Report

Current Biology

# Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi

## Highlights

- New HTRs are found in cheese fungi
- HTRs are flanked by specific transposable elements
- HTRs have spread in cheese-associated fungi through recent selective sweeps
- Experiments link two HTRs to growth and competitive advantages on cheese

## Authors

Jeanne Ropars, Ricardo C. Rodríguez de la Vega, Manuela López-Villavicencio, ..., Joëlle Dupont, Antoine Branca, Tatiana Giraud

## Correspondence

antoine.branca@u-psud.fr (A.B.), tatiana.giraud@u-psud.fr (T.G.)

## In Brief

Ropars et al. report newly discovered horizontally transferred regions, flanked by specific transposable elements that allow cheese-making fungi and food spoilers to grow faster and be better competitors on cheese. These findings have industrial and food safety implications and also improve our understanding of adaptation processes.

## Accession Numbers

HG813308–HG813531
HG813601–HG814182
HG814183–HG815135
HG815136–HG815288
HG815290–HG816004
HG816029–HG818118

Ropars et al., 2015, Current Biology *25*, 2562–2569
October 5, 2015 
http://dx.doi.org/10.1016/j.cub.2015.08.025

CellPress

CellPress

Current Biology
Report

# Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi

Jeanne Ropars,[1,2,8] Ricardo C. Rodríguez de la Vega,[1,2,8] Manuela López-Villavicencio,[3] Jérôme Gouzy,[4,5] Erika Sallet,[4,5] Émilie Dumas,[1,2] Sandrine Lacoste,[3] Robert Debuchy,[6,7] Joëlle Dupont,[3] Antoine Branca,[1,2,9,*] and Tatiana Giraud[1,2,9,*]
[1]Ecologie, Systématique et Evolution, UMR8079, Univ. Paris-Sud, 91405 Orsay, France
[2]Ecologie, Systématique et Evolution, UMR8079, CNRS, 91405 Orsay, France
[3]Institut de Systématique, Evolution, Biodiversité, UMR 7205 CNRS-MNHN-UPMC-EPHE, Muséum national d'Histoire naturelle, Sorbonne Université, CP39, 57 Rue Cuvier, 75231 Paris Cedex 05, France
[4]Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, INRA, Castanet-Tolosan 31326, France
[5]Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, CNRS, Castanet-Tolosan 31326, France
[6]Institut de Génétique et Microbiologie, UMR8621, Univ. Paris-Sud, 91405 Orsay, France
[7]Institut de Génétique et Microbiologie, UMR8621, CNRS, 91405 Orsay, France
[8]Co-first author
[9]Co-senior author
*Correspondence: antoine.branca@u-psud.fr (A.B.), tatiana.giraud@u-psud.fr (T.G.)
http://dx.doi.org/10.1016/j.cub.2015.08.025


## SUMMARY

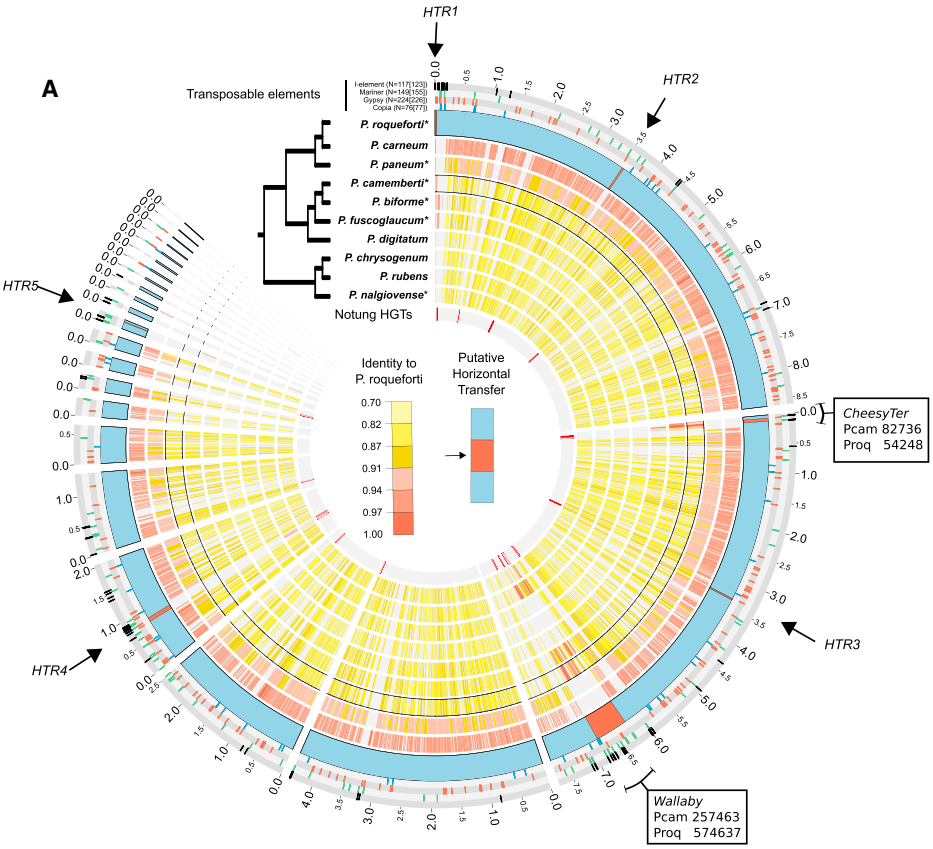
Domestication is an excellent model for studies of adaptation because it involves recent and strong selection on a few, identified traits [1–5]. Few studies have focused on the domestication of fungi, with notable exceptions [6–11], despite their importance to bioindustry [12] and to a general understanding of adaptation in eukaryotes [5]. *Penicillium* fungi are ubiquitous molds among which two distantly related species have been independently selected for cheese making—*P. roqueforti* for blue cheeses like Roquefort and *P. camemberti* for soft cheeses like Camembert. The selected traits include morphology, aromatic profile, lipolytic and proteolytic activities, and ability to grow at low temperatures, in a matrix containing bacterial and fungal competitors [13–15]. By comparing the genomes of ten *Penicillium* species, we show that adaptation to cheese was associated with multiple recent horizontal transfers of large genomic regions carrying crucial metabolic genes. We identified seven horizontally transferred regions (HTRs) spanning more than 10 kb each, flanked by specific transposable elements, and displaying nearly 100% identity between distant *Penicillium* species. Two HTRs carried genes with functions involved in the utilization of cheese nutrients or competition and were found nearly identical in multiple strains and species of cheese-associated *Penicillium* fungi, indicating recent selective sweeps; they were experimentally associated with faster growth and greater competitiveness on cheese and contained genes highly expressed in the early stage of cheese maturation. These findings have industrial and food safety implications and improve our understanding of the processes of adaptation to rapid environmental changes.

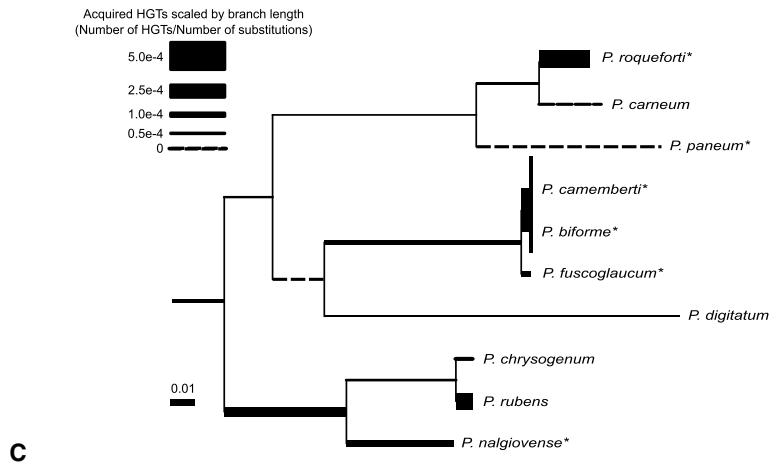
## RESULTS AND DISCUSSION

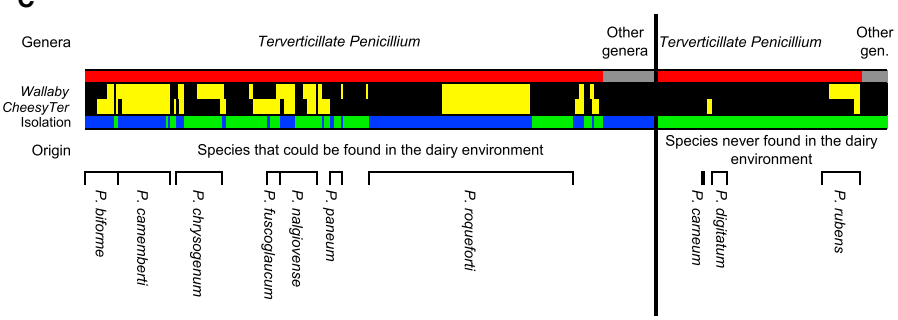
### Multiple Recent Horizontal Gene Transfers between Distant *Penicillium* Species, Flanked by Specific Retrotransposons

We report here five newly sequenced and assembled *Penicillium* genomes, which we compared with previously available data [16–19]. The full dataset included the genome sequences of ten *Penicillium* species, six of which are either used as industrial inocula for cheese making (*Penicillium roqueforti* and *Penicillium camemberti*) or occur as contaminants in cheeses (Figure 1; Table S1). *P. camemberti* is only found in cheese and is thought to include a single clonal lineage originating from selection programs at the end of the 19$^{th}$ century from the blue-gray cheese molds used at that time, i.e., *Penicillium biforme* and *Penicillium fuscoglaucum* [20, 21]. By contrast, *P. roqueforti* also occurs in habitats other than cheese, such as silage or wood, and displays substantial genetic diversity [22, 23]. For reconstructing a rooted phylogeny of these ten *Penicillium* species, we used four *Aspergillus* species as an outgroup. We concatenated alignments of 3,089 single-copy genes shared by at least ten species and reconstructed a fully resolved and well-supported maximum-likelihood phylogeny (Figures 1A and 1B).

We used this rooted phylogeny to investigate the occurrence of horizontal gene transfers (HGTs) between *Penicillium* species. As HGTs (also known as xenology [24]) result in incongruences between gene genealogies and species trees, we compared all individual gene genealogies with the species tree. For this goal, we used the Notung software [25–27] to infer the number of duplication, loss, and HGT events that reconciled the gene genealogies with the species tree. Notung is conservative regarding the inference of HGTs because it tests their temporal feasibility, assumes that HGTs occur with a low probability, and forces the poorly supported nodes to follow the species




(legend on next page)

tree. Only orthologous groups with at least one homolog in at least eight genomes were analyzed, further rendering our estimates of HGTs a lower bound. We found overall 104 HGTs between *Penicillium* species, distributed among 77 orthologous groups. Notung inferred the highest number of HGTs relative to branch length in the clade encompassing *P. camemberti*, *P. biforme*, and their common ancestor (Figure 1B). *P. roqueforti* also acquired many xenologs relative to its branch length. 8 of the 21 horizontally acquired genes detected in *P. roqueforti* were inferred to come from *P. camemberti*, *P. biforme*, or their common ancestor, indicating recent transfers from species sharing the same ecological niche (Figure S1). Only five of these eight genes could be assigned putative functions, i.e., a protein kinase, two transcription factors, a cation transporter, and an integrase-like protein. Cation transport seems particularly relevant for growth in cheeses as several ions are limiting in this medium (e.g., iron ions), and pH drastically drops during cheese maturation [28].

Another line of evidence for the abundance of HGTs in *Penicillium* fungi came from the finding of multiple large genomic islands that were almost 100% identical at the nucleotidic level between distant species, while being absent from closely related species (Figure 1A). The only substitutions detected in these genomic islands corresponded to repeat-induced point mutations, i.e., C-to-T transitions induced by a specific fungal defense mechanism against transposable elements (TEs) that can substitute multiple base pairs in a single meiosis [29]. In *P. roqueforti*, for example, seven genomic islands larger than 10 kb and displaying above 97% nucleotide identity with multiple other species were found. Only the largest region had previously been identified and was called *Wallaby* [16]. Such a high level of identity suggests that these genomic islands correspond to recent horizontally transferred regions (HTRs), although they could alternatively be recent introgressions. Two lines of evidence, however, support the HTR hypothesis rather than introgression: (1) the presence of several of these regions at non-homologous locations in the different *Penicillium* genomes (Figure S2; [16]) and (2) the low mean genome sequence identity between the *Penicillium* species sharing these regions, being less than 90%, an identity level at which no successful interspecific crosses have ever been reported in fungi [30]. Interestingly, these HTRs were flanked in *P. roqueforti* by copies of TEs from a particular family that were rare elsewhere in the genomes (Figures 1A and S3), the *i* non-LTR retrotransposons [6]. The other abundant TEs (e.g., *mariner* DNA transposons and *copia* retrotransposons) were in contrast scattered in the genomes (Figure 1A). The genes present in these genomic islands of high sequence similarity partially overlapped with the HGTs detected by Notung. In *P. roqueforti* for example, 17% of the HGTs inferred by Notung were located in the seven large HTRs (Figure 1A). The fact that not all genes in HTRs were detected by Notung mainly results from the filter of this analysis where we used only orthologous groups with homologs in at least eight species. Further, 11% of the inferred HGTs in *P. roqueforti* clustered within 50 kb of the HTRs (Figure 1A), suggesting that these genomic regions may be prone to integrate foreign DNA. This is consistent with the previous finding that the genomic region where *Wallaby* is inserted in *P. roqueforti* carries other species-specific genes in each of *P. camemberti*, *P. rubens*, and *P. roqueforti* [16].

The identification of multiple very recent HTRs, with almost 100% identity in multiple species (Figure 1A), together with Notung inferences of horizontally transferred genes occurring also elsewhere in the genomes (at least 104 in total among *Penicillium* fungi), indicates that HGTs occur frequently among *Penicillium*. The clustering of *i* elements in the flanking regions of HTRs suggests that they may be involved in the mechanism of horizontal transfers. It has been shown that TEs can pass across species boundaries [31] and that they promote genomic rearrangements and recombination [32–34]. The capacity for mycelia fusions may also facilitate the exchange of genetic material in fungi [35].

## Two Horizontally Transferred Genomic Regions Are Likely Involved in Cheese Adaptation and Have Spread in Cheese-Associated *Penicillium* through Recent Selective Sweeps

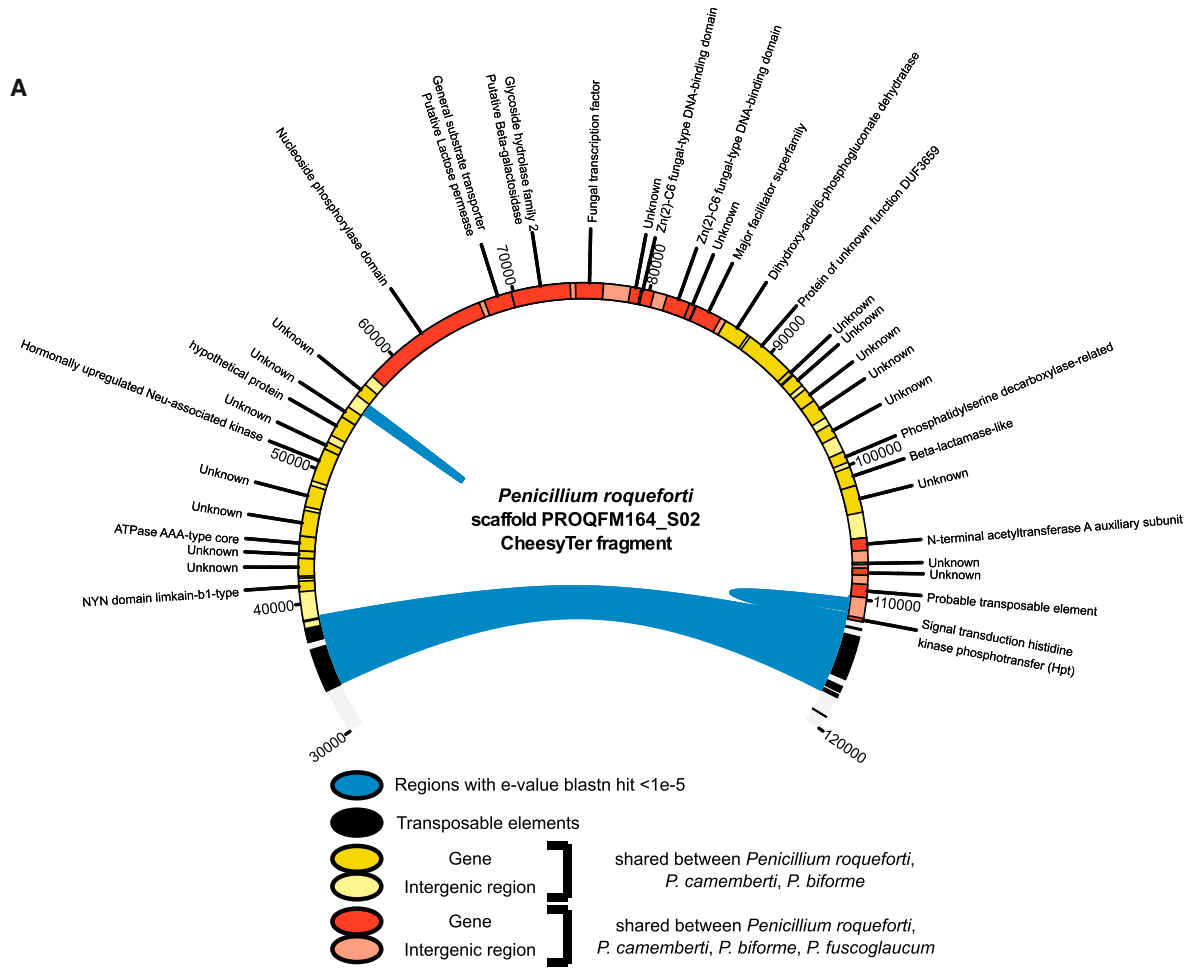
Five of the seven large HTRs detected in *P. roqueforti* were shared between *Penicillium* strains isolated from cheese, *P. camemberti* carrying four of them (Figure 1A). Two HTRs appeared of special relevance for cheese adaptation. *Wallaby* [16] carries a gene encoding an antifungal protein, known for inhibiting the growth of competitors. The second largest HTR was found at terminal edges of scaffolds and was therefore named *CheesyTer*. This 80-kb region, found as a single block in all the genomes studied, carried 37 putative genes, among which two had relevant putative functions for adaptation to cheese, i.e., lactose permease and beta-galactosidase (Figure 2A). Lactose

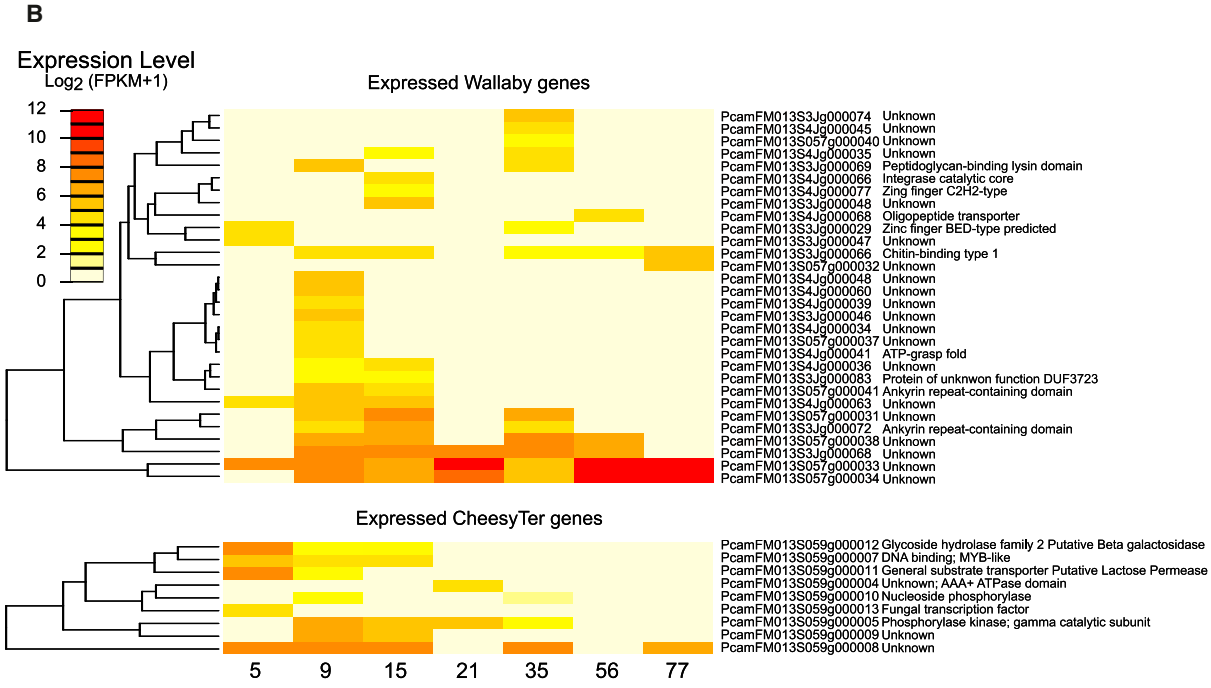
**Figure 1. Horizontal Gene Transfers between *Penicillium* Fungi**

(A) The syntenic blocks in *Penicillium* genomes larger than 10,000 bp are shown as heat-colored circles, aligned against the 23 *P. roqueforti* scaffolds that are larger than 10,000 bp (outer blue circle). The percentage of identity to the *P. roqueforti* genome is indicated by heat colors, from yellow for low identity level to red for high identity level. The seven large regions represented in red on the *P. roqueforti* outer circle (blue otherwise) display levels of identity above 97% between several distantly related species, while being absent from others, and are indicated as horizontally transferred regions (HTRs) numbers 1 to 5, *Wallaby* and *CheesyTer*, respectively. These regions are also characterized by the clustering of *i* LINE retrotransposons at their edges; the four most abundant transposable element families are shown on the four outermost gray circles. The red bars in the inner circle indicate the location of the horizontal gene transfer (HGT) events inferred by Notung. The topology of the species phylogeny obtained with 100% bootstrap support based on 3,089 single-copy genes is represented. Asterisks indicate the strains collected from cheese.

(B) Phylogenetic tree of *Penicillium* fungi based on the 3,089 single-copy genes, with branch lengths as estimated by RAxML. The asterisks are as in (A). Branch widths in the tree are proportional to HGT acquisition rate, i.e., number of genes acquired by HGT as inferred by Notung divided by the number of substitutions along the branch.

(C) Presence of *Wallaby* and *CheesyTer* in 416 strains from 65 fungal species. Each column represents a strain; positive (for at least one primer pair) and negative (for all three primer pairs) PCR amplifications are indicated in yellow and black, respectively. Species are ordered by origin (i.e., dairy environment above the blue line versus other environments above the green line) and by taxon (i.e., terverticillate *Penicillium* below the red lines versus other genera below gray lines).




(legend on next page)

is present during the first few days of cheese maturation, and it acts as a primary carbon source, being rapidly consumed by lactic acid bacteria [37]. The two lactose metabolism genes present in *CheesyTer* were among the most strongly expressed in *P. camemberti* during the first step of cheese rind maturation, during which lactose is available (Figure 2B; Table S2; Supplemental Experimental Procedures), indicating a role in the use of the cheese substrate.

We investigated the presence of *CheesyTer* and *Wallaby* by PCR in 416 strains from 65 fungal species from various environments (Figure 1C; Table S3; Supplemental Experimental Procedures). The presence of the transfers was found highly significantly associated with dairy environment both among species ($\chi^2 = 55.7$; degrees of freedom [df] = 1; p value = 8.571e−14) and among strains within species ($\chi 2 = 45.2$; df = 1, p value = 1.774e−11). Amplicons were actually obtained only for *Penicillium* species that are frequently isolated in the dairy environment, with the only exception of *P. rubens*, the penicillin-producer fungus, in which 16 strains out of 20 carried either one of the two HTRs. The *CheesyTer* and *Wallaby* fragments obtained by PCR showed zero substitution among all strains from all species, including synonymous sites and non-coding regions, as previously found for *Wallaby* [16]. This result confirms that the presence of *CheesyTer* and *Wallaby* is not ancestral in *Penicillium* species and that these genomic islands have instead been acquired very recently. This also indicates that the two HTRs have spread in several species through recent selective sweeps. *P. roqueforti* was found polymorphic for the presence of *Wallaby* and *CheesyTer*, with all tested strains carrying either both of these regions or neither of them (Table S3). Within *P. roqueforti*, these regions were present only in strains isolated from the cheese environment, suggesting a role in adaptation to the cheese environment.

## The *Wallaby* and *CheesyTer* Horizontally Transferred Regions Are Experimentally Associated with Faster Growth and Greater Competitiveness on Cheese

We therefore investigated whether strains carrying *Wallaby* and *CheesyTer* showed higher fitness in terms of growth on cheese substrate or for competitiveness. We set up three experiments, focusing on *P. roqueforti*, because a large collection of strains was available, isolated from various environments, and including strains carrying both *Wallaby* and *CheesyTer* (hereafter named W+C+) and strains lacking them (hereafter named W−C−).

We first compared the growth of 50 *P. roqueforti* strains on a cheese medium and on a minimal medium (26 W+C+ and 24 W−C−; Table S4, tab a). Neither the presence of *Wallaby* and *CheesyTer*, as a main effect independent of the medium, nor the origin of the strain (i.e., cheese versus other environments) significantly influenced the growth of *P. roqueforti* (Table S1). By contrast, the effect on growth of the medium and its interaction with the presence of the two genomic islands were significant (Figure 3; Table S1): W+C+ strains had a growth advantage on cheese medium but a slower growth on minimal medium.

Second, we investigated whether *P. roqueforti* strains carrying *Wallaby* and *CheesyTer* had a higher ability to exclude competitors. We measured the growth of three fungal strains belonging to species commonly found in cheese but lacking *Wallaby* and *CheesyTer* (*P. nalgiovense* FM193, *P. biforme* LCP05529, and *Geotrichum candidum* FM074) on plates covered with lawns of *P. roqueforti* either W+C+ (n = 11) or W−C− (n = 12) strains (Table S4, tab b). These experiments were carried out on minimal, cheese, and malt agar media. No difference in growth was detected for the yeast *G. candidum* between lawns of W+C+ or W−C− *P. roqueforti* strains (Table S1). By contrast, W+C+ *P. roqueforti* strains significantly impaired the growth of the two *Penicillium* challengers on the cheese and malt media (Figure 3B). This was not the case on the minimal medium: the interaction between the presence of the transfers and the medium was significant (Table S1). Using the same experimental design, we then investigated the effect of the two genomic islands when present in the challengers. For this goal, we inoculated on *P. roqueforti* lawns (W+C+, n = 2, or W−C−, n = 2), on cheese medium, different strains of species displaying a polymorphism in the *Wallaby* and/or *CheesyTer* presence (Table S4, tab c). We used as challengers different strains of *P. camemberti* (W+C−, n = 1, or W+C+, n = 3), *P. biforme* (W−C−, n = 2, or W+C+, n = 2), and *P. rubens* (W−C−, n = 1, or W+C−, n = 3). For all three species, we found that the *P. roqueforti* lawns significantly inhibited the growth of challengers and significantly more so when the *P. roqueforti* lawn carried *Wallaby* and *CheesyTer*. Interestingly, the presence of either *CheesyTer* or *Wallaby* in the challengers allowed better growth on W−C− *P. roqueforti* lawns while neither had significant effect on the growth on W+C+ *P. roqueforti* lawns (Table S1).

Third, we investigated competition among *P. roqueforti* strains carrying (W+C+, n = 8) or lacking (W−C−, n = 11) *Wallaby* and *CheesyTer*. We grew *P. roqueforti* strains on cheese medium as pairwise face-to-face confrontations, and we measured the deviations from symmetrical growth (Table S4, tab d; Figure 3C; Supplemental Experimental Procedures). For the W+C+ versus W+C+ confrontations, the mean growth deviation from a boundary in the exact middle of the Petri dish was not significantly different from zero (t test, t = 1.5; df = 36; p value = 0.14). Similar results were obtained for the W−C− versus W−C− confrontations (t test, t = 0.25; df = 70; p value = 0.80). For the W+C+ versus W−C− confrontations, deviations were measured by

**Figure 2. Structure of *CheesyTer* and Gene Expression of *Wallaby* and *CheesyTer* in *P. camemberti***

(A) Structure of the *CheesyTer* island in *P. roqueforti* (scaffold PROQFM164_S02 from 30,000 bp to 120,000 bp). *CheesyTer* is entirely shared and syntenic between *P. roqueforti*, *P. biforme*, and *P. camemberti*. The region in *P. fuscoglaucum* lacks some fragments (shown in yellow) but is syntenic otherwise (fragments shown in red). The putative functions of the genes are shown. *CheesyTer* is flanked by *i* transposable elements, represented in black, showing a high level of identity (e value < 1e−5). This suggests that they are recently duplicated copies.

(B) Expression of *Wallaby* and *CheesyTer* genes in *P. camemberti* during the first 77 days of cheese rind maturation in industrial Camembert, represented as a heatmap of $\log_2$(FPKM+1), a measure of transcript abundance (fragments per kilobase of exon per million fragments mapped). These data were generated in a previous study [36]. The putative functions of the genes are shown; the two genes of *CheesyTer* whose functions are likely involved in lactose metabolism, i.e., the putative lactose permease and beta-galactosidase, are highlighted in gray.

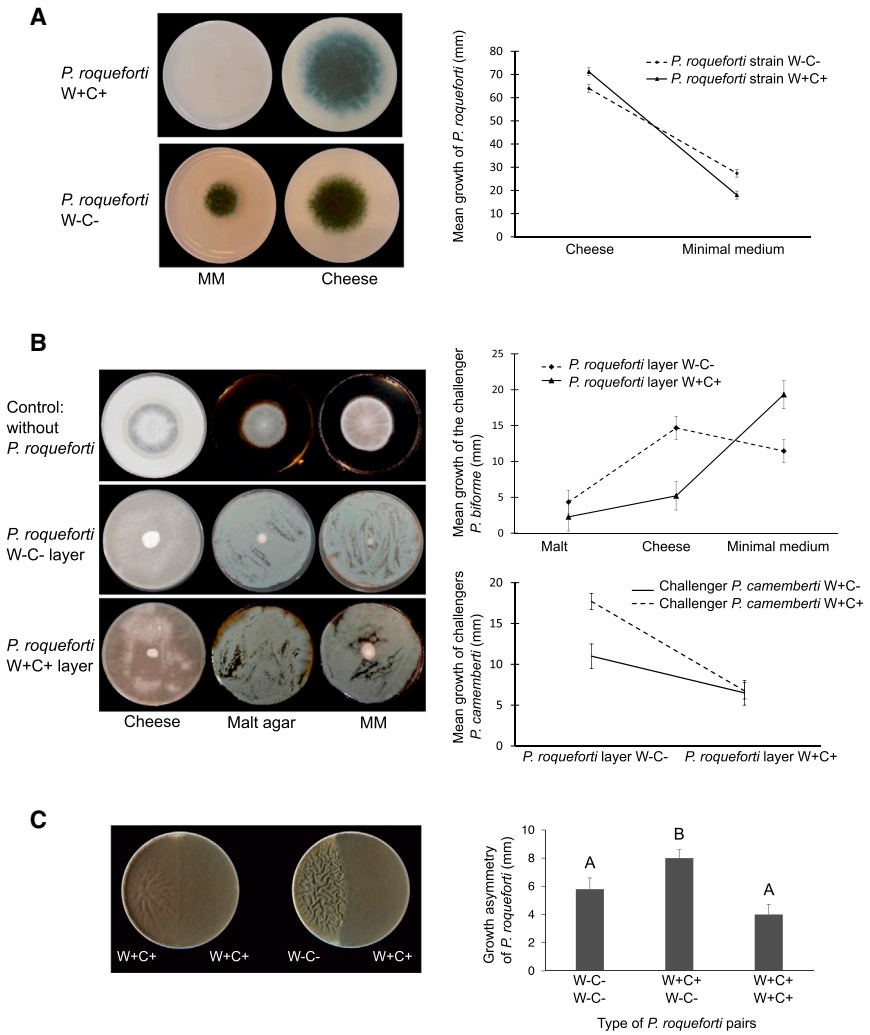



**Figure 3. Fitness Advantages of *P. roqueforti* Carrying *Wallaby* and *CheesyTer*, for the Use of the Cheese Substrate and for Outgrowing Competitors**

(A) Left: pictures of two *P. roqueforti* strains, with (LCP06166, top) and without (LCP06040, bottom) *Wallaby* and *CheesyTer* on minimal medium and cheese medium. Right: mean growth ± SE (in mm) of *P. roqueforti* strains with and without *Wallaby* and *CheesyTer* on the two media.

(B) Left: pictures of a *P. biforme* challenger (LCP05529, without *Wallaby* and *CheesyTer*) on two different *P. roqueforti* lawns, on cheese, malt agar, and minimal media (bottom: strain LCP06149 with *Wallaby* and *CheesyTer*; top: strain LCP05885 without the genomic islands; the first line is a control, i.e., with no *P. roqueforti* lawn). Right: mean growth ± SE (in mm) of a *P. biforme* (top) or a *P. camemberti* (bottom) challenger on *P. roqueforti* strain lawns with or without *Wallaby* and *CheesyTer*.

(C) Growth asymmetry (mean ± SE in mm of deviations from the middle of the Petri dish) in pairwise confrontations of *P. roqueforti* strains with (W+C+) or without (W−C−) *Wallaby* and *CheesyTer*, on cheese medium, for the three types of possible pairs. The A and B letters correspond to significantly different means according to a Tukey-Kramer test. The picture shows examples of confrontations, at left LCP06271 (W+C+) against LCP06157 (W+C+) and at right LCP00148 (W+C+) against LCP06157 (W−C−).

taking the W+C+ strain as the focal strain; the mean growth deviation was significantly different from zero and positive, the W+C+ strains thus growing farther than the W−C− strain (t test, t = 12.32; df = 90; p value < 0.0001). The mean deviations were significantly higher in the W−C− versus W+C+ confrontations than in the W+C+ versus W+C+ or W−C− versus W−C− confrontations (Tukey-Kramer test, p value < 0.0001), while the means between these two latter were not significantly different. This experiment shows that the competitive advantage of W+C+ strains against W−C− strains also holds within the species *P. roqueforti*. Altogether, these experimental results strongly support the existence of fitness advantages for the *Penicillium* strains carrying the horizontally transferred genomic islands, both in the use of cheese substrate and in competition with fungal competitors.

## Conclusions

Our present study on domesticated fungi shows how adaptation can occur rapidly in eukaryotes. The two cheese species studied here underwent parallel adaptation to the cheese medium, and this involved the transfers of identical regions across species boundaries. HGT events have been reported in fungi [35, 38, 39], particularly in environments created by humans, in domesticated yeasts and fungal pathogens of crops [10, 40, 41]. The extent and timing of gene transfers and the number of species having received the same HTRs are here particularly striking. Furthermore, we provide experimental evidence of fitness advantages for strains carrying these HTRs on a human-made medium. These findings altogether are potentially useful for guiding modern strain improvement programs. Indeed, together with the protocol for inducing sex in *P. roqueforti* [22, 42], the identification here of several key candidate genes important for cheese metabolism and competition may allow further selecting interesting traits for cheese industry using the great genetic variability present in *P. roqueforti* strains without *Wallaby* or *CheesyTer* [22]. In addition, our results suggest that caution is required concerning the introduction of genes into microorganisms, as these genes could readily be transferred to other species in the food environment. Indeed, the rapid spread of *Wallaby* and *CheesyTer* into many species of the dairy environment, even when occurring only as contaminants, indicates that transgenes may readily cross species boundaries in the food chain. Finally, the findings here of rapid adaptation through frequent horizontal gene transfers among distant species under selection in novel, human-made media contribute to our understanding of the evolutionary genomic mechanisms allowing rapid adaptation to environmental changes in eukaryotes.

## ACCESSION NUMBERS

The accession numbers for the *Penicillium* genome sequences reported in this paper are GenBank: HG813601–HG814182 for *P. biforme* FM169; HG816029–HG818118 for *P. carneum* LCP05634; HG814183–HG815135 for *P. fuscoglaucum* FM041; HG815136–HG815288 and HG815290–HG816004 for *P. nalgiovense* FM193; and HG813308–HG813531 for *P. paneum* FM227.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2015.08.025.

## AUTHOR CONTRIBUTIONS

Conceptualization, T.G., J.D., and J.R.; Methodology, A.B., T.G., M.L.-V., R.C.R.d.l.V., and J.R.; Software, A.B. and R.C.R.d.l.V.; Validation, É.D. and S.L.; Formal Analysis, A.B., T.G., M.L.-V., R.C.R.d.l.V., and J.R.; Investigation, A.B., M.L.-V., R.C.R.d.l.V., and J.R.; Resources, R.D., J.D., and J.G.; Data Curation, A.B., J.G., R.C.R.d.l.V., and E.S.; Writing – Original Draft, A.B., T.G., R.C.R.d.l.V., and J.R.; Writing – Review & Editing, A.B., T.G., M.L.-V., R.C.R.d.l.V., and J.R.; Visualization, A.B., M.L.-V., R.C.R.d.l.V., and J.R.; Supervision, A.B. and T.G.; Project Administration, A.B. and T.G.; Funding Acquisition, A.B., J.D., T.G., J.G., and R.C.R.d.l.V.

## ACKNOWLEDGMENTS

This work was supported by the ANR FROMA-GEN grant (ANR-12-PDOC-0030) awarded to A.B., an "*attractivite*" grant from Paris-Sud University to A.B., the ERC starting grant GenomeFun 309403 awarded to T.G., a Marie Curie postdoctoral fellowship to R.C.R.d.l.V. (FP7 COFUND PRES-SUD No. 246556), and the ANR grant "Food Microbiomes" (ANR-08-ALIA-007-02) awarded to J.D. We thank Marco van den Berg for sharing raw data for the *Penicillium chrysogenum* Wisconsin 54-1255 strain (recently renamed *P. rubens*).

Received: January 29, 2015
Revised: July 9, 2015
Accepted: August 11, 2015
Published: September 24, 2015

## REFERENCES

1. Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z., Lun, Y., Li, S., Wang, X., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. *46*, 1220–1226.
2. Bosse, M., Megens, H.J., Frantz, L.A.F., Madsen, O., Larson, G., Paudel, Y., Duijvesteijn, N., Harlizius, B., Hagemeijer, Y., Crooijmans, R.P., and Groenen, M.A. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nat. Commun. *5*, 4392.
3. Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., et al. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. *32*, 656–662.
4. Tian, F., Stevens, N.M., and Buckler, E.S., 4th. (2009). Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proc. Natl. Acad. Sci. USA *106* (*Suppl 1*), 9979–9986.
5. Gladieux, P., Ropars, J., Badouin, H., Branca, A., Aguileta, G., de Vienne, D.M., Rodríguez de la Vega, R.C., Branco, S., and Giraud, T. (2014). Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. Mol. Ecol. *23*, 753–773.
6. Galagan, J.E., Calvo, S.E., Cuomo, C., Ma, L.J., Wortman, J.R., Batzoglou, S., Lee, S.I., Baştürkmen, M., Spevak, C.C., Clutterbuck, J., et al. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature *438*, 1105–1115.
7. Borneman, A.R., Desany, B.A., Riches, D., Affourtit, J.P., Forgan, A.H., Pretorius, I.S., Egholm, M., and Chambers, P.J. (2011). Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet. *7*, e1001287.
8. 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. *22*, 1403–1409.
9. Fay, J.C., and Benavides, J.A. (2005). Evidence for domesticated and wild populations of Saccharomyces cerevisiae. PLoS Genet. *1*, 66–71.
10. Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.L., Wincker, P., Casaregola, S., and Dequin, S. (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc. Natl. Acad. Sci. USA *106*, 16333–16338.
11. 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.
12. Stajich, J.E., Berbee, M.L., Blackwell, M., Hibbett, D.S., James, T.Y., Spatafora, J.W., and Taylor, J.W. (2009). The fungi. Curr. Biol. *19*, R840–R845.
13. McSweeney, P.L.H. (2004). Biochemistry of cheese ripening. Int. J. Dairy Technol. *57*, 127–144.
14. Wolfe, B.E., Button, J.E., Santarelli, M., and Dutton, R.J. (2014). Cheese rind communities provide tractable systems for *in situ* and *in vitro* studies of microbial diversity. Cell *158*, 422–433.
15. Irlinger, F., and Mounier, J. (2009). Microbial interactions in cheese: implications for cheese quality and safety. Curr. Opin. Biotechnol. *20*, 142–148.
16. 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.
17. Marcet-Houben, M., Ballester, A.R., de la Fuente, B., Harries, E., Marcos, J.F., González-Candelas, L., and Gabaldón, T. (2012). Genome sequence of the necrotrophic fungus *Penicillium digitatum*, the main postharvest pathogen of citrus. BMC Genomics *13*, 646.
18. Cerqueira, G.C., Arnaud, M.B., Inglis, D.O., Skrzypek, M.S., Binkley, G., Simison, M., Miyasato, S.R., Binkley, J., Orvis, J., Shah, P., et al. (2014). The Aspergillus Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. Nucleic Acids Res. *42*, D705–D710.
19. van den Berg, M.A., Albang, R., Albermann, K., Badger, J.H., Daran, J.M., Driessen, A.J.M., Garcia-Estrada, C., Fedorova, N.D., Harris, D.M., Heijne, W.H.M., et al. (2008). Genome sequencing and analysis of the filamentous fungus Penicillium chrysogenum. Nat. Biotechnol. *26*, 1161–1168.
20. 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.
21. Pitt, J.I., Cruickshank, R.H., and Leistner, L. (1986). *Penicillium commune, P. camemberti,* the origin of white cheese moulds, and the production of cyclopiazonic acid. Food Microbiol. *3*, 363–371.
22. 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.
23. Gillot, G., Jany, J.-L., Coton, M., Le Floch, G., Debaets, S., Ropars, J., López-Villavicencio, M., Dupont, J., Branca, A., Giraud, T., and Coton, E. (2015). Insights into Penicillium roqueforti morphological and genetic diversity. PLoS ONE *10*, e0129849.

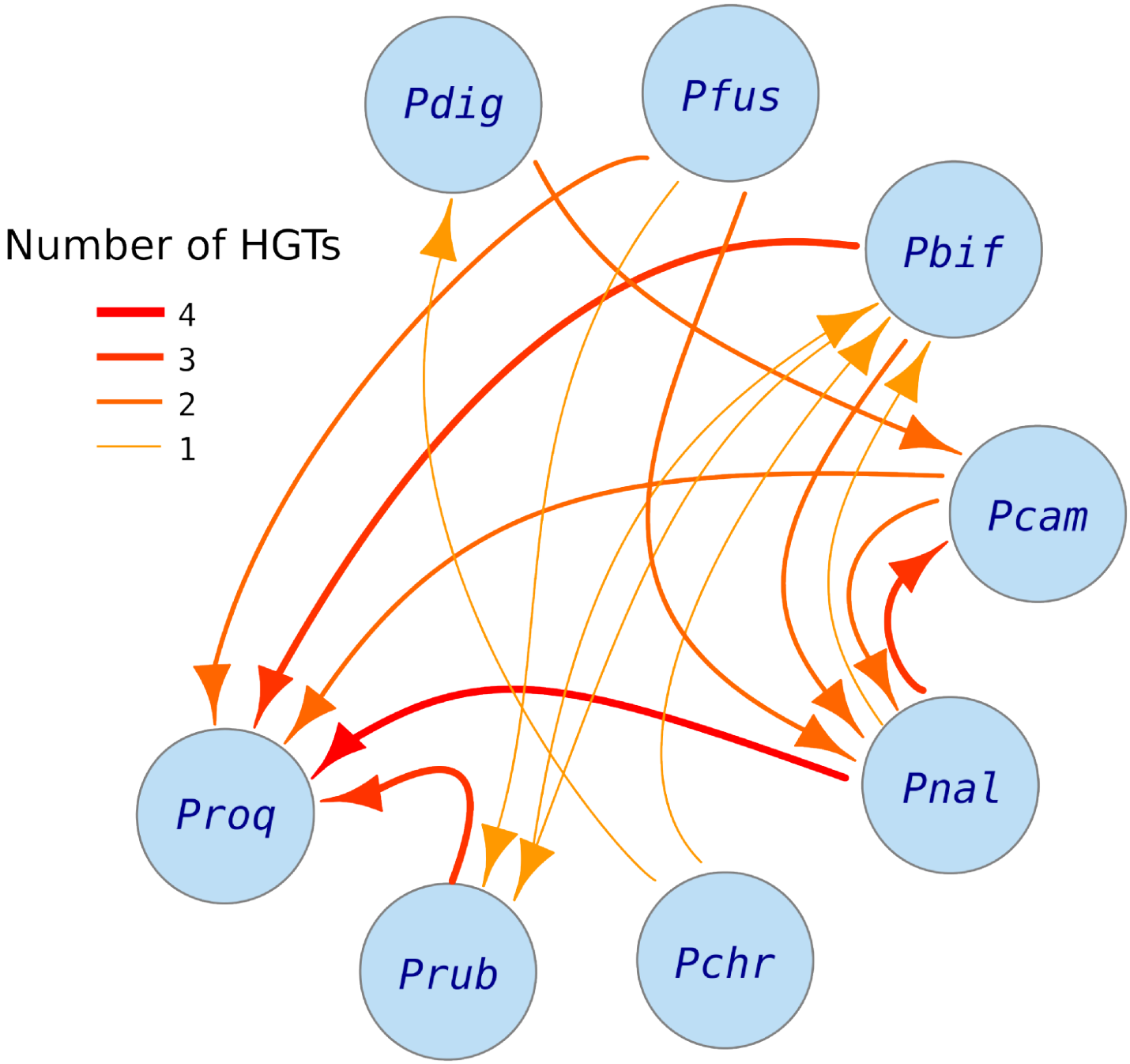
24. Koonin, E.V. (2005). Orthologs, paralogs, and evolutionary genomics. Annu. Rev. Genet. *39*, 309–338.
25. Durand, D., Halldórsson, B.V., and Vernot, B. (2006). A hybrid micro-macroevolutionary approach to gene tree reconstruction. J. Comput. Biol. *13*, 320–335.
26. Vernot, B., Stolzer, M., Goldman, A., and Durand, D. (2008). Reconciliation with non-binary species trees. J. Comput. Biol. *15*, 981–1006.
27. Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics *28*, i409–i415.
28. Monnet, C., Landaud, S., Bonnarme, P., and Swennen, D. (2015). Growth and adaptation of microorganisms on the cheese surface. FEMS Microbiol. Lett. *362*, 1–9.
29. Galagan, J.E., and Selker, E.U. (2004). RIP: the evolutionary cost of genome defense. Trends Genet. *20*, 417–423.
30. Le Gac, M., and Giraud, T. (2008). Existence of a pattern of reproductive character displacement in Homobasidiomycota but not in Ascomycota. J. Evol. Biol. *21*, 761–772.
31. 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.
32. Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., and Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. Genes Dev. *23*, 2478–2483.
33. Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F., and Craig, N.L. (2004). Transposition of hAT elements links transposable elements and V(D)J recombination. Nature *432*, 995–1001.
34. Rajaei, N., Chiruvella, K.K., Lin, F., and Aström, S.U. (2014). Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. Proc. Natl. Acad. Sci. USA *111*, 15491–15496.
35. Wisecaver, J.H., and Rokas, A. (2015). Fungal metabolic gene clusters-caravans traveling across genomes and environments. Front Microbiol *6*, 161.
36. Lessard, M.H., Viel, C., Boyle, B., St-Gelais, D., and Labrie, S. (2014). Metatranscriptome analysis of fungal strains *Penicillium camemberti* and *Geotrichum candidum* reveal cheese matrix breakdown and potential development of sensory properties of ripened Camembert-type cheese. BMC Genomics *15*, 235.
37. Leclercq-Perlat, M.N., Buono, F., Lambert, D., Latrille, E., Spinnler, H.E., and Corrieu, G. (2004). Controlled production of Camembert-type cheeses. Part I: Microbiological and physicochemical evolutions. J. Dairy Res. *71*, 346–354.
38. Marcet-Houben, M., and Gabaldón, T. (2010). Acquisition of prokaryotic genes by fungal genomes. Trends Genet. *26*, 5–8.
39. Wisecaver, J.H., Slot, J.C., and Rokas, A. (2014). The evolution of fungal metabolic pathways. PLoS Genet. *10*, e1004816.
40. Friesen, T.L., Stukenbrock, E.H., Liu, Z., Meinhardt, S., Ling, H., Faris, J.D., Rasmussen, J.B., Solomon, P.S., McDonald, B.A., and Oliver, R.P. (2006). Emergence of a new disease as a result of interspecific virulence gene transfer. Nat. Genet. *38*, 953–956.
41. Coelho, M.A., Gonçalves, C., Sampaio, J.P., and Gonçalves, P. (2013). Extensive intra-kingdom horizontal gene transfer converging on a fungal fructose transporter gene. PLoS Genet. *9*, e1003587.
42. Goarin, A., Silar, P., and Malagnac, F. (2015). Gene replacement in *Penicillium roqueforti*. Curr. Genet. *61*, 203–210.

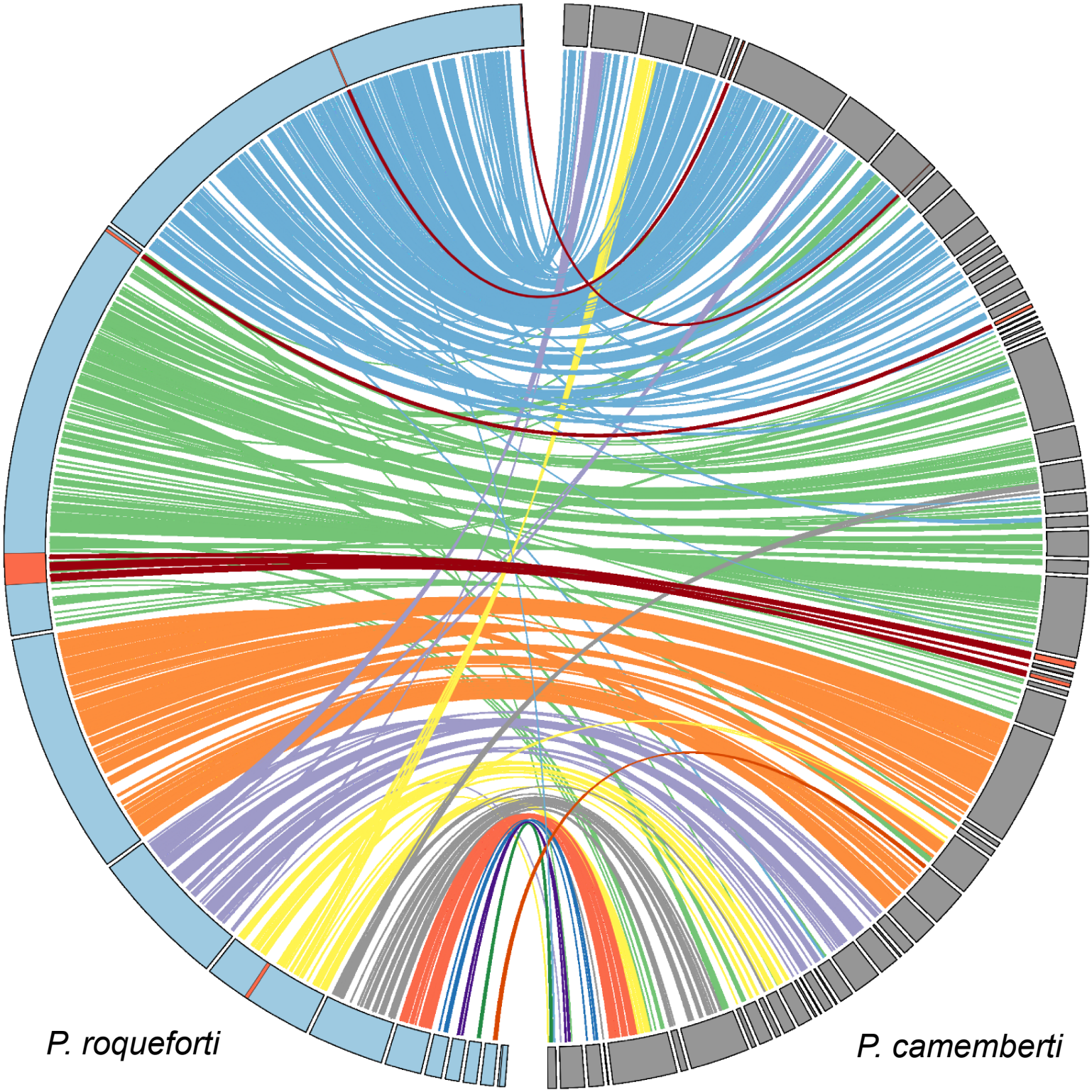
**Current Biology**

**Supplemental Information**

# Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi

**Jeanne Ropars, Ricardo C. Rodríguez de la Vega, Manuela López-Villavicencio, Jérôme Gouzy, Erika Sallet, Émilie Dumas, Sandrine Lacoste, Robert Debuchy, Joëlle Dupont, Antoine Branca, and Tatiana Giraud**


Number of HGTs
4
3
2
1
Pdig
Pfus
Pbif
Pcam
Proq
Pnal
Prub
Pchr


P. roqueforti
P. camemberti

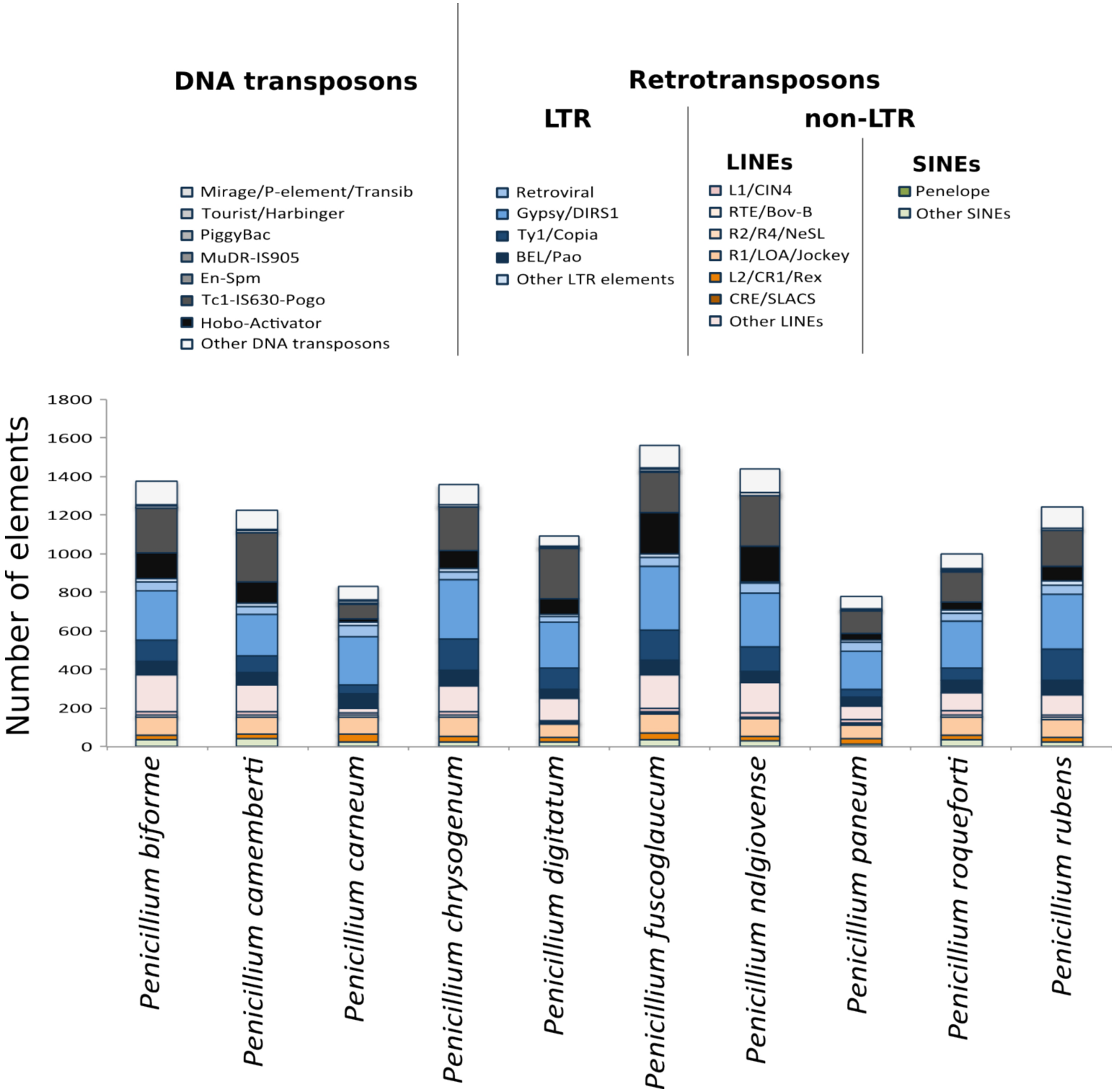

DNA transposons
Mirage/P-element/Transib
Tourist/Harbinger
PiggyBac
MuDR-IS905
En-Spm
Tc1-IS630-Pogo
Hobo-Activator
Other DNA transposons
Retrotransposons
LTR
Retroviral
Gypsy/DIRS1
Ty1/Copia
BEL/Pao
Other LTR elements
non-LTR
LINEs
L1/CIN4
RTE/Bov-B
R2/R4/NeSL
R1/LOA/Jockey
L2/CR1/Rex
CRE/SLACS
Other LINEs
SINEs
Penelope
Other SINEs
Number of elements
0
200
400
600
800
1000
1200
1400
1600
1800
Penicillium biforme
Penicillium camemberti
Penicillium carneum
Penicillium chrysogenum
Penicillium digitatum
Penicillium fuscoglaucum
Penicillium nalgiovense
Penicillium paneum
Penicillium roqueforti
Penicillium rubens

**Figure S1 (related to figure 1):** Oriented graph showing the direction and number of horizontal gene transfers between *Penicillium* species as inferred by the Notung software. The direction of each event is represented by the direction of the arrow, and the number of events by the width of the arrow and its color (the redder the more events). The figure was generated with the R package igraph [S1].

**Figure S2 (related to figure 1)**: Synteny between the *Penicillium roqueforti* and *P. camemberti* genomes. Ortholog blocks larger than 10 kilobases are linked between the two genomes. Each color represents a specific *P. roqueforti* scaffold. Only the scaffolds 1 to 12 are represented for *P. roqueforti* as no matches of more than 10 kilobases have been found for the 36 other scaffolds. The red bands on the circle linked by dark red lines correspond to the horizontally transferred regions.

**Figure S3 (related to figure 1)**: Classification and number of transposable elements in each *Penicillium* genome. In most of the genomes, *mariner* elements constitute the largest group of DNA transposons, while *gypsy* and *copia* elements represent most of the LTR retrotransposons, and *i* elements most of the LINE elements. Very few SINE elements were detected.

| | *P. biforme* | *P. camemberti* | *P. fuscoglaucum* | *P.carneum* | *P. paneum* | *P. roqueforti* | *P. digitatum* | *P. rubens* | *P. chrysogenum* | *P. nalgiovense* |
|---|---|---|---|---|---|---|---|---|---|---|
| | FM169 | FM013 | FM041 | LCP05634T | FM227 | FM164 | PHI26 | Wisconsin54-1255 | | FM193 |
| Substrate | Cheese environment | Cheese environment | Cheese environment | Mouldy rye bread | Cheese environment | Cheese environment | Citrus contaminant | Mouldy cantaloupe | Contaminant in the Postia Placenta MAD 698R | Cheese environment |
| Accession number | Genbank: HG813601-HG814182 | Genbank: HG793134-HG793313 | Genbank: HG814183-HG815135 | Genbank: HG816029-HG818118 | Genbank: HG813308-HG813531 | Genbank: HG792015-HG792062 | Genbank: JH993687-JH993786 | Genbank: AM920416-AM920464 | www.jgi.doe.gov | Genbank: HG815136-HG815288, HG815290-HG816004 |
| Authors | This study | Cheeseman *et al.*, 2014* | This study | This study | This study | Cheeseman *et al.*, 2014* | Marcet-Houben *et al.*, 2012† | Van den Berg *et al.*, 2008‡ | Department of Energy Joint Genome Institute | This study |
| Number of scaffolds | 582 | 180 | 953 | 2,090 | 224 | 48 | 100 | 49 | 27 | 869 |
| Total length of scaffolds | 34,874,608 | 35,011,981 | 36,085,495 | 25,936,550 | 26,579,303 | 29,010,044 | 25,956,151 | 32,223,735 | 31,340,922 | 31,937,463 |
| N50 (bp) | 277,312 | 947,716 | 245,012 | 24,373 | 416,003 | 7,752,559 | 878,909 | 3,889,175 | 5,340,701 | 306,303 |
| Maximum length (bp) | 1,223,141 | 2,612,452 | 1,273,777 | 107,387 | 1,211,662 | 8,605,340 | 4,553,353 | 6,387,817 | 7,361,319 | 930.736 |
| Minimum length (bp) | 1,008 | 1,054 | 1,000 | 2,000 | 1,046 | 2,101 | 964 | 1,032 | 2,067 | 1,002 |
| GC% | 48.09 | 48.18 | 47.74 | 48.73 | 49.01 | 48.64 | 48.9 | 48.96 | 48.78 | 48.48 |
| Nb of protein coding genes | 15,246 | 14,525 | 15,615 | 13,412 | 12,201 | 12,319 | 9,133 | 12,591 | 11,396 | 14,503 |
| Mean gene length (bp) | 1264.8 | 1294.0 | 1253.2 | 1128.4 | 1243.9 | 1243.2 | 1390.1 | 1340.7 | 1407 | 1202.3 |
| Transposable Elements (bp) | 446,173 | 357,683 | 525,762 | 90,938 | 154,502 | 298,483 | 287,550 | 370,483 | 445,385 | 545,631 |

*Cheeseman, K. et al. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat. Commun. 5, 2876 (2014).
†Marcet-Houben, M. et al. Genome sequence of the necrotrophic fungus Penicillium digitatum, the main postharvest pathogen of citrus. BMC Genomics 13, 646 (2012).
‡van den Berg, M. a et al. Genome sequencing and analysis of the filamentous fungus Penicillium chrysogenum. Nat Biotechnol 26, 1161–1168 (2008).

**Table S1 (related to figure 1):** *Penicillium* genome sequencing, assembly, and annotation summary.

**Table S2 (related to figure 2):** Time-wise expression of *Penicillium camemberti* genes during cheese rind maturation; these data were generated in a previous study [42].

**Table S3 (related to figures 1 and 2)**: Presence/absence of the horizontally-transferred regions *Wallaby* and *CheesyTer* in 416 different isolates from diverse origins belonging to 65 fungal species. Strains isolated from the cheese environment are shown in purple and positive PCR amplifications are represented in yellow. Type, neotype and epitype strains are indicated with respectively a "T", a "NeoT" and a "EpiT" after the strains number. All strains included in this study are available in the public fungal strain collection of the National Museum of Natural History in Paris, except the strains labeled "FM", corresponding to a private collection provided by French anonymous stakeholders. Only the six FM strains which genomes are available (*e.g.,* FM013, FM041, FM164, FM169, FM193 and FM227, indicated with an asterisk) have LCP numbers and are publicly available.

**Table S4 (related to figure 3):** Growth measures (in mm) obtained in the fitness experiments. **a)** Experiments assessing the ability of using the cheese substrate: growth measures of 50 *Penicillium roqueforti* strains on minimal medium (MM) and cheese medium. **b)** Experiments assessing competitive ability: growth measures of three challengers, i.e., *P. biforme*, *P. nalgiovense* and *Geotrichum candidum*, inoculated in the middle of Petri dishes on *P. roqueforti* lawns (N=23). **c)** Experiments assessing competitive ability: growth measures of 12 challengers belonging to *P. biforme*, *P. camemberti* or *P. rubens*, inoculated on *P. roqueforti* lawns (2 W+C+ and 2 W-C-). **d)** Experiments assessing competitive ability: growth measures of 19 *P. roqueforti* strains on cheese medium, inoculated pairwise in two opposite points of Petri dishes.

## Supplemental Experimental Procedures

### DNA extraction, sequencing and annotation

DNA was extracted as previously described [16]. The genomes of the strains *Penicillium biforme* FM 169, *P. fuscoglaucum* FM 041, *P. nalgiovense* FM 193, *P. paneum* FM 227 and *P. carneum* LCP05634 (CBS number: 112297) were sequenced at the BGI (Beijing Genomics Institute, Beijing, China), with Illumina HiSeq2000 Solexa sequencing technology. Paired-end genomic DNA libraries were prepared with an insert size of either 500 bp (*P. biforme* FM 169, *P. fuscoglaucum* FM 041, *P. nalgiovense* FM 193) or 700 bp (*P. paneum* FM 227, *P. carneum* LCP05634), and reads of 90 bp were obtained. Sequences were assembled at the BGI, with SOAPdenovo v1.05 [S2]. This software generated contigs, for kmers of 27, 35, 43, 47 and 51 for *P. carneum* LCP05634, *P. paneum* FM 227, *P. fuscoglaucum* FM 041, *P. biforme* FM 169, *P. nalgiovense* FM 193, respectively. Gap filling and single base correction were performed with SOAPaligner.

Gene models were predicted with EuGene [S3], a highly integrative eukaryotic protein-coding gene prediction platform previously trained for *Penicillium* [16]. Similarities to sequences present in a GenBank database of all Eurotiales proteins (release 2011/02/11), the proteome of *Saccharomyces cerevisiae* and the proteomes of three *Penicillium* genomes (*P. camemberti* FM013*, P. roqueforti* FM164 and *P. chrysogenum* Wisconsin 54-1255), were used as backbones to guide EuGene predictions. Initial functional annotations of protein domains on draft genomes were generated by applying InterProScan version v4.8 [S4] to the InterPro database (release 43.1). Functional and pathway annotations were obtained by directly mapping the InterPro

identifiers (InterPro release 48) onto GO. Information about the genomes is provided in Table S1.

**Ortholog detection and multiple sequence alignments**

Whole protein-coding gene sets from 10 *Penicillium* and four *Aspergillus* genomes (178,235 sequences in total) were blasted against each other with BLASTP and clustered with OrthoMCL V1.4 software and the default parameters [S5]. The *Aspergillus* ORF coding versions were retrieved directly from AspGD [18]: *A. fumigatus* Af 293 s03-m04-r21, *A. nidulans* FGSC_A4 s10-m03-r04, *A. niger* CBS_513_88 s01-m06-r09 and *A. oryzae* RIB40 s01-m08-r20. Almost 89% of the sequences clustered into one of the 18,536 orthologous groups (hereafter referred to as orthologous groups). There was a median of seven genes per group. The genes not assigned to clusters were considered to be species-specific. MUSCLE [S6] was then used to align the sequences of the proteins of each orthologous group, and the resulting amino-acid sequence alignments were used as templates for codon-based nucleotide alignments, generated with TranslatorX [S7].

**Annotation and analyses of interspersed repeats**

Transposable elements (TEs) were detected and classified with RepeatMasker version 4.0.3, using the rmblast algorithm [S8] and the repbase RepeatMasker libraries (01/31/2014 release). Total interspersed repeat content was low, ranging from 0.35% to 1.71% of the entire genome assembly for the 10 genomes. The three sister species, *P. carneum*, *P. paneum* and *P. roqueforti,* had genomes consisting of less than 1% TEs, this proportion being the lowest among the

*Penicillium* genomes. The most widespread classes of detected TEs were LTR retroelements and DNA transposons and, to a lesser extent, non-LTR retroelements (Figure S3). The most abundant elements were *mariner* DNA transposons of the *Tc-IS360-Pogo* type, which accounted for 16 to 34 % of the genome sequences attributed to TEs. The second most widespread elements were non-LTR *i* elements, corresponding to 10 to 40% of the total genome sequence attributed to TEs. The cheese fungi *P. camemberti* and *P. roqueforti* had the highest proportions of *i* elements relative to the total length of TE sequences, 39 and 36%, respectively, the mean value for *Penicillium* genomes being 25%. The third and fourth most widespread elements were *gypsy* and *copia* LTR retroelements, which accounted for 12 to 24% and 3 to 15 % of the genome sequences attributed to TEs, respectively. Together, the *mariner*, *i*, *gypsy* and *copia* elements accounted for 62 to 83% of the genome sequences attributed to TEs. We detected 37 TE families expanded only in the *P. digitatum* lineage. This represents a relatively large number of families, given that fewer than 10 TE families were specifically expanded in each of the other lineages. Nevertheless, *Penicillium* genomes generally had few interspersed repeats, slightly fewer, even, than the closely related *Aspergillus* [6].

**Species and gene genealogy reconstruction**

Species trees were reconstructed from the concatenation of alignments of single-copy genes present in at least 10 species, including either or both *P. roqueforti* and *P. camemberti* orthologs, and producing congruent phylogenetic trees. We first constructed maximum likelihood (ML) phylogenetic trees independently of all 6,275 single-copy ortholog alignments, using RAxML

[S9]. We constructed a meta-alignment of the concatenated 3,089 orthologous group alignments (176,4879 codons in total), to obtain a single maximum likelihood phylogenetic tree with RAxML [S9]. For each of the species missing from the 1,030 incomplete alignments, we completed the missing sequence with gap characters (-). Branch support was determined by a bootstrap analysis of 100 resampled data sets. All nodes in the tree based on the concatenated alignment had 100% bootstrap support. ML trees were also reconstructed for two subsets of orthologous groups containing more than one gene in either or both of *P. camemberti* and *P. roqueforti*.

**Horizontal gene transfers detection through gene-species tree reconciliation**

Gene genealogies were reconciled with the fully resolved species phylogeny using Notung in order to reliably infer HGTs [25-27]. First, to obtain the most conservative bifurcating gene genealogies, the root and the poorly supported nodes (Bootstrap values below 95%) in individual gene genealogies were resolved according to the species tree. Then, the Notung reconciliation mode was run for each gene genealogy with a duplication cost of 1.5, a loss cost of 1 and a transfer cost of 4.5 (thus assuming low HGT probability) and with the prune option enabled. Notung reconstructs each gene history by minimizing the total cost and also tests for the temporal feasibility of HGT events. Finally, the inferred transfer events that may have actually resulted from incomplete lineage sorting were filtered out, *i.e.,* we did not consider inferred transfers between lineages separated by a single node in the tree.

**Macrosynteny**

We aligned genomes pairwise with *P. roqueforti* genome, using YASS software [S10] and the following parameters: -O 100000 -C 5,-4,-3,-4 -E 10 -G -16,-4 -W 20,40000 -X 25 -c 1 -d 1 -e 2.8 -i 10 -m 25 -p "#@##-#---##---#--@####,####-#@#@#--##-##" -r 2 -s 70 -w 4. Only orthologous blocks of more than 10 kilobases were retained. The graphical representation of pairwise identity between orthologous blocks was generated with Circos [S11].

***Penicillium camemberti*** **gene expression in cheese rind**

Transcriptome data for Camembert cheese rind were recovered from the NCBI short-read archive, for project SRP030470 samples SRR1002975 to SRR1002981 [42]. The first 25 positions of the 454 reads were trimmed, to remove adapter sequences. The reads were also trimmed on the basis of quality, by removing all 10 bp windows with a minimum aggregate quality of less than 20 from the reads. The reads were mapped onto the *P. camemberti* FM013 genome with Tophat2 [S12], and transcript abundances were estimated with cufflinks v2.2.0 [S13]. The following parameters were used for mapping: all positions, alignment of reads to all possible locations, maximum mismatch threshold of 15%, minimum alignment threshold of 0.5%, 8 processors, Smith-Waterman bandwidth of 235, alignment candidate threshold of 55 bp, hash position threshold of 200 and a homopolymer gap open penalty of 4. The levels of expression of the genes were estimated with cufflinks [S13] as Fragment per Kilobase per Millions of Reads (FPKM). Transcripts were clustered according to their expression level by hierarchical clustering

using the hclust() function of R.

An analysis of gene expression data from a previous study [42] showed that few genes were expressed constitutively during the 77 days of rind maturation. Over the full gene set of *P. camemberti*, we found 4,912 genes expressed on at least one day of the experiment (Table S2). We then focused on expressed genes encoded within the *Wallaby* and *CheesyTer* regions (Figure 2B).

Two genes located in *CheesyTer* were strongly expressed only at day 5, corresponding to the beginning of rind maturation: PcamFM013S059g000011, encoding a putative general substrate transporter, and PcamFM013S059g000012, encoding a putative β-galactosidase. During cheese maturation, lactose is present only during the first few days, during which it constitutes a primary source of carbon that is rapidly consumed by lactic acid bacteria, yeast and molds [36]. The possession of genes involved in lactose metabolism and transport, and the strong expression of these genes early in rind maturation, should be advantageous in the cheese environment. Two highly expressed genes from *Wallaby* were clustered next to each other, namely PcamFM013S057g00033 and PcamFM013S057g00034. None of these genes displayed significant similarity to sequences in public databases, and they were both found to be strongly expressed at later stages of rind maturation (days 21, 56 and 77). The patterns of expression of *Wallaby* and *CheesyTer* genes suggest that these regions could be involved in different stages of cheese maturation, reflecting changes in the substrate or in the organisms with which *Penicillium* fungi interact at different stages [14].

Finally, we found seven expressed genes in HTR2 and none in HTR1, the two other HTRs shared

between *P. camemberti* and *P. roqueforti*. Among the seven expressed genes of HTR2, PcamFM013S067g000021 was highly expressed at day 5. However no putative function could be assigned to this gene.

**DNA extraction for PCR amplification, PCR amplification and sequencing of amplicons**

Genomic DNA was extracted from fresh mycelium of the isolates listed in Table S3. Mycelium was obtained after 3–5 days on malt agar for *Aspergillus, Colletotrichum, Eupenicillium, Penicillium*, *Scopulariopsis* and *Fusarium* strains and on a confidential medium provided by starter producers for *Sporendonema casei*. The Qiagen DNeasy Plant Mini Kit (Qiagen Crawley, UK) was used for DNA extraction.

PCR was performed in a volume of 50 µL, containing 25 µL template DNA, 1.25 U AmpliTaq DNA polymerase (Roche Molecular Systems, Branchburg, NJ, USA), 5 µL 10 × *Taq* DNA polymerase buffer, 2 µL 5 mM dNTPs, 2 µL of each 10 µM primer. Primer sets designed to detect the presence of *Wallaby* (primers named Wallaby 1, 2 and 3) [16] and *CheesyTer* (primers named *CheesyTer* 1, 2 and 3) are shown in the section below. For *Wallaby,* results had been generated in our previous study [16]. For *CheesyTer,* amplifications were performed using a touchdown program with 20 cycles of 5 min at 94°C, a decreasing of 0.5°C every 30 sec from 60 to 50°C, and 30 sec at 72°C, followed by 35 additional cycles of 30 sec at 94°C, 30 s at 50°C, 30 sec at 72°C; the PCR program was followed by a final 30 s extension step at 72°C. PCR products were purified and sequenced by Eurofins (France). Further primer pairs were designed to confirm the *Wallaby* presence (as above) or absence (primer pair named *Wallaby* Ø) in 12 isolates of *P. roqueforti* (symbols in red in Table S3. PCR amplifications validating *Wallaby* absence were run with 30 cycles of 30 s at 95 ºC, 30 s at 60 ºC and 30 s at 72 ºC followed by a final extension of 1.5 min at 72 ºC.

**Primer sets used to test the presence of *Wallaby* and *CheesyTer* in the fungal collection**

| | |
|---|---|
| Wallaby 1 | Forward: 5'-GGCTCCTGCAGTAGACCAAG-3' |
| | Reverse: 5'-TTTGCCTTGGCTGAGTCTTT-3' |
| | Gene included: PROQFM164_S02g002809 |
| Wallaby 2 | Forward: 5'-CGCTATAACGGAAGCAGGAG-3' |
| | Reverse: 5'-CCCGTCCTTTTTAGACACCA-3' |
| | Gene included: PROQFM164_S02g002870 |
| Wallaby 3 | Forward: 5'-GCAACAGATCCAGCAGTTCA-3' |
| | Reverse: 5'-CGTTTCTCCTGAGTCCTTGC-3' |
| | Gene included: PROQFM164_S02g002951 |
| Wallaby Ø | Forward: 5'-AATGATCCAGTGTACGGGCT-3' |
| | Reverse: 5'-AAGAGACAAGCAAAGGAGGC-3' |
| CheesyTer 1 | Forward: 5'-ATCTCGACGACATAGGCAGG-3' |
| | Reverse: 5'-GGAGCGTGCAGATGAGAAAG-3' |
| | Gene included: PROQFM164_S02g000030 |
| CheesyTer 2 | Forward: 5'-CGTACGTCAAGCCTGGAATG-3' |
| | Reverse: 5'-AAGGCCATCATCATCGACCT-3' |
| | Gene included: PROQFM164_S02g000037 |
| CheesyTer 3 | Forward: 5'-CCCAAAGCCATGACCAACAA-3' |

Reverse: 5'-ACGTCTTCCTGTCCGACAAT-3'

Gene included: PROQFM164_S02g000032

**Fitness experiments: competition and use of the cheese substrate**

Experiments were performed in duplicate, in 8.5 cm Petri dishes. Goat's cheese medium was prepared as described in a previous study [22]. We first compared the growth rates of 50 *P. roqueforti* strains (26 W+C+ and 24 W-C-, Table S4a). Experiments were performed on a minimal medium, in which sucrose was the only available carbon source, and the cheese medium. Plates were placed in an incubator, at 25°C, in the dark, for seven days. We then measured two perpendicular diameters for each colony, to assess colony size.

We then measured the impact of several *P. roqueforti* strains, carrying or lacking *Wallaby* and *CheesyTer* on the growth of three fungal strains: *P. biforme* LCP05529 (W-C-), *P. nalgiovense* FM193 (W-C-) and *Geotrichum candidum* FM074 (W-C-). We performed competition experiments on cheese, malt agar and minimal media. We first prepared a suspension of fungal spores from cultures of each of 23 *P. roqueforti* strains (11 W+C+, 12 W-C-). This last class included eight isolates from cheeses and five isolates from wild environments (see Table S4b for the list of strains used). Spores of *P. roqueforti* were collected from the surface of five-day cultures and vortexed in a solution of water and Tween. We then poured 50 µl of spore suspension for each *P. roqueforti* strain into Petri dishes containing cheese, malt agar and minimal media (two replicates for each treatment). The three competitors were inoculated at a

single point at the center of the plates (5 µL each), 24 h after the plating of *P. roqueforti.* We also inoculated the three different media with the three competitors in the absence of *P. roqueforti* lawns. Plates were incubated in a refrigerated and ventilated FOC 225I incubator at 15°C in the dark (mimicking cheese cellar conditions) for 11 days. Colony diameters were measured after 12 days (two perpendicular measurements per colony). In a second set of this experiment, we inoculated and measured, as described above, using four *P. roqueforti* strains as lawns (2 W+C+ and 2 W-C-), on cheese medium, different strains of species displaying a polymorphism in the *Wallaby* and/or *CheesyTer* presence. We used as challengers different strains of *P. camemberti* (W+C- or W+C+, N=4), of *P. biforme* (W-C- or W+C+, N=4) and of *P. rubens* (W-C- or W+C-, N=4); see Table S4c for the list of strains used.

In the third experiment, we compared the growth of two *P. roqueforti* strains in face-to-face confrontations, after inoculations in opposite points of Petri dishes. We used 11 W-C- strains and 8 W+C+ strains (Table S4d) and performed all pairwise face-to-face confrontations on cheese medium. Plates were incubated at 25°C in the dark and observed daily. Colony growths were measured after 8 days, when the mycelia of the two strains met in the Petri dishes. We measured the deviations from symmetrical growth that would lead to a boundary between strains exactly in the middle of the Petri dish. Therefore, deviation is defined as:

$$Deviation = \frac{D_{focal\ strain} - D_{other\ strain}}{2}$$

where $D_{focal\ strain}$ is the distance between the inoculation point and the edge of the growth of the focal strain and $D_{other\ strain}$ is the distance between the inoculation point and the edge of the growth of the competing strain.

For the W+C+ *versus* W+C+ and W-C- *versus* W-C- confrontations, deviations were measured as positive or negative values, the focal strain being chosen at random. For the W+C+ *versus* W-C- confrontations, relative deviations were measured as positive or negative values, taking the W+C+ strain as the focal strain. The distributions of the deviations did not significantly depart from normal distributions (Shapiro Wilk tests W-C- *versus* W-C-: W=0.99, *p*-value=0.96; W-C- *versus* W+C+: W=0.98, *p*-value=0.36; W+C+ *versus* W+C+: W=0.98, *p*-value=0.88).

**Statistical analyses of competition and fitness experiments**

All statistical analyses were performed with JMP, version 9 (SAS Institute). ANOVAs were carried out by first running full models including all factors and interactions. We then sequentially removed the non-significant interactions.

# Supplemental References


S1. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research, http://igraph.org. InterJournal, Complex Systems 1695.

S2. 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.

S3. Foissac, S., Gouzy, J., Rombauts, S., Mathe, C., Amselem, J., Sterck, L., Van de Peer, Y., Rouze, P., and Schiex, T. (2008). Genome annotation in plants and fungi: EuGene as a model platform. Current Bioinformatics *3*, 87-97.

S4. Zdobnov, E., and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847-848.

S5. Li, L., Stoeckert, C.J., and Roos, D. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research *13*, 2178-2189.

S6. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. *32*, 1792-1797.

S7. Abascal, F., Zardoya, R., and Telford, M. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*,

W7-13.
S8. Smit, A., Hubley, R., and Green, P. (2013-2015 ). RepeatMasker Open-4.0 http://www.repeatmasker.org
S9. Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics *21*, 456-463.
S10. Noé, L., and Kucherov, G. (2005). YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Research *33*, W540-W543.
S11. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Research *19*, 1639-1645.
S12. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.
S13. Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics *27*, 2325-2329.