bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Footprints of domestication in dry-cured meat *Penicillium* fungi: convergent specific phenotypes and horizontal gene transfers

Ying-Chu Lo[1]†, Jade Bruxaux†[1], Ricardo C Rodríguez de la Vega[1], Alodie Snirc[1], Monika Coton[2], Mélanie Le Piver[3], Stéphanie Le Prieur[1], Daniel Roueyre[3], Joëlle Dupont[4], Jos Houbraken[5], Robert Debuchy[6], Jeanne Ropars[1], Tatiana Giraud* [1] and Antoine Branca* [1]

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

2Univ Brest, Laboratoire Universitaire de Biodiversité et Ecologie Microbienne, 29280 Plouzané, France

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

[4]Origine, Structure, Evolution de la Biodiversité, UMR 7205 CNRS-MNHN, Muséum National d'Histoire Naturelle, CP39, 57 rue Cuvier, 75231 Paris Cedex 05, France

[5]Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands

[6]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.

†These authors contributed equally:

*These authors jointly supervised the study

Corresponding author: Jade Bruxaux bruxaux.jade@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Short title: Footprints of domestication in dry-cured meat *Penicillium* fungi

Keywords: fungi, parallel adaptation, convergence, sausage, domestication

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract: 246 words

Many fungi have been domesticated for food production, with genetic differentiation between populations from food and wild environments, and food populations often acquiring beneficial traits through horizontal gene transfers. We studied the population structures and phenotypes of two distantly related *Penicillium* species used for dry-cured meat production, *P. nalgiovense*, the most common species in the dry-cured meat food industry, and *P. salamii*, used locally by farms. Both species displayed low genetic diversity, with no differentiation between strains isolated from dry-cured meat and those from other environments. Nevertheless, the strains collected from dry-cured meat within each species displayed slower proteolysis and lipolysis than their wild-type conspecifics, and those of *P. nalgiovense* were whiter. The phenotypes of the non-dry-cured meat strains were more similar to their sister species than to their conspecific dry-cured meat strains, indicating an evolution of specific phenotypes in dry-cured meat strains. A comparison of available *Penicillium* genomes from various environments revealed evidence of multiple horizontal gene transfers, particularly between *P. nalgiovense* and *P. salamii*. Some horizontal gene transfers involving *P. biforme*, also found in dry-cured meat products, were also detected. We also detected positive and purifying selection based on amino-acid changes. Our genetic and phenotypic findings suggest that human selection has shaped the *P. salamii* and *P. nalgiovense* populations used for dry-cured meat production, which constitutes domestication. Several genetic and phenotypic changes were similar in *P. salamii*, *P. nalgiovense*, and *P. biforme*, providing an interesting case of convergent adaptation to the same human-made environment.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

The ways in which populations adapt to their environment is a fundamental issue in biology. Domestication is widely used as a model for understanding adaptation as it corresponds to a recent adaptation process caused by strong, human-driven selection [1,2]. Humans have grown and bred various organisms, choosing individuals with the most interesting features for their particular uses or for consumption. For instance, many morphological traits differ between domesticated maize and its wild ancestor, including floral morphology, seed size, dispersal ability, stem number and length [3,4]. Charles Darwin used domesticated pigeons as a model for investigating adaptation and natural selection processes [1,2]. The parallel adaptation of distantly related species to the same environment provides an excellent study model for investigating whether evolution is repeatable — i.e., whether independent adaptation events to the same ecological niche involve similar phenotypic and genomic changes — a situation known as convergence [5–7]. In addition to adaptive changes, domesticated populations may suffer from bottlenecks, in which a drastic loss of genetic diversity occurs due to the strong selection exerted by humans. Such bottlenecks may jeopardize the ability of the domesticated populations to adapt [8,9].

Fungi are good models for studies of adaptation using domestication. Their experimental tractability makes it possible to perform fitness tests in controlled environments, and their small and compact genomes facilitate genomic studies to elucidate the genetic basis of adaptation [10–15]. Many fungi have been domesticated for food production [16], the chief examples being *Saccharomyces, Aspergillus* and *Penicillium*. *Saccharomyces cerevisiae* is the most studied domesticated fungal species, with different populations used for winemaking, baking and brewing [12,17–20]. Fungi are also used for cheese ripening. For example, *P. roqueforti* is used for blue-veined cheeses and *P. camemberti* is used for soft cheeses, such as

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Brie and Camembert. Both fungi have recently emerged as useful models for adaptation and domestication studies [13,21–23]. The *Saccharomyces cerevisiae*, *P. camemberti* and *P. roqueforti* populations used in the food industry have specific traits that are beneficial for food production and enable them to thrive in their human-made environment to a much greater extent than other populations [21,22,24], providing evidence of genuine domestication. The genomic basis of several of these adaptive traits has been determined in *Saccharomyces* and *Penicillium* food populations, with the identification of gene duplications, interspecific hybridizations and horizontal gene transfers [13,17,19,23,25–27]. Several adaptive horizontal gene transfers have occurred between distantly related cheese *Penicillium* fungi, leading to rapid convergent adaptation through the very same genes and genomic mechanisms [13,23]. As expected from a domestication process, the fungal populations used for food production have lost much of their genetic variability, as demonstrated for *Saccharomyces* from wine and beer and *Penicillium* fungi from cheese [12,21,22]. Most fungi can be cultured clonally, leading to a particularly large loss of genetic variability in domesticated fungi [21,22].

Like cheese fungi, *Penicillium nalgiovense* is an excellent model for studying domestication and adaptation. This species has been used to inoculate the surface of most dry-cured meat products for decades, to enhance the ripening process [28–31]. Another recently described fungus, *Penicillium salamii*, also occurs in dry-cured meats [32] and may be included in starter cultures or form part of the natural mycobiota at local production sites [31,33]. Several traits have probably been subject to similar selection processes in both these fungi, either deliberately through human intervention or unwittingly, to enable these fungi to thrive in dry-cured meats and produce high quality products, with an appealing visual appearance and aroma, and a better shelf life and product safety. The fungi colonize the surface, thereby helping to preserve the dry-cured meat and to protect it against excessive desiccation, irradiation by light, oxygen

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and undesirable microorganisms, such as contaminating toxigenic molds [31,34,35]. These fungal species also help to improve the aroma of the dry-cured meat, through the metabolites they produce with their lipase, protease, transaminase and dehydrogenase activities [30,36]. However, if proteolytic and lipolytic activities are too high, this can result in an undesirable texture, bitterness and a rancid taste [37]. Another important trait is the color of *Penicillium* fungi, which is also under selection, to meet consumer preferences [38]. Salt is one of the ingredients added to dry-cured meats during production; it helps to exclude undesirable microorganisms by reducing water activity, one of the key mechanisms of preservation in these products. The *Penicillium* fungi used to inoculate dry-cured meats may, therefore, have evolved a high degree of salt tolerance. *Penicillium nalgiovense* and *P. salamii* have also been isolated from environments other than dry-cured meats, such as the soil, plants, sewage and dog bones [39]. However, it remains unknown whether the dry-cured meat *Penicillium* fungi display particular phenotypes and/or form populations that are genetically isolated from those found in other environments. *Penicillium nalgiovense* and *P. salamii* lineages have diverged at least 30 million years ago within the genus *Penicillium* [32,40,41], which corresponds, approximately, to the divergence between humans and Cercopithecoids (macaques, langurs…) [42,43]; a convergent adaptation to dry-cured meat, if observed, would therefore constitute an interesting case of parallel adaptation.

The aim of this study was, therefore, to determine whether, within these two *Penicillium* species, the strains used in dry-cured meat production had features typical of domesticated organisms, by investigating whether dry-cured meat strains 1) were genetically different from the strains collected from other environments, 2) displayed a genome-wide reduction of genetic diversity potentially indicative of strong population reduction and/or clonality, 3) had specific traits that would make them better suited to human needs or preferences than other

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

strains, 4) displayed footprints of genomic adaptation, and 5) displayed convergent phenotypic and/or genomic changes. We addressed these questions by isolating strains from dry-cured meat, sequencing the genomes of multiple *P. nalgiovense* and *P. salamii* strains collected from dry-cured meat and other environments, and searching for genetic differentiation, recombination footprints, horizontal gene transfers, specific genome arrangements and traces of positive selection. Phenotypic differences between *Penicillium* fungi from dry-cured meat environments and other species/populations were also investigated by comparing their growth rates at different salt concentrations, color, lipid and protein degradation rates and their ability to produce harmful toxins. For interpretation of the phenotypic differences between strains from dry-cured meat and other environments in terms of domestication, we further compared the phenotypes of the non-dry-cured meat strains to those of their respective sister species (*P. chrysogenum* for *P. nalgiovense* and *P. olsonii* for *P. salamii*). Under the hypothesis that dry-cured meat populations are domesticated, we would expect traits essential for dry-cured meat products to have evolved specifically in dry-cured meat strains, and to be absent in strains from other environments and in the corresponding sister species. In addition, because we found horizontally-transferred genomic regions shared between *P. nalgiovense*, *P. salamii* and *P. biforme*, the last of these species also being used to inoculate dry-cured meats, albeit rarely, we also compared phenotypes between strains from dry-cured meat and other environments for *P. biforme* and its sister species, *P. palitans*.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

Genetic diversity and population structure in *P. salamii* and *P. nalgiovense*

We collected 140 dry-cured meat casings, from which we isolated 134 monospore-derived strains. The most frequent *Penicillium* species isolated from dry-cured meat was *P. nalgiovense* (*n*=70) followed by *P. salamii* (*n*=16). We also found *P. biforme* (*n*=5), *P. chrysogenum* (*n*=5), *P. solitum* (*n*=5) and *P. camemberti* (*n*=5) and, sporadically, other *Penicillium* species and other genera (Table 1).

We obtained high-quality genome assemblies for *P. salamii* LCP06525 and *P. nalgiovense* ESE00252 with long-read (PacBio) sequencing technology and we sequenced the genomes of 55 strains with Illumina technology (Table S1). We identified 1,395,641 SNPs by mapping reads from the 21 Illumina-sequenced *P. salamii* genomes onto the *P. salamii* reference genome. A population tree based on a 2,977,098 bp alignment after the addition of the outgroup (Fig. 1A) revealed genetic differentiation by sampling environment in *P. salamii*. Two lineages contained mostly dry-cured meat strains, whereas the other two lineages contained mostly non-dry-cured meat strains. The frequencies of the two collection environments differed significantly between the three clades with more than one strain (Chi^2= 7.6723; df = 2; *p*-value = 0.02). No association with geographic origin was observed for *P. salamii* strains; the least diverse clade included only strains from Europe, but was also a clade containing mostly dry-cured meat strains (Fig. 1A).

For the 28 *P. nalgiovense* genomes obtained by Illumina sequencing, reads were mapped onto the *P. nalgiovense* reference genome (LCP06099=FM193), leading to the identification of only 380,884 SNPs (381,996 when the other available PacBio genome ESE00252 was added),

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

indicating a very low level of genetic diversity. The *P. nalgiovense* population tree, based on a 2,367,990 bp alignment with the outgroup (Fig. 1B), had very short branches for most strains, suggesting that these strains represent a clonal lineage. No strong genetic differentiation was observed, with no effect of sampling environment or geographic location in particular. Only two strains displayed some genetic differences from the main lineage.

We analyzed 1,766 distant SNPs for *P. nalgiovense* and 7,685 for *P. salamii*, to investigate the population structures of these two species. For *P. salamii*, the STRUCTURE analysis supported four well-delimited genetic clusters, consistent with the population tree (Fig. 1A, Fig. S2A). For *P. nalgiovense*, the STRUCTURE analysis was also consistent with the population tree, with two well-delimited genetic clusters (Fig. 1B, Fig. S2B); the other genetic clusters corresponded only to admixed strains, and were not, therefore, considered to correspond to genuine clusters. PCA (Fig. 1) recovered patterns similar to those revealed by the population trees and STRUCTURE analyses: four clusters for *P. salamii*, with a tendency towards separation by geographic origin, and weaker structure for *P. nalgiovense*, with the most differentiated strain being from North America.

We identified six different genomic architecture types among *P. salamii* strains (Fig. 1A, Fig S1A-B). The genomes of *P. nalgiovense* strains were highly collinear, except for two strains displaying large rearrangements relative to the other strains (Fig. 1B). These collinearity groups in both *P. salamii* and *P. nalgiovense* corresponded to the genetic clusters detected above (Fig. 1). In *P. salamii*, genomic rearrangements differentiated the four strains obtained outside of Europe (USA and Middle East).

Nucleotide and genetic diversities were much higher in *P. salamii* (π = 0.0111 bp^{-1}, standard error 1 x 10^{-5}; θ = 0.0286 bp^{-1}, standard error = 0.1580) than in *P. nalgiovense* (π = 0.0009 bp^{-}

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[1], standard error 1 x 10^{-6}; θ = 0.0089 bp^{-1}, standard error = 0.0194), which had diversity levels similar to those in *P. roqueforti* (π = 0.0011 bp^{-1}; θ = 0.0007 bp^{-1}) [22] and *P. biforme* (π = 0.0011 bp^{-1}; θ = 0.0011 bp^{-1}) [21]. We detected both mating types (MAT1-1 and MAT1-2) in *P. salamii*, in balanced proportions and with a distribution across strains corresponding to the genetic structure and genomic architecture types (Fig. 1A). In *P. nalgiovense*, only the two most divergent strains were identified as MAT1-1, all the other strains being identified as MAT1-2, supporting the hypothesis of the use of a single clonal lineage for dry-cured meat production (Fig. 1B).

Overall, *P. nalgiovense* displayed high levels of genome-wide linkage disequilibrium (LD), with no detectable decay over physical distance (Fig. S1C). This finding is consistent with a complete absence of recombination due to exclusively clonal multiplication in this lineage. In *P. salamii*, we observed a sharp decay in LD within 200 bp, but LD level remained high. Analysis of LD within each of the two *P. salamii* populations with sufficient numbers of strains and both mating types revealed the same LD pattern as in *P. nalgiovense* (Fig. S1C). This LD pattern in *P. salamii* is consistent with the clonal multiplication of two lineages of opposite mating types, with recombination events occurring in the ancestral population, before the clonal multiplication of these two lineages. The neighbor-net networks obtained further support the population structure identified and recent clonality in *P. salamii* and *P. nalgiovense* (Fig. 1).

Detection of horizontal gene transfers between *P. nalgiovense* and *P. salamii*

We aimed to detect very recent horizontal gene transfers (HGTs) occurring during domestication by humans that could potentially be involved in parallel adaptation to the dry-cured meat environment. The genomic regions resulting from such recent HGTs would be

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

expected to be almost identical between the distantly related *P. salamii* and *P. nalgiovense* species, would not necessarily be present in all strains and would be absent from many non-dry-cured meat species (e.g. [23]).

We searched for putative horizontal gene transfers between *P. nalgiovense* and *P. salamii*, by identifying thresholds above which regions could be considered to be more similar across a longer length than expected from the phylogenetic distance between the species. For comparisons of genomic sequences of more than 1,000 bp harboring blocks of 65 identical base pairs separated by less than 90 bp between the dry-cured meat species and their respective sister species (i.e., comparisons of *P. nalgiovense* with *P. chrysogenum* and *P. rubens*, two species formerly considered as a single species, and of *P. salamii* with *P. olsonii*), 98.21% was the highest level of similarity (Fig. S3A). We therefore set the identity threshold for identifying potential horizontally transferred regions (HTRs) at 98%, because lower levels of similarity might be expected to occur by chance for such long sequences between sister species. The longest of the nearly identical sequences was 11,414 bp long (Fig. S3B). However, as the second longest sequence was only 6,972 bp long, we decided to set the threshold length for identifying potential HTRs to 7,000 bp, because shorter fragments with such high levels of similarity might occur by chance between sister species. These thresholds are conservative for the identification of putative HTRs between *P. nalgiovense* and *P. salamii*, given that we took the maximal values and these two dry-cured meat species are as distantly related as the sister species considered above.

Our analysis of genomic regions of more than 7,000 bp in length with more than 98% similarity between pairs of *P. nalgiovense* and *P. salamii* strains identified 26 putative HTRs. These putative HTRs were between 7,224 bp and 496,686 bp long (mean 55,656 bp; median

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

30,091 bp; standard deviation 96,347 bp; Table S4). The cumulative total length of the putative HTRs was 1,447,064 bp, including 145,640 bp assigned to transposable elements (TEs) of known families and 558,247 bp assigned to TEs of unknown families (10% and 38.6%, respectively; Table S4). We identified 34 (65%) of the 52 (2x26) consensus HTR extremities (the very last bases) as belonging to a TE, and 87% of the extremities had a detected TE within 1,000 bp, suggesting a possible role for TEs in the horizontal transfer of some of these regions. After the removal of all TEs of known and unknown families identified in the *P. nalgiovense* and *P. salamii* reference genomes ("repeat-masked" analyses), three putative HTRs had lengths below 500 bp and were therefore discarded for subsequent analyses (Table S4).

We identified 283 coding genes in the remaining 23 putative HTRs, 199 of which yielded hits in one of the egg-NOG databases, mostly with the fungal egg-NOG database (87.4%, 173 genes), suggesting that most of the putative HTRs identified originated from fungi. The other 25 genes had best hits in the eukaryote egg-NOG database (12.1%) or the principal egg-NOG database (0.5%). They could, therefore, also be of fungal origin, but better annotated in more phylogenetically distant groups (e.g. mammals), resulting in a best match at a higher taxonomical level.

We describe below only the presence/absence and content of the putative HTRs in the repeat-masked case; we provide the results for the unmasked and TE-masked datasets in Figs. S4-5, in which highly similar patterns are observed. Most *P. nalgiovense* strains contained the same putative HTRs, whereas the diversity was greater in *P. salamii* strains, with some strains even carrying none of the putative HTRs (Fig. 2A; Figs. S4-5). The number and length of the putative HTRs largely accounted for the differences in genome size between *P. salamii* strains

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(adjusted R^2: 0.67; Fig. S3C). The length proportions of each putative HTRs differed between strains (Fig. 2A) but, when present, the sequence was always very similar to that of the reference putative HTR (>95% identity; Fig. 2B; Figs. S4-5). Only two regions (HTR 8 and HTR 21) in six *P. salamii* strains displayed lower levels of similarity to the reference sequence. Twelve of the putative HTRs were found to be split into different parts within the genome in some strains (Table S4). The split often occurred in TE-rich regions. A comparison of the genomic regions flanking the putative HTRs between strains showed these regions to be located at different sites in the two species, and sometimes also in the different groups of *P. salamii* with different genomic arrangements, and, more rarely, in different *P. nalgiovense* strains. These differences in genomic sites between strains, and in terms of the presence/absence of HTRs, provide strong support for the view that these regions are genuine HTRs, and not introgressions or vertically inherited fragments with a high degree of sequence conservation.

Furthermore, the 23 putative HTRs were absent from most of the 68 genomes from other *Penicillium*, *Aspergillus* or *Monascus* species analyzed (Fig. 3A, Fig. S6-7). Only 25 species carried parts of at least one putative HTR, and 19 species carried parts of more than one putative HTR. The species carrying the largest number of the 23 putative HTRs belonged to genus *Penicillium* (Fig. 3A, Fig. S6-7). Tow *Penicillium* species had large fragments of seven putative HTRs: *P. biforme*, *P. camemberti* var. *camemberti* and *P. camemberti* var. *caseifulvum* (Fig. 3A, Fig. S6-7). Interestingly, these two species belong to a species complex found in dry-cured meat and sometimes even used for inoculation for dry-cured meat maturation, but mostly used for cheese production. The two varieties of *P. camemberti,* var. *camemberti* and var. *caseifulvum,* are two closely related clonal lineages displaying very little variability, having diverged recently from *P. biforme,* which harbors higher levels of genetic

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and phenotypic variability [21]. Most of the other *Penicillium* species harbored only small parts of the putative HTRs. HTR 6 and HTR 15 were present in more species than the other putative HTRs and had similar distributions across species (Fig. 3A, Fig. S6-7). The similarity between the putative HTRs present in the different species and the reference putative HTR was very high, despite the phylogenetic distances between the species harboring them (typically more than 95%, Fig 3B, Fig. S6-7), providing further support for the view that they result from horizontal gene transfers.

We, thus, identified 23 genomic regions that were almost identical between distantly related species, absent in more closely related species, rich in TEs (accounting for a mean of 44% of the regions, ranging from 9 to 91%), including some firmly linked to known TE families, and located in different regions of the genomes depending on the species, or even the population. These findings provide compelling evidence that these regions resulted from horizontal gene transfers. Given the very nature of the putative HTRs, and their complete absence in many strains and species, it is not possible to use genealogies to test for horizontal gene transfer, but the evidence provided above is even stronger.

We constructed the phylogenetic species tree with 1102 single-copy orthologous genes. The resulting species tree (Fig. 3) was consistent with previous studies [32,41,44] and was well supported. Given the distance separating species in our tree, the *P. nalgiovense* – *P. salamii* pair appeared to have a particularly high shared HTR content, in terms of total length, with only the *P. nalgiovense* – *P. biforme* pair having a longer total length of HTRs (Fig. S3D). More generally, all the species that were phylogenetically distantly related but had a large number of putative HTRs in common were known to be present and sometimes used in either cheese or dry-cured meat production (Fig. S3D).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

An analysis of functional categories in clusters of orthologous groups (COG) present in the HTRs identified eight categories present in most strains and four categories that were less present (Table S5). None of the *P. nalgiovense* or *P. salamii* strains displayed an enrichment in certain functions in the HTRs relative to the rest of the genome. The seven HTRs present in *P. biforme* and the two varieties of *P. camemberti* added between 42 and 47 genes to the rest of the genome, but only seven to 20 of the HTRs in each strain had a predicted function. The remaining genes in the HTRs were involved in the cytoskeleton, transport and metabolism of carbohydrates, inorganic ions, lipids or secondary metabolites, posttranslational modifications, or intracellular trafficking (Table S5). The two regions present in the largest number of species (HTR6 and HTR15) carried 62 and 11 genes, respectively, 45 and 10 of which, respectively, were annotated, and only 24 and seven, respectively, with predicted functions, belonging to 11 different COGs. The carbohydrate transport and metabolism category was the only one containing five genes, three being the maximum for the other categories.

Detection of positive selection

For *P. salamii*, we found no evidence of genome-wide positive or negative selection or a recent change in demography (Tajima's D = 0.148, not significantly different from zero, 95% confidence interval [-1.657; 2.153]; p-value: 0.432). By contrast, for *P. nalgiovense*, we found evidence of widespread selective sweeps, background selection or recent population expansion (Tajima's D = -2.834, 95% confidence interval [-4.641; -0.814]; p-value: 6e-08).

We looked for footprints of selection outside of HTR regions with a Bayesian generalization of the McDonald and Kreitman test [45], which identifies genes in which amino-acid changes are more or less frequent than expected under neutrality (inferring positive or purifying selection, respectively), by detecting higher or lower non-synonymous divergence from an

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

outgroup than expected based on observed synonymous and nonsynonymous SNPs in the in-group. We analyzed 7,155 genes present in all, or all but one of the strains of *P. nalgiovense*, and 7,010 genes present in all, or all but one of the *P. salamii* strains. We found that 1,572 of these genes in *P. nalgiovense* and 466 in *P. salamii* were evolving under purifying selection, and that 71 of these genes were common to the two species (Table S6). We identified 751 genes in *P. salamii* and 30 in *P. nalgiovense* evolving under positive selection, with only three of these genes common to the two species (one with no predicted function, one predicted to encode a phosphotransferase and the third predicted to be involved in meiosis; Table S6). The genes under positive or purifying selection displayed no specific clustering within genomes. We detected a significant excess of positive selection in *P. nalgiovense* for genes involved in secondary metabolite biosynthesis, transport, and catabolism, and in *P. salamii* for genes involved in transcription, posttranslational modification, protein turnover, chaperones, signal transduction and defense mechanisms (Table S6). We detected a significant excess of functions under purifying selection for functions involved in the transport and metabolism of carbohydrates and amino acids in *P. nalgiovense* and for functions involved in energy production and conversion, and in secondary metabolite biosynthesis, transport, and catabolism in *P. salamii* (Table S6).

Phenotypic experiments to test the hypothesis of domestication

We analyzed several phenotypic traits, to test the hypothesis that dry-cured meat fungi are domesticated: growth and spore production on malt extract agar medium at different salt concentrations, colony color on malt extract agar medium, and lipolytic and proteolytic activities. If dry-cured meat fungi have been domesticated, then we would expect to see: 1) significant differences between dry-cured meat and non-dry-cured meat strains within each

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

species, and 2) no significant differences between the non-dry-cured meat strains and the sister species not used for dry-cured meat production. These findings would indicate that only the fungi used for dry-cured meat production had evolved specific phenotypes. We found that *P. nalgiovense* and *P. salamii* contained horizontally transferred regions common to the *P. biforme* complex, which can also be used to inoculate dry-cured meat. We therefore also included *P. biforme* in phenotypic assays. We investigated the possibility that there were phenotypic differences between dry-cured meat and non-dry-cured meat strains within each of the species used for dry-cured meat maturation: *P. nalgiovense, P. salamii* and *P. biforme.* We also investigated whether there were differences between the non-dry-cured meat strains (the dry-cured meat strains were excluded from this analysis) and sister species not used for dry-cured meat production (i.e., *P. olsonii* for *P. salamii, P. chrysogenum* for *P. nalgiovense* and *P. palitans* for *P. biforme*). For the comparison of sister species, we included a "clade" effect in this test, with clades corresponding to sister-species pairs, to control for phylogenetic relatedness in the analysis.

Salt tolerance: effect of salt concentration on growth and spore number

In comparisons of growth between non-dry-cured meat populations and their sister species in the three clades under various salt concentrations, we found significant effects of salt concentration, clade and an interaction between clade and salt concentration, but no effect of species identity within a clade (Table 2, Fig. 4). Thus, the non-dry-cured meat strains and their sister species had similar responses to salt. All pairs of species grew faster in the presence of 2% salt than without salt, but growth rates rapidly decreased with increasing salt concentration beyond 2% (Table 2, Fig. 4).

In the analysis including dry-cured meat strains, we found no difference in growth between

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

populations from dry-cured meat and populations from other environments (Table 2, Fig. 4). We found a significant interaction between collection environment and salt concentration, a significant interaction between species and salt concentration and separate independent effects for salt concentration and species. This suggests that salt tolerance levels may differ between dry-cured meat and non-dry-cured meat strains (Table 2, Fig. 4). However, no post-hoc test of the impact of collection environments was significant, in any species. The strongest difference between different collection environments was for *P. salamii*, in which dry-cured meat strains grew better than non-dry-cured-meat strains in the absence of salt, although this difference was not significant in a post-hoc test (df =327; t=-3.285; p-value =0.1533).

We also measured spore production after 17 days of growth on malt extract agar medium containing various concentrations of salt. Spore numbers were very low on medium containing 18% salt. We therefore removed this category from the analysis. An analysis including only non-dry-cured meat strains and their sister species revealed a significant and strong effect of salt, with higher concentrations impairing spore production (Table 2, Fig. 4); we also found a weaker, but nevertheless significant within-clade species effect (Table 2, Fig. 4). In the analysis limited to dry-cured meat species, we found no difference between populations from dry-cured meat and those from other environments (Table 2, Fig. 4). We found a significant interaction between species identity and salt concentration (Table 2, Fig. 4). Significantly higher levels of sporulation were observed for *P. nalgiovense* than for *P. salamii*, but only in the total absence of salt (df=113, t=3.416, p-value=0.0240).

Proteolysis and lipolysis activity

We investigated whether the dry-cured meat strains had evolved specific proteolysis and/or lipolysis activities in *P. nalgiovense, P. salamii* and *P. biforme.* A comparison of species

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

without dry-cured meat strains revealed a significant effect of clade on both proteolysis and lipolysis (Table 2, Fig. 4). We detected no within-clade effect of species: the non-dry-cured meat populations and their sister species had similar levels of activity (Table 2, Fig. 4).

In the analysis including both dry-cured meat and non-dry-cured meat strains, the within-species effect of the collection environment was significant for both proteolysis and lipolysis, and there was also a significant species effect (Table 2, Fig. 4). Post-hoc tests showed that, in *P. biforme* and *P. salamii*, proteolysis was significantly slower in strains from dry-cured meat than in strains from other environments (Table 2). We observed a similar tendency in *P. nalgiovense*, but it was not significant. Lipolysis rates were also lower in strains from dry-cured meat than in those from other environments in *P. nalgiovense* and in *P. salamii*, but no such effect was detected in *P. biforme* (Table 2).

The differences between dry-cured meat and non-dry-cured meat strains, together with the lack of difference between the non-dry-cured meat populations and their sister species suggest that dry-cured meat strains have evolved specific catabolic abilities, constituting domestication (Table 2, Fig. 4).

Brightness and color

We tested whether colony color was closer to white in dry-cured meat populations, as this may have been a criterion of selection by humans. We found no significant species or clade effect in the analysis including only non-dry-cured meat strains and their sister species (Table 2, Fig. 5). In the analysis including all strains of the dry-cured meat species, we found a significant effect of species and of collection environment within species on the distance from whiteness in the RGB (red, green and blue) space (Table 2, Fig. 5). Post-hoc tests showed that

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

only *P. nalgiovense* colonies from dry-cured meat were significantly closer to white than colonies from other environments (Table 2). Visual inspection, indeed, suggested that the colonies on malt extract agar medium were typically white with a smooth surface in *P. nalgiovense* strains from dry-cured meat, whereas the colonies were more greenish, with a smooth or cotton-like surface in *P. salamii* (Fig. 5). These results indicate that colony color in the dry-cured meat population in *P. nalgiovense* has evolved toward greater whiteness.

We then investigated whether the dry-cured meat strains had evolved specific colony colors in terms of their red, green, and blue (RGB) components. In analyses limited to non-dry-cured meat strains and their sister species, the colors of fungal colonies were significantly affected by clade identity but not by the species within a clade (Table 2, Fig. 5). In analyses limited to dry-cured meat species, both collection environment and species had significant effects on colony color (Table 2, Fig. 5). Post-hoc tests showed that *P. biforme* dry-cured meat strains were redder and/or less blue (Table 2), whereas the differences were not significant in the other two species. Thus, *P. biforme* dry-cured meat strains have evolved to have less blue colonies, probably under human selection.

Mycotoxins and penicillin production

We detected none of the mycotoxins or penicillin assessed (andrastin A, ermefortins A & B, (iso)-fumigaclavin A, meleagrin, mycophenolic acid, ochratoxin A, penicillic acid, penicillin G, penitrem A, roquefortine C and sterigmatocystin) in any of the *P. salamii* and *P. nalgiovense* dry-cured meat strains tested (Table S7). None of these extrolites were detected in *P. biforme* in a previous study [21]. All the sequenced strains from both *P. salamii* and *P. nalgiovense* harbored genes for the entire pathway for penicillin production in their genomes (*PCBAB, PCBC, penDE* and *phl* genes, and the *RFX* transcription factor gene).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

None of the extrolites were detected in *P. olsonii*. We observed two distinct extrolite production profiles in *P. chrysogenum*, the sister species of *P. nalgiovense*. Five strains produced low levels of roquefortine C and no other extrolite (Table S7). The other nine *P. chrysogenum* strains produced high levels of andrastin A, meleagrin, roquefortine C and penicillin G.

Discussion

The use of dry-curing as a method for conserving a perishable raw meat material can be traced back to Greek and Roman times [46]. However, it remains unclear when the inoculation of dry-cured meat casings with *Penicillium* fungi began. *Penicillium nalgiovense* is the commonest species in the dry-cured meat food industry [30,47,48]. *Penicillium salamii* was recently identified as an alternative ripening agent and has only been used or recovered from local dry-cured meat production sites, often producing naturally fermented meat products [30,31,49]. This difference in usage was reflected in our samples from dry-cured meats, for which *P. nalgiovense* was the most frequently isolated species, followed by *P. salamii*. The wider use of *P. nalgiovense* may explain the difference in genetic diversity and the differentiation between the two species: the very low level of genetic diversity in *P. nalgiovense* may result from the selection of a single clonal lineage widely used in commercial dry-cured meat production, resulting in a strong bottleneck. Alternatively, the low level of genetic diversity may be explained by a possible relatively recent introduction of this species into Europe. Indeed, we had only four strains from outside Europe (USA and the Middle East), two of which were genetically different from the main lineage. *Penicillium nalgiovense* strains from non-food environments displayed no specific clustering, and most originated from human-related environments, such as the car upholstery, cheese, and dog bone. One of the potential "feral" strains (i.e. closely related to dry-cured meat strains) tested for all the phenotypic traits

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

did, however, present phenotypes close to those of non-dry-cured meat strains, suggesting that it was a genuine wild strain rather than a recent escape from an industrial environment. For the other potentially feral strains tested, it was not possible to rule out a feral origin entirely. *Penicillium salamii* had a higher level of genetic diversity, consistent with the lack of wide-scale commercial use for this species [32,33]. Genetic differentiation was observed in *P. salamii*, with clusters including different proportions of strains from dry-cured meats or other environments (e.g., sewage plant, tea leaves or soil), indicating that migration between dry-cured meat and other environments is still occurring.

We found no evidence of recombination in *P. nalgiovense*, with little linkage disequilibrium decay, and only one mating type among the dry-cured meat strains, consistent with the use of asexual cultures of commercial starters. In the cheese fungi *P. camemberti* and *P. roqueforti*, cheese populations also have low levels of genetic diversity and clonal population structures [21,22]. Non-cheese *P. roqueforti* populations have much higher levels of genetic diversity, with a larger number of recombination footprints, and higher levels of sexual fertility [50,51], contrasting with our findings for *P. nalgiovense*. However, the alternative mating type was detected in the non-dry-cured meat strains. In *P. salamii*, we identified both mating types among the dry-cured meat strains, suggesting that sexual reproduction may occur in this environment. However, the linkage disequilibrium patterns and the presence of only one mating type in each *P. salamii* population suggests an absence of recent sexual reproduction. The detection of both mating types in the two dry-cured meat species nevertheless opens up possibilities for future strain breeding.

Despite the lack of strong genetic differentiation between dry-cured meat strains and other strains, we found domestication footprints when comparing phenotypes relevant to dry-cured

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

meat production. Indeed, dry-cured meat strains of both *P. salamii* and *P. nalgiovense* displayed slower proteolysis and lipolysis than strains from other environments. *Penicillium biforme* presented the same pattern for proteolysis. Excessive protein degradation is detrimental for dry-cured meat products, potentially rendering the product less firm and imparting bitter and metallic tastes [37], whereas excessive lipolysis would dry the product [52] and have a negative impact on flavor. We found no major difference in lipolysis or proteolysis between the non-dry-cured meat strains of *P. salamii*, *P. nalgiovense* and *P. biforme* and their respective sister species, indicating that the dry-cured meat strains had evolved specific and convergent phenotypes, probably under human selection.

Fungal colonization of the surface of the meat casing has a direct impact on the observed color of the product, and this fungal coating has probably been subject to human selection for dry-cured meat fungal strains. Indeed, the visual appearance of the product is crucial for consumers, and a white color typically makes such products more attractive. *Penicillium nalgiovense* strains sampled from dry-cured meat were whiter than strains from other environments, consistent with the selection of dry-cured meat starters on the basis of color [53]. Dry-cured meat strains of *P. biforme* were less blue than non-dry-cured meat strains of this fungus. *Penicillium biforme* is widely used for cheese-making and is whiter than the closely related species *P. fuscoglaucum* on cheese media [21]. Whiter *P. biforme* strains may have been selected for both cheese and dry-cured meat production, especially as the strains isolated from these two types of product belong to the same population [21]. By contrast, we observed no difference in color between *P. salamii* strains from different environments, probably due to weaker or more recent human selection, possibly even with colonization of the product principally from the environment rather than through active inoculation [32], or due to contrasting selection pressures. Demand for local and traditional products is increasing,

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

including that for dry-cured meats [54], and this may have fostered an interest in less homogeneous and less white fungi on sausage casings. In the analysis limited to non-dry-cured meat strains, the lack of major color differences between the species used for dry-cured meat and their sister species indicated that the changes occurred specifically in dry-cured meat strains, further supporting the occurrence of domestication.

Phenotypic tests thus suggest that there may have been selection in *Penicillium* fungi for certain strains relevant to the production of dry-cured meats, with convergent evolution toward similar phenotypes, especially for lipolysis and proteolysis. By contrast, we detected no specific traits relating to the response to salt concentration in dry-cured meat strains. *Penicillium* fungi, particularly those present in fermented food, are salt-tolerant and, therefore, able to grow in the salt-rich environments found in dry-cured meats and cheeses [55]; there may have been no further advantage of higher levels of salt tolerance or evolutionary constraints may have prevented the development of this tolerance.

None of the targeted extrolites, including mycotoxins and penicillin, were detected in *P. nalgiovense* or *P. salamii* dry-cured meat strains [31]. However, the production of penicillin has been reported in some *P. nalgiovense* strains [48] and *P. chrysogenum*, its sister species, is known to produce mycotoxins [56,57]. The lack of antibiotic or mycotoxin production in dry-cured meat strains may be due to relaxed selection in a human-made environment containing fewer competitors or to active selection for safe products. In *Penicillium* cheese fungi, some of the mycotoxin biosynthesis pathway genes harbor deletions, impairing these functions specifically in cheese populations and indicating selection for safe food production or relaxed selection [21,58,59]. The absence of mycotoxin production in both *P. salamii* and its sister species *P. olsonii* suggests that this feature did not result from human selection in this clade. However,

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the culture medium is known to affect toxin production in *Penicillium* fungi [56].

We then investigated genomic footprints potentially associated with adaptation to dry-cured meat production. Given the clonal structure of the dry-cured meat strains, especially in *P. nalgiovense*, it was not possible to perform genome-wide scans for the detection of selection based on either local decreases in genetic diversity or local genomic differentiation between populations. Indeed, under clonal reproduction, the selection of a beneficial variant results in a hitchhiking of the whole genome without the generation of typical local selective sweeps or islands of genomic differentiation. We therefore looked for footprints of horizontal gene transfers, which have been shown to be an important and frequent mechanism of rapid adaptation under human selection in fungi [13,23,26,60,61]. By comparing available *Penicillium* genomes from various environments, we found evidence of multiple horizontal gene transfers, and, specifically, of horizontal transfers between *P. nalgiovense* and *P. salamii* despite the relatively large genetic distance between these species. We could not generate genealogical trees of these regions to support the inference of horizontal transfer because most species completely lack these regions, but these absences, together with the very high levels of identity between these regions and their different locations within the genome, provide even stronger support for the inference of horizontal transfer. By contrast to findings for cheese fungi, the HTRs were not specific to the dry-cured meat strains [13,23], which may be due to some non-dry-cured meat strains being feral strains or to gene flow between dry-cured meat and non-dry-cured meat populations. Several HTRs also appeared to be present in other species, mostly in *P. camemberti* and *P. biforme*, which are also used commercially in the food industry, mostly for the inoculation of cheeses, but also in dry sausages (this study [13,21,23]). The abundance of shared HTRs was higher in dry-cured meat species than in available *Penicillium* genome pairs, suggesting their possible selection in this rich, human-made

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

environment, as previously reported for cheese *Penicillium* fungi [13,23] and *Saccharomyces* yeasts used in wine-making [26]. We were unable to identify any particular functions overrepresented in these regions relative to the rest of the genome, possibly due to the low proportion of genes with predicted functions or to the diversity of functions under selection. Despite the low level of genetic differentiation between strains from dry-cured meat and other environments, we were able to detect phenotypic differences, even in *P. nalgiovense*, in which overall genetic diversity was low. However, a few SNPs may have a strong phenotypic impact, and the insertion of transposable elements may also modify gene regulation.

We ran tests to detect non-synonymous substitutions that were more or less frequent than would be expected under neutrality. *Penicillium nalgiovense* had few genes under positive selection and many under purifying selection, consistent with a high frequency of clonal reproduction in this species and, thus, probable genome-wide hitchhiking effects and background selection. The number of genes under positive selection was 25 times greater in *P. salamii* than in *P. nalgiovense*, and the genes concerned were involved in different functions. This finding may reflect the lower adaptation ability of this species, with low levels of genetic polymorphism. A few categories of genes seemed to be overrepresented among the genes under positive or negative selection, including some that could be linked to the dry-cured meat environment (carbohydrate metabolism and transport, defense mechanisms and secondary metabolite production).

Overall, our findings suggest that human selection has induced convergent phenotypic changes in *P. biforme*, *P. salamii* and *P. nalgiovense*, and that *Penicillium* dry-cured meat fungi have acquired beneficial traits for sausage production. These fungi have, therefore, been domesticated. The convergence in the evolution of dry-cured meat populations was most

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

striking for proteolysis, with lower levels of proteolysis in dry-cured meat populations for *P. nalgiovense*, *P. salamii* and *P. biforme*. Our findings, thus, add to the growing evidence of domestication in food fungi, such as in beer and wine *Saccharomyces* yeast, koji *Aspergillus oryzae* [10], and cheese *Penicillium* fungi [18,21,22,62,63]. Genomic analyses in *Saccharomyces cerevisiae* have revealed a differentiation of populations between wild and food environments, and even specific populations corresponding to different uses of yeast, for the fermentation of bread, beer, wine or cheese [18,63]. The different clades have contrasting fermentation traits, such as the capacity to use malt, aroma production and ethanol production. In the blue cheese fungus *P. roqueforti*, population differentiation was also observed between cheese and other environments, with a more rapid growth of cheese strains on cheese medium, higher levels of lipolytic activity, more efficient cheese cavity colonization, more diverse and positive aromas, and a higher salt tolerance [13,22,64]. In both *S. cerevisiae* and *Penicillium* cheese-making fungi, HGT events have been implicated in adaptation [13,23].

In conclusion, we found clear convergent adaptation in two distantly related dry-cured meat species of fungi, in terms of proteolysis and lipolysis, and in terms of shared horizontal gene transfers. Such studies on parallel adaptation to the same ecological niche are important for understanding the process of evolution and whether it is repeatable. However, only a few studies on a handful of models have investigated parallel adaptation. Studies on natural populations have shown that adaptation to similar ecological niches has led to the convergent evolution of phenotypes, such as size, color and form, in *Anolis* lizards, three-spined sticklebacks and cichlid fishes in African rift lakes, for example [6,7,65–67]. Dry-cured meat species of *Penicillium* therefore appear to be good models for investigating general questions about adaptation. We also found a substantial loss of diversity in dry-cured meat populations, as previously reported for cheese fungi [21,22]. This lack of diversity may jeopardize future strain improvement and lead to degeneration.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Material and Methods

Strain collection and DNA extraction

We used 94 strains for experiments and genome analyses including 22 isolated from dry-cured meat samples in this study (see below; Table S1), 33 from the CBS collection housed at the Westerdijk Fungal Biodiversity Institute (Utrecht, the Netherlands), 33 from the public collection of the national museum of natural history in Paris (LCP) and six from previous studies. For strain isolation, dry-cured meat casing samples were collected from Europe, Asia, and America. To avoid mite contamination, we first put dry-cured meat casing samples in the freezer (-20°C) for three days. After freezing treatment, we deposited a piece of dry-cured meat casing onto a malt extract agar (MEA) medium (16g malt, 16g agar, 800 mL H20) for two to four days until fungal colonies were observed. For each Petri dish, we scraped the entire fungal material and placed it into 1 mL of 0.05% of Tween 20 (Sigma-Aldrich); we spread 100 µL of a diluted suspension (10^{-3} dilution) on MEA medium. This dilution step was used to isolate single-spore colonies on Petri dishes. We then selected colonies with *Penicillium*-like morphologies (Fig. 5) and purified them separately on MEA medium. All cultures were stored in 10% glycerol at -80°C. The Nagoya protocol on biodiversity and shared benefits did not apply to our strain collection, as all strains were isolated prior to law publications in their respective countries of collection.

For DNA extraction, monospore-derived *Penicillium* strains were cultured on MEA medium for seven days. Then 3-5 ml of 0.05 % Tween 20 was added to the Petri dishes to collect fungal tissues. DNA was extracted using the NucleoSpin soil kit for genomic DNA (Macherey-Nagel) following manufacturer instructions. DNA quality was assessed by measuring 260/230 and 280/260 nm ratios with a NanoDrop 2000 spectrophotometer (Thermo

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fisher Scientific Inc), and DNA concentration was measured with a Qubit 2.0 fluorometer (Thermo Fisher Scientific Inc).

To identify the species isolated from dry-cured meat casings, *bt2a/bt2b* [68] primers were used to amplify a fragment of the β-tubulin gene, which is recognized as a discriminant marker for this purpose [39,69]. In addition, to identify the closely related species *P. biforme* and *P. fuscoglaucum*, previously grouped under the name *P. commune*, we used the PC4 and PC9 markers [70]. We distinguished the closely related species *P. chrysogenum* and *P. rubens* [71], by using the RPB2 (DNA-directed RNA polymerase II) marker [39]. PCRs were performed using a Bio-Rad thermal cycler with an initial denaturation of 2 min at 94°C, 40 cycles of 30 s denaturing at 94°C, 30 s annealing at 54°C and 40 s extension at 72°C with a final extension step at 72°C for 5 min. PCR fragments were sequenced at GENEWIZ Europe and sequences blasted (Blastn) against public databases (nucleotide collection nr/nt) for species identifications [72,73].

Genome sequencing and long-read assembly

For short-read sequencing, DNA libraries and sequencing were performed by the INRAE GenoToul platform (Toulouse, France). Paired-end libraries of 2x150 bp fragments were prepared with TruSeq Nano DNA Library Prep Kits (Illumina), and sequencing was performed on a HiSeq3000 Illumina sequencer, with a mean final coverage of 38x and a median of 10x (genomes of 27 to 35 Mb). We sequenced the genomes of 55 strains collected from either dry-cured meat or other environments (Table S1).

We used the P6/C4 Pacific Biosciences SMRT technology to obtain a high-quality reference genome sequence of the LCP06525 *P. salamii* strain and the ESE00252 *P. nalgiovense* strain, both from dry-cured meat. Nuclear genomic DNA was extracted with the cetyltrimethylammonium bromide (CTAB) extraction procedure described previously [23].

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Sequencing was performed by the UCSD IGM Genomics Facility La Jolla, CA, USA. The complete genome assemblies were performed using the SMRT analysis system v2.3.0 (Pacific Biosciences). Genome sequences were *de novo* assembled using the HGAP v3 [74] protocol, with a minimum seed read length of 1000, genome size of 30 Mb, target coverage of 20, and an overlap error rate of 0.03. Polished contigs were further error corrected using Quiver v1. A summary of raw data and assembly statistics is reported in Table S1. The protein-coding gene models were predicted with EuGene v4.2a trained for *Penicillium* [23,75,76] and their putative function with eggNOG-mapper online [77] and the eggNOG 4.5.1. general database [78].

Short-read assembly

Short-read sequencing data were checked using FastQC (v0.11.6) [79]. No adapter sequence could be detected. A first assembly round was performed using SOAPdenovo v2.04 [80] and a k-mer value of 71. To confirm species identity, we recovered part of the calmodulin and β-tubulin genes and of the ITS regions using blastn v2.2.29+ [73] and the corresponding *P. nalgiovense* genes as references (accessions JX996974.1, KX928936.1 and KC009791.1), then identified species by comparing the partial genes obtained with the NCBI nucleotide database (nt) using blastn v2.2.29+ [72,73]. We then compared all the assembled contigs and scaffolds to the NCBI nucleotide database (nt) using blastn v2.2.29+ [72,73] to identify those contaminated by other species.

Among the 55 genomes sequenced, we detected three *P. nalgiovense* samples (ESE00267, ESE00268 and ESE00262) as contaminated by *Debaryomyces hansenii*, a yeast almost systemically inoculated into dry-cured meats during production [81]. Among *P. salamii* genomic data, one strain (LCP06522) was contaminated by *Geotrichum candidum*, and another one (CCFL11ab2.1) by a fungal species belonging to Phanerochaetaceae. By comparing each strain assembly with itself, we did not detect any large-scale duplication or assembly errors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We removed the reads corresponding to contaminants identified in the previous step with BBsplit from the BBmap package v38.22 [82], keeping reads mapping on both the expected reference genome and the identified contaminant one. We also kept reads mapping on neither the reference genome nor the contaminant one to detect regions that would be genuinely present in some *Penicillium* genomes but absent from the reference genomes, such as horizontally transferred regions that would be specific to some strains. It is a rather conservative approach, as the annotation of the potential horizontally transferred regions identified in this study did not match the contaminants identified at this step (see Results). With the reads cleaned from contaminants, we did a second assembly round with SPAdes v3.11.0 [83] and default k-mer parameters (21, 33, 55, 77), and then removed scaffolds shorter than 2,000 bp. We checked assembly quality with QUAST v5.0.2 [84]. We performed pairwise genome alignments with NUCmer v3.1 and mummerplot v3.5 from the MUMmer3 package [85] with default parameters to detect large genomic rearrangements between strains within each species, and duplications within genomes. We predicted the genes as previously described for long-read assemblies. All the genomes were deposited on ENA under the study accession number PRJEB44534 (Table S1).

Population genetic analyses

We mapped cleaned sequencing reads to the reference genomes LCP06525 (*P. salamii*) and LCP06099 (FM193; *P. nalgiovense*; accession numbers HG815136–HG815288 and HG815290–HG816004 [13]) using Stampy with default parameters (v1.0.32) [86]. SNPs were called using the genome analysis toolkit GATK (v4.1.0.0) [87] with the *RealignerTargetCreator*, *IndelRealigner* and *HaplotypeCaller* functions using default parameters except that we used a ploidy of one. We filtered out variant call formats (VCF) files as follows: quality (with low qual-tag), quality by depth (QD < 2.0), mapping quality (MQ < 40), Fisher strand bias (FS >

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

60; a statistic to detect strand bias in variant calling in reverse and forward strands), mapping quality rank sum (MQRankSum < 12.5; to test read quality differences between reference and alternative alleles) and read position rank (ReadPosRank < 8.0; based on the distance from the end of reads compared with alternative allele). The filter-tags were prepared using vcftools with PASS/filtered tag and were further filtered using GATK with the *VariantFiltration* function. Indels were filtered out from the original VCF files using vcftools (v0.1.16) [88]. SNP data were used to analyze population structure and linkage disequilibrium. To produce an alignment for a phylogenetic analysis, we had to include the *P. nalgiovense* PacBio sequence (ESE00252) and the respective outgroups which were available as fasta files and therefore absent in the vcf file. We thus used NUCmer v3.1 from the MUMmer3 package [85] to align those sequences against the two reference genomes (LCP06099 for *P. nalgiovense* and LCP06525 for *P. salamii*) and called SNPs. We then merged those SNPs with the previous vcf files in R v3.6.2 [89]. We built population trees using RAxML (v8.2.7) [90], run with the general time-reversible model of nucleotide substitution with the gamma model of rate heterogeneity (GTRGAMMA) and 100 bootstraps.

In order to analyze population subdivision, we filtered our SNP dataset with bcftools (v1.7) [91] to keep only distant SNPs to avoid too strong LD (r^2 lower than 0.2 in a window of 1kbp), with a minor allele present in at least two individuals per species (minimal allele frequency of 0.05). We then ran STRUCTURE (v2.3.4) [92] with 20,000 iterations after 10,000 iterations as burn-in, for a number of populations (K) varying from 2 to 10, with 10 independent runs for each K. We combined the results with the R package pophelper v2.3.0 [93]. We used the full fasta alignment built to estimate the population tree to run a principal component analysis following the code used in Konishi et al. [94] and ran it in R v3.6.2 [89]. We performed neighbor-net analyses with the *phangorn* R package [95]. The substitution model used for building the distance matrix was JC69 [91,96]. The two most divergent *P. nalgiovense* strains, CBS297.97

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and DTO 204-F1, were removed from the neighbor-net analyses to better see the remaining relationships.

We calculated Tajima's D [97] using the *strataG* R package (v2.0.2) [98] and tested neutral evolution by comparing the 95% confidence interval and zero. We obtained nucleotide diversity indexes (π, the average number of pairwise nucleotide differences per site; and θ_h, the observed nucleotide diversity calculated from heterozygosity) using the same package and scaling to the length of the reference genome. We used vcftools to measure linkage disequilibrium (LD), in windows of 15,000 bp and with a minor allele frequency cutoff of 0.2 [88].

We searched for the mating-type locus in the genomes and identified the alleles as MAT1-1 or MAT1-2 with tblastx v2.6.0 and the *P. chrysogenum* strain ATCC02889 MAT1-1 (AM904544.1) and strain NRRL1249B21 MAT1-2 (AM904545.1) sequences as references. The presence of the two different mating types in a population could potentially allow sexual reproduction and recombination [50,99], while the presence of a single mating type suggests strict clonal reproduction.

Detection of horizontal gene transfers between *P. salamii* and *P. nalgiovense*

We aimed at detecting very recent horizontal gene transfers (HGTs) that could be involved in parallel adaptation to the dry-cured meat environment, having therefore occurred during domestication and thus human times. Genomic regions resulting from such recent HGTs were expected to be nearly identical between the two distant species *P. salamii* and *P. nalgiovense* in contrast to the genomic background, not necessarily present in all strains, and lacking in many non-dry-cured meat species (e.g. [23]). By the very nature of such recent HGTs, the genomic regions are lacking in many strains, which prevents using phylogenetic incongruence between different regions in the genomes to detect or validate HGTs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In order to identify HGTs, we first searched for thresholds above which we could consider that regions were more similar across a longer length than expected given the species phylogenetic distance. Given that the phylogenetic distance between *P. olsonii* and *P. chrysogenum* or *P. rubens* (two species formerly considered as one) is the same as the distance between *P. nalgiovense* and *P. salamii*, sequences that would be more similar and longer than the ones typically detected between *P. olsonii* and *P. chrysogenum* or *P. rubens* would not be expected given the phylogenetic distance separating *P. nalgiovense* from *P. salamii*, and would thus more likely correspond to recent gene flow, i.e. HGT or introgression. We therefore looked for blocks of 65-mer that were shared between sister species (i.e., 65 contiguous identical base pairs), which were later combined into longer regions if several 65 bp identical blocks were not separated by more than 90bp, and we retained such regions when they were at least 1,000 bp. We characterized the length and identity level of these genomic fragments that were nearly identical between sister species and obtained two distributions, one for the identity percentage and one for the length, by pooling the values of the comparisons between *P. olsonii* and *P. chrysogenum* on the one hand and *P. olsonii* and *P. rubens* on the other hand. We considered the maximal identity percentage as the identity threshold for identifying HGTs. Because the longest sequence in the distribution was two times longer than the second longest one, we used the second maximal value of sequence length as the length threshold for identifying HGTs. These two threshold values for length and identity did not have to come from the same sequence and were considered independently. When comparing *P. salamii* and *P. nalgiovense* genomes, genomic regions longer and more similar than the two threshold cut-off values were therefore unexpected given their phylogenetic distance. The thresholds were highly conservative, to avoid false positives as much as possible, as we took the most extreme values of the distributions. All the pairwise comparisons were performed

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

using NUCmer v3.1 from the MUMmer3 package [85], keeping all anchored matches even in cases of multiple matches (--maxmatch option).

We then compared the genomes of each *P. nalgiovense* strain to each *P. salamii* strain pairwise with the same method as above and identified sequences shared between strains of the two species meeting the length and identity criteria, which we considered as putative horizontally-transferred regions (HTRs). The sequences of the putative HTRs retrieved from the different genomes were mapped against each other using the "map to reference" algorithm in Geneious v9.1.8 (Biomatters Ltd., Auckland, New Zealand) and a consensus sequence was kept when they overlapped to obtain a set of reference HTRs.

As transposable elements (TEs) could be involved in horizontal gene transfers [100], we checked whether particular transposable elements were present in the putative HTRs identified, and in particular at their margins. For this purpose, we identified TE present in *P. salamii* and *P. nalgiovense* reference genomes (LCP06525 and LCP06099-FM193) with RepeatModeler v1.0.11 [101]. This software identifies TE families based on high frequency patterns but also relationship between the different elements and compare them to the Dfam library (v3.0)[102], allowing some of the families to be formerly associated with known TE families in the five species present in the library (human, mouse, zebrafish, fruit fly, nematode), like LINE, LTR-Copia… We analyzed the putative HTRs with two databases: one based on all the TE identified in the *P. salamii* and *P. nalgiovense* reference genomes, and one based only on repeats associated to known TE families. Given the absence of fungi reference in the Dfam library, it is highly probable that the TE identified but not associated with any known family are indeed TE, simply not annotated, but we cannot entirely rule out the possibility of any other duplicated sequence. We therefore looked at the 76 genes identified in these regions, none of them could be annotated. To take this small uncertainty into consideration, we either used the entire HTRs or masked all the TE present in the first database with RepeatMasker

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

v4.0.7 [103] for analyses not specifically focusing on TEs and mentioned those HTRs as "repeat-masked". When the second database was used, masked sequences of the putative HTRs were identified as "TE-masked".

We then searched for the presence of the reference set of HTRs in all *P. nalgiovense* and *P. salamii* genomes using blastn v2.2.29+ [72], as well as in 68 other genomes from 54 species (four *Aspergillus* spp., *Monascus purpureus* and 49 other *Penicillium* spp.), using two different genomes for each species whenever possible (Table S1-S2). One genome of each of *P. biforme* and *P. egyptiacum*, and five genomes of *P. olsonii* were assembled from Illumina HiSeq reads for this analysis, with the method described above (Table S1). The other genomes were retrieved from public databases (Table S2). When hits were obtained with previously identified HTRs, we checked the proportion of each putative HTR present in the genome, its percentage of identity with the reference HTR sequence, its position in the genome by looking at flanking regions and whether it was fragmented on different scaffolds. To confirm the split and avoid confusion with potential assembly errors, we considered a region to be split only when the region was present on two different scaffolds and distant by at least 10% more base pairs than the distance separating the two parts in the reference sequence.

To assess whether horizontal gene transfers were more frequent between *P. nalgiovense* and *P. salamii* than between two random pairs of *Penicillium* species, we compared the genomes of the 54 other species studied above in a pairwise fashion to identify HTRs using the same method as above. To obtain the phylogenetic distances between species pairs, we kept a single genome for each of the 54 species (Table S2), we extracted the coding sequences and grouped them into orthologous groups. To identify orthologs, we compared their protein sequences with blastp v2.6.0+ [72] and grouped them by similarity using orthAgogue [104], choosing the markov clustering inflation parameter values that maximized the number of clusters containing 54 sequences (i.e one gene per species, I=2). We aligned the gene nucleotide

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sequences of single-copy orthologs using translatorX v1.1 and default parameters (alignment of the translated sequences with Muscle and cleaning with GBlocks and default parameters)[105]. We then built a phylogenetic tree for each ortholog with RAxML v8.2.11, using the GTRGAMMAI model, starting from 20 independent trees and evaluated through 200 bootstraps [90]. We used these gene trees as input to build a species tree with Astral v5.6.3. [106]. We used the function cophenetic in the package *ape* [107] run on R v3.6.2 [89] to extract the distance between species pairs.

We plotted the cumulative lengths of the putative HTRs identified between species pairs as a function of the phylogenetic distance between species with the packages *ggplot* v3.3.0 and *ggrepel* v0.8.2 on R v3.6.2 [89].

To identify functions that could be enriched in the putative HTRs shared by *P. nalgiovense* and *P. salamii* compared to the rest of the genome, and therefore be potentially beneficial in dry-cured meat, the common method would be to perform an enrichment test in gene ontology (GO) function. However, because most genes present in these regions could not be annotated with GO terms, we used COG functional categories (clusters of orthologous groups) from eggNOG [77,78] and performed comparisons through Fisher's exact tests given the small number of functions considered; we tested unilaterally for the over-representation of a particular function in HTR gene sets. To do so, we used the function fisher.test() in the package *stats* in R v3.6.2 [89]3/25/2022 7:56:00 AM.

To illustrate the presence, length, and similarity of HTRs in genomes, we clustered the regions based on the Euclidean distance between the length of the regions across strains, based on the default complete linkage clustering method. We plotted heatmaps using the heatmap.2 function of the *gplots* v3.0.1.1 package [108] in R [89].

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Detection of selection: SnIPRE analyses

Because the two species studied here are used in industrial contexts [28–30,33], and therefore potentially cultured in high quantities after an initial bottleneck, we cannot assume a constant effective population size through time. Even more importantly, these fungi are cultivated asexually to be sold as starters in the industry, therefore constituting clonal lineages (see results). Most of the classic frameworks for detecting genes under positive selection are therefore unsuitable. We decided to use SnIPRE [109], a Bayesian generalization of the McDonald and Kreitman analyses [45], which does not make any assumption on demography or reproduction mode. This method detects genes in which amino-acid changes are more or less frequent than expected under neutrality (inferring positive or purifying selection, respectively), by identifying genes in which non-synonymous divergence to an outgroup is higher or lower than expected based on observed synonymous and nonsynonymous SNPs in the in-group, thus accounting for gene-specific mutation rates. We ran the tests for *P. nalgiovense* using *P. chrysogenum* as an outgroup and for *P. salamii* using *P. olsonii* as an outgroup. For this goal, we applied the method described above for the phylogeny reconstruction to obtain single-copy genes present in all the *P. nalgiovense* strains and five *P. chrysogenum* on one dataset, and in all the *P. salamii* strains and five *P. olsonii* on a second dataset, allowing one missing sequence per alignment. From these orthologous gene alignments, we recovered the number of synonymous and non-synonymous sites, and among each category the number of fixed and polymorphic sites in each of *P. salamii* and *P. nalgiovense* gene with the MKtest program of libsequence [110]. We estimated the mean number of synonymous and polymorphic sites by performing a weighted average on the polydNdS output, these two functions belonging to the package “analysis” of libsequence [110]. We ran the SnIPRE analysis in its Bayesian implementation and ran it for 25,000 iterations and remove a burn-in of 10,000 iterations,

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

with a thinning value of four. To test whether the proportion of genes under positive, negative, or neutral selection was different between one specific COG functional category and the rest of the genome, we used Fisher's exact tests as previously described. As strains from dry-cured meat did not cluster together neither in the phylogenetic nor in the population genomic analyses (see results), we did not compare this population with others within each species.

Phenotypic tests: mycelium growth, salt tolerance, lipolysis, proteolysis, color, and spore numbers

We compared various phenotypic traits between species and populations from different environments: growth rate at different salt concentrations, spore production, lipolysis, and proteolysis abilities. We first studied *P. nalgiovense* and *P. salamii*, the most common species inoculated for dry-cured meat production, with their respective sister species *P. chrysogenum* and *P. olsonii*. However, we found shared horizontally-transferred regions between these dry-cured meat species and the *P. biforme* complex, including the two clonal lineages *P. camemberti* var. *camemberti* and *P. camemberti* var. *caseifulvum* [21], and these can also be, even if more rarely, inoculated in dry-cured meat. We therefore also analyzed phenotypes in *P. biforme*, that display genetic and phenotypic variability, in contrast to the cheese clonal *P. camemberti* varieties [21], and compared them to phenotypes of their sister species *P. palitans.* For all the experiments below, we used all available *P. nalgiovense*, *P. salamii* and *P. biforme* strains that we could find in public databases, from non-food environments, and similar numbers of strains, chosen at random, from dry-cured meat and from their closely related outgroup species. The Table S1 gives the number and identity of strains used for each species, type of environment and experiment.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

For assessing salt tolerance, we used seven *P. chrysogenum* strains, 14 *P. nalgiovense* strains, eight *P. olsonii* strains, 17 *P. salamii* strains, 13 *P. biforme* strains and five *P. palitans* strains. Because the salt content of dry-cured meat products usually ranges between 2% to 5% of dry weight and can be much higher near the surface [111,112], we measured growth rates on malt extract agar medium with various salt concentrations: 2% (2g NaCl/100 mL), 10% (10g NaCl/100 mL) and 18% (18g NaCl/100 mL). Malt extract agar medium was prepared with 16 g malt extract powder (Difal), 16g agar dissolved in 800 mL ddH_2O and adjusted to pH 5.7 which is in the pH range of dry-cured meat products (pH 4.5 – 6)[112]. We poured 30 mL of medium into Petri dishes (150 mm diameter). For inoculation, *Penicillium* spore suspensions were calibrated by counting spores in a Malassez hemocytometer cell using six squares (0.2 mm, two repeats) under the microscope. Spore number estimates were highly correlated between repetitions ($r^2 = 0.94$, df = 63, p-value < 2.2e-16). We then diluted spores to a concentration of 6 $x10^5$ $spores.mL^{-1}$ and deposited 10 µL in the center of each Petri dish before incubating plates at room temperature. We recorded growth by taking photos under constant conditions in terms of light and position of Petri dishes, at days 8 and 17. We used the Image J package Fuji [69,113] to measure colony area on the pictures using Trainable Weka Segmentation v3.2.23 [114]. The software was trained to recognize background, medium and colony areas on the picture. The area estimates in pixel units were transformed into square centimeters using the size of the Petri dish. In the last step of the salt tolerance experiment, we collected all fungal material from each Petri dish in an Eppendorf tube filled with 0.05 % Tween® 20 water, which was then diluted (dilution 1:100). Spore numbers were estimated from these solutions under a microscope (Nikon ECLIPSE TS100, 40X) with a Malassez hemocytometer using the same protocol as above after 17 days of growth.

The lipolytic and proteolytic activities were tested with the same *Penicillium* strains as salt tolerance tests. Lysis activities were measured as follows: a suspension adjusted to 2,500

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

spores for each strain was inoculated on the top of a test tube containing tributyrin (triglyceride) agar for lipolytic activity measurements (10 $mL.L^{-1}$, ACROS Organics, Belgium) or semi-skimmed milk for proteolytic activity measurements (40 $g.L^{-1}$, from large retailers). Lipolytic and proteolytic activities were estimated on days 7, 14, 21 and 28 by determining extent of compound degradation as the media changes from opaque to translucent. Experiments were divided into three batches (N=26, 30 and 8 strains respectively for each test period).

To evaluate fungal colony color, spore suspensions of all available *Penicillium* strains (n=78; *P. chrysogenum* n=7, *P. nalgiovense* n=27, *P. olsonii* n=8, *P. salamii* n=22, *P. biforme* n=9 and *P. palitans* n=5; Table S1) were inoculated onto malt extract agar medium (70164 Sigma-Aldrich) and calibrated photos were taken after 18 days growth at 25°C using a ScanStation 100 (Interscience, Saint Nom, France). Images were analyzed using Image J packageFuji [69,113] to record the color decomposition among red, green and blue colors (RGB system). Levels of red, green, and blue are represented by a range of integers from 0 to 255 (256 levels for each color). White is 255,255,255 and black 0,0,0. Color was measured in three different circles on Petri dishes, each circle being of 0.5 cm diameter, located respectively in the center, edge of the colony and midway between these two points. RGB color values were averaged among pixels within each circle. As colony whiteness potentially represents a trait under selection by humans, we computed and analyzed the degree of whiteness by computing the distance to white, *i.e.*, the Euclidean distance in the 3-D orthogonal coordinates composed of red, blue and green values. We tested whether the colors of *P. biforme, P. nalgiovense* and *P. salamii* were closer to white than that of their sister species not used for dry-cured meat production, and whether, within each species, dry-cured meat strains were whiter than non-dry-cured meat strains. Both strains and the position of the measured circle were considered as random factors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Statistical analyses of the phenotypic measures

We analyzed several phenotypic traits to test our hypothesis that dry-cured meat fungi are domesticated: growth and spore production at different salt concentrations using malt extract agar medium, colony color on malt extract agar medium, and lipolytic and proteolytic activities. If dry-cured meat fungi have been domesticated, we expected that: 1) there would be significant differences between dry-cured meat and non-dry-cured meat strains within each species, and 2) there would be no significant differences between the non-dry-cured meat strains and the sister species not used for dry-cured meat production; this would indeed suggest that only the fungi used for dry-cured meat production have evolved specific phenotypes. Because we found that *P. nalgiovense* and *P. salamii* shared horizontally transferred regions with the *P. biforme* complex, which can also be inoculated in dry-cured meat, we included *P. biforme* in phenotypic assays. We therefore tested first whether there were significant phenotypic differences between dry-cured meat and non-dry-cured meat strains within each of the species used for dry-cured meat maturation, i.e., *P. nalgiovense*, *P. salamii* and *P. biforme*. Then, we tested whether there were significant differences between the non-dry-cured meat strains (excluding dry-cured meat strains from this analysis) and the sister species populations, not used for dry-cured meat production (i.e., *P. olsonii* for *P. salamii*, *P. chrysogenum* for *P. nalgiovense* and *P. palitans* for *P. biforme*). To be able to compare sister species, we included in the tests a "clade" effect, clades corresponding to sister species pairs, to control for the phylogenetic relatedness in the analysis.

For all analyses, fixed effects were added sequentially to test for their significance through likelihood ratio tests until reaching the full model with all possible fixed effects. We thereafter only introduce the full model. All post-hoc tests were conducted using the emmeans function of package emmeans [115] with Tukey correction for pairwise mean comparison. Post-hoc tests

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

were run when we found a significant effect either between sister species or between populations from dry-cured meat and non-dry-cured meat.

For analyzing growth at different salt concentrations, we fitted two linear mixed models on the log-transformed colony area, keeping the day of measurement as a random variable and the salt concentration as fixed effects in each of them. In the first one, testing the differences between strains from dry-cured meat and other environments, we further included as fixed effects the species and its interaction with salt concentration, as well as the collection environment (dry-cured meat or not) nested within species and its interaction with salt concentration. In the second one, where we tested the differences between non-dry-cured meat strains and the sister species, we included as fixed effects the clade and its interaction with salt concentration, as well as the species nested within the clade and its interactions with salt concentration.

For quantifying spore number after growth at various salt concentrations, as too few spores were recovered at 18% salt concentration, we only analyzed growth data for 0%, 2% and 10% salt concentration. As spores were only counted once (after the 17-day growth experiment), we did not include any random effect in the model. We therefore fitted two linear models using as a dependent variable the log-transformed number of spores divided by the colony area to normalize the effect of growth on sporulation.

For lipolysis and proteolysis analyses, we fitted two linear mixed models on the distance to the lysis zone using the day of measurement and the batch number as random effects. To test for differences between dry-cured meat and other environment strains, we used as fixed effects species and collection environment (dry-cured meat or not) nested within species. To test for differences between non-dry-cured meat populations and the sister species, we included as fixed effects the clade and the species nested within the clade.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

For colony whiteness, we fitted two linear mixed models using the same fixed effects as for the two models used for proteolysis and lipolysis tests. As three color measurements per strain were done, we added the measured circle location and strain ID as random effects. For colony color, because red and green values were strongly correlated (r=0.998), we only used red and blue components in RGB space to fit generalized linear mixed models with a binomial response. We considered as random effects the measured circle location and strain ID. Each model used the same fixed effects as those for proteolysis and lipolysis tests.

All linear mixed models were fitted using the lmer() function of *lme4* R package [89,116] and glmer() for the generalized linear mixed models. Simple linear models were fitted with lm() function [89]. Likelihood ratio tests between models were performed using anova() function in R [89]. For all tests, we first checked that assumptions were met, i.e., homoscedasticity, independence of residuals, and normal and identical distribution of residuals.

Mycotoxins and penicillin production

For mycotoxin production, we grew fungal cultures in 24-well sterile microplates containing 2 mL agar medium per well. For each strain, 1 µL of a calibrated spore suspension (10^6 spores.mL^{-1}) prepared from a 7-day culture was inoculated in the center of yeast extract sucrose (YES) agar medium buffered at pH 4.5 with phosphate-citrate buffer and characterized by a high C/N ratio to favor mycotoxin production as already described by Gillot et al [58]. Six replicates per strain were performed, three for mycotoxin analyses and three for fungal dry-weight measurements. In the latter case, growth was performed on cellophane disks to collect fungal mycelium. The plates were incubated at 25°C in the dark for seven days and then stored at -20°C until mycotoxin analysis.

For mycotoxin extractions, an optimized “high-throughput” extraction method based on the one described by Gillot et al [58] was used. Briefly, 2g-aliquots (the entire YES culture obtained

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

from a well) were homogenized after thawing samples with a sterile flat spatula then 12.5 mL of acetonitrile (ACN) supplemented with 0.1% formic acid (v/v) was added, samples were vortexed for 30 sec followed by 15 min sonication. Then, extracts were again vortexed before being centrifuged for 10 min at 5000g at 4°C. The supernatants were directly collected and filtered through 0.45 µm PTFE membrane filters (GE Healthcare Life Sciences, UK) into amber vials and stored at -20°C until analyses.

Mycotoxin detection and quantification were performed using an Agilent 6530 Accurate-Mass Quadropole Time-of-Flight (Q-TOF) LC/MS system equipped with a Binary pump 1260 and degasser, well plate autosampler set to 10°C and a thermostated column compartment. Filtered 2µL aliquots were injected into a ZORBAX Extend C-18 column (2.1x50 mm and 1.8 µm, 600 bar) maintained at 35°C with a flow rate set to 0.3 $mL.min^{-1}$. The mobile phase A contained milli-Q water + 0.1% formic acid (v/v) and 0.1% ammonium formate (v/v) while mobile phase B was ACN + 0.1% formic acid. Mobile phase B was maintained at 10% for 4 min followed by a gradient from 10 to 100% for 18 min. Then, mobile phase B was maintained at 100% for 5 min before a 5-min post-time. Samples were ionized in both positive (ESI+) and negative (ESI-) electrospray ionization modes in the mass spectrometer with the following parameters: capillary voltage 4 kV, source temperature 325°C, nebulizer pressure 50 psig, drying gas 12 $L.min^{-1}$, ion range 100-1000 m/z. Target extrolite characteristics used for quantifications are given in Table S3 and included commercially available extrolites produced by *Penicillium* species: andrastin A, ermefortins A & B, (iso)-fumigaclavin A, meleagrin, mycophenolic acid, ochratoxin A, penicillic acid, penicillin G, penitrem A, roquefortin C, sterigmatocystin. Andrastin A, ermefortins A & B and (iso)-fumigaclavin A standards were obtained from Bioviotica (Goettingen, Germany), penitrem A from Cfm Oskar Tropitzsch (Marktredwitz, Germany), while all others were from

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Sigma-Aldrich (St Louis, MO, USA). All stock solutions were prepared in dimethyl sulfoxide (DMSO) at 1 $mg.mL^{-1}$ in amber vials.

For these analyses, metabolite identification was performed using both the mean retention time ± 1 min and the corresponding ions listed in Table S3. We used a matrix matched calibration curve (R^2 >0.99 for all extrolites except 2 >0.96) for reliable mycotoxin quantification with final concentrations ranging from 1 to 10000 $ng.mL^{-1}$ according to the target metabolite and method performance was carried out as previously described by Gillot et al [58]. Mycotoxin concentrations were calculated from the equation y = mx+b, as determined by weighted ($1/x^2$) linear regression of the matrix-matched calibration data and correlated to calculated fungal growth areas. Specific mycotoxin production was expressed as ng per g of fungal dry weight.

Data accessibility

All sequencing data have been deposited to the European Nucleotide Archive under the accession PRJEB44534. The HTR sequences and their annotations are available under the publicly available gitlab project https://gitlab.com/abranca/consensus_htr.

Acknowledgments

This work was supported by the ERC starting grant GenomeFun 309403 and Louis D foundation award to TG, a CNRS PEPS adaptation grant to AB and the FUNGADAPT ANR□19□CE20□0002 ANR grant to TG, AB and MC. YCL acknowledges a PhD grant from the France–Taiwan joint program Campus France. We thank Alice Feurtey, Fanny Hartmann and Fantin Carpentier for help with genome analyses. We thank MNHN for

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

granting access to strains with the help of Manuela Lopez-Villavicencio and Sandrine Lacoste. We thank all the people who sent dry-cured meat casings from around the world. We thank the GenoToul platform for sequencing. We are also grateful to the GenoToul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi: 10.15454/1.5572369328961167E12) for providing some computing resources. PacBio sequencing was conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA. We thank Federico Laich and Giancarlo Perronne for providing *P. salamii* strains. Some sequence data (Table S2) were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community.

Author contributions

TG and AB acquired funding, designed and supervised the study. JD and JH provided strains. RD, YCL and AS produced the genomes. JB, YCL, AB and RCdlV analyzed the genomes. JR produced the splitstree and pictures. YCL, AS and SLP isolated strains. YCL, AS, MLP and MC performed the phenotype experiments. YCL, JB, AB and TG analyzed the data from the experiments. YCL, JB and TG wrote the manuscript with contributions from the other authors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References:

1. Darwin, C. (1868). The Variation of Animals and Plants Under Domestication (London, UK: John Murray).

2. Darwin, C. (1859). The origin of species (London, UK: John Murray).

3. Hufford, M.B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R.A., Elishire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppler, S.M., et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. *44*, 808–811.

4. Stitzer, M.C., and Ross-Ibarra, J. (2018). Maize domestication and gene interaction. New Phytol. *220*, 395–408.

5. Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. Science *327*, 302.

6. Colosimo, P.F., Peichel, C.L., Nereng, K., Blackman, B.K., Shapiro, M.D., Schluter, D., and Kingsley, D.M. (2004). The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. PLOS Biol. *2*, e109.

7. Duboué, E.R., Keene, A.C., and Borowsky, R.L. (2011). Evolutionary Convergence on Sleep Loss in Cavefish Populations. Curr. Biol. *21*, 671–676.

8. Maruyama, T., and Fuerst, P.A. (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by recent bottleneck. Genetics *111*, 675.

9. Nei, M., Maruyama, T., and Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. Evolution *29*, 1–10.

10. 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., and Rokas, A. (2012). The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. Curr. Biol. CB *22*, 1403–1409.

11. Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguileta, G., M., V.D., C., R. de la V.R., Sara, B., and Tatiana, G. (2013). Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. Mol. Ecol. *23*, 753–773.

12. Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., et al. (2018). Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. Mol. Biol. Evol. *35*, 1712–1727.

13. Ropars, J., Rodríguez de la Vega, R.C., López-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, É., Lacoste, S., Debuchy, R., Dupont, J., Branca, A., et al. (2015). Adaptive

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. Curr. Biol. *25*, 2562–2569.

14. Wolfe, B.E., and Dutton, R.J. (2015). Fermented Foods as Experimentally Tractable Microbial Ecosystems. Cell *161*, 49–55.

15. Watarai, N., Yamamoto, N., Sawada, K., and Yamada, T. (2019). Evolution of *Aspergillus oryzae* before and after domestication inferred by large-scale comparative genomic analysis. DNA Res. *26*, 465–472.

16. Dupont, J., Dequin, S., Giraud, T., Le Tacon, F., Marsit, S., Ropars, J., Richard, F., and Selosse, M.-A. (2017). Fungi as a Source of Food. Microbiol. Spectr. *5*.

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

18. Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., et al. (2016). Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. Cell *166*, 1397-1410.e16.

19. Gonçalves, C., Wisecaver, J.H., Kominek, J., Oom, M.S., Leandro, M.J., Shen, X.-X., Opulente, D.A., Zhou, X., Peris, D., Kurtzman, C.P., et al. (2018). Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. eLife *7*, e33034.

20. Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature *556*, 339–344.

21. Ropars, J., Didiot, E., Rodríguez de la Vega, R.C., Bennetot, B., Coton, M., Poirier, E., Coton, E., Snirc, A., Le Prieur, S., and Giraud, T. (2020). Domestication of the emblematic white cheese-making fungus *Penicillium camemberti* and its diversification into two varieties. Curr. Biol. *30*, 4441–4453.

22. 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., et al. (2020). Independent domestication events in the blue-cheese fungus *Penicillium roqueforti*. Mol. Ecol. *29*, 2639–2660.

23. Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A.-L., Ceppi, M., Conseiller, E., Debuchy, R., et al. (2014). Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat. Commun. *5*, 2876.

24. Marsit, S., and Dequin, S. (2015). Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. FEMS Yeast Res. *15*, fov067.

25. Guillamón, J.M., and Barrio, E. (2017). Genetic Polymorphism in Wine Yeasts: Mechanisms and Methods for Its Detection. Front. Microbiol. *8*, 806.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

26. Marsit, S., Mena, A., Bigey, F., Sauvage, F.-X., Couloux, A., Guy, J., Legras, J.-L., Barrio, E., Dequin, S., and Galeote, V. (2015). Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. Mol. Biol. Evol. *32*, 1695–1707.

27. Marsit, S., Leducq, J.-B., Durand, É., Marchant, A., Filteau, M., and Landry, C.R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. Nat. Rev. Genet. *18*, 581.

28. Garcia, M.L., Casas, C., Toledo, V.M., and Selgas, M.D. (2001). Effect of selected mould strains on the sensory properties of dry fermented sausages. Eur. Food Res. Technol. *212*, 287–291.

29. Ludemann, V., Pose, G., Pollio, M.L., and Segura, J. (2004). Determination of growth characteristics and lipolytic and proteolytic activities of *Penicillium* strains isolated from Argentinean salami. Int. J. Food Microbiol. *96*, 13–18.

30. Sunesen, L.O., and Stahnke, L.H. (2003). Mould starter cultures for dry sausages—selection, application and effects. Meat Sci. *65*, 935–948.

31. Coton, M., Deniel, F., Mounier, J., Joubrel, R., Robieu, E., Pawtowski, A., Jeuge, S., Taminiau, B., Daube, G., Coton, E., et al. (2021). Microbial Ecology of French Dry Fermented Sausages and Mycotoxin Risk Evaluation During Storage. Front. Microbiol. *12*, 737140–737140.

32. Perrone, G., Samson, R.A., Frisvad, J.C., Susca, A., Gunde-Cimerman, N., Epifani, F., and Houbraken, J. (2015). *Penicillium salamii*, a new species occurring during seasoning of dry-cured meat. Int. J. Food Microbiol. *193*, 91–98.

33. Magistà, D., Ferrara, M., Del Nobile, M.A., Gammariello, D., Conte, A., and Perrone, G. (2016). *Penicillium salamii* strain ITEM 15302: A new promising fungal starter for salami production. Int. J. Food Microbiol. *231*, 33–41.

34. Chávez, R., Fierro, F., García-Rico, R.O., and Laich, F. (2011). Mold-fermented foods: *Penicillium* spp. as ripening agents in the elaboration of cheese and meat products.

35. Nielsen, M.S., Frisvad, J.C., and Nielsen, P.V. (1998). Protection by fungal starters against growth and secondary metabolite production of fungal spoilers of cheese. Int. J. Food Microbiol. *42*, 91–99.

36. Leroy, F., Verluyten, J., and De Vuyst, L. (2006). Functional meat starter cultures for improved sausage fermentation. Int. J. Food Microbiol. *106*, 270–285.

37. Toldrá, F. (1998). Proteolysis and lipolysis in flavour development of dry-cured meat products. Meat Sci. *49*, S101–S110.

38. Filtenborg, O., Frisvad, J.C., and Thrane, U. (1996). Moulds in food spoilage. Specif. Spoilage Org. *33*, 85–102.

39. Visagie, C.M., Houbraken, J., Frisvad, J.C., Hong, S.-B., Klaassen, C.H.W., Perrone, G., Seifert, K.A., Varga, J., Yaguchi, T., and Samson, R.A. (2014). Identification and nomenclature of the genus *Penicillium*. Stud. Mycol. *78*, 343–371.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

40. Houbraken, J., Kocsubé, S., Visagie, C.M., Yilmaz, N., Wang, X.-C., Meijer, M., Kraak, B., Hubka, V., Bensch, K., Samson, R.A., et al. (2020). Classification of *Aspergillus*, *Penicillium*, *Talaromyces* and related genera (*Eurotiales*): An overview of families, genera, subgenera, sections, series and species. Genera Hyphomycetes Tribute Keith Seifert *95*, 5–169.

41. Steenwyk, J.L., Shen, X.-X., Lind, A.L., Goldman, G.H., and Rokas, A. (2019). A Robust Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the Genera *Aspergillus* and *Penicillium*. mBio *10*, e00925-19.

42. Pozzi, L., Hodgson, J.A., Burrell, A.S., Sterner, K.N., Raaum, R.L., and Disotell, T.R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. *75*, 165–183.

43. Reis, M.D., Gunnell, G.F., Barba-Montoya, J., Wilkins, A., Yang, Z., and Yoder, A.D. (2018). Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. Syst. Biol. *67*, 594–615.

44. Houbraken, J., de Vries, R.P., and Samson, R.A. (2014). Chapter Four - Modern Taxonomy of Biotechnologically Important *Aspergillus* and *Penicillium* Species. In Advances in Applied Microbiology, S. S. and G. G.M., eds. (Academic Press), pp. 199–249.

45. McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. Nature *351*, 652–654.

46. Frost, F.J. (1999). Sausage and Meat Preservation in Antiquity. Greek Roman Byzantine Stud. *40*, 241–252.

47. Larsen, T.O., and Breinholt, J. (1999). Dichlorodiaportin, Diaportinol, and Diaportinic Acid: Three Novel Isocoumarins from *Penicillium nalgiovense*. J. Nat. Prod. *62*, 1182–1184.

48. Andersen, S.J., and Frisvad, J.C. (1994). Penicillin production by *Penicillium nalgiovense*. Lett. Appl. Microbiol. *19*, 486–488.

49. Sinch, B.J., and Dincho, D. (1994). Molds as Protective Cultures for Raw Dry Sausages. J. Food Prot. *57*, 928–930.

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

51. Ropars, J., Rodríguez de la Vega, R.C., López-Villavicencio, M., Gouzy, J., Dupont, J., Swennen, D., Dumas, E., Giraud, T., and Branca, A. (2016). Diversity and Mechanisms of Genomic Adaptation in *Penicillium*. In *Aspergillus* and *Penicillium* in the post-genomic era, de Vries R.P., Benoit Gelber I., Rørdam Andersen M.., eds. (Caister Academic Press), 978-1-910190-39-5.

52. Demyer, D., Hoozee, J., and Mesdom, H. (1974). Specificity of lipolysis during dry sausage ripening. J. Food Sci. *39*, 293–296.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

53. Leistner, L. (1990). Mould fermented foods: Recent developments. Food Biotechnol. *4*, 433–441.

54. Iaccarino, T., Monaco, R. di, Mincione, A., Cavella, S., and Masi, P. (2006). Influence of information on origin and technology on the consumer response: the case of soppressata salami. Food Qual. Prefer. *17*, 76–84.

55. Yadav, A.N., Verma, Dr.P., Kumar, V., Sangwan, P., Mishra, S., Panjiar, N., Gupta, V.K., and Saxena, A. (2018). Biodiversity of the Genus *Penicillium* in Different Habitats. In New and Future Developments in Microbial Biotechnology and Bioengineering (Elsevier), pp. 3–18.

56. Frisvad, J., Smedsgaard, J., Larsen, T.O., and Samson, R. (2004). Mycotoxins, drugs and other extrolites produced by species in *Penicillium subgenus Penicillium*. Stud Mycol *49*, 201–241.

57. Coton, M., Auffret, A., Poirier, E., Debaets, S., Coton, E., and Dantigny, P. (2019). Production and migration of ochratoxin A and citrinin in Comté cheese by an isolate of *Penicillium verrucosum* selected among *Penicillium* spp. mycotoxin producers in YES medium. Food Microbiol. *82*, 551–559.

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

59. Coton, E., Coton, M., Hymery, N., Mounier, J., and Jany, J.-L. (2020). *Penicillium roqueforti*: an overview of its genetics, physiology, metabolism and biotechnological applications. Fungal Biol. Rev. *34*, 59–73.

60. Novo, M., Mangado, A., Quirós, M., Morales, P., Salvadó, Z., and Gonzalez, R. (2013). Genome-Wide Study of the Adaptation of *Saccharomyces cerevisiae* to the Early Stages of Wine Fermentation. PLOS ONE *8*, e74086.

61. Boothby, T.C., Tenlen, J.R., Smith, F.W., Wang, J.R., Patanella, K.A., Osborne Nishimura, E., Tintori, S.C., Li, Q., Jones, C.D., Yandell, M., et al. (2015). Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proc. Natl. Acad. Sci. *112*, 15976.

62. Sicard, D., and Legras, J.-L. (2011). Bread, beer and wine: Yeast domestication in the *Saccharomyces sensu stricto* complex. C. R. Biol. *334*, 229–236.

63. Margarida Gonçalves, Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., and Sampaio, J.P. (2016). Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts. Curr. Biol. *26*, 2750–2761.

64. Caron, T., Le Piver, M., Péron, A.-C., Lieben, P., Lavigne, R., Brunel, S., Roueyre, D., Place, M., Bonnarme, P., Giraud, T., et al. (2020). Strong effect of *Penicillium roqueforti* populations on volatile and metabolic compounds responsible for aromas, flavour and texture in blue cheeses. bioRxiv, 2020.03.02.974352.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

65. Mahler, D.L., Ingram, T., Revell, L.J., and Losos, J.B. (2013). Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. Science *341*, 292.

66. Salzburger, W. (2009). The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. *18*, 169–185.

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

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

69. Samson, R.A., Seifert, K.A., Kuijpers, A.F.A., Houbraken, J.A.M.P., and Frisvad, J.C. (2004). Phylogenetic analysis of *Penicillium subgenus Penicillium* using partial β-tubulin sequences. Stud. Mycol. *49*, 175–200.

70. Giraud, F., Giraud, T., Aguileta, G., Fournier, E., Samson, R., Cruaud, C., Lacoste, S., Ropars, J., Tellier, A., and Dupont, J. (2010). Microsatellite loci to recognize species for the cheese starter and contaminating strains associated with cheese manufacturing. Int. J. Food Microbiol. *137*, 204–213.

71. Houbraken, J., Frisvad, J.C., and Samson, R.A. (2011). Fleming's penicillin producing strain is not *Penicillium chrysogenum* but *P. rubens*. IMA Fungus Glob. Mycol. J. *2*, 87–95.

72. Sayers, E.W., Agarwala, R., Bolton, E.E., Brister, J.R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T., et al. (2019). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *47*, D23–D28.

73. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

74. Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569.

75. Foissac, S., Gouzy, J., Mathe, C., Sterck, L., Yves Van de Peer, Rouze, P., and Schiex, T. (2008). Genome Annotation in Plants and Fungi: EuGene as a Model Platform. Curr. Bioinforma. *3*, 87–97.

76. Sallet, E., Gouzy, J., and Schiex, T. (2019). EuGene: An Automated Integrative Gene Finder for Eukaryotes and Prokaryotes. In Gene Prediction: Methods and Protocols, M. Kollmar, ed. (Springer New York), pp. 97–120.

77. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Bork, P., von Mering, C., Szklarczyk, D., and Jensen, L.J. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol. Biol. Evol. *34*, 2115–2122.

78. Huerta-Cepas, J., Forslund, K., Sunagawa, S., Bork, P., Szklarczyk, D., Heller, D., von Mering, C., Cook, H., Jensen, L.J., Walter, M.C., et al. (2015). eggNOG 4.5: a

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. *44*, D286–D293.

79. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at:http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

80. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience *1*, 18–18.

81. Ravyts, F., Vuyst, L.D., and Leroy, F. (2012). Bacterial diversity and functionalities in food fermentations. Eng. Life Sci. *12*, 356–367.

82. Bushnell, B. (2018). BBMap – short-read aligner for DNA and RNA-seq data [online]. Available online at: http://www.sourceforge.net/projects/bbmap/.

83. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. J. Comput. Mol. Cell Biol. *19*, 455–477.

84. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinforma. Oxf. Engl. *29*, 1072–1075.

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

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

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

88. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

89. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

90. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

91. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

92. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.

93. Francis, R.M. (2017). pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. *17*, 27–32.

94. Konishi, T., Matsukuma, S., Fuji, H., Nakamura, D., Satou, N., and Okano, K. (2019). Principal Component Analysis applied directly to Sequence Matrix. Sci. Rep. *9*, 19297.

95. Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics *27*, 592–593.

96. Jukes, T., and Cantor, C. (1969). Evolution of protein molecules. Mamm. Protein Metab., 21–132.

97. Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123(3)*, 585–595.

98. Archer, F.I., Adams, P.E., and Schneiders, B.B. (2017). stratag: An r package for manipulating, summarizing and analysing population genetic data. Mol. Ecol. Resour. *17*, 5–11.

99. Böhm, J., Dahlmann, T.A., Gümüşer, H., and Kück, U. (2015). A MAT1-2 wild-type strain from *Penicillium chrysogenum*: functional mating-type locus characterization, genome sequencing and mating with an industrial penicillin-producing strain. Mol. Microbiol. *95*, 859–874.

100. Schaack, S., Gilbert, C., and Feschotte, C. (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol. Evol. *25*, 537–546.

101. Smit, A., and Hubley, R. (2017). RepeatModeler. [online]. Available online at: http://www.repeatmasker.org.

102. Storer, J., Hubley, R., Rosen, J., Wheeler, T.J., and Smit, A.F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob. DNA *12*, 2.

103. Smit, A., and Hubley, R. (2017). RepeatMasker. [online]. Available online at: http://www.repeatmasker.org.

104. Ekseth, O.K., Kuiper, M., and Mironov, V. (2014). orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics *30*, 734–736.

105. Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7–W13.

106. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19*, 153.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

107. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.

108. Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., et al. (2019). gplots: Various R Programming Tools for Plotting Data.

109. Eilertson, K.E., Booth, J.G., and Bustamante, C.D. (2012). SnIPRE: Selection Inference Using a Poisson Random Effects Model. PLOS Comput. Biol. *8*, e1002806.

110. Thornton, K. (2003). libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics *19*, 2325–2327.

111. Corral, S., Salvador, A., and Flores, M. (2013). Salt reduction in slow fermented sausages affects the generation of aroma active compounds. Meat Sci. *93*, 776–785.

112. Aaslyng, M.D., Vestergaard, C., and Koch, A.G. (2014). The effect of salt reduction on sensory quality and microbial growth in hotdog sausages, bacon, ham and salami. Meat Sci. *96*, 47–55.

113. Guzmán, C., Bagga, M., Kaur, A., Westermarck, J., and Abankwa, D. (2014). ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. PloS One *9*, e92444.

114. Arganda-Carreras, I., Kaynig, V., Rueden, C., Schindelin, J., Eliceiri, K.W., Cardona, A., and Seung, H.S. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics *33*, 2424–2426.

115. Russell V., L., Paul, B., Maxime, H., Jonathon, L., Hannes, R., and Henrik, S. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means version 1.7.0 from CRAN.

116. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. *67*, 1–48.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure legends

Figure 1. Rooted maximum likelihood population trees, principal component analysis (PCA), STRUCTURE and splitstree results based on genome-wide single nucleotide polymorphisms (SNPs) of *Penicillium salamii* (A) and *P. nalgiovense* (B), with information on genomic architecture types and mating-types. Bootstraps are indicated for the well-supported nodes (i.e. bootstraps higher than 70) and the scale indicates branch length (substitution rate per polymorphic site). The strain geographical origins are shown with colors (blue: Europe, purple: America, yellow: Australia, light green: Africa, green: Middle East). The color circles indicate environment of collection, with dry-cured meat in red and other environments in blue. For the STRUCTURE plot, each bar represents an individual, with its estimated proportions of genetic information assigned to different population clusters, represented by colors, for K = 4 for *P. salamii* and K = 2 for *P. nalgiovense*. See Fig. S2 for barplots at other K values, giving similar patterns. In the PCA, the shape indicates the environment of collection, and the color follows the STRUCTURE results. For each PCA axis, the percentage of variance explained is indicated between brackets. For the splitstrees, the bubbles colors follow the STRUCTURE results. For *P. nalgiovense*, the two most divergent strains CBS297.97 and DTO204.F1 were removed from the left splitstree.

Figure 2. Presence (dark blue rectangles) or absence (light blue rectangles) of putative horizontally-transferred regions (HTRs) in *Penicillium nalgiovense* and *P. salamii* strains. Transposable elements were masked in the HTRs reference ("repeat-masked" scenario). The blue darkness indicating the length percentage of HTRs present in each strain compared to the length of the reference consensus region (A), or the percentage of identity between the sequence in the focal strain and the reference consensus region (B). The clustering of the regions was performed by Euclidian distance between coverage of the regions in each strain without masking (Fig. S4). The lengths between brackets at the bottom correspond to the total lengths of the regions (without masking).

Figure 3. Presence (dark blue rectangles) or absence (light blue rectangles) of putative horizontally-transferred regions (HTRs) in *Aspergillus*, *Monascus* and *Penicillium* strains. Transposable elements were masked in the reference HTRs ("repeat-masked" scenario). The blue darkness indicating the length percentage of HTRs present in each strain compared to the length of the reference consensus region (A), or the percentage of identity between the sequence in the focal strain and the reference consensus region (B). The lengths between brackets correspond to the total lengths of the regions (without masking). The unrooted phylogenetic tree was built from 1102 single-copy genes with ASTRAL and the distance in the internal branches is expressed in coalescent units. Stars on the right end of each branch correspond to a full support of the relationship between the four leaves surrounding this branch.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 4. Proteolysis (A) and lipolysis (B) dynamics, and growth (C) and spore production (D) in response to different salt concentrations in *Penicillium biforme, P. nalgiovense* and *P. salamii* and their respective sister species, *P. palitans, P. chrysogenum* and *P. olsonii*. (A) and (B): The x-axis represents the number of days since the start of the experiment, the y-axis represents the front of lysis for proteolysis or lipolysis in mm in measure tubes. The colored area around each logistic line represents the standard error, and colors correspond to the environment of collection (red=dry-cured meat, blue and turquoise=other environments excluding dry-cured meat). Stars indicate significant post-hoc tests (p-value < 0.05), squares marginally significant ones (0.05 < p-value < 0.1). (C) Growth rates represented as colony area in cm^2 at days 8 (left panel) and 17 (right panel) post-inoculation at different salt concentrations (X axis). (D) Spore production is estimated at day 17 for the same species and the same salt concentrations. In (C) and (D), circles indicate species used for dry-cured meat production, i.e., *P. biforme*, *P. nalgiovense* and *P. salamii*, and squares indicate their respective sister species not used for producing dry-cured meat (*P. palitans*, *P. chrysogenum* and *P. olsonii*, respectively). The strains were collected either from dry-cured meat (red) or from other environments excluding dry-cured meat (blue). The error bars correspond to the standard error. We did not detect any significant difference between dry-cured meat and non-dry-cured meat samples in each species with post-hoc tests.

Figure 5. Degree of white, color composition and pictures on malt extract agar medium of *Penicillium nalgiovense* and *P. chrysogenum* (A), *P. salamii* and *P. olsonii* (B), *P. biforme* and *P. palitans* (C), separated by their environment of collection. The left panel plots show the degree of white on a scale from 50 to 550 (represented as dark grey to light grey). The middle panel shows the distribution among the red, blue and green values in the color of each *Penicillium* strain. The colors and shapes of the points correspond to the environment of collection (red circle=dry-cured meat, blue circle=non-dry-cured meat, blue triangle=non-dry-cured meat sister species). Stars indicate significant post-hoc tests (p-value < 0.05). The right panel shows pictures of *P. nalgiovense, P. salamii, P. biforme* and their respective sister species on malt extract agar medium: the first row corresponds to strains collected on dry-cured meat, the second (and the sister species) to strains isolated from other environments.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 1. Species identified among the 134 monospore-derived strains isolated from dry-cured meat casings.

Species	Number of identified monospores
Total *Penicillium*	124
P. nalgiovense	70
P. salamii	16
P. biforme	5
P. camemberti	5
P. chrysogenum	5
P. solitum	5
P. polonicum	4
P. brevicompactum	2
P. chermesinum	2
P. crustosum	2
P. nordicum	2
P. expansum	1
P. glabrum	1
P. italicum	1
P. olsonii	1
P. roqueforti	1
P. tularense	1
Total other genera	10
***Aspergillus* sp.**	2
Fusarium proliferatum	2
***Scopulariopsis* sp.**	2
***Alternaria* sp.**	1
Cladosporium salinae	1
Cladosporium spinulosum	1
Geotrichum candidum	1

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

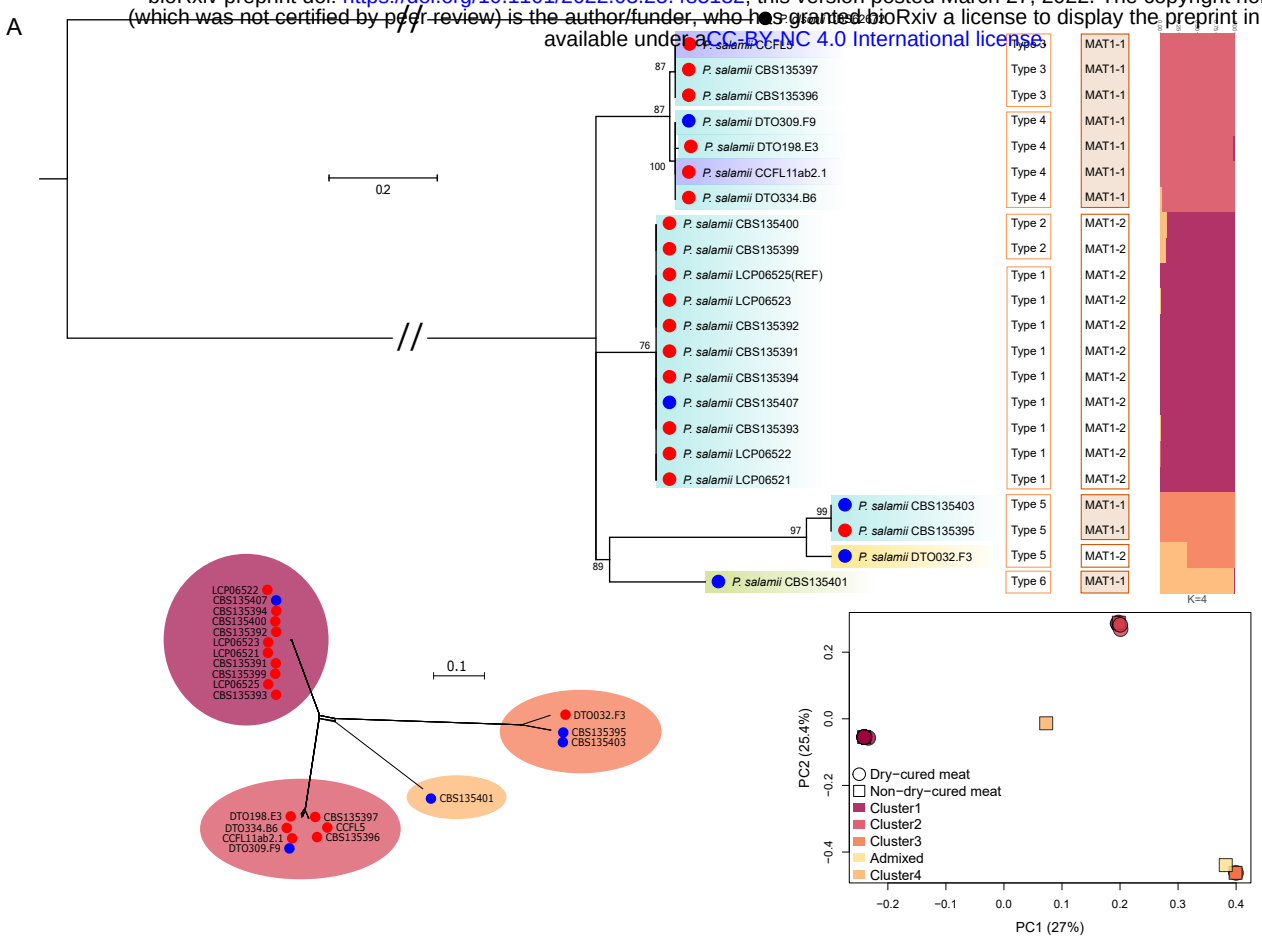
Table 2. Statistics on the phenotypes measured in this study. Model A corresponds to the model testing whether non-dry-cured meat population in species used for dry-cured meat harbored different phenotypes than their sister species and Model B aimed at testing whether non-dry-cured samples showed significant differences relative to dry-cured meat samples**.** A clade corresponds to the species used in dry-cured-meat production and their sister species. Environment effect corresponds to either dry-cured meat or non-dry-cured meat. Some variables being non-independent (a species belong to a clade, and an environment is characteristic of a species), they are included as nested parameters and shown with squared brackets. Interactions are represented by colons. Cells with red bold font represent the main focus of the test, that is whether there is a difference between dry-cured meat and non-dry-cured meat populations. Cells with yellow background highlight significant differences with p-value < 0.05. Stars after the p-values correspond to a value between 0.05 and 0.01, between 0.01 and 0.001, and below 0.001 for one, two and three stars respectively.

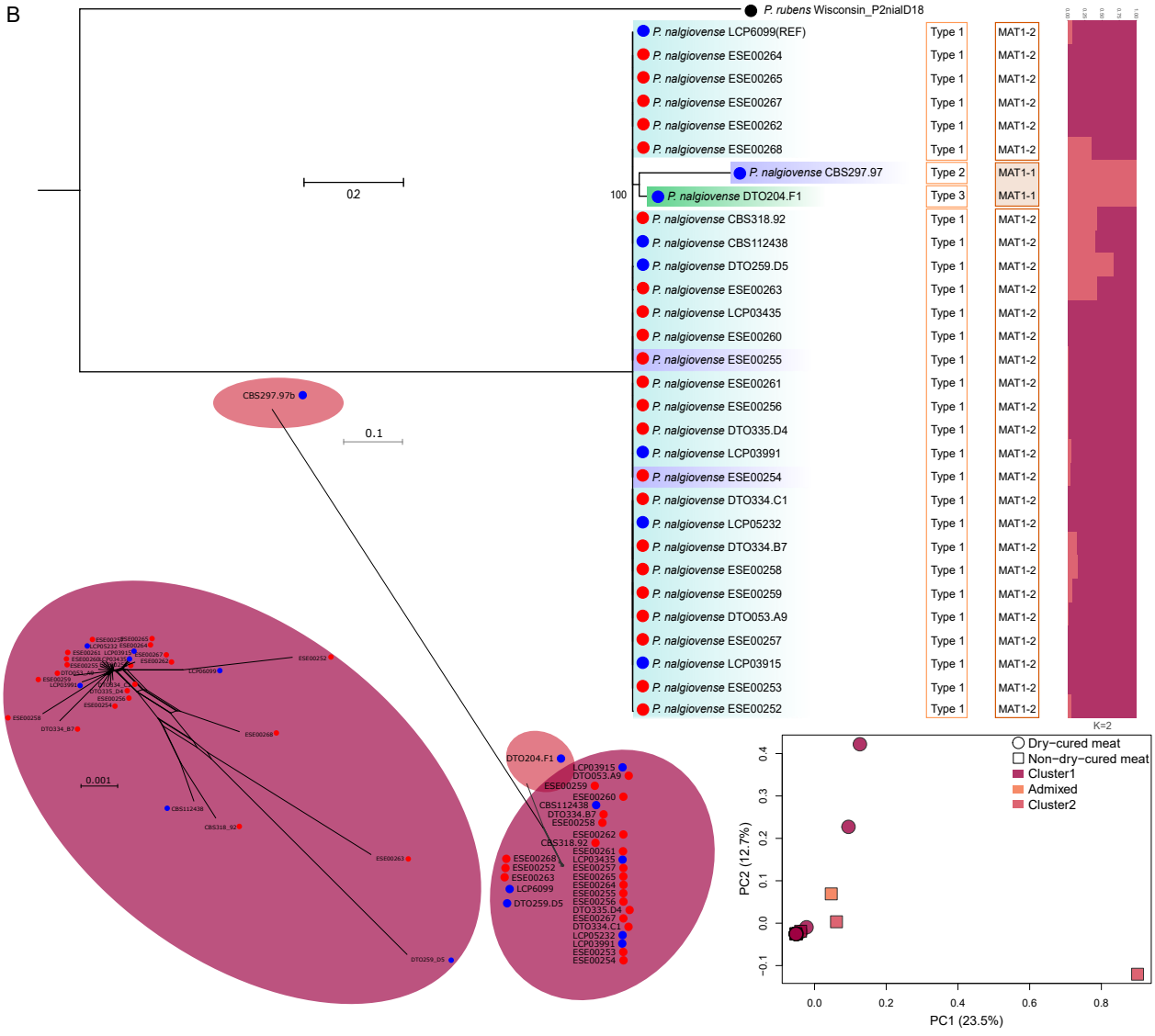
	Dependent variable ->	Growth on Salt			Sporulation on Salt				Proteolysis			Lipolysis			Color (Red/Blue)			Whiteness		
	Explanatory variables **V**	Chi Df	Chi Sq	*p*-value	Df	SumOfSq	*F*-value	*p*-value	Chi Df	ChiSq	*p*-value	Chi Df	ChiSq	*p*-value	Chi Df	ChiSq	*p*-value	Chi Df	ChiSq	*p*-value
Model A	**Clade**	**2**	7.084	0.0290 *	2	0.321	0.1974	0.8211	2	33.618	<0.001 ***	2	35.7256	<0.001 ***	2	6.0635	0.0482 *	2	1.3504	0.5091
	Species[Clade]	**3**	7.0541	0.0702	3	7.772	3.1869	0.0266 *	3	4.6392	0.2002	3	1.4852	0.6857	3	3.9425	0.2677	3	3.1148	0.3743
	Salt Concentration	3	546.1269	<0.001 ***	2	18.291	11.2509	<0.001 ***												
	Clade:Salt Concentration	6	46.0191	<0.001 ***	4	4.170	1.2823	0.2814												
	Species[Clade]:Salt Concentration	9	11.5286	0.2412	6	0.785	0.1609	0.9864												
Model B	**Species**	**2**	15.6135	<0.001 ***	2	11.823	9.5065	<0.001 ***	2	14.164	<0.001 ***	2	9.1309	0.0104 *	2	28.708	<0.001 ***	2	13.526	0.0012 **
	Environment[Species]	**3**	**0.6689**	**0.8805**	**3**	**1.663**	**0.8912**	**0.4481**	**3**	**63.867**	**<0.001 ***[1]**	**3**	**53.2360**	**<0.001 ***[2]**	**3**	**21.779**	**<0.001 ***[3]**	**3**	**14.073**	**0.0028 **[4]**
	Salt Concentration	3	547.9128	<0.001 ***	2	10.188	8.1918	<0.001 ***												
	Species:Salt Concentration	6	38.4008	<0.001 ***	4	7.652	3.0761	0.0191 *												
	Environment[Species]:Salt Concentration	**9**	**17.2739**	**0.0446 ***	**6**	**4.250**	**1.1391**	**0.3444**												

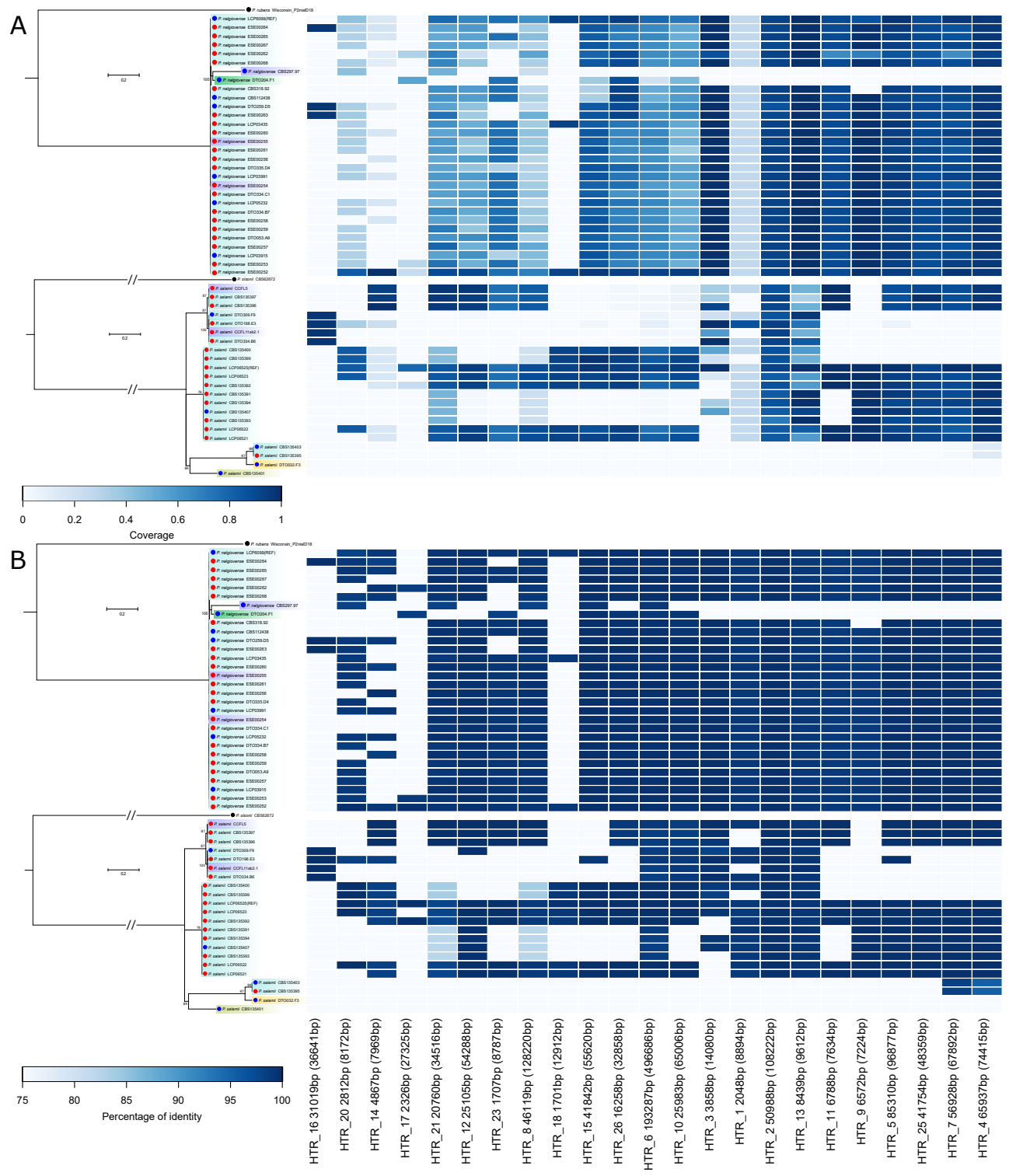
[1] Post-hoc tests for *P. biforme*: df=165, *t*=4.397, *p*-value=0.0003; *P. nalgiovense*: df=165, *t*=2.670, *p*-value=0.0869 and *P. salamii*: df=167, *t*=7.606, *p*-value<0.0001.
[2] Post-hoc tests for *P. biforme*: df=154.17, *t*=-1.113, *p*-value=0.8752; *P. nalgiovense*: df=165.34, *t*=5.283, *p*-value<0.0001 and *P. salamii*: df=61.68, *t*=5.154, *p*-value<0.0001.
[3] Post-hoc tests for *P. biforme*: *z*=4.586, *p*-value=0.0001; *P. nalgiovense*: *z*=0.794, *p*-value=0.9687 and *P. salamii*: *z*=0.034, *p*-value=1.
[4] Post-hoc tests for *P. biforme*: df=51, *t*=-0.294, *p*-value=0.997; *P. nalgiovense*: df=51, *t*=-3.779, *p*-value=0.0052 and *P. salamii*: df=51, *t*=-0.136, *p*-value=1.

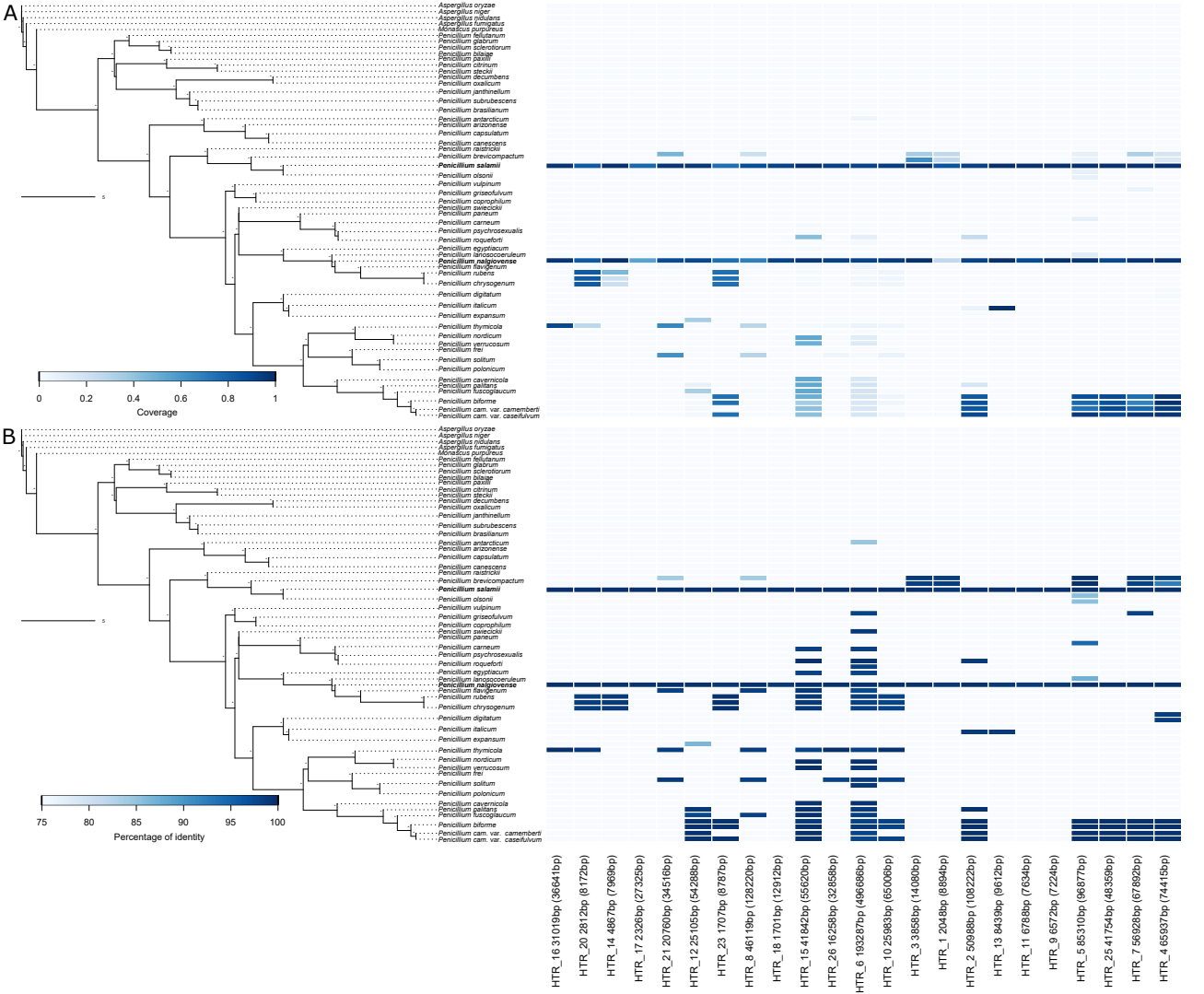
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.25.485132; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

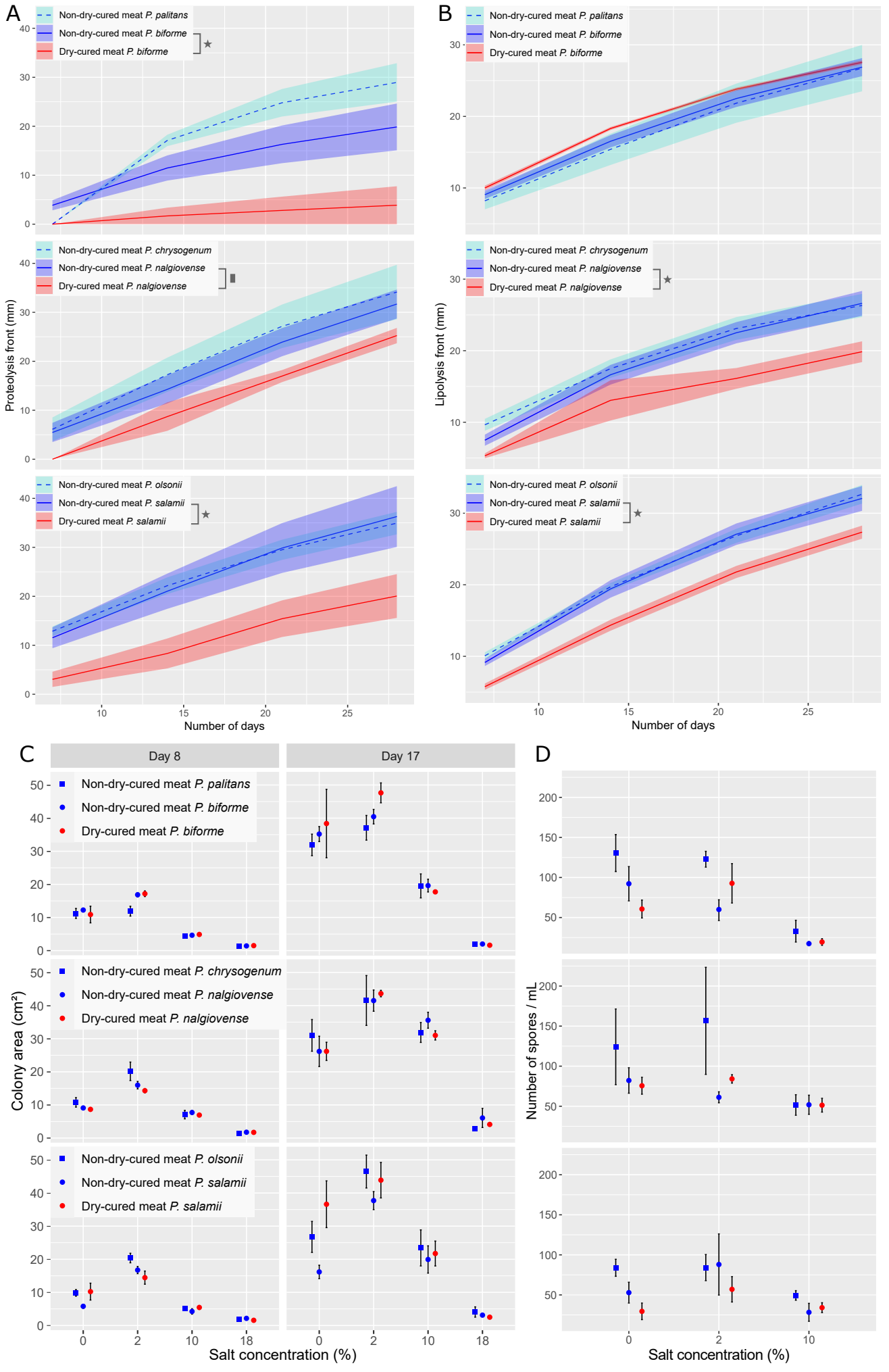
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.15.484102; this version posted March 27, 2022. The copyright holder (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perp available under a CC-BY-NC 4.0 International license.

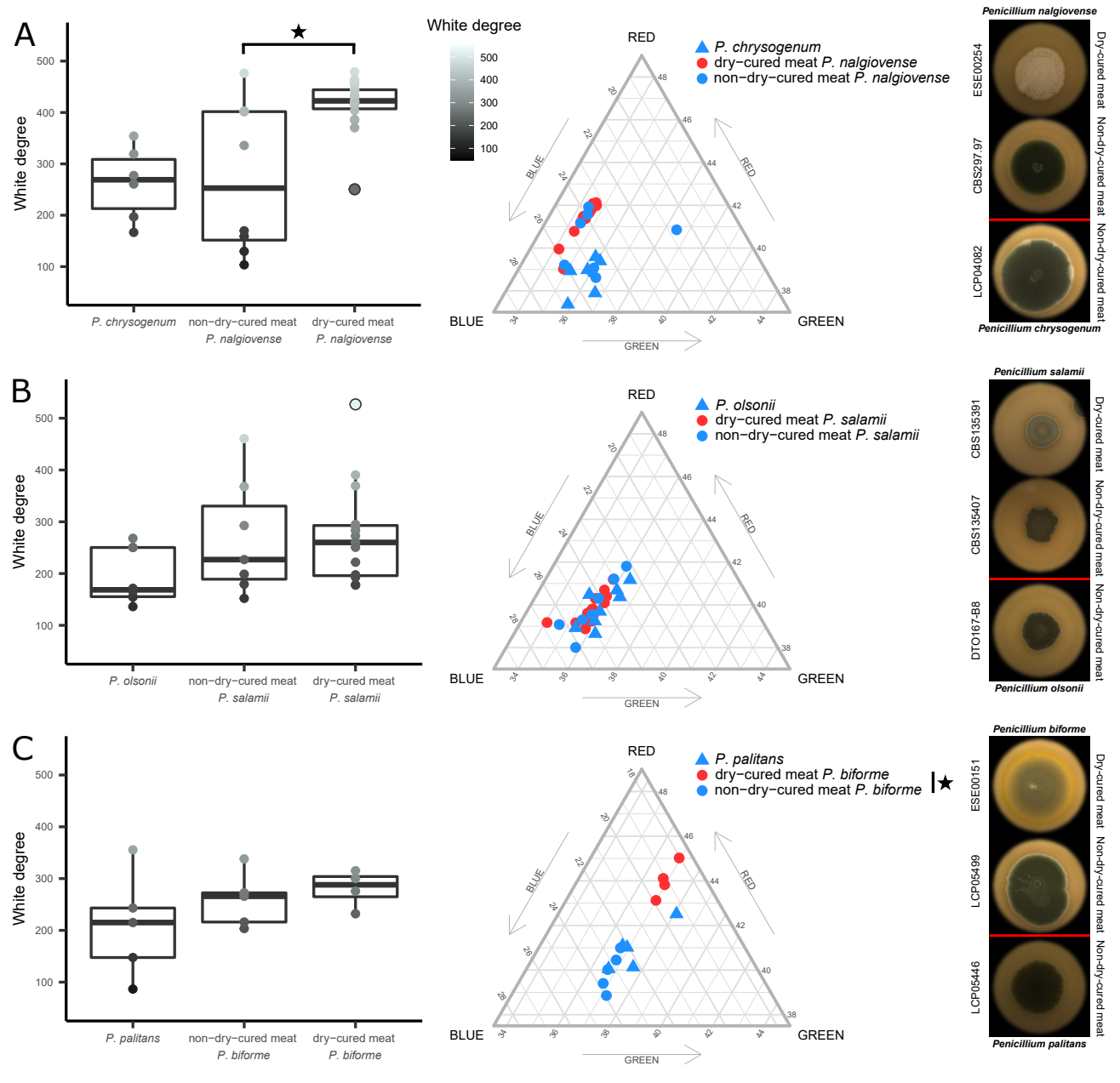





A
B
0
0.2
0.4
0.6
0.8
1
Coverage
75
80
85
90
95
100
Percentage of identity
HTR_16 31019bp (366641bp)
HTR_20 2812bp (81720bp)
HTR_14 4867bp (79690bp)
HTR_17 2326bp (27325bp)
HTR_21 20760bp (34516bp)
HTR_12 25105bp (54288bp)
HTR_23 1707bp (8787bp)
HTR_8 46119bp (128220bp)
HTR_18 1701bp (12912bp)
HTR_15 41842bp (55620bp)
HTR_26 16258bp (32858bp)
HTR_6 193287bp (496686bp)
HTR_10 25983bp (65006bp)
HTR_3 3858bp (14080bp)
HTR_1 2048bp (8894bp)
HTR_2 50988bp (108222bp)
HTR_13 8439bp (9612bp)
HTR_11 6788bp (7634bp)
HTR_9 6572bp (7224bp)
HTR_5 85310bp (96877bp)
HTR_25 41754bp (48359bp)
HTR_7 56928bp (67892bp)
HTR_4 65937bp (74415bp)


A
Aspergillus oryzae
Aspergillus niger
Aspergillus nidulans
Aspergillus fumigatus
Monascus purpureus
Penicillium fellutanum
Penicillium glabrum
Penicillium sclerotiorum
Penicillium bilaiae
Penicillium paxilli
Penicillium citrinum
Penicillium steckii
Penicillium decumbens
Penicillium oxalicum
Penicillium janthinellum
Penicillium subrubescens
Penicillium brasilianum
Penicillium antarcticum
Penicillium arizonense
Penicillium capsulatum
Penicillium canescens
Penicillium raistrickii
Penicillium brevicompactum
Penicillium salamii
Penicillium olsonii
Penicillium vulpinum
Penicillium griseofulvum
Penicillium coprophilum
Penicillium swiecickii
Penicillium paneum
Penicillium carneum
Penicillium psychrosexualis
Penicillium roqueforti
Penicillium egyptiacum
Penicillium lanosocoeruleum
Penicillium nalgiovense
Penicillium flavigenum
Penicillium rubens
Penicillium chrysogenum
Penicillium digitatum
Penicillium italicum
Penicillium expansum
Penicillium thymicola
Penicillium nordicum
Penicillium verrucosum
Penicillium freii
Penicillium solitum
Penicillium polonicum
Penicillium cavernicola
Penicillium paltans
Penicillium fuscoglaucum
Penicillium biforme
Penicillium cam. var. camemberti
Penicillium cam. var. casefulvum
0
0.2
0.4
0.6
0.8
1
Coverage
B
75
80
85
90
95
100
Percentage of identity
HTR_16 31019bp (36641bp)
HTR_20 2812bp (8172bp)
HTR_14 4867bp (7969bp)
HTR_17 2326bp (27325bp)
HTR_21 20760bp (34516bp)
HTR_12 25105bp (54288bp)
HTR_23 1707bp (8787bp)
HTR_8 46119bp (128220bp)
HTR_18 1701bp (12912bp)
HTR_15 41842bp (55620bp)
HTR_26 16258bp (32658bp)
HTR_6 193287bp (496686bp)
HTR_10 25983bp (65006bp)
HTR_3 3858bp (14080bp)
HTR_1 2048bp (8894bp)
HTR_2 50988bp (108222bp)
HTR_13 8439bp (9612bp)
HTR_11 6788bp (7634bp)
HTR_9 6572bp (7224bp)
HTR_5 85310bp (96877bp)
HTR_25 41754bp (48359bp)
HTR_7 56928bp (67892bp)
HTR_4 65937bp (74415bp)


A
B
C
D
Non-dry-cured meat P. palitans
Non-dry-cured meat P. biforme
Dry-cured meat P. biforme
Non-dry-cured meat P. chrysogenum
Non-dry-cured meat P. nalgiovense
Dry-cured meat P. nalgiovense
Non-dry-cured meat P. olsonii
Non-dry-cured meat P. salamii
Dry-cured meat P. salamii
Proteolysis front (mm)
Lipolysis front (mm)
Number of days
Day 8
Day 17
Colony area (cm²)
Number of spores / mL
Salt concentration (%)


A
White degree
500
400
300
200
100
P. chrysogenum
non−dry−cured meat P. nalgiovense
dry−cured meat P. nalgiovense
White degree
RED
BLUE
GREEN
P. chrysogenum
dry−cured meat P. nalgiovense
non−dry−cured meat P. nalgiovense
Penicillium nalgiovense
ESE00254
CBS297.97
LCP04082
Dry−cured meat
Non−dry−cured meat
Non−dry−cured meat
Penicillium chrysogenum
B
P. olsonii
non−dry−cured meat P. salamii
dry−cured meat P. salamii
P. olsonii
dry−cured meat P. salamii
non−dry−cured meat P. salamii
Penicillium salamii
CBS135391
CBS135407
DTO167-B8
Penicillium olsonii
C
P. palitans
non−dry−cured meat P. biforme
dry−cured meat P. biforme
P. palitans
dry−cured meat P. biforme
non−dry−cured meat P. biforme
Penicillium biforme
ESE00151
LCP05499
LCP05446
Penicillium palitans