic.

Some cheese strains could not be assigned to any genetic cluster with the admixture analysis and were placed on the PCA between the three well-delimited cheese genetic clusters (Figure 1C-D), suggesting that they resulted from admixture events. To test this hypothesis, we investigated whether these strains had mosaic genomes, with different genomic regions assigned to distinct clusters. We calculated pairwise identity between unassigned strains and the other strains, computing mean identity to the different genetic clusters along scaffolds using sliding windows. For all unassigned strains in the cheese clade, we observed shifts in identity values along scaffolds, confirming that these strains are the results of admixture between clusters (Figure S2). In contrast, the three unassigned strains outside of the cheese clade did not show changes in similarity level to the different clusters along their genome; these strains may belong to yet additional genetic clusters that could not be distinguished by the analyses because too few strains belonged to these clusters in the sampling (Figure 1B).

We found no particular cheese type distribution among the three cheese populations, except that all strains from goat cheeses clustered in the Cheese_1 population. The wild clade was the most differentiated population from all other *G. candidum* populations with F_{ST} values above 0.70 (Table S2). Its differentiation level was similar to that found between the domesticated *P. camemberti* mold and its wild closest relative species, *P. fuscoglaucum* (F_{ST} = 0.83; d_{xy} = 6E-03), and much higher than the differentiation between the cheese Roquefort population and the lumber/food spoiler population in *P. roqueforti* (F_{ST} = 0.27). The percentage of private SNPs in the five populations was also high (Table S3). F3 tests based on the number of shared sites (Table S4) supported the differentiation between these populations.

High nucleotide diversity within cheese populations and footprints of recombination

The overall diversity in the cheese clade (π=2.82E-3) was higher than in the wild population (π=2.12E-03). Each of the three cheese populations of *G. candidum* had however reduced nucleotide diversities compared to wild and mixed-origin populations (Table S5), by at least a factor of two. The Cheese_2 population showed the lowest genetic diversity (π=4.82E-04), by a factor of four compared to the two other cheese populations (Table S5), being of the same order of magnitude as in the Roquefort *P. roqueforti* population (Dumas et al., 2020). The Cheese_1 population had a nucleotide diversity (π=1.26E-03) similar to that in the domesticated cheese species *P. biforme* (π=1.09E-03), whereas the Cheese_3 population (π=1.92E-03) was genetically as diverse as *P. fuscoglaucum* (π=1.93E-03), the closest wild relative of the clonal lineage *P. camemberti* (Ropars et al., 2020b).

Geotrichum candidum is a heterothallic fungus, meaning that sexual reproduction can only occur between two haploid cells carrying different mating types. Two mating types have been described in *G. candidum* (Morel et al., 2015): MATA, encoded by a HMG box gene homolog to the MATA2 *Kluyveromyces lactis* allele, present in CLIB 918 (sequence id: HF558448.1), and MATB, encoded by an alphabox gene homolog to the MATα1 *S. cerevisiae* allele, present in the strain CBS 615.84 (sequence id: HF558449.1). In the Cheese_2 population, we found a significant departure from the 1:1 mating-type ratio expected under regular sexual reproduction*;* all the 12 strains carried the MATB allele, suggesting that this population is at least partly clonal (Table S6). The absence of linkage disequilibrium decay with physical distance between two SNPs (Figure S3), together with the absence of reticulation in the neighbor-net (Figure 1), are also consistent with a lack of recombination in the Cheese_2 population. However, pairwise homology index (PHI) tests, testing with permutations the null hypothesis of no recombination by looking at the genealogical association among adjacent sites, were significant in all the *G. candidum* populations (Table S7); this indicates that recombination did occur at least in a recent past in the Cheese_2 population.

We did not detect any accumulation of nonsense or missense mutations in any population compared to silent mutations (Felsenstein, 1974; Table S8), while degeneration can be expected to be particularly strong in clonally replicated populations as recombination allows more efficient selection. None of the genes that presented nonsense mutation had predicted functions that could be detected as specific either to the wild or cheese environments (Figure S4). This also suggest that the absence of sexual reproduction in the Cheese_2 population may be too recent to observe degeneration.

In contrast to the Cheese_2 population, we found both mating-type alleles in balanced proportions in both Cheese_1 and Cheese_3 populations (Table S6) and we observed sharp decays in linkage disequilibrium (LD) with genomic distance, although LD levels remained higher than in the mixed-origin and wild populations (Figure S3). We observed reticulations in the neighbor-net network within populations and, to a lesser extent, between populations (Figure 1A).

As previously mentioned, the 16 commercial starter strains in our *G. candidum* dataset were scattered in the maximum likelihood tree (in yellow, Figure 1B.a.) and we detected above footprints of recombination in the Cheese_1 and Cheese_3 populations (Figure 1A, Figure S3). We nevertheless detected a few groups of clonemates, by the lack of branches in the trees and the presence of fewer than 1,200 SNPs between strains (Figure 1; Table S1, clonal group column). As strains within these clonal lineages were isolated from different cheeses, it indicates that some lineages may be clonally cultivated for cheese-making; some of the commercial starter strains indeed clustered within these clonal groups ("$" symbol on Figure 1B). Within the admixed cheese strains, 19 out of 23 were part of clonal lineages, suggesting that hybrid lineages may have been selected for beneficial traits for cheese-making, as in other domesticated fungi (e.g. *S. pastorianus*; Gallone et al., 2018).

Copy number variation: loss of two tandem beta lactamase-like genes in the cheese populations and repeat expansion

Expansions of gene families involved in specific metabolism pathways, of transposable elements and loss of genes no longer required in the new environment can be involved in adaptation to new environments. For example, variations in gene copy number were associated with the adaptation of *S. cerevisiae* to beer making, with duplications of genes involved in maltose and maltotriose metabolism specifically in domesticated beer strains (Gallone et al., 2016; Giannakou et al., 2020; Gonçalves et al., 2016). We therefore looked for gene copy-number variation (CNV) that differentiated wild and cheese populations, using 500 bp sliding windows and two reference genomes, belonging to the Cheese_3 and the wild populations, respectively (Table S9). Using the Cheese_3 reference, we found 61 CNV regions (mean length of 1515 bp and 45 non-genic CNVs), encompassing in total 16 genes, half having predicted functions, none being obviously related to cheese adaptation (e.g. methylglyoxal reductase and tRNAs, Table S9). Using the wild genome reference, we found 132 CNV regions (mean length of 1664 bp and 105 non genic CNVs), encompassing 29 genes (seven with unknown functions).

One of these regions, 20 kb long, included only two genes, both matching the Pfam hidden markov model for beta-lactamases; these two genes (g5112 and g5113) were absent from all cheese populations, and were present in most wild strains (except one that had partially the region) and in four strains belonging to the mixed-origin population (Figure 2). The nucleotide identity between the two beta lactamase-like genes was 93%. A third beta lactamase-like gene (g5111) was found immediately next to this CNV region in all *G. candidum* strains, and displayed a nucleotide identity of 87% with the two other beta lactamase-like genes within the CNV. Surrounding these different genes, we found several Tc1/*mariner*, a LINE/Tad1 and other DNA transposons, that may have contributed to the beta-lactamase-like gene deletion (Figure 2). Fungal beta lactamase-like genes are known to contribute to hydrolysis of microbial and plant xenobiotics, and thus may be important in the wild environment to compete with other microorganisms (Gao et al., 2017). The cheese populations may have lost these two copies of the beta-lactamase genes due to relaxed selection; indeed, these functions may not be useful in the cheese environment if *G. candidum* is inoculated in high quantity compared to potential competitors.

De novo detection of repeats using the wild strain LMA-244 yielded a library containing 107 types of repeated elements (including 15 types of DNA transposons and 11 of retroelements and 3 rolling-circles). We identified 14 types of repeated elements present in at least one other *G. candidum* genome with five times more copies than in the LMA-244 wild strain (this threshold was set based on the fat tail of the distribution; Figure S5). Among these 14 types of repeated elements, several DNA transposons of the Tc1/*mariner* repeat family showed a cheese-clade specific expansion (Table S10, Figure 3). Several unidentified, *Tad1* and *Helitron* repeat types also showed expansions in the cheese clade, alongside a milder expansion in the mixed origin clade. Such transposable element expansions in the cheese clade could be due to relaxed selection (Baduel et al., 2019).

Genomic footprints of adaptation: genomic islands of differentiation in cheese populations and genes under positive selection

We looked for genomic regions with a greater differentiation or a lower genetic diversity than the genomic background when comparing each of the three cheese populations to the wild population, to detect footprints of divergent selection and recent selective sweeps, respectively. We scanned the whole genome using-non-overlapping windows and explored the windows with the 1% highest d_{xy} (high differentiation) or 5% lowest π (low diversity) values. Regions of high differentiation appeared as outliers in the distribution of d_{xy} values, representing a small peak of high d_{xy} values (Figure S6), and were often located in non-genic or in low gene-density regions (Figure S7). We however detected 69 genes in the high-differentiation regions across the three cheese populations, including 26 genes with predicted functions. Two of them encoded proteases, one of them being an ADAM metalloprotease which was an enriched function in the high-differentiation regions when compared to the rest of the genome , and the only one (Table S11). Proteases are important in cheese-making as the breakdown of milk caseins greatly contributes to cheese texture and decreases water activity by degrading proteins into molecules with free carboxyl and amino groups (McSweeney, 2004). *Geotrichum candidum* is prevalent during the amino-acid catabolism ripening step of Pelardon fresh cheese (Penland et al., 2021), suggesting that *G. candidum* plays an important role in proteolysis in cheese-making.

In the 198 windows representing the pooled set of the 5% lowest π values in the three cheese populations, we detected 497 genes, 323 of which had predicted functions Among the 323 annotated genes, five predicted proteases or lipases. Although these functions were not enriched compared to the rest of the genome (Table S11), this could still represent footprints of selection on some of these individual genes. Lipases are key enzymes for cheese flavor as they enable the breakdown of milk fats through fatty acid production (Collins et al., 2003), and are thus crucial for cheese-making.

We searched for genes evolving under positive selection in terms of high rates of non-synonymous substitutions by performing McDonald and Kreitman (MK) tests (Table S12), comparing the mixed-origin population to each cheese population and to the cheese clade as a whole. We detected 25 genes as evolving under positive selection in at least one cheese population (9 for Cheese_1, 18 for Cheese_2, two in Cheese_3 and one in all three cheese populations at once; Table S12). Among them, a metalloendopeptidase evolved under positive selection in all three cheese populations, likely playing a role in casein degradation through cell lysis (Dugat-Bony et al., 2015; KUMURA et al., 2002). A Glucan 1,3-beta-glucosidase was also detected as evolving under positive selection in the Cheese_2 population; this enzyme could be involved in fungal inhibition through fungal cell degradation (Adams, 2004). The other genes under positive selection had either no predicted function or putative functions that could not be related to cheese adaptation (Table S12).

Phenotypic differentiation between cheese and wild populations

Denser mycelial growth and/or faster proteolysis in cheese populations of *Geotrichum candidum*

Strains selected by humans are expected to display specific traits beneficial for cheese-making, such as faster growth in cheese at cave temperature or colonies of attractive aspect or color. For example, the *P. camemberti* strains used for soft cheese production were selected for their white and fluffy aspect, to make cheeses more attractive to consumers compared to the blue-grey crust produced by the *P. camemberti* ancestor (Pitt et al., 1986). In contrast, the ability to grow in harsh conditions may have been lost in cheese strains due to relaxed selection, as often reported for unused traits in human-made environments in domesticated organisms (Gallone et al., 2018; Price, 2002; Ropars et al., 2015).

We therefore measured colony radial growth of 31 strains from the five *G. candidum* populations on different agar media (cheese, rich and poor media) at different temperatures. Wild populations grew faster than cheese populations on all media at all temperatures, with a more pronounced difference at 25°C (Table S13, Figure S8).

This may result from trade-offs with other traits, such as a fluffier mycelium, i.e. more vertical growth at the expense of less radial growth.

To test whether cheese populations had a denser mycelium or had become whiter and/or fluffier, we compared the opacity of populations on cheese agar at cave temperature (10°C), which integrates the brightness and fluffiness of a colony. The Cheese_1 and Cheese_3 populations were not more opaque than wild populations (Figure 4B). The Cheese_2 population had a significantly higher opacity than all other *G. candidum* populations, except the mixed-origin population (post-hoc Tukey test in Table S13). This represents a convergence with *P. camemberti* var. *camemberti*, with independent evolution of similar phenotypes in two distantly related cheese fungi.

Lipolysis and proteolysis are crucial biochemical processes during cheese ripening, that influences flavor and texture of the final product; lipolysis and proteolysis contribute to energy and nutrient uptake, and they affect the production of volatile compounds, which are key flavor factors in cheeses (McSweeney, 2004). All populations of *G. candidum* had similar lipolysis rates. The wild and mixed-origin populations had degraded a significantly higher amount of proteins than the cheese populations and we did not detect any proteolysis in the Cheese_2 population in our experiment (Figure S9; Table S14).

No adaptation to high salt concentration or milk origin in cheese populations

Cheese is a salty medium, with the percentage of salt varying from 0.5 g / 100 g for Emmental to 3 g / 100 g for Roquefort. Salt is added on the surface of cheeses to prevent the growth of contaminants, and cheese populations of *G. candidum* may thus have adapted to high salt concentrations. Cheeses display a wide range of salt concentrations so we tested four cheese media: unsalted, 1% salt as St Nectaire and cream cheeses, 2% as Camembert and goat cheeses and 4% as Roquefort blue cheeses. Wild populations grew faster than cheese populations in all salt concentrations tested, as on YPD and minimal media (Figure S10A ;Table S13).

Because all strains sampled from goat cheeses belonged to the Cheese_1 population, we tested whether this population was able to grow faster on goat cheese medium (1% salt) compared to other populations. We however found no significant interaction between population and media on radial growth effects, i.e. no specific adaptation to any particular kind of milk by the different populations (Figure S10B).

Contrasting volatile compound production between wild and cheese populations

Cheese ripening fungi, including *G. candidum*, contribute to cheese flavor through the production of volatile compounds (McSweeney, 2004). Flavor is a crucial criterion for cheese consumers and the cheese populations may have been selected for desirable and specific volatile compounds. We grew 14 *G. candidum* strains on a sterilized Camembert curd for 21 days at 10°C, *i.e.,* the ripening conditions of a Camembert. On average across compounds, the wild population produced five times higher quantities of volatiles than cheese populations. In order to compare the relative proportions of the different compounds, which is also an important aspect for flavor, we standardized the values by dividing all compound quantities by the total quantity of volatiles per sample. The PCA indicated a differentiation between wild and cheese strains in terms of volatile relative proportions (Figure 4D). The wild population thus produced combinations of volatile compounds different from cheese populations, with a high proportion of ethyl esters and ethyl acetates (Figure S10C), known to be key compounds in fermented beverages such as wine and beer. However, the impact of ethyl acetate on flavor is rather negative because it brings solvent type notes. In cheese, these esters are never predominant (Liu et al., 2004; Urbach, 1997). Ethyl esters are involved in anaerobic metabolism and may be important for survival in the wild. By contrast, cheese strains produced many alcohols, ketones, aldehydes and sulfur compounds (Figure S10C), known for producing attractive flavors such as buttery, cheesy, fermented and aldehydic notes (Curioni and Bosset, 2002). These volatile compounds, attractive in cheese, were present in similar absolute quantities in wild strains but were in minor proportions compared to other volatile compounds (Table S13), suggesting that cheese populations evolved a lower production of undesirable and unused volatiles. The overall balance between different volatile compounds is as important as volatile absolute quantities for flavor perception (Liu et al., 2004). The dimethyl sulfone, a compound previously reported to be produced during the catabolism of L-methionine in *G. candidum*, is actually specifically produced by the cheese populations (Bonnarme et al., 2001; Penland et al., 2021).

Cheese populations inhibit more the growth of food spoilers than wild populations

Cheese is a protein- and fat-rich medium, where many microorganisms, including desired microbes, but also spoilers, can thrive and thus compete for nutrients; for example, iron is limiting in cheese (Mayo et al., 2021; Monnet et al., 2015, 2012). Cheese *G. candidum* populations may have been selected for excluding competitors by inhibiting their growth (Boutrou and Gueguen, 2005). This fungus is known to inhibit fungal and bacterial food spoilers, such as *Aspergillus* species and *Listeria monocytogenes,* but these inhibitory activities have only been investigated in cheese *G. candidum* strains so far (Dieuleveux et al., 1998; Nielsen et al., 1998; Omeike et al., 2021). We therefore tested whether cheese populations displayed better growth inhibition abilities than the wild population, using common fungal food spoilers as competitors: *Debaryomyces hansenii, Penicillium biforme, P. roqueforti* and *Scopulariopsis asperula*. We also tested whether growth inhibition of challengers occurred via secreted and/or volatile compounds.

Inhibition by a mycelium lawn - In the first experiment, we grew challengers in a central spot for 24h, alone or after spreading out *G. candidum* to let it grow as a lawn; growth inhibition could occur in this setting by secreted molecules in the medium, volatile compounds and/or a physical barrier to reach nutrients and grow. The growth of *D. hansenii* was completely inhibited by all populations of *G. candidum*. *Penicillium roqueforti* was strongly inhibited by *G. candidum*, in particular by the Cheese_2 population that completely prevented *P. roqueforti* growth (Figure 5; Table S13). The growth of *Scopulariopsis asperula* and *P. biforme* was also inhibited by *G. candidum,* with a significant difference between competitor growth when spread alone or on a *G. candidum* lawn; the Cheese_2 population again inhibited better competitors than any other population (Figure 5; Table S13). The Cheese_2 population was the most opaque population on cheese and had a beta-glucanase gene under positive selection, suggesting that challenger inhibition would be due to either mycelium density as a physical barrier or degradation of competitor cell wall.

Inhibition by volatile compounds - In a second experiment, we used splitted Petri dishes (the two parts being separated by a plastic barrier) to test whether cheese populations inhibited competitors to a greater extent than the wild population when only volatile compounds can reach challengers. No significant growth difference was observed between the growth alone and at the side of *G. candidum* for neither *S. asperula* nor *P. roqueforti* (Figure 5B, Table S13). Only *P. biforme* showed a significant growth inhibition by *G. candidum* in this setup (Table S13)*;* such a growth inhibition by *G. candidum* from an isolated Petri dish compartment indicates that volatile compounds produced by *G. candidum* are able to impair the growth of some competitors.

The two sets of experiments enabled us to assess by which mechanism *G. candidum* can inhibit competitors: *P. biforme* growth was inhibited by *G. candidum* in the splitted Petri dishes, suggesting that volatile compounds are able to impair its growth. On the contrary, *P. roqueforti* and *S. asperula* were only inhibited by *G. candidum* when molecules could diffuse in their medium and *G. candidum* mycelium could form a physical barrier. The Cheese_2 population had a stronger inhibition ability than the other *G. candidum* populations only when molecules could diffuse in the medium and the mycelium could act as a barrier.

Discussion

Analyzing the genomes of 98 *G. candidum* strains isolated from different kinds of cheeses, other food substrates and other environments revealed three monophyletic clades, corresponding to strains isolated from cheese, mixed-origins (dairy and other environments) and the wild, respectively. Within the cheese clade, we found three distinct clusters and several admixed strains. In terms of genetic diversity, the mixed-origin clade contained the highest diversity level, followed by the wild clade and then the distinct cheese populations. However, the nucleotide diversity within each of the cheese populations was still relatively high compared to other cheese fungi. Indeed, the Cheese_2 population, while being four times less diverse than the two other cheese populations, was as diverse as the Roquefort *P. roqueforti* cheese population. The low diversity, the presence of a single mating type, a high level of linkage disequilibrium and the absence of reticulation in the neighbor-net network indicated a lack of recombination in the Cheese_2 population, that may thus correspond to a clonally cultivated line for cheese-making. Additional, less widespread clonal lineages may be cultivated for cheese-making, as we found clonemates in all cheese populations, even in the clusters of intra-specific hybrids, and including some commercial starter strains. This presence of commercial starter strains in the clusters of intra-specific hybrids suggests that hybrid lineages may have been selected for beneficial traits for cheese-making, as in other domesticated fungi (e.g. *Saccharomyces pastorianus* used for the production of lager beer) but this needs further investigation (Rainieri et al., 2006). The Cheese_1 population was as diverse as the domesticated cheese species *P. biforme* and the Cheese_3 population as its wild relative *P. fuscoglaucum.* The genetic diversity of the cheese clade as a whole was even higher than that in the wild population, which may be due to the relatively low number of wild strains available and to the diversification of the cheese clade into three varieties.

The genetic relationships between *G. candidum* populations and their contrasting levels of diversity suggest that domestication occurred in several steps, with an ancient domestication event separating the mixed-origin and the wild clades, then the cheese and the mixed-origin clades, and yet more recently the three cheese clusters. The domestication of *P. camemberti* similarly occurred in several steps, the last steps involving the selection of a white and fluffy clonal lineage (Ropars et al., 2020b). Considering the genetic diversity, the situation in *G. candidum* is however very different from that of *P. camemberti* and *P. roqueforti,* for which a single or a few clonal lineages are sold by spore producers for all kinds of cheeses (Ropars et al., 2020b). The domestication of *G. candidum* did not involve strong bottlenecks that occurred in other domesticated cheese fungi, such as *P. camemberti, P. roqueforti* and *S. cerevisiae*, perhaps because it is more abundant spontaneously in raw milk and because there has been a diversification into three genetically differentiated varieties. It may also be that the domestication of the three *G. candidum* varieties is more recent or has not involved a selection as strong as in other cheese fungi, except perhaps in the Chhese_2 lineage. Sampling further cheese types, geographic regions and wild environnements may reveal further genetic clusters.

We found evidence of phenotypic adaptation to cheese-making in *G. candidum* cheese populations, with a slower growth on all media, even cheese, a prominent production of attractive cheese flavors and a lower proteolytic activity compared to the wild population. The slower growth and proteolysis activity may allow to prevent a too fast degradation of products during maturation, as found in the Roquefort *P.*

roqueforti population (Dumas et al. 2020). The lack of adaptation to salt was also found in the Roquefort *P. roqueforti* population and in dry-cured meat *Penicillium* fungi and may be due to evolutionary constraints (Lo et al., 2022). We also found genomic footprints of adaptation to cheese, with the presence of genomic islands of differentiation in cheese populations and the loss of genes no longer required in the human-made environment, i.e. tandem beta lactamase-like genes. This may correspond to a first step of domestication.

The Cheese_2 population appeared to represent a more advanced state of domestication than the other cheese populations, with much lower genetic diversity, a fluffier mycelium, a higher competitive ability and a complete lack of proteolysis activity. Denser mycelial growth leading to a fluffy aspect at the expense of less rapid radial growth has also been selected in *P. camemberti* var. *camemberti* (Ropars et al., 2020b), thus representing a convergent phenotype between two distantly related cheese fungi. *Geotrichum candidum* is increasingly inoculated in milk in the place of *P. camemberti* for industrial soft cheese production, as it provides the fluffy desired aspect without the disadvantage of *P. camemberti* that browns the surface of Camembert cheeses at the end of the ripening process (Carreira et al., 2002). Proteolysis activity was also found lower in the Roquefort *P. roqueforti* population, which may be beneficial to obtain not too degraded cheeses (Dumas et al. 2020). The volatile proportions produced by cheese strains corresponded to attractive flavor for cheese-making, in contrast to wild strains, as also documented in P. roqueforti (Caron et al., 2021; Dumas et al., 2020). *Geotrichum candidum* is able to efficiently inhibit the growth of common food spoilers, in particular *P. biforme* and *P. roqueforti,* and the clonal Cheese_2 population is the most efficient competitor.

Our study shows that it is of fundamental importance to study further domestication in various cheese fungi as it allows assessing whether independent adaptation events to similar media and usage lead to evolutionary convergence, as this is an important question in evolutionary biology (Alberto et al., 2018; Cresko et al., 2004; Dyer et al., 2012; Elmer et al., 2014, 2010; Lin et al., 2012; Macías et al., 2021; O'Quin et al., 2010; Thorpe et al., 2015). We found here both similarities (convergence) and differences in the adaptation of *G. candidum* to cheese compared to other cheese fungi. One of the most striking convergence was the evolution of a fluffy and white mycelium as in *P. camemberti* with a trade-off with radial growth

(Ropars et al., 2020b). The higher competitive ability and lower proteolysis activity, as well as the greater production of positive volatile compounds, also represent interesting convergence events between multiple very distant fungal lineages, indicating that evolution can be repeatable.

Our findings also have industrial implications, as they reveal high genetic diversity and subdivision in a fungus widely used in the cheese industry, and the existence of different varieties, *i.e.,* genetically and phenotypically different populations used for cheese-making, with specific and contrasted traits beneficial for cheese-making. The most fluffy and most competitive cheese population corresponded to a clonal lineage which may represent the most recent selection event. The occurrence of recombination between cheese strains is highly relevant for cheese producers as it opens possibilities for further improvement for the agrofood sector. It is crucial to maintain the larger genetic diversity in cheese *G. candidum* populations as genetic diversity is essential in domesticated organisms for variety improvement and diversification and to avoid degeneration (Harlan et al., 2012).

Material and Methods

Sampling

We isolated 53 strains from different kinds of cheeses (e.g. Camembert, Brie, Saint Nectaire, Ossau Iraty, comté, bleu de chèvre) from five European countries, Canada and the USA. Cheese crusts were left in the freezer for 24h to kill acarians. Then, we diluted a piece of each crust in sterile water and spread 50 uL of the suspension on a malt agar Petri dish. When colonies appeared on the Petri dish, typically after three days, we isolated the different morphotypes with a sterile toothpick and inoculated them on new Petri dishes. After seven days of growth, we performed monospore isolation by several dilution steps, in order to obtain separated colonies arising each from a single spore. We identified the species of these pure strains after DNA extraction by sequencing the 5’ end of the nuclear ribosomal large subunit (LSU rDNA) using the LROR/LR6 oligonucleotide primers (Vilgalys and Hester, 1990). We also gathered 24 strains from INRAE, isolated from cheeses but also other environments (e.g. sand, hay, rainforest) and 15 strains from a French spore seller. We gathered all the wild strains available in public collections. For each strain, single spore cultures were generated to ensure the presence of a single genotype before DNA extraction.

The LMA-244 strain were inoculated on Yeast Extract Glucose (YEG) agar plates (10 g.L-1 of yeast extract (Fischer Scientific), 10 g.L-1 of D-glucose (EMD Chemicals) and 15 g.L-1 of Bacto agar (BD Diagnostics)) directly from 15% glycerol (v/v) stock cultures stored at -80°C. The plates were incubated in the dark for five days at 25°C.

DNA extraction, genome sequencing, assembly, annotation and mapping

We used the Nucleospin Soil Kit (Macherey-Nagel, Düren, Germany) to extract DNA from 88 *G. candidum* strains cultured for five days on malt agar. Sequencing was performed with Illumina HiSeq 3000 paired-end technology (Illumina Inc.), 2x150 bp. For the eight LMA strains, sequencing was performed using the Illumina HiSeq paired-end technology.

All Illumina reads were trimmed and adapters cleaned with Trimmomatic v0.36 (Bolger et al., 2014). Leading or trailing low quality or N bases below a quality score of three were removed. For each read, only parts that had an average quality score higher than 20 on a four base window are kept. After these steps, only reads with a length of at least 36 bp were kept.

Cleaned Illumina reads were assembled with SPAdes v3.15.3 not using unpaired reads with “--careful” parameter.

For the LMA-244 strain, Genomic DNA was extracted using the Fungi/Yeast Genomic DNA Isolation Kit (Norgen Biotek Corp.) with the following modifications. Thirty milligrams of frozen grounded mycelium were thawed and homogenized in 500 µL of a 0.9% NaCl solution. The elution buffer was replaced by a Tris 10 mM buffer (pH 8). Following the extraction step, gDNA suspensions were purified and concentrated using Agencourt AMPure XP magnetic beads (Beckman-Coulter), according to the manufacturer’s protocol.

DNA concentration and purity were measured using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, U.S.A.) and a Qubit Fluorometer 3.0 (Thermo Fisher Scientific Inc., Wilmington, U.S.A.). The DNA library was prepared following the Pacific Biosciences 20 kb template preparation using BluePippin Size-Selection System protocol and the Pacific Biosciences Procedure & Checklist – Preparing Multiplexed Microbial Libraries Using SMRTbell Express Template Prep Kit 2.0 protocol. No DNA shearing was performed. The DNA damage repair, end repair and SMRT bell ligation steps were performed as described in the template preparation protocol with the SMRTbell Template Prep Kit 1.0 reagents and the SMRTbell Express Template Prep Kit 2.0 reagents (Pacific Biosciences, Menlo Park, CA, USA). The DNA library was size selected on a BluePippin system (Sage Science Inc., Beverly, MA, USA) using a cut-off range of 10 kb to 50 kb. The sequencing primer was annealed at a final concentration of 0.83 nM and the P6 v2 polymerase was bound at 0.50 nM while the sequencing primer was annealed with sequencing primer v4 at a final concentration of 1 nM and the Sequel 3.0 polymerase was bound at 0.5 nM.. The libraries were sequenced on a PacBio RS II instrument at a loading concentration (on-plate) of 160 pM using the MagBead OneCellPerWell loading protocol, DNA sequencing kit 4.0 v2, SMRT cells v3 and 4 hours movies.

Raw PacBio reads were corrected using Illumina reads already available and described in a previous article (Perkins et al., 2020), with the default parameters of the LoRDEC software and trimmed with Canu v1.6 (Koren et al., 2017; Salmela and Rivals, 2014). Corrected and trimmed PacBio reads were then assembled using Canu v1.6. Illumina polishing of the Canu assembly was performed using Pilon v1.22 (Walker et al., 2014). A final assembly step was then performed with the hybrid assembler SPAdes v3.11.1 using the trimmed PacBio reads, the Illumina reads and the Pilon corrected assembly as trusted contigs (Antipov et al., 2016; Prjibelski et al., 2020) . Additionally, the CLIB 918 assembly (Bioproject PRJEB5752) was used as a reference in the SPAdes script for the assembly of each *G. candidum* genome (Morel et al., 2015). Scaffolds were filtered using the khmer software with a length cut-off of 1,000 bp (Crusoe et al., 2015).

The LMA-244 PacBio assembly and reads have been deposited in GenBank: nbPROJECT. To annotate short read assemblies and the LMA-244 genome, gene prediction was performed using Augustus v3.4.0 (Stanke et al., 2008). The training annotation file “saccharomyces” was used, with parameters as follows: “--gff3=on”, “--protein=on”, “--codingseq=on”, “--exonnames=on”, “--cds=on” and “--uniqueGeneId=true”. The output of Augustus and the CLIB 918 gff was provided to Funannotate v1.8.9 (ref DOI:10.5281/zenodo.4054262) for functional annotation. InterProscan was used under Funannotate pipeline locally (Blum et al., 2021). Funannotate then searched in the Pfam database v34.0 and dbCAN database version 10.0 with Hmmer v3.3.2 (Eddy, 2011; Huang et al., 2018; Mistry et al., 2021), in database UniProt version 2021_03 and database MEROPS version 12.0 with diamond blastp v2.0.11 (Rawlings et al., 2018; The UniProt Consortium et al., 2021), eggNOG-mapper v2 on the database eggNOG 5.0 (Cantalapiedra et al., n.d.; Huerta-Cepas et al., 2019).

Cleaned reads were mapped on the reference genomes CLIB 918 and LMA-244 using Bowtie2 v2.4.2 (Langmead and Salzberg, 2012). Maximum fragment length was set to 1000 and the preset “very-sensitive-local” was used.

SAMtools v1.7 (Li et al., 2009) was used to filter out duplicate reads and reads with a mapping quality score above ten for SNP calling and above one for CNV analyses.

In total, we have a dataset of 98 genomes, 88 being sequenced, eight from the University of Laval (LMA strains: Bioproject PRJNA482576, PRJNA482605, PRJNA482610, PRJNA482613, PRJNA482616, PRJNA482619, PRJNA490507, PRJNA490528), one strain CLIB 918 from the Collection de Levures d'Intérêt Biotechnologique (Bioproject PRJEB5752), and one of the strain Phaff72-186 from the 1000 Fungal Genomes project (Bioproject PRJNA334358 NCBI).

SNP calling

Single nucleotide polymorphisms (SNPs) were called using GATK v4.1.2.0 HaplotypeCaller, which provides one gVCF per strain (option -ERC GVCF). GVCFs were combined using GATK CombineGVCFs, genotypes with GATK GenotypeGVCFs, SNPs were selected using GATK SelectVariants (option -select-type SNP). SNPs were filtered using GATK VariantFiltration and options QUAL < 30, DP < 10, QD < 2.0, FS > 60.0, MQ < 40.0, SOR > 3.0, QRankSum < -12.5, ReadPosRankSum < -8.0. All processes from cleaning to variant calling were performed with Snakemake v5.3.0 (script available at https://github.com/BastienBennetot/Article_Geotrichum_2022).

Phylogenetic analysis

We inferred phylogenetic relationships among the 98 isolates using the dataset of 699,755 SNPs in a maximum likelihood framework using IQ-Tree2 v2.1.1 (Minh et al., 2020). The tree has been midpoint rooted. The best-fit model chosen according to Bayesian information criterion (BIC) was TVMe+R2 . Branch supports are ultrafast bootstrap support (1000 bootstrap replicates, Minh et al., 2013).

Genetic structure

We used the dataset of 699,755 SNPs to infer population structure based on the mapping on the CLIB 918 reference genome. We used Splitstree v4.16.2 (Huson and Bryant, 2006) for the neighbor-net analysis. We used the R package *Ade4* (Bougeard and Dray, 2018; Chessel et al., 2004; Dray et al., 2007; Dray and Dufour, 2007; Thioulouse et al., 2018) for principal component analyses (PCA, centered and unscaled). We used NGSadmix v.33 (Jørsboe et al., 2017) from the ANGSD (Korneliussen et al., 2014) package (version 0.933-110-g6921bc6) to infer individual ancestry from genotype likelihoods based on realigned reads, by assuming a given number of populations. A Beagle file was first prepared from bam using ANGSD with the following parameters: “-uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -GL 1 -doMajorMinor 1 -doMaf 1 -doGlf 2 -SNP_pval 1e-6”. The Beagle file was used to run NGSadmix with 4 as the minimum number of informative individuals. Given the high number of strains genetically highly similar among cheese strains (that may represent clonal lineages), we randomly sampled one of the individuals for each group of clonemates identified on the ML tree as having fewer than 90,000 SNPs and filtered out the other strains (N=64 strains kept) to avoid biasing the analysis. The analysis was run for different K values, ranging from 2 to 10. A hundred independent runs were carried out for each number of clusters (K).

The nucleotide diversity π (Nei’s Pi; Hudson et al., 1992; Nei and Li, 1979), the Watterson’s θ (Watterson, 1975), the fixation index F_{ST} (Hudson et al., 1992) and the absolute divergence d_{XY} (Nei and Li, 1979) were calculated using the *popgenome* package in R (Pfeifer et al., 2014). Fixed, private and shared sites were counted using custom scripts available at https://github.com/BastienBennetot/fixed_shared_private_count, with bcftools version 1.11 (using htslib 1.13+ds). F3 tests were computed using the *admixr* package v0.9.1. The pairwise homology index (PHI) test was performed using PhiPack v1.1 and CLIB 918 genome as reference.

Linkage disequilibrium was calculated using vcftools v0.1.17 with the --hap-r2 parameter and a minimum distance between SNPs of 15,000 bp. Values were averaged when SNPs had the same distance.

Pairwise identity between an admixed strain and each non-admixed strain was calculated using overlapping sliding windows of 30 kb span and 5 kb step. Admixed clusters are indicated in Table S1. The custom script is available on https://github.com/BastienBennetot/Article_Geotrichum_2022

Copy number variation and identification of premature stop codons in CDS

Copy number variation (CNV) was analyzed using Control-FREEC v11.6 with the following parameters: ploidy was set to 1, non-overlapping windows of 500 bp, telomeric and centromeric regions were excluded, expected GC content was set between 0.25 and 0.55, minimum of consecutive windows to call a CNV set to 1. This analysis was performed using as references the CLIB 918 (cheese_3) and LMA-244 (wild) genome sequences. CNVs were classified in different groups when the median of copy number was different between populations. We defined three groups: regions for which copy number was different between wild and cheese populations, between mixed-origin and cheese populations and when at least one cheese population differed from another population. For each InterPro term present in these regions, we performed enrichment tests, i.e., a fisher exact test comparing the number of a particular InterPro term found in these regions and the whole genome (Table S9).

We used snpeff (Cingolani et al., 2012) to assess how each SNP affected the coding sequence of predicted proteins, in the vcf file containing all SNPs and all genomes of our dataset. We detected premature stop codons in the 7,150 CDS of the CLIB 918 genome and the 5,576 CDS of the LMA-244 genome using a custom script and bcftools v1.11.

Analyzing the repeat landscape

In order to *de novo* detect repeats within *G. candidum*, RepeatModeler (v2.0.2; Flynn et al., 2020), using the ncbi engine (-engine ncbi) and the option -LTRStruct, was run on the pacbio genome assembly of LMA 244 generating a library of 176 repeats. The repeat redundancy was reduced using cd-hit-est, as described in Goubert et al., giving a final library of 108 repeats (Goubert et al., 2022). To estimate the per strain copy number of each repeat, illumina reads were aligned using bwa mem (v0.7.17; Li, 2013) to the repeat library and the median coverage for each repeat was then normalized by the LMA 244 genome wide median coverage.

Detecting positive selection

The assemblies LMA-317, LMA-77 and LMA-563 have been excluded for this analysis because of a N50 under ten kb. All the 437441 predicted protein sequences from the 66 genomes of all cheese clades and mixed-origin clade were searched against each other with BLASTP using diamond v0.9.36 and clustered into orthologous groups using Orthagog v1.0.3 (Ekseth et al., 2014). For these analyses, we only kept single-copy orthologs shared between two populations. We compared the mixed origin population to each cheese population and the cheese clade. Multiple nucleotide sequence alignments with predicted gene sequences were then constructed using MACSE v2.0.3 with default parameters (Ranwez et al., 2018). We performed an approximative MacDonald Kreitman tests using the R package PopGenome (Pfeifer et al., 2014). The approximation comes from the fact that only codons with a single SNP are examined. The assumption of this version of the test is that the probability that two SNPs will appear in the same codon is very low. To identify genes evolving under positive selection in *G. candidum* genomes, α, i.e. the representation of the proportion of substitutions driven by positive selection was used. Genes with an alpha under 0 were filtered out. Of these genes, only those with a Fisher's test p-value under 0.05 were kept.

Phenotypic characterization

Sampling and strain calibration

We used 36 *Geotrichum candidum* strains for laboratory experiments: seven from the Cheese_1 population, five from the Cheese_2 population, eleven from the Cheese_3 population, eight from the mixed-origin population and five from the wild population (Table S1). This set encompassed 26 strains isolated from dairies, one from other food environments and nine isolated from environments other than food. Experiments were initiated with spore suspensions calibrated to 1.10^6 spores/mL with a hemocytometer.

Media preparation

All media were sterilized in an autoclave at 121°C for 20 minutes except those with cheese or milk for which the autoclave was run at 110°C for 15 minutes to avoid curdling. Each 94mm-diameter Petri dish was filled with 25mL of the appropriate medium. Cheese medium was prepared as follows for 800mL: 300g of unsalted cream cheese from La Doudou farm in Cheptainville, 16g agar, 8g NaCl dissolved in 200mL of deionized water. Deionized water was added to reach 800mL. pH was adjusted to 6.5 and drops of blue food dyes were added to enable fungal colony measures (white medium and white colonies are not distinguable). Yeast Peptone Dextrose (YPD) medium was prepared as follows for 1L: 10g Yeast extract, 10g Bacto Peptone, 10g glucose, 14g agar powder. Minimal medium was prepared as described in “Improved protocols for *Aspergillus* minimal medium: trace element and minimal medium salt stock solutions”, Terry W. Hill, Rhodes College, Etta Kafer, Simon Fraser University. Tributyrin agar was prepared as follows: Tributyrin medium 33 g/L, neutral Tributyrin 10 g/L, Bacto Agar 15 g/L. Ingredients were bought at Nutri-Bact company, Québec, Canada. Caseinate agar was prepared according to Frazier and Rupp, modified as follows: Calcium caseinate medium 37.2 g/L, Bacto Agar 15 g/L. Ingredients were bought at Nutri-Bact company, Québec, Canada. For yogurt media we used three different types of raw milk, i.e. sheep, goat and cow milks, coming from d’Armenon farm near Les Molières (Esonne, France), Noue farm in Celle les Bordes and Coubertin farm in Saint-Rémy-lès-Chevreuse respectively. Each medium was prepared following the same procedure: 1L of milk was mixed with 62.5g of Danone brand yogurt, heated for 5 hour at 43°C and stored in a fridge before use. A subset of 300g of this preparation was used with 16g of agar powder, 8g of NaCl, 4 drops of blue food dye and filled up with deionized water to reach 800mL.

Growth in different conditions and different media

Petri dishes were inoculated with 10µL of the 1.10^6 cells/mL in a 10% glycerol solution. Inoculated Petri dishes were wrapped with plastic film before letting them grow in the dark. A millimeter rule was used to measure two opposite diameters of fungal colonies to estimate their growth. Means of these two measures were used for statistical analyses.

To test media and temperature effect on growth, *G. candidum* strains were grown on minimal, YPD and cheese media. We took pictures and measured their growth at seven, 11 and 14 days for minimal, YPD and cheese media at 10°C (ripening cellar temperature), at seven and 11 days for the cheese medium at 15°C and at seven days for minimal, YPD and cheese media at 25°C (Figure S8).

To test salt tolerance, *G. candidum* strains were grown at 10°C on cheese media of different salt concentrations: unsalted media, 1% salt as St Nectaire and cream cheeses, 2% as Camembert and goat cheeses and 4% as Roquefort blue cheeses. We took pictures and measured colony diameters after 14 days of growth.

To test adaptation of *G. candidum* populations to different milk origins, growth was measured on different yogurt media made from goat, sheep and cow raw milk for seven days at 25°C.

To test lipolytic and proteolytic activities of *G. candidum* populations, we grew strains on tributyrin agar and caseinate agar, respectively. Each strain was inoculated in triplicate Petri dishes that were let grown at 25°C for 14 days. The radius of lysis was measured and the mean between triplicates was used for the analysis.

Pictures were taken using a Scan 1200 (Interscience). Petri dishes grown on cheese were analyzed using IRIS (Kritikos et al., 2017) which measured Integral opacity scores, defined as the sum of the brightness values for all the pixels within the colony bounds.

Volatile compounds analysis using Gas-chromatography mass-spectrometry (GC-MS)

Volatile compounds produced by *G. candidum* were analyzed using gas-chromatography mass-spectrometry (GC-MS). Compounds were extracted and concentrated by using a dynamic headspace (DHS) combined with a thermal desorption unit (TDU). Strains were grown for 21 days at 10°C (minimum Camembert ripening time) on a cheese agar medium made with Camembert-type curds. After 21 days, each Petri dish content, with its medium and *G. candidum* mycelium, was mixed with a fork for one minute, gathered in vials and immediately frozen in liquid nitrogen. For each sample, three grams of frozen cultured media were weighted and stored in vials with septum caps at -80°C. Sixteen hours before analysis, samples were stored at 4°C. The Cheese_2 population was not tested in this experiment because population delineation was not known at this time.

Dynamic headspace (DHS) conditions were as follows: Inert gas: He; Incubation: 30°C for 3min; Needle temperature: 120°C; Trap: nature tenax, 30°C, 450 mL He;

He flow: 30 mL/min; Dry purge : temperature 30°C, 850 mL He, He flow 50 mL/ min. Thermal Desorption Unit (TDU) conditions were as follows: inert gas : He; Initial temperature: 30°C, then 60°C/min until 290°C kept for 7 minutes; Transfer temperature: 300°C. Cool Injection System (CIS) conditions were as follows: inert gas : He; Initial temperature: -100°C, then 12°C/s until 270°C kept for 5 minutes. Gas chromatograph (brand Agilent 7890B) was used with a polyethylene glycol (PEG) type polar phase column (HP-Innowax, ref. Agilent 19091N-116I, 60mx0.32mm, 0.25µm film thickness). Helium flow was set at 1.6mL/min. Samples were injected in splitless mode with a holding time of 1 minute. To optimize separation of compounds, a specific program of the gas chromatography oven was used, with initial temperature at 40°C for 5 minutes, rising temperature from 40°C to 155°C with a slope of 4°C/min, rising temperature from 155°C to 250°C with a slope of 20°C/min and then temperature was kept at 250°C for 5 minutes. A single quadrupole mass spectrometer was used to determine m/z of sample molecules (Agilent, référence 5977B MSD). Molecules were identified using NIST libraries (NIST 2017 Mass Spectral Library).

Competition experiments

To test the abilities of *G. candidum* populations to exclude other fungi by secreting molecules or volatile compounds, we compared the growth of competitors when grown alone and on a lawn of an already grown *G. candidum* mycelium. We inoculated a cheese medium with 150µL of a *G. candidum* calibrated spore solution (1.10^6 spores/mL), spread evenly on the Petri dish. After 24h of growth, we inoculated 10µL of a competitor spore solution (1.10^6 spores/mL) in a single spot, in the middle of the Petri dish. We used as competitors the following species and strains: *Penicillium biforme* (ESE00018, ESE00023, ESE00125, ESE00222), *Penicillium roqueforti* (ESE00645, ESE00925, LCP06040), *Scopulariopsis asperula* (ESE00044, ESE00102, ESE00835, ESE01287, ESE01324) and *Debaryomyces hansenii* (ESE00284, ESE00561, ESE00576; Table S15). For each competitor, two Petri dishes were inoculated without any *G. candidum* as controls for measuring growth without a lawn.

We took pictures of the Petri dishes at 6 days, when the competitor mycelium grown alone was near the Petri dishes border; we measured colony size at 7 days for *P. biforme* and *P. roqueforti* and at 19 days for *D. hansenii*, which grows more slowly.

To test the abilities of *G. candidum* populations to exclude other microorganisms by producing volatile compounds, we set up an experiment with splitted Petri dishes where only air can be shared between the two parts. In one part of the Petri dish, we spread 75µL of a *G. candidum* spore solution (1.10^6 spores/mL) and let it grow during 24 hours before adding on the other part of the Petri dish a drop of 5µL of a competitor spore solution (1.10^6 spores/mL). For each competitor, two Petri dishes were inoculated without any *G. candidum* as controls. We used as competitors the following species and strains: *Penicillium biforme* (ESE00018, ESE00023, ESE00125, ESE00222, ESE00423)*, Penicillium roqueforti* (ESE00250, ESE00631, ESE00640, ESE00925) and *Scopulariopsis asperula*(ESE00044, ESE00102, ESE00835, ESE01287, ESE01324; Table S15). Petri dishes were grown at 10°C, measured and pictured at 11 days for *P. biforme* and *P. roqueforti* and 19 days for *Scopulariopsis asperula*.

Graphics and statistical analyses

Plots and statistical analyses were made using *ggplot2* (Wickham, 2016), *rstatix* and *ggpubr* packages in the R environment. For ANOVAs, we used standard linear models in which all explanatory variables were discrete, with explained variables being radial growth for growth conditions (for media, temperature, salt content and adaptation to milk experiments), integral opacity score (for opacity experiment), relative proportions of volatiles compounds (for volatile compounds experiment) and radial growth of the competitor (for competition experiments). The explanatory variable common for all analyses was the ‘population’ of *G. candidum*. The variables ‘medium’, ‘day’ and ‘temperature’ were explanatory variables specific to the growth analysis. The ‘competitor species’ variable was specific to competition analyses. All variables and all interactions between them were implemented in the ANOVA and non-significant interactions were subsequently removed before performing post-ANOVA Tukey’s honest significant difference (HSD) tests. The data normality of residuals was checked; when residues deviated from normality (only for the opacity experiment), we also ran non-parametric tests (Wilcoxon ranking tests) using R. Radius of lysis for lipolytic and proteolytic activities experiments was often discrete, strains either showing lytic activity or not at all. This is why we decided to transform these data into qualitative discrete data in order to fit a generalized linear model with a binomial function as logit. Growth time (7, 14 and 21 days) and temperature (15 and 25°C) were taken as random variables because no fit could be achieved with little data and we wanted to test for population effect. Tukey contrasts were used to compare population means of populations when population effect was significant.

Acknowledgments

We thank everyone who sent cheese crusts and Riwanon Lemee for providing 15 strains from Laboratoires STANDA, Caen, France. We thank Laura Prugneau, Jérémy Raynaud, Thomas Mari, Véronique Bougie and Jules Larouche for their help in the laboratory experiments. This work was funded by the Artifice ANR-19-CE20-0006-01 ANR grant to J.R. In addition, a NSERC-Discovery Grant supported the scholarship of V.P and data generation, and is held by S. Labrie. (RGPIN-2017-06388).

Author contributions

J.R. designed and supervised the study, and obtained funding. B.B., V.P., J.R. and A.S. generated the data. C.G. provided strains from the CIRM-Levures INRAE collection. B.B., J.-P.V., R.C.d.I.V. and S.O. analyzed the genomes. B.B., S.H., J.R. and A.S. performed the experiments. B.B. analyzed the data from laboratory experiments. M.H.L. and St.L. supervised the lipolysis and proteolysis analyses. B.B., So.L. and A.-C.P. performed the volatile compound experiment. T.G. contributed to interpretation and writing; B.B. and J.R. wrote the manuscript, with contributions from all the authors.

References

https://doi.org/10.1016/j.cub.2015.08.025

Adams, D.J., 2004. Fungal cell wall chitinases and glucanases. Microbiology 150, 2029–2035. https://doi.org/10.1099/mic.0.26980-0

Alberto, F.J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., Barbato, M., Zamani, W., Alberti, A., Engelen, S., Stella, A., Joost, S., Ajmone-Marsan, P., Negrini, R., Orlando, L., Rezaei, H.R., Naderi, S., Clarke, L., Flicek, P., Wincker, P., Coissac, E., Kijas, J., Tosser-Klopp, G., Chikhi, A., Bruford, M.W., Taberlet, P., Pompanon, F., 2018. Convergent genomic signatures of domestication in sheep and goats. Nat. Commun. 9, 813. https://doi.org/10.1038/s41467-018-03206-y

Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., Albertin, W., Durrens, P., Sherman, D.J., Marullo, P., Todd Hittinger, C., Gonçalves, P., Sampaio, J.P., 2014. A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum. Nat. Commun. 5, 4044. https://doi.org/10.1038/ncomms5044

Alper, I., Frenette, M., Labrie, S., 2013. Genetic diversity of dairy Geotrichum candidum strains revealed by multilocus sequence typing. Appl. Microbiol. Biotechnol. 97, 5907–5920. https://doi.org/10.1007/s00253-013-4776-2

Antipov, D., Korobeynikov, A., McLean, J.S., Pevzner, P.A., 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32, 1009–1015. https://doi.org/10.1093/bioinformatics/btv688

Baduel, P., Quadrana, L., Hunter, B., Bomblies, K., Colot, V., 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat. Commun. 10, 5818. https://doi.org/10.1038/s41467-019-13730-0

Bai, F.-Y., Han, D.-Y., Duan, S.-F., Wang, Q.-M., 2022. The Ecology and Evolution of the Baker's Yeast Saccharomyces cerevisiae. Genes 13, 230. https://doi.org/10.3390/genes13020230

Barros Lopes, M., Bellon, J.R., Shirley, N.J., Ganter, P.F., 2002. Evidence for multiple interspecific hybridization in *Saccharomyces* sensu stricto species. FEMS Yeast Res. 1, 323–331. https://doi.org/10.1111/j.1567-1364.2002.tb00051.x

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580. https://doi.org/10.1093/nar/27.2.573

Bigey, F., Segond, D., Friedrich, A., Guezenec, S., Bourgais, A., Huyghe, L., Agier, N., Nidelet, T., Sicard, D., 2021. Evidence for Two Main Domestication Trajectories in Saccharomyces cerevisiae Linked to Distinct Bread-Making Processes. Curr. Biol. 31, 722-732.e5. https://doi.org/10.1016/j.cub.2020.11.016

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., Finn, R.D., 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–D354. https://doi.org/10.1093/nar/gkaa977

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonnarme, P., Arfi, K., Dury, C., Helinck, S., Yvon, M., Spinnler, H.-E., 2001. Sulfur compound production by Geotrichum candidum from l-methionine: importance of the transamination step. FEMS Microbiol. Lett. 205, 247–252. https://doi.org/10.1111/j.1574-6968.2001.tb10956.x

Borneman, A.R., Forgan, A.H., Kolouchova, R., Fraser, J.A., Schmidt, S.A., 2016. Whole Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy Across the Spectrum of Commercial Wine Strains of *Saccharomyces cerevisiae*. G3 GenesGenomesGenetics 6, 957–971. https://doi.org/10.1534/g3.115.025692

Bougeard, S., Dray, S., 2018. Supervised Multiblock Analysis in R with the ade4 Package. J. Stat. Softw. 86, 1–17. https://doi.org/10.18637/jss.v086.i01

Boutrou, R., Gueguen, M., 2005. Interests in Geotrichum candidum for cheese technology. Int. J. Food Microbiol. 102, 1–20. https://doi.org/10.1016/j.ijfoodmicro.2004.12.028

Bruen, T.C., Philippe, H., Bryant, D., 2006. A Simple and Robust Statistical Test for Detecting the Presence of Recombination. Genetics 172, 2665–2681. https://doi.org/10.1534/genetics.105.048975

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., n.d. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale 7.

Caron, T., Piver, M.L., Péron, A.-C., Lieben, P., Lavigne, R., Brunel, S., Roueyre, D., Place, M., Bonnarme, P., Giraud, T., Branca, A., Landaud, S., Chassard, C., 2021. Strong effect of Penicillium roqueforti populations on volatile and metabolic compounds responsible for aromas, flavor and texture in blue cheeses. Int. J. Food Microbiol. 354, 109174. https://doi.org/10.1016/j.ijfoodmicro.2021.109174

Carreira, A., Dillinger, K., Eliskases-Lechner, F., Loureiro, V., Ginzinger, W., Rohm, H., 2002. Influence of selected factors on browning of Camembert cheese. J. Dairy Res. 69, 281–292. https://doi.org/10.1017/S0022029902005393

Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A.-L., Ceppi, M., Conseiller, E., Debuchy, R., Malagnac, F., Goarin, A., Silar, P., Lacoste, S., Sallet, E., Bensimon, A., Giraud, T., Brygoo, Y., 2014. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat. Commun. 5, 1–9. https://doi.org/10.1038/ncomms3876

Chessel, D., Dufour, A.-B., Thioulouse, J., 2004. The ade4 Package – I: One-Table Methods. R News 4, 5–10.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118}; iso-2; iso-3. Fly (Austin) 6, 80–92. https://doi.org/10.4161/fly.19695

Collins, Y.F., McSweeney, P.L.H., Wilkinson, M.G., 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. Int. Dairy J. 13, 841–866. https://doi.org/10.1016/S0958-6946(03)00109-2

Cornille, A., Giraud, T., Smulders, M.J.M., Roldán-Ruiz, I., Gladieux, P., 2014. The domestication and evolutionary ecology of apples. Trends Genet. 30, 57–65. https://doi.org/10.1016/j.tig.2013.10.002

Cresko, W.A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M.A., Kimmel, C.B., Postlethwait, J.H., 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc. Natl. Acad. Sci. 101, 6050–6055. https://doi.org/10.1073/pnas.0308479101

Crusoe, M.R., Alameldin, H.F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., Charbonneau, A., Constantinides, B., Edvenson, G., Fay, S., Fenton, J., Fenzl, T., Fish, J., Garcia-Gutierrez, L., Garland, P., Gluck, J., González, I., Guermond, S., Guo, J., Gupta, A., Herr, J.R., Howe, A., Hyer, A., Härpfer, A., Irber, L., Kidd, R., Lin, D., Lippi, J., Mansour, T., McA'Nulty, P., McDonald, E., Mizzi, J., Murray, K.D., Nahum, J.R., Nanlohy, K., Nederbragt, A.J., Ortiz-Zuazaga, H., Ory, J., Pell, J., Pepe-Ranney, C., Russ, Z.N., Schwarz, E., Scott, C., Seaman, J., Sievert, S., Simpson, J., Skennerton, C.T., Spencer, J., Srinivasan, R., Standage, D., Stapleton, J.A., Steinman, S.R., Stein, J., Taylor, B., Trimble, W., Wiencko, H.L., Wright, M., Wyss, B., Zhang, Q., zyme, en, Brown, C.T., 2015. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Research 4, 900. https://doi.org/10.12688/f1000research.6924.1

Curioni, P.M.G., Bosset, J.O., 2002. Key odorants in various cheese types as determined by gas chromatography-olfactometry. Int. Dairy J. 12, 959–984. https://doi.org/10.1016/S0958-6946(02)00124-3

Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. Nature 418, 700–707. https://doi.org/10.1038/nature01019

Dieuleveux, V., Van Der Pyl, D., Chataud, J., Gueguen, M., 1998. Purification and characterization of anti-Listeria compounds produced by Geotrichum candidum. Appl. Environ. Microbiol. 64, 800–803. https://doi.org/10.1128/AEM.64.2.800-803.1998

Dray, S., Dufour, A.-B., 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. J. Stat. Softw. 22, 1–20. https://doi.org/10.18637/jss.v022.i04

Dray, S., Dufour, A.-B., Chessel, D., 2007. The ade4 Package – II: Two-Table and K-Table Methods. R News 7, 47–52.

Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., Bai, F.-Y., 2018. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nat. Commun. 9, 2690. https://doi.org/10.1038/s41467-018-05106-7

Dugat-Bony, E., Straub, C., Teissandier, A., Onésime, D., Loux, V., Monnet, C., Irlinger, F., Landaud, S., Leclercq-Perlat, M.-N., Bento, P., Fraud, S., Gibrat, J.-F., Aubert, J., Fer, F., Guédon, E., Pons, N., Kennedy, S., Beckerich, J.-M., Swennen, D., Bonnarme, P., 2015. Overview of a Surface-Ripened Cheese Community Functioning by Meta-Omics Analyses. PLOS ONE 10, e0124360. https://doi.org/10.1371/journal.pone.0124360

Dumas, E., Feurtey, A., Rodríguez de la Vega, R.C., Le Prieur, S., Snirc, A., Coton, M., Thierry, A., Coton, E., Le Piver, M., Roueyre, D., Ropars, J., Branca, A., Giraud, T., 2020. Independent domestication events in the blue-cheese fungus *Penicillium roqueforti*. Mol. Ecol. 29, 2639–2660. https://doi.org/10.1111/mec.15359

Dyer, A.G., Boyd-Gerny, S., McLoughlin, S., Rosa, M.G.P., Simonov, V., Wong, B.B.M., 2012. Parallel evolution of angiosperm colour signals: common evolutionary pressures linked to hymenopteran vision. Proc. R. Soc. B Biol. Sci. 279, 3606–3615. https://doi.org/10.1098/rspb.2012.0827

Eddy, S.R., 2011. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7, e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Ekseth, O.K., Kuiper, M., Mironov, V., 2014. orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics 30, 734–736. https://doi.org/10.1093/bioinformatics/btt582

Elmer, K.R., Fan, S., Kusche, H., Spreitzer, M.L., Kautt, A.F., Franchini, P., Meyer, A., 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat. Commun. 5, 1–8. https://doi.org/10.1038/ncomms6168

Elmer, K.R., Kusche, H., Lehtonen, T.K., Meyer, A., 2010. Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. Philos. Trans. R. Soc. B Biol. Sci. 365, 1763–1782. https://doi.org/10.1098/rstb.2009.0271

Fay, J.C., Benavides, J.A., 2005. Evidence for Domesticated and Wild Populations of Saccharomyces cerevisiae. PLoS Genet. 1, e5. https://doi.org/10.1371/journal.pgen.0010005

Felsenstein, J., 1974. The evolutionary advantage of recombination. Genetics 78, 737–756. https://doi.org/10.1093/genetics/78.2.737

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Frantz, L.A.F., Schraiber, J.G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., Paudel, Y., Crooijmans, R.P.M.A., Larson, G., Groenen, M.A.M., 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat. Genet. 47, 1141–1148. https://doi.org/10.1038/ng.3394

Galagan, J.E., Calvo, S.E., Cuomo, C., Ma, L.-J., Wortman, J.R., Batzoglou, S., Lee, S.-I., Baştürkmen, M., Spevak, C.C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G.H., Draht, O., Busch, S., D'Enfert, C., Bouchier, C., Goldman, G.H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J.H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E.U., Archer, D.B., Peñalva, M.Á., Oakley, B.R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W.C., Denning, D.W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M.S., Osmani, S.A., Birren, B.W., 2005. Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature 438, 1105–1115. https://doi.org/10.1038/nature04341

Gallone, B., Mertens, S., Gordon, J.L., Maere, S., Verstrepen, K.J., Steensels, J., 2018. Origins, evolution, domestication and diversity of Saccharomyces beer yeasts. Curr. Opin. Biotechnol. 49, 148–155. https://doi.org/10.1016/j.copbio.2017.08.005

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., Verstrepen, K.J., 2016. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. Cell 166, 1397-1410.e16. https://doi.org/10.1016/j.cell.2016.08.020

Gao, M., Glenn, A.E., Blacutt, A.A., Gold, S.E., 2017. Fungal Lactamases: Their Occurrence and Function. Front. Microbiol. 8.

Giannakou, K., Cotterrell, M., Delneri, D., 2020. Genomic Adaptation of Saccharomyces Species to Industrial Environments. Front. Genet. 11.

Gibbons, J.G., Salichos, L., Slot, J.C., Rinker, D.C., McGary, K.L., King, J.G., Klich, M.A., Tabb, D.L., McDonald, W.H., Rokas, A., 2012. The Evolutionary Imprint of Domestication on Genome Variation and Function of the Filamentous Fungus Aspergillus oryzae. Curr. Biol. 22, 1403–1409. https://doi.org/10.1016/j.cub.2012.05.033

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. https://doi.org/10.1016/j.ijfoodmicro.2016.10.001

Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguileta, G., Vienne, D.M., Rodríguez de la Vega, R.C., Branco, S., Giraud, T., 2014. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. Mol. Ecol. 23, 753–773. https://doi.org/10.1111/mec.12631

Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., Sampaio, J.P., 2016. Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts. Curr. Biol. CB 26, 2750–2761. https://doi.org/10.1016/j.cub.2016.08.040

Goubert, C., Craig, R.J., Bilat, A.F., Peona, V., Vogan, A.A., Protasio, A.V., 2022. A beginner’s guide to manual curation of transposable elements. Mob. DNA 13, 7. https://doi.org/10.1186/s13100-021-00259-7

Harlan, J.R., Gepts, P., Famula, T.R., Bettinger, R.L., Brush, S.B., Damania, A.B., McGuire, P.E., Qualset, C.O., 2012. Biodiversity in agriculture: domestication, evolution, and sustainability. Cambridge University Press.

Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., Yi, H., Yang, Z., Yin, Y., 2018. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res. 46, D516–D521. https://doi.org/10.1093/nar/gkx894

Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589. https://doi.org/10.1093/genetics/132.2.583

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314. https://doi.org/10.1093/nar/gky1085

Hufford, M.B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppler, S.M., Lai, J., Morrell, P.L., Shannon, L.M., Song, C., Springer, N.M., Swanson-Wagner, R.A., Tiffin, P., Wang, J., Zhang, G., Doebley, J., McMullen, M.D., Ware, D., Buckler, E.S., Yang, S., Ross-Ibarra, J., 2012. Comparative population genomics of maize domestication and improvement. Nat. Genet. 44, 808–811. https://doi.org/10.1038/ng.2309

Huson, D.H., Bryant, D., 2006. Application of Phylogenetic Networks in Evolutionary Studies. Mol. Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030

Jacques, N., Mallet, S., Laaghouiti, F., Tinsley, C.R., Casaregola, S., 2017. Specific populations of the yeast Geotrichum candidum revealed by molecular typing.

Yeast 34, 165–178. https://doi.org/10.1002/yea.3223

Jørsboe, E., Hanghøj, K., Albrechtsen, A., 2017. fastNGSadmix: admixture proportions and principal component analysis of a single NGS sample. Bioinformatics 33, 3148–3150. https://doi.org/10.1093/bioinformatics/btx474

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. Genome Res. 27, 722–736. https://doi.org/10.1101/gr.215087.116

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 356. https://doi.org/10.1186/s12859-014-0356-4

Kritikos, G., Banzhaf, M., Herrera-Dominguez, L., Koumoutsi, A., Wartel, M., Zietek, M., Typas, A., 2017. A tool named Iris for versatile high-throughput phenotyping in microorganisms 24.

KUMURA, H., TAKAGAKI, K., SONE, T., TSUKAHARA, M., TANAKA, T., SHIMAZAKI, K., 2002. Casein Digestion by Debaryomyces hansenii Isolated from Cheese. Biosci. Biotechnol. Biochem. 66, 1370–1373. https://doi.org/10.1271/bbb.66.1370

Lahue, C., Madden, A., Dunn, R., Smukowski Heil, C., 2020. History and Domestication of Saccharomyces cerevisiae in Bread Baking. Front. Genet. 11.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldon, T., Schuller, D., Sampaio, J.P., Dequin, S., 2018. Adaptation of S. cerevisiae to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. Mol. Biol. Evol. 35, 1712–1727. https://doi.org/10.1093/molbev/msy066

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (No. arXiv:1303.3997). arXiv. https://doi.org/10.48550/arXiv.1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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

Lin, Z., Li, X., Shannon, L.M., Yeh, C.T., Wang, M.L., Bai, G., Peng, Z., Li, J., Trick, H.N., Clemente, T.E., Doebley, J., Schnable, P.S., Tuinstra, M.R., Tesso, T.T., White, F., Yu, J., 2012. Parallel domestication of the Shattering1 genes in cereals. Nat. Genet. 44, 720–724. https://doi.org/10.1038/ng.2281

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., Tsai, I.J., Bergman, C.M., Bensasson, D., O'Kelly, M.J.T., van Oudenaarden, A., Barton, D.B.H., Bailes, E., Nguyen, A.N., Jones, M., Quail, M.A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., Louis, E.J., 2009. Population genomics of domestic and wild yeasts. Nature 458, 337–341. https://doi.org/10.1038/nature07743

Liu, S.-Q., Holland, R., Crow, V.L., 2004. Esters and their biosynthesis in fermented dairy products: a review. Int. Dairy J. 14, 923–945. https://doi.org/10.1016/j.idairyj.2004.02.010

Liu, X., Zhang, Y., Liu, W., Li, Y., Pan, J., Pu, Y., Han, J., Orlando, L., Ma, Y., Jiang, L., 2022. A single-nucleotide mutation within the TBX3 enhancer increased body size in Chinese horses. Curr. Biol. 32, 480-487.e6. https://doi.org/10.1016/j.cub.2021.11.052

Lo, Y.-C., Bruxaux, J., Vega, R.C.R. de la, Snirc, A., Coton, M., Piver, M.L., Prieur, S.L., Roueyre, D., Dupont, J., Houbraken, J., Debuchy, R., Ropars, J., Giraud, T., Branca, A., 2022. Domestication in dry-cured meat Penicillium fungi: convergent specific phenotypes and horizontal gene transfers without strong genetic subdivision. https://doi.org/10.1101/2022.03.25.485132

Mabry, M.E., Turner-Hissong, S.D., Gallagher, E.Y., McAlvay, A.C., An, H., Edger, P.P., Moore, J.D., Pink, D.A.C., Teakle, G.R., Stevens, C.J., Barker, G., Labate, J., Fuller, D.Q., Allaby, R.G., Beissinger, T., Decker, J.E., Gore, M.A., Pires, J.C., 2021. The Evolutionary History of Wild, Domesticated, and Feral Brassica oleracea (Brassicaceae). Mol. Biol. Evol. 38, 4419–4434. https://doi.org/10.1093/molbev/msab183

Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusumoto, K.-I., Arima, T., Akita, O., Kashiwagi, Y., Abe, K., Gomi, K., Horiuchi, H., Kitamoto, K., Kobayashi, T., Takeuchi, M., Denning, D.W., Galagan, J.E., Nierman, W.C., Yu, J., Archer, D.B., Bennett, J.W., Bhatnagar, D., Cleveland, T.E., Fedorova, N.D., Gotoh, O., Horikawa, H., Hosoyama, A., Ichinomiya, M., Igarashi, R., Iwashita, K., Juvvadi, P.R., Kato, M., Kato, Y., Kin, T., Kokubun, A., Maeda, H., Maeyama, N., Maruyama, J., Nagasaki, H., Nakajima, T., Oda, K., Okada, K., Paulsen, I., Sakamoto, K., Sawano, T., Takahashi, M., Takase, K., Terabayashi, Y., Wortman, J.R., Yamada, O., Yamagata, Y., Anazawa, H., Hata, Y., Koide, Y., Komori, T., Koyama, Y., Minetoki, T., Suharnan, S., Tanaka, A., Isono, K., Kuhara, S., Ogasawara, N., Kikuchi, H., 2005. Genome sequencing and analysis of Aspergillus oryzae. Nature 438, 1157–1161. https://doi.org/10.1038/nature04300

Macías, L.G., Flores, M.G., Adam, A.C., Rodríguez, M.E., Querol, A., Barrio, E., Lopes, C.A., Pérez-Torrado, R., 2021. Convergent adaptation of Saccharomyces uvarum to sulfite, an antimicrobial preservative widely used in human-driven fermentations. PLOS Genet. 17, e1009872. https://doi.org/10.1371/journal.pgen.1009872

Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K., Lohmueller, K.E., 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc. Natl. Acad. Sci. 113, 152–157. https://doi.org/10.1073/pnas.1512501113

Mayo, B., Rodríguez, J., Vázquez, L., Flórez, A.B., 2021. Microbial Interactions within the Cheese Ecosystem and Their Application to Improve Quality and Safety. Foods 10, 602. https://doi.org/10.3390/foods10030602

McSweeney, P.L.H., 2004. Biochemistry of cheese ripening. Int. J. Dairy Technol. 57, 127–144. https://doi.org/10.1111/j.1471-0307.2004.00147.x

Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast Approximation for Phylogenetic Bootstrap. Mol. Biol. Evol. 30, 1188–1195. https://doi.org/10.1093/molbev/mst024

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von

Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: The protein families database in 2021. Nucleic Acids Res. 49, D412–D419. https://doi.org/10.1093/nar/gkaa913

Monnet, C., Back, A., Irlinger, F., 2012. Growth of Aerobic Ripening Bacteria at the Cheese Surface Is Limited by the Availability of Iron. Appl. Environ. Microbiol. 78, 3185–3192. https://doi.org/10.1128/AEM.00085-12

Monnet, C., Landaud, S., Bonnarme, P., Swennen, D., 2015. Growth and adaptation of microorganisms on the cheese surface. FEMS Microbiol. Lett. 362, 1–9. https://doi.org/10.1093/femsle/fnu025

Morel, G., Sterck, L., Swennen, D., Marcet-Houben, M., Onesime, D., Levasseur, A., Jacques, N., Mallet, S., Couloux, A., Labadie, K., Amselem, J., Beckerich, J.-M., Henrissat, B., Van de Peer, Y., Wincker, P., Souciet, J.-L., Gabaldón, T., Tinsley, C.R., Casaregola, S., 2015. Differential gene retention as an evolutionary mechanism to generate biodiversity and adaptation in yeasts. Sci. Rep. 5, 11571. https://doi.org/10.1038/srep11571

Naumova, E.S., Naumov, G.I., Masneuf-Pomarède, I., Aigle, M., Dubourdieu, D., 2005. Molecular genetic study of introgression between Saccharomyces bayanus and S. cerevisiae. Yeast 22, 1099–1115. https://doi.org/10.1002/yea.1298

Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 76, 5269–5273.

Nielsen, M.S., Frisvad, J.C., Nielsen, P.V., 1998. Colony interaction and secondary metabolite production of cheese-related fungi in dual culture. J. Food Prot. 61, 1023–1029. https://doi.org/10.4315/0362-028X-61.8.1023

Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.-L., Wincker, P., Casaregola, S., Dequin, 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. https://doi.org/10.1073/pnas.0904673106

Omeike, S.O., Kareem, S.O., Nandanwar, H., Lasisi, A.A., Oluwafemi, F., Jangra, M., 2021. Purification, De Novo Characterization and Antibacterial Properties of a Novel, Narrow-Spectrum Bacteriostatic Tripeptide from Geotrichum candidum OMON-1. Arab. J. Sci. Eng. 46, 5275–5283. https://doi.org/10.1007/s13369-020-05024-1

O'Quin, K.E., Hofmann, C.M., Hofmann, H.A., Carleton, K.L., 2010. Parallel Evolution of Opsin Gene Expression in African Cichlid Fishes. Mol. Biol. Evol. 27, 2839–2854. https://doi.org/10.1093/molbev/msq171

Peng, J.H., Sun, D., Nevo, E., 2011. Domestication evolution, genetics and genomics in wheat. Mol. Breed. 28, 281. https://doi.org/10.1007/s11032-011-9608-4

Penland, M., Falentin, H., Parayre, S., Pawtowski, A., Maillard, M.-B., Thierry, A., Mounier, J., Coton, M., Deutsch, S.-M., 2021. Linking Pélardon artisanal goat cheese microbial communities to aroma compounds during cheese-making and ripening. Int. J. Food Microbiol. 345, 109130. https://doi.org/10.1016/j.ijfoodmicro.2021.109130

Perkins, V., Vignola, S., Lessard, M.-H., Plante, P.-L., Corbeil, J., Dugat-Bony, E., Frenette, M., Labrie, S., 2020. Phenotypic and Genetic Characterization of the Cheese Ripening Yeast Geotrichum candidum. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.00737

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Petersen, J.L., Mickelson, J.R., Rendahl, A.K., Valberg, S.J., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., Brama, P., da Câmara Machado, A., Capomaccio, S., Cappelli, K., Cothran, E.G., Distl, O., Fox-Clipsham, L., Graves, K.T., Guérin, G., Haase, B., Hasegawa, T., Hemmann, K., Hill, E.W., Leeb, T., Lindgren, G., Lohi, H., Lopes, M.S., McGivney, B.A., Mikko, S., Orr, N., Penedo, M.C.T., Piercy, R.J., Raekallio, M., Rieder, S., Røed, K.H., Swinburne, J., Tozaki, T., Vaudin, M., Wade, C.M., McCue, M.E., 2013. Genome-Wide Analysis Reveals Selection for Important Traits in Domestic Horse Breeds. PLoS Genet. 9, e1003211. https://doi.org/10.1371/journal.pgen.1003211

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S.E., Lercher, M.J., 2014. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. Mol. Biol. Evol. 31, 1929–1936. https://doi.org/10.1093/molbev/msu136

Pitt, J.I., Cruickshank, R.H., Leistner, L., 1986. Penicillium commune, P. camembertii, the origin of white cheese moulds, and the production of cyclopiazonic acid. Food Microbiol. 3, 363–371. https://doi.org/10.1016/0740-0020(86)90022-5

Plassais, J., vonHoldt, B.M., Parker, H.G., Carmagnini, A., Dubos, N., Papa, I., Bevant, K., Derrien, T., Hennelly, L.M., Whitaker, D.T., Harris, A.C., Hogan, A.N., Huson, H.J., Zaibert, V.F., Linderholm, A., Haile, J., Fest, T., Habib, B., Sacks, B.N., Benecke, N., Outram, A.K., Sablin, M.V., Germonpré, M., Larson, G., Frantz, L., Ostrander, E.A., 2022. Natural and human-driven selection of a single non-coding body size variant in ancient and modern canids. Curr. Biol. 32, 889-897.e9. https://doi.org/10.1016/j.cub.2021.12.036

Price, E.O. (Ed.), 2002. Animal domestication and behavior. CABI, Wallingford. https://doi.org/10.1079/9780851995977.0000

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., Korobeynikov, A., 2020. Using SPAdes De Novo Assembler. Curr. Protoc. Bioinforma. 70, e102. https://doi.org/10.1002/cpbi.102

Purugganan, M.D., 2019. Evolutionary Insights into the Nature of Plant Domestication. Curr. Biol. 29, R705–R714. https://doi.org/10.1016/j.cub.2019.05.053

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T.M., Fries, R., Nielsen, R., Simianer, H., 2014. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. PLOS Genet. 10, e1004148. https://doi.org/10.1371/journal.pgen.1004148

Rainieri, S., Kodama, Y., Kaneko, Y., Mikata, K., Nakao, Y., Ashikari, T., 2006. Pure and Mixed Genetic Lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and Their Contribution to the Lager Brewing Strain Genome. Appl. Environ. Microbiol. 72, 3968–3974. https://doi.org/10.1128/AEM.02769-05

Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., Delsuc, F., 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. Mol. Biol. Evol. 35, 2582–2584. https://doi.org/10.1093/molbev/msy159

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., Finn, R.D., 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 46, D624–D632. https://doi.org/10.1093/nar/gkx1134

Ropars, J., Caron, T., Lo, Y.-C., Bennetot, B., Giraud, T., 2020a. La domestication des champignons Penicillium du fromage. C. R. Biol. 343, 155–176. https://doi.org/10.5802/crbiol.15

Ropars, J., de la Vega, R.C.R., Lopez-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, E., Lacoste, S., Debuchy, R., Dupont, J., Branca, A., Giraud, T., 2015. Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. Curr. Biol. 25, 2562–2569. https://doi.org/10.1016/j.cub.2015.08.025

Ropars, J., Didiot, E., Rodríguez de la Vega, R.C., Bennetot, B., Coton, M., Poirier, E., Coton, E., Snirc, A., Le Prieur, S., Giraud, T., 2020b. Domestication of the Emblematic White Cheese-Making Fungus Penicillium camemberti and Its Diversification into Two Varieties. Curr. Biol. https://doi.org/10.1016/j.cub.2020.08.082

Salmela, L., Rivals, E., 2014. LoRDEC: accurate and efficient long read error correction. Bioinformatics 30, 3506–3514. https://doi.org/10.1093/bioinformatics/btu538

Stanke, M., Diekhans, M., Baertsch, R., Haussler, D., 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24, 637–644. https://doi.org/10.1093/bioinformatics/btn013

Steensels, J., Gallone, B., Verstrepen, K.J., 2021. Interspecific hybridization as a driver of fungal evolution and adaptation. Nat. Rev. Microbiol. 1–16. https://doi.org/10.1038/s41579-021-00537-4

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.-C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.-S., Zhang, J., Ruch, P., Teodoro, D., 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49, D480–D489. https://doi.org/10.1093/nar/gkaa1100

Thioulouse, J., Dray, S., Dufour, A.-B., Siberchicot, A., Jombart, T., Pavoine, S., 2018. Multivariate Analysis of Ecological Data with ade4. Springer. https://doi.org/10.1007/978-1-4939-8850-1

Thorpe, R.S., Barlow, A., Malhotra, A., Surget-Groba, Y., 2015. Widespread parallel population adaptation to climate variation across a radiation: implications for adaptation to climate change. Mol. Ecol. 24, 1019–1030. https://doi.org/10.1111/mec.13093

Tinsley, C.R., Jacques, N., Lucas, M., Grondin, C., Legras, J.-L., Casaregola, S., 2022. Molecular Genetic Analysis with Microsatellite-like Loci Reveals Specific Dairy-Associated and Environmental Populations of the Yeast Geotrichum candidum. Microorganisms 10, 103. https://doi.org/10.3390/microorganisms10010103

Urbach, G., 1997. The flavour of milk and dairy products: II. Cheese: contribution of volatile compounds. Int. J. Dairy Technol. 50, 79–89. https://doi.org/10.1111/j.1471-0307.1997.tb01743.x

Vilgalys, R., Hester, M., 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several Cryptococcus species. J. Bacteriol. 172, 4238–4246. https://doi.org/10.1128/jb.172.8.4238-4246.1990

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE 9, e112963. https://doi.org/10.1371/journal.pone.0112963

Warmuth, V., Eriksson, A., Bower, M.A., Cañon, J., Cothran, G., Distl, O., Glowatzki-Mullis, M.-L., Hunt, H., Luís, C., do Mar Oom, M., Yupanqui, I.T., Ząbek, T., Manica, A., 2011. European Domestic Horses Originated in Two Holocene Refugia. PLoS ONE 6, e18194. https://doi.org/10.1371/journal.pone.0018194

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7, 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Wickham, H., 2016. Getting Started with ggplot2, in: Wickham, H. (Ed.), Ggplot2: Elegant Graphics for Data Analysis, Use R! Springer International Publishing, Cham, pp. 11–31. https://doi.org/10.1007/978-3-319-24277-4_2

Zhu, Q., Zheng, X., Luo, J., Gaut, B.S., Ge, S., 2007. Multilocus Analysis of Nucleotide Variation of Oryza sativa and Its Wild Relatives: Severe Bottleneck during Domestication of Rice. Mol. Biol. Evol. 24, 875–888. https://doi.org/10.1093/molbev/msm005

Figure legends

Figure main:

Figure 1: Phylogenetic relationships and population structure of 98 strains of *Geotrichum candidum,* based on whole-genome data

(A) Neighbor-net analysis based on a single nucleotide polymorphism (SNP) distance matrix. The scale bar represents 0.01 substitutions per site for branch lengths.

(B) Genetic relationships among strains and population structure in *Geotrichum candidum* based on 699,755 SNPs*.* a) Maximum likelihood tree showing genetic relationships among the 98 isolates used in this study. All nodes are supported by bootstrap support >98% (bootstrap analysis with 1000 resampled datasets). The scale bar represents 0.05 substitutions per site for branch lengths. We used the midpoint rooting method to root the tree. The “$” symbol pinpoints commercial starter strains and “*” the PacBio sequences. Genomes used as reference are written in bold. b) Population subdivision inferred for K = 5. Colored bars represent the coefficients of membership in the five gene pools based on SNP data.

(C) Principal component analysis (PCA) based on 699,755 SNPs and 98 strains. Genetic clusters are represented by the same colors on all panels: light blue for Cheese_1, dark blue for Cheese_2, pink for Cheese_3, light grey for the mixed-origin population and dark grey for the wild population. Borders of points were colored in red when multiple points were overlapping due to clonal lineages (with a threshold set at <90,000 SNP for defining clonal strains). The shape of points represents the environment from which strains were sampled: circle for cheese/dairy, square for food and triangle for wild environment.

(D) PCA based on the 323,385 SNPs when the dataset was restricted to the 78 strains from the cheese clade

Figure 2: Synteny of the beta lactamase-like genes lost in the cheese clade of *Geotrichum candidum*

Synteny of the scaffold QQZM01000080.1 of the LMA-244 (wild) assembly against the scaffold CCBN010000010.1 of the cheese CLIB 918 (Cheese_3) assembly. The two scaffolds were shortened according to the range of nucleotide position on the right of each sequence. Beta-lactamase-like genes are annotated in red while other genes are displayed in blue. Black triangles indicate positions where repeated sequences were detected by tandem repeats finder (Benson, 1999). All strains from the cheese clades and 5 from the mixed-origin populations (LMA-317, ESE00274, MUCL8652 and ESE00540) carried the 20 kb deletion containing the g5112 and g5113 genes, both encoding for beta-lactamase like.

Figure 3: Heatmap of different repeats that expanded in the cheese populations

From the repeat database based on the LMA 244 (wild) assembly, the copy number of each repeated element was estimated by aligning illumina reads of each strain. Only repeated elements that were at least in a 5-fold copy relative to the LMA 244 genome were kept on the heatmap. The total number of copies is written in the center of each cell and filled with different shadings of grey to red depending on the relative expansion from the smallest copy number for each type of repeat. The maximum likelihood (ML) tree from the figure 1 (without admixed strains) was plotted below strains name to highlight the population's delineation.

Figure 4: Differences in growth, opacity and volatile compounds among the five populations of *Geotrichum candidum*

Each point represents a strain, horizontal dotted lines and vertical lines represent the mean and the standard deviation of the phenotype in the population, respectively. The number *n* at the bottom of plots indicates the number of strains used per population for measuring the corresponding phenotypes. The pairwise Tukey tests performed to assess whether there were mean differences between populations are indicated with brackets and their p-values are given.

(A) Mean radial growth of the three cheeses, mixed-origin and wild populations on cheese (1% salt), yeast peptone dextrose (YPD) and minimal media at 25°C.

(B) Differences in opacity between the three cheese populations, mixed-origin and wild populations on cheese medium (1% salt) at 10°C. Integral opacity is defined as the sum of the brightness values for all the pixels within the fungal colony bounds and measures the whiteness and density of the mycelium.

(C) Petri dish of a strain (ESE00182) from the Cheese_2 population showing the fluffiness of the colony

(D) Principal component analysis (PCA) of *Geotrichum candidum* strains based on their relative proportions of the different volatile compounds produced. Strains are plotted using the first two PCA axes.

(E) Contribution of each volatile compound to the first two PCA axes. Compounds contributing the most to the differentiation were colored in red and labeled (i.e., those distant from 0 by an Euclidean distance >=0.1).

Figure 5: Competitive abilities of the different populations of *Geotrichum candidum* against *Penicillium biforme, P. roqueforti* and *Scopulariopsis asperula* challengers.

(A) At the top, radial growth abilities of the competitors on lawns of *Geotrichum candidum* belonging to different populations (the three cheese, the mixed-origin and the wild populations). Each point represents a combination of the growth of a competitor strain on a lawn of a *G. candidum* strain. Horizontal dotted lines and vertical lines represent the mean and the standard deviation of the competitor growth in the population, respectively. The number *n* at the bottom of plots indicates the number of combinations of competitor-mat used per population. The competitor was inoculated in a central point 24h later on a lawn of *G. candidum.* At the bottom, from left to right, are shown pictures of *P. biforme* ESE00023 on a *G. candidum* ESE00186 lawn and without any lawn, *P. roqueforti* ESE00645 on a *G. candidum* ESE00186 lawn and without any lawn, and *S. asperula* ESE01324 on a *G. candidum* ESE00198 lawn and without any lawn, all on a salted cheese medium.

B: At the top, radial growth abilities of competitors, with various *G. candidum* strains belonging to different populations being grown on the other side of splitted Petri dishes. The competitor was inoculated in a central spot on one side and the *G. candidum* strain was spread on the other side of the splitted Petri dish (a picture is shown as example below the figure). Media is not contiguous between the two sides of Petri dishes, so that inhibition can only occur by volatile compounds. *Penicillium biforme* and *P. roqueforti* were grown for 11 days while *S. asperula* was grown for 19 days at 25°C.

Each point represents a combination of the growth of a competitor strain on a *G. candidum* strain. Horizontal dotted lines and vertical lines represent the mean and the standard deviation of the competitor growth in the population, respectively. The number *n* at the bottom of plots indicates the number of combinations of competitor-mat used per population.

Supplementary Figure:

Figure S1: Population structure of *Geotrichum candidum*.

Population subdivision inferred for *K* population ranging from two to six. Colored bars represent the coefficients of membership in the K gene pools based on genomic data. Each bar represents a strain, its name being indicated at the bottom of the figure. The new color for each *K* increment is indicated on the right part. The second order rate of change in the likelihood (ΔK) peaked at *K*=6. However the additional population distinguished at *K*=6 only encompassed two strains that were not that differentiated in the splistree (MUCL 14462 and CBS 9194; ; FIGURE 1).

Therefore, we chose to set the number of populations to five, the *K* value at which the structure was the strongest and the clearest, with three cheese populations, several admixed or hybrid cheese strains, a population of mixed origins and a wild population. These populations are indicated on the last rows.

Below the admixture plot, two rows are indicating milk from which the strains was sampled and the population delineation.

Figure S2: Pairwise identity between admixed and other strains, averaged by population of *Geotrichum candidum.*

In order to see traces of introgression from different populations of *Geotrichum candidum* in some hybrid strains we computed the pairwise identity along the genome. Only the first scaffold of the CLIB 918 genome is shown in this figure. Values were averaged by population and when admixed strains had the same genetic background (name of strains are above each subplot) across 30 kb overlapping sliding windows with 5 kb steps. If no introgression happened, we expect the admixed strain to be equally distant to the different populations along the genome. Introgression results in genomic regions being atypically closer to a single population. All admixed strains within the cheese clades had introgression imprints while the other three strains (CBS 9194^T, MUCL 14462 and ESE01080) did not, meaning that they were either from different genetic backgrounds or introgressed with genetic backgrounds different from the five populations of *G. candidum*.

Figure S3: Linkage disequilibrium against distance between SNPs for the five *Geotrichum candidum* populations.

The r² (square of the correlation coefficient between two indicator variables) varies between 0 when two markers are in perfect equilibrium and 1 when they provide identical information. Under recombination, we expect r² to decrease exponentially with the distance between two SNP while in non recombining lineages linkage disequilibrium remains flat. All *Geotrichum candidum* populations r² decay curve behaved as recombining populations except the Cheese_2 population.

Figure S4: SNPs inducing a premature stop codon for each *Geotrichum candidum* strain.

In order to keep genes that carried a single nucleotide polymorphism (SNP) inducing a premature STOP codon in most of a population, we only showed STOP-inducing SNP that were at least in more than three strains. Columns represent strains and strains are ordered following the maximum likelihood (ML) tree. Cells are colored in black when the corresponding SNP induces a premature stop codon, white when there is no substitution for this site compared to the reference genome and grey when the SNP status could not be attributed during SNP calling, substitution that induces other effects on the protein sequence were not present in this subset of STOP-inducing sites. Each row is a site, when multiple SNPs induced stop codons in the same gene, the corresponding rows were grouped and separated from other genes by black lines. The analysis was done using either the CLIB 918 (Cheese_3) genome (A) or the LMA-244 (Wild) genome (B) as a reference.

Figure S5: Density of transposable elements copy number relative to the LMA-244 strain

To better show the fat tail distribution, the y-axis (density) was cut at 25%. A red dashed line indicates the threshold of five times more copy number than the LMA-244 strain. This threshold was set to identify repeats expansion related to small peaks on the density curve

Figure S6: Distribution of absolute divergence d_{xy} values for different pairwise populations tested.

Distribution of absolute divergence (d_{xy}) values for each pairwise population comparison from the genomic scan analysis. Density is given as an overlapping window number for a specific value of d_{xy}, each window being 7.5 kb wide with a step of 5 kb (optimal values based on variants densities). A black vertical line indicates the threshold of 1% highest values kept for the enrichment test.

Figure S7: Genomic scan of within-population genetic diversity and between-population differentiation in *Geotrichum candidum*.

Genomic of the nucleotide diversity π, watterson's theta θ_w, absolute divergence d_{xy} and fixation index F_{ST} (Hudson et al., 1992) along the scaffold 1 of the CLIB 918 reference genome. At the bottom a guide indicates genic regions in black and non genic regions in white. 7.5 kb overlapping windows with a step of 5 kb (optimal values based on variant densities). On the bottom of the figure, genic regions are shown in black (not positively selected) or red (positively selected in the MK test analysis) rectangle. On the first panel (nucleotide diversity π), 5% lowest π values in the three cheese populations were highlighted by black dots. On the third panel (absolute divergence d_{xy}), 1% highest values of d_{xy} of each pairwise comparison were highlighted by back dots. These outliers were checked for gene presence and functions that could be involved in cheese adaptation and subsequently tested for enrichment within this subset of outliers compared to the whole genome

Figure S8: Differences in growth among the five populations of *Geotrichum candidum* populations for different media and temperature

Mean radial growth of the three cheese populations, mixed-origin and wild populations on cheese (1% salt), yeast peptone dextrose (YPD) and minimal media at 10, 15 and 25°C for 7, 11 and 14 days.

Each point represents a strain, horizontal dotted lines and vertical lines represent the mean and the standard deviation of the population respectively. The number *n* at the bottom of plots indicates the number of strains per population used for measuring these phenotypes. To assess difference in means between populations, significant pairwise Tukey tests are indicated with brackets and p-values.

Figure S9: Differences in lipolytic and proteolytic activity among the five populations of *Geotrichum candidum* populations for different growing time and temperature

A: Lipolytic activity of *Geotrichum candidum* at 15 and 25°C and grown for 7, 14 and 21 days.

B: Proteolytic activity of *G. candidum* at 15 and 25°C and grown for 7, 14 and 21 days.

Each point represents a strain, horizontal dotted lines and vertical lines represent the mean and the standard deviation of the population respectively. The number *n* at the bottom of plots indicates the number of strains per population used for measuring these phenotypes. To assess difference in means between populations, significant pairwise Tukey tests are indicated with brackets and p-values. Length of lysis was measured as a radius between the center of the colony and the limit of the lysis area when lysis happened under the colony or measured as the thickness of the lysis area when there was no lysis under the colony.

Figure S10: Differences in salt tolerance, growth on different milk and volatile compounds in detail among the five populations of *Geotrichum candidum.*

Each point represents a strain, horizontal dotted lines and vertical lines represent the mean and the standard deviation of the phenotype in the population, respectively. The number *n* at the bottom of plots indicates the number of strains used per population for measuring the corresponding phenotypes. The pairwise Tukey tests performed to assess whether there were mean differences between populations are indicated with brackets and their p-values are given.

(A) Mean radial growth at 10°C of the three cheese, mixed-origin and wild populations of *Geotrichum candidum* on cheese agar medium with different salt concentrations: unsalted, 1% salt for mimicking St Nectaire and cream cheeses, 2% salt for mimicking Camembert and goat cheeses and 4% salt for mimicking blue cheeses.

(B) Mean radial growth at 10°C of the three cheese populations, the mixed-origin and the wild populations on yogurt agar media made with raw cow, goat and sheep milks.

(C) Relative proportions of major volatile compounds in Cheese_1, Cheese_3, mixed-origin or wild populations of *G; candidum.* The volatile compounds shown were those contributing the most to the two first PCA axes or that are important for cheese ripening. For each compound, the related corresponding descriptor from thegoodscentscompany.com was added.

Supplementary Table:

Table S1: Description of the origin, phylogenetic assignation, sequencing statistics, and phenotype tested of *Geotrichum candidum* strains used in this study.

Strains: Name of the strains; ESE collection number: collection number of strains kept at the author laboratory; Population: population attributed to the strains in this study; Clonal group: strains with the same number are clonal (less than 1,200 SNP between them); Species name: species identified either based on public collection or from ITS identification for collected strains in this study; Environment of sampling simplified: broader categories for source of sampling; Milk type: if extracted from dairy, indicates species from which was made the dairy; Location: Region or country of origin; Mating type: mating type identified in *Geotrichum candidum* either MATA or MATB (Morel et al., 2015).

Table S2: Population genetics statistics estimating genetic differentiation (F_{st}, d_{xy}) in the five Geotrichum candidum populations and among the identified population within each of two other fungal species (*Penicillium camemberti* and *Penicillium roqueforti*) (Dumas et al., 2020; Ropars et al., 2020b)

Table of F_{st} (fixation index) and d_{xy} (absolute divergence) (Hudson et al., 1992; Nei and Li, 1979). Values for *Geotrichum candidum*, *Penicillium roqueforti* (Dumas et al., 2020), *Penicillium camemberti* (Ropars et al., 2020b) are indicated for each population. Cells are colored from the lowest (white) and the highest (red) value of each indices.

Table S3: Proportions of fixed, shared and private SNPs for each pair of populations of *Geotrichum candidum* population and for each pair of populations within each of two other fungal species (*Penicillium camemberti* and *Penicillium roqueforti*) (Dumas et al., 2020; Ropars et al., 2020b).

Percent and number of fixed, shared or private single nucleotide polymorphisms (SNPs) for *Geotrichum candidum*, *Penicillium roqueforti* (Dumas et al., 2020), *Penicillium camemberti* (Ropars et al., 2020b) are indicated for each populations. The method of attributions of SNPs to different categories are available on https://github.com/BastienBennetot/fixed_shared_private_count.

Table S4: F3 test performed on each trio of populations of *Geotrichum candidum* populations.

The F3 test is a test between three populations. It tests whether a target population (C) is admixed between two source populations (A and B) and gives a measure of shared drift between two test populations (A and B) from an outgroup (C). In case of introgression, we expect negative F3 values. A Z-score is computed based on F3 value and the standard error to assess the deviation from zero of the F3 value. If the Z-score is inferior to minus three then we can conclude a significant rejection of the Null hypothesis (F3 value is not negative).

Table S5: Population genetics statistics estimating genetic diversity (π, watterson's θ) in the five *Geotrichum candidum* populations and among the identified population of three other fungal species (*Penicillium roqueforti, Penicillium camemberti* and *Saccharomyces cerevisiae*) (Dumas et al., 2020; Peter et al., 2018; Ropars et al., 2020b).

Nucleotide diversity statistics (π and watterson's θ) are indicated for *Geotrichum candidum*, *Penicillium roqueforti* (Dumas et al., 2020), *Penicillium camemberti* (Ropars et al., 2020b) and *Saccharomyces cerevisiae* (Peter et al., 2018). Cells are colored from the lowest (white) and the highest (red) value of each indices.

Table S6: Distribution of the two mating types in each population and proportion test of the deviation from 1:1 ratio.

Table S7: Phi test of each population of *Geotrichum candidum* using the first scaffold of CLIB 918 assembly.

Pairwise homoplasy index (PHI) test helps to discriminate between the presence or absence of recombination between a population (Bruen et al., 2006). It tests with permutations the null hypothesis of no recombination by looking at the genealogical association among adjacent sites. The PHI test was performed using PhiPack v1.1 and the first scaffold of the CLIB 918 assembly as reference.

Table S8: Number and percentage of SNPs classified by impact, functional class, effect and genomic regions for each *Geotrichum candidum* populations.

Based on the 7,150 CDS of the CLIB 918 assembly, the effect of all variants was assessed. Each variant is categorized in different functional effects on the protein sequence. Results are meaned by population and shown either as the mean number of SNP or percentage. Results were computed using snpeff (Cingolani et al., 2012).

(A) Total number of single nucleotide polymorphism (SNP) within each *G. candidum* populations.

(B) Putative variant impact prediction. Different impacts categories are defined in Snpeff manual in ‘Input & output files’ section

(C) Protein sequence effect of SNPs. Variants can either not change amino acid sequence (silent), change the amino acid (missense) or induce stop codons (nonsense)

(D) Functional effect of SNPs defined in Snpeff manual in ‘Input & output files’ section

(E) Position of the SNPs compared to genes

Table S9: Number of copy number variants windows that differentiated G*eotrichum candidum* populations and test for enrichment of gene ontologies contained within these windows.

Copy number windows (non-overlapping windows of 500 bp) were subsetted and classified in three comparison when the median copy number between the two clades compared was different: wild against all other populations (mixed origin and cheese populations), mixed-origin against cheese clade, and any pairwise difference within the different cheese populations (subtable A and B). For each comparison, when subsetted windows contained genes, gene ontologies (GO) of these genes were used to perform an enrichment test compared to the rest of the genome (subtable C and D). The same methodology was used for LMA 244 (Wild) or CLIB 918 (Cheese_3) as reference assembly ensuring we see regions unique to the Cheese_3 and the Wild populations.

(A) Number of windows subsetted in each clade comparison using LMA 244 assembly as reference

(B) Number of windows subsetted in each clade comparison using CLIB 918 assembly as reference

(C) Enrichment test on gene ontologies (GO) present within the subsets of windows in each clade comparison using CLIB 918 assembly as reference

(D) Enrichment test on gene ontologies (GO) present within the subsets of windows in each clade comparison using LMA 244 assembly as reference

Table S10: Repeat copy number for the different strains of *Geotrichum candidum.*

In order to de novo detect repeats within *Geotrichum candidum*, RepeatModeler v2.0.2 (Flynn et al., 2020) was run on the pacbio genome assembly of LMA 244 generating a library of 176 repeats. The repeat redundancy was reduced using cd-hit-est, as described in Goubert et al., giving a final library of 108 repeats (presented in column “Clustering of repeat family”) (Goubert et al., 2022). Sometimes the type and family of these repeats was inferred (column type of repeat and repeat family). To estimate the per strain copy number of each repeat, illumina reads were aligned using bwa mem (v0.7.17; Li, 2013) to the repeat library and the median coverage for each repeat was then normalized by the LMA 244 genome wide median coverage. Strain: name of the strains studied; relative median coverage: Coverage of all genomic reads mapped to the repeated sequence normalized by genome wide coverage before taking the median of all nucleotides of the repeated sequence (gives an idea of the copy number of the repeated element genome-wide); Copy number relative to LMA-244 strains: relative median coverage of the strain divided by the one of LMA-244 (Wild) strain to better emphasize repeat expansion within the cheese clade.

Table S11: Test for enrichment of gene functions that showed footprints of divergent selection and recent selective sweeps.

We tested gene function enrichment that were detected either by keeping 1% highest of absolute divergence (d_{xy}) between cheese and wild strains (A) or 5% lowest nucleotide diversity (π) in the cheese populations but not in the wild population (B). Only functions related to lactose, lipid, protease were kept. Windows were 7.5 kb wide and overlapping windows with a step of 5 kb.

(A) Test for enrichment of gene functions within the 1% highest Dxy windows when comparing a cheese population to the Wild population

The subsetted windows for each test are based on the 1% highest Dxy windows for a specific comparison between a cheese and the wild population

(B) Test for enrichment of gene functions within the 5% lowest Pi windows in cheese populations. Windows are excluded when they are also in the 5% lowest Pi of the wild population.

The subsetted windows for each test are based on the 5% lowest Pi windows for a specific cheese population.

Table S12: Results of the McDonald and Kreitman (MK) test for positive selection.

Genes evolving under positive selection were assessed using McDonald and Kreitman (MK) tests. Using contrasting levels of polymorphism and divergence at neutral and functional sites, MK test estimates the fraction of substitutions at the functional sites that were driven by positive selection. When fisher.P.value was lower than 0.05, we considered that the gene was under positive selection. Only genes under positive selection were shown in the table. The mixed-origin population was compared to the cheese clade (A), Cheese_1 (B), Cheese_2 (C) and Cheese_3 (D). Ortho_id: identifier of the orthologous gene; P1_nonsyn: the number of non-synonymous polymorphisms in the first population; P2_nonsyn: the number of non-synonymous polymorphisms in the second population; P1_syn: the number of synonymous polymorphisms in the first population; P2_syn: the number of synonymous polymorphisms in the second population; D_nonsyn: the number of non-synonymous substitutions; D_syn: the number of synonymous substitutions; neutrality.index: quantifies the degree of departure from neutrality; alpha: the proportion of substitutions driven by positive selection; fisher.P.value: P-value of the MK test; GeneID: gene identifier in the CLIB 918 assembly annotation; Contig Start: Start of the gene sequence; Stop: Stop of the gene sequence; Name: Name of the closest orthologous annotated genes; Product: Function of the protein; PFAM: pfam database annotation; InterPro: InterPro database annotation

Table S13: Anova table of all phenotypic linear models and post-hoc test table.

All outputs of analysis of variance (ANOVA) (A) and post-hoc test (B) based on different linear models that were used for phenotypic analyses.Linear models tested: Effect of media, temperature on radial growth (1); effect of media and population on radial growth of *Geotrichum candidum* at 25°C (2); effect of salt content and population on radial growth (3); effect of milk origin and population on radial growth (4); effect of media and population on opacity (5); effect of population of *Geotrichum candidum* on competitor growth (6); competition abilities by volatiles on splitted Petri dishes (7); effect of population on relative proportions of volatile compounds (8).

In ANOVA table, columns correspond to degree of freedom (Df), sum of squares (Sum Sq), mean square (Mean Sq), the F statistic (value) and the p-value of this test.

Columns in post-hoc tables before the “term column” gives the condition kept for each single test. term: variable used for pairwise comparison; group1: the group that we compare against group2; group2: a second group that we compare to group1; null.value: value of group1 - group2 under the null.hypothesis; estimate: value of group1 - group2 using the data; conf.low: Lower value of the confidence interval; conf.high: higher value of the confidence interval; p.adj: adjusted p-value; p.ad.signif: significance of the adjusted p-value (p-value >0.05:n.s.; <0.05:*; <0.01:**; <0.001:***, <1E-04:****). Post hoc test for testing the effect of salt content and population on radial growth of *G. candidum*.

Table S14: Anova table and post-hoc test of lipolysis and proteolysis analyses.

All outputs of analysis of variance (ANOVA) (A) and post-hoc test (B) based on different linear models that were used for lipolysis (1) and proteolysis (2) analyses. Radius of lysis for lipolytic and proteolytic activities experiments was often discrete, strains either showing lytic activity or not at all. Thus, data were transformed into qualitative discrete data and a generalized linear model with a binomial function as logit was fitted. No post-hoc tests were computed for the lipolysis analysis because there was no population effect.

In ANOVA table we find columns $\sigma 2$: mean random effect variance of the model; *τ00*: The random intercept variance of a given variable, or between-subject variance; ICC: the intraclass correlation coefficient; N variable: the number of categories of the given variable; Observations: sample size of the model; marginal R^2 : represents the variance explained by the fixed effects; conditional R^2: interpreted as the variance explained by the entire model (i.e. the fixed and random effects).

In post hoc table we find columns Linear Hypotheses: null hypothesis considered for each post-hoc test; Estimate: Measured value; Std. Error: standard error of this measure; z value: Z statistic of this test.

Table S15: Description of the origin and species of strains used in the competition experiments.

ESE collection number: Author collection number for this strain; Previous name in public collection: name of the strain in other public collection; Species: Species of the strain; Origin of sampling: Environment of sampling for this strain; Milk (if cheese): species from which was collected the milk that was used to make the dairy where the strain was sample; Location of sampling; Location of the sampling; Cheese shop: where the cheese was bought for cheese strains; Information on sampling date: Sampling date when known for strains sampled years before the study

Figure 1: Phylogenetic relationships and population structure of 98 strains of *Geotrichum candidum,* based on whole-genome data

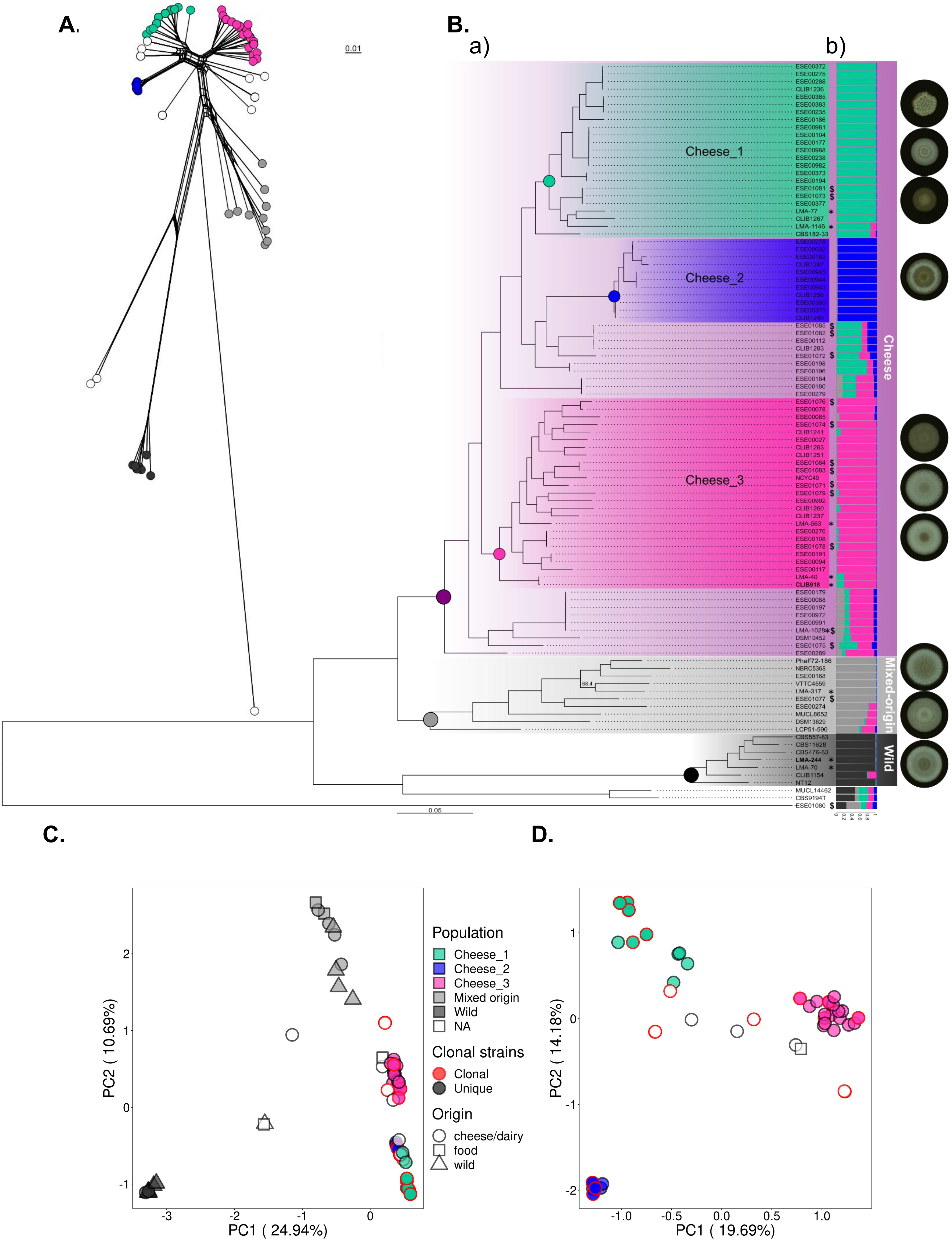

Figure 2: Synteny of the beta lactamase-like genes lost in the cheese clade of *G. candidum*

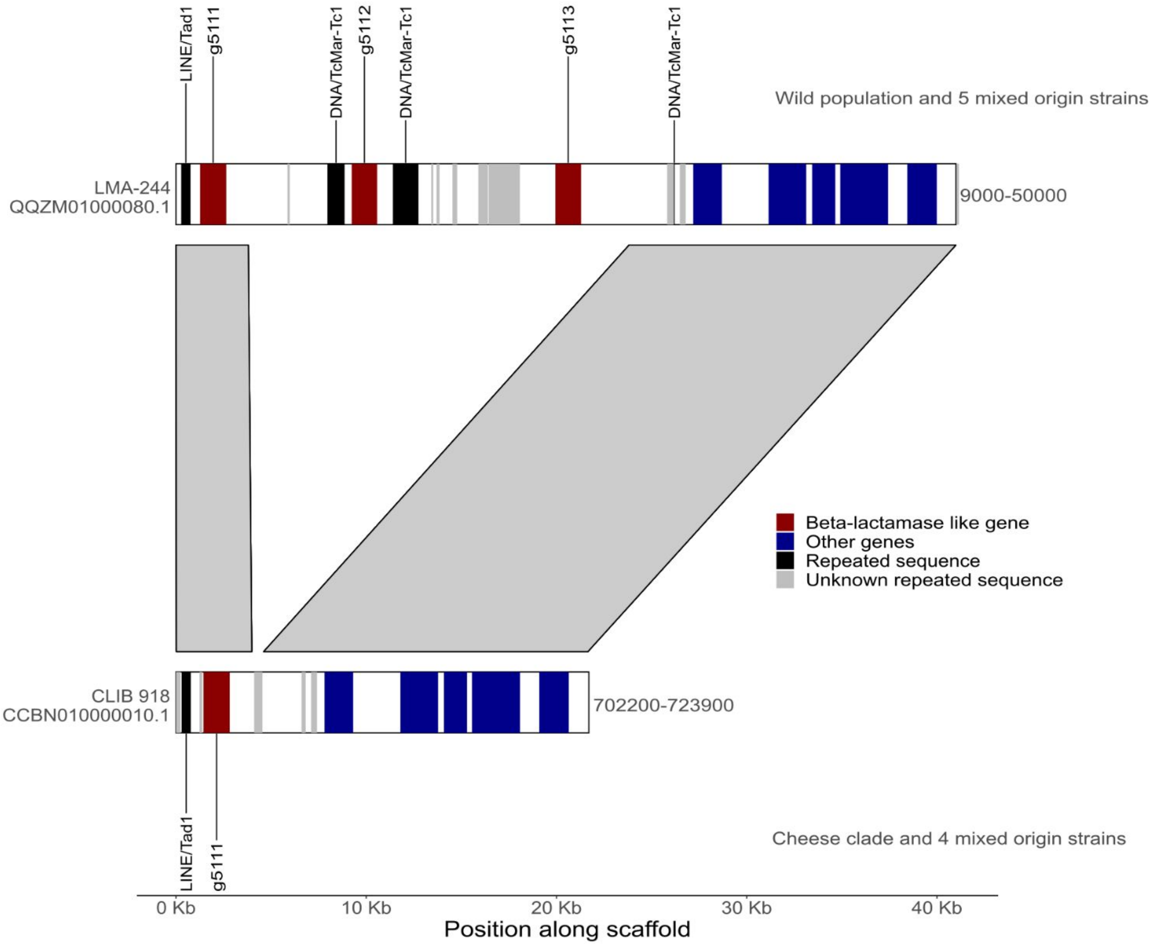

Figure 3: Heatmap of different repeats that expanded in the cheese populations

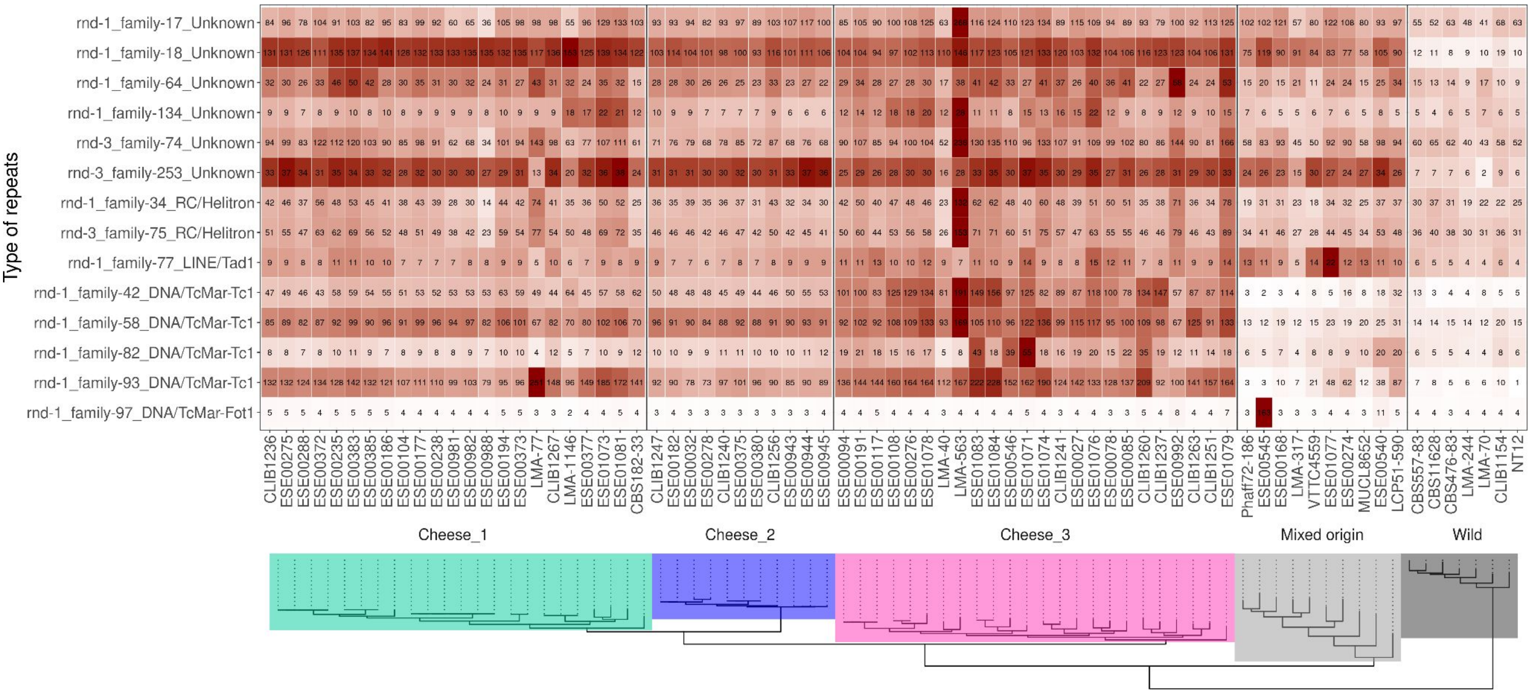

Figure 4: Differences in growth, opacity and volatile compounds among the five populations of *Geotrichum candidum*

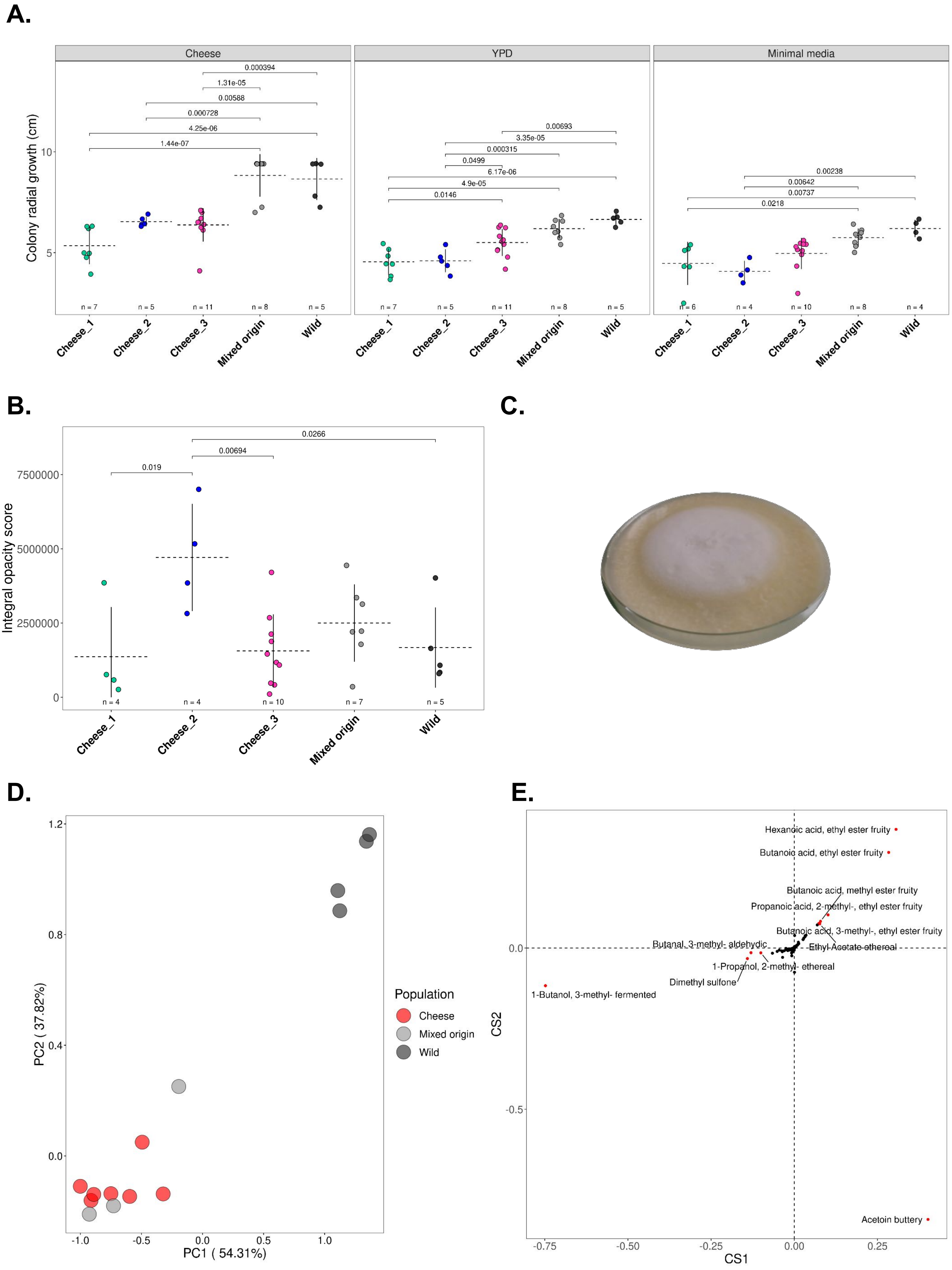

Figure 5: Competitive abilities of the different populations of *G. candidum* against *Penicillium biforme, P. roqueforti* and *Scopulariopsis asperula* challengers.