Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts

Souhir Marsit,1,2,3 Adriana Mena,4 Frédéric Bigey,1,2,3 François-Xavier Sauvage,1,2,3 Arnaud Couloux,5 Julie Guy,5 Jean-Luc Legras,1,2,3 Eladio Barrio,4 Sylvie Dequin,1,2,3 and Virginie Galeote*,1,2,3
1INRA, UMR1083, SPO, F-34060 Montpellier, France
2Montpellier SupAgro, UMR1083, SPO, F-34060 Montpellier, France
3Montpellier University, UMR1083, SPO, F-34060 Montpellier, France
4Department of Genetics, University of Valencia, and Department of Biotechnology, Institute of Agrochemistry and Food Technology, CSIC, Valencia, Spain
5CEA, Institut de Génomique, Genoscope, Centre National de Séquençage, Evry, France
*Corresponding author: E-mail: galeote@supagro.inra.fr.

Abstract

Although an increasing number of horizontal gene transfers have been reported in eukaryotes, experimental evidence for their adaptive value is lacking. Here, we report the recent transfer of a 158-kb genomic region between Torulaspora microellipsoides and Saccharomyces cerevisiae wine yeasts or closely related strains. This genomic region has undergone several rearrangements in S. cerevisiae strains, including gene loss and gene conversion between two tandemly duplicated FOT genes encoding oligopeptide transporters. We show that FOT genes confer a strong competitive advantage during grape must fermentation by increasing the number and diversity of oligopeptides that yeast can utilize as a source of nitrogen, thereby improving biomass formation, fermentation efficiency, and cell viability. Thus, the acquisition of FOT genes has favored yeast adaptation to the nitrogen-limited wine fermentation environment. This finding indicates that anthropic environments offer substantial ecological opportunity for evolutionary diversification through gene exchange between distant yeast species.

Key words: HGT, domestication, competition, nitrogen, oligopeptides, FOT genes.

Introduction

Horizontal gene transfer (HGT), the acquisition of new genes from distantly related species, has been established as a dominant force for prokaryotic evolution (Koonin and Wolf 2008) and is increasingly recognized as a prominent source of adaptation in eukaryotes. With the availability of sequence data for numerous eukaryotic genomes, many examples of HGT from prokaryotes to eukaryotes and, increasingly, from eukaryotes to eukaryotes have been reported, in particular among fungi (Keeling and Palmer 2008; Andersson 2009; Marcit-Houben and Gabaldon 2009; Gladieux et al. 2014).

HGT is a means by which organisms can rapidly acquire new biological functions (Keeling and Palmer 2008), facilitating their specialization and adaptation to new environments. An analysis of the potential function of 66 genes identified in nine reported cases of HGT between fungi showed a high representation of genes involved in nitorgen, amino acid, nucleic acid, protein, carbohydrate, and lipid metabolism, with cellular transport being the most prevalent putative function (Richards 2011). The acquisition of transporter-encoding genes is likely to be advantageous, as these genes may function to allow fungi to colonize new environments and utilize previously inaccessible nutrient sources.

Multiple recent HGTs have been reported in man-made environments and food processing chains (Borneman et al. 2008; Novo et al. 2009; Fitzpatrick 2011; Cheeseman et al. 2014). These ecological niches impose novel and stressful environments that exert strong selective pressure on microorganisms such as fungal pathogens, food contaminants, and the domesticated fungi used in beverages and food. The recent transfer of a huge (575 kb) genomic island, called Wallaby, between cheese-associated Penicillium species has recently been reported (Cheeseman et al. 2014). The function of the predicted 250 genes transferred suggests that some of them are involved in competition between microorganisms. Another striking example of HGT associated with domestication is the occurrence of multiple gene transfers from evolutionarily distant yeasts in the genome of Saccharomyces cerevisiae wine yeast (Novo et al. 2009). Three large genomic regions encompassing 39 genes (including five pseudogenes) and encoding functions potentially important for winemaking, such as sugar and nitrogen metabolism, have been found in S. cerevisiae (Novo et al. 2009). One of these regions was acquired from Zygosaccharomyces bailii, a distantly related species frequently found in wine environments as a contaminant. In many strains of S. cerevisiae, several copies of this region were found in different chromosomal locations, suggesting that a circular episome is involved in the transfer from Z. bailii to a recipient S. cerevisiae strain (Borneman et al. 2011; Galeote et al. 2011). Another independent event led to the acquisition, from an as yet unidentified yeast species, of a...
65-kb subtelomeric region in EC1118 chromosome XV called region C (Novo et al. 2009). This region carries 19 genes, including FSY1, which encodes a high-affinity fructose/H+ symporter that may be advantageous at the end of grape fermentation, when fructose is the most abundant sugar (Galeote et al. 2010). Two tandem duplicated genes—FOT1–2, encoding oligopeptide transporters—were also acquired, increasing considerably the number of di/tripeptides transported by this strain (Damon et al. 2011). However, the functional and ecological impacts of these events have not yet been demonstrated.

Despite a high number of documented HGT cases, there is a lack of functional evidence demonstrating the evolutionary advantage and biological relevance of HGT, preventing the evaluation of the importance of this process in the evolution of eukaryotic genomes (Ros and Hurst 2009; Schonknecht et al. 2014). In very few instances, the adaptive advantage of HGT genes in eukaryotes has been experimentally demonstrated, in most cases for single gene transfers of prokaryotic origin. A well-known example is the HGT of a bacterial gene encoding an oxygen-independent form of dihydroorotate dehydrogenase (DHOD) involved in the synthesis of uracil in Saccharomycetaceae, whereas the ancestral gene encoding a strict aerobic mitochondrial enzyme was lost in Saccharomyces sensu stricto species. Thus, this HGT was an important step in allowing these yeasts to grow in low-oxygen environments (Gojkovic et al. 2004; Hall et al. 2005).

Here, we report that Torulaspora microellipsoides is the donor of a large genomic region (region C) of between 65 and 80 kb in length found in S. cerevisiae wine yeast and closely related strains. Although the structure and number of genes present on this genomic island vary between wine strains, two tandem-duplicated FOT1–2 genes encoding oligopeptide transporters are strongly conserved. We show that these genes confer a competitive advantage to wine yeast by increasing the amount and nature of nitrogen sources (oligopeptides) transported from grape must. These results highlight the existence of recurring eukaryote–eukaryote transfers between species sharing an ecological niche and constitute the first clear evidence of HGT-mediated adaptation of wine yeast to the nutritional challenges faced during wine fermentation.

**Results**

**On the Origin of Region C**

The origin of region C, a large genomic region of 65 kb in size initially found in the genome of the commercial S. cerevisiae wine strain EC1118 and carrying 19 genes (supplementary table S1, Supplementary Material online), has not been elucidated (Novo et al. 2009). We assayed for the presence of region C by polymerase chain reaction (PCR) amplification and sequencing of genomic segments of 187 strains from different genera belonging to the Saccharomyces complex (Kurtzman 2003). An analysis of these sequences indicated that T. microellipsoides might be the donor of region C. We therefore performed the sequencing and assembly of the whole genome of the T. microellipsoides type strain CLIB830T. Mapping of the EC1118 region C to this genome enabled the identification of one scaffold of T. microellipsoides (fig. 1). This scaffold exhibits 99.5% shared identity with region C, which reaches 99.9% if we exclude two small regions of 1,400 and 2,150 bp with substantially lower identity (97.3% and 91.7% respectively) than the rest of region C (supplementary fig. S1, Supplementary Material online). Several hypotheses, exposed later, may explain the higher diversity of these two regions, encompassing FOT1 gene (6502g) and on the other hand ATO3-like (6568g).

We found that the entire region C was present in T. microellipsoides. Moreover, T. microellipsoides carries an additional 93-kb region containing 40 genes between the first and the second genes of region C (as described in EC1118). The first gene of the EC1118 region C was a truncated, 500-bp ARB1-like gene (6469g), whereas in T. microellipsoides, a complete 1.8-kb ARB1-like gene was identified. This finding suggests an initial insertion of a 158-kb genomic fragment from T. microellipsoides into the S. cerevisiae genome followed by a loss of the 93-kb region. Phylogenetic analysis of each protein encoded by genes of the EC1118 region C (supplementary fig. S2A–O, Supplementary Material online) shows that most proteins cluster together with homologs from the Zygosaccharomyces/Torulaspora clade, supporting a transfer from T. microellipsoides to S. cerevisiae. However, few proteins (Hxt13, Gal10, Ato3, Fre7, and Sor1) clustered more closely than expected to Saccharomyces homologs. In these latter cases, they clustered with homologs from only few
Saccharomyces species which refutes the hypothesis that these genes are originated from this clade. The observed topology could be explained by a complex evolutionary history of these genes, as suggested by the higher divergence observed above between the EC1118 and T. microellipsoides ATO3 gene which may attest that this gene is evolving rapidly. Alternatively, this topology could be explained by other but more ancient HGT events associated with a pattern of gain and loss across species leading to a complex phylogeny. Such a turbulent evolutionary history has recently been described for FSY1 found in region C (Coelho et al. 2013). The lack of genome sequences of other species closer to T. microellipsoides could be also another plausible explanation to the obtained topologies.

Evolutionary History of Region C in S. cerevisiae

We previously reported that regions A, B (from Z. bailii), and C (from T. microellipsoides, this study) are frequently found in wine yeasts and mosaic strains but only rarely in other lineages, suggesting that their acquisition is related to the early stages of wine domestication (Novo et al. 2009). In a previous study based on a limited number of available low-coverage genomes (39), region C was suggested to be incomplete in several wine strains (Novo et al. 2009). We therefore systematically analyzed 134 available genomes of S. cerevisiae from different ecological niches (wine, flor, rum, sake, palm wine, oak, bread, laboratory, clinical isolate, etc.; see supplementary table S2, Supplementary Material online, for strain list) for the presence of the 19 region C genes using BLASTN. This survey identified 36 strains containing at least one gene from region C, including 23 of 45 wine strains.

We selected 25 high-coverage genomes carrying this region for further analysis. Region C was fully conserved in 50% of the strains and partially conserved in the other half (fig. 2). Most strains that have a partially conserved region C have conserved only the four first genes, putatively encoding an ATPase (ARBI-like, 6469g), a transcription factor (6480g), an allantoate permease (SEO1-like, 6491g), and one copy of FOT. Interestingly, three strains (VIN13, K1-28, and 20B2) have lost all genes from this region except for one FOT copy and a truncated SEO1-like gene. These data show a robust conservation of FOT, suggesting that this gene plays an important role during the winemaking process.

Sequence alignment of Fot1 and Fot2 proteins and of the single Fot copy (present in strains K1-28, 20B2, VIN13, MO1A, 22A4, M22, DB1373, Lava32_15, L1414, and D47-6) revealed a novel Fot protein that we named Fot3. A comparative analysis of the different FOT genes found in S. cerevisiae and T. microellipsoides shows 99.7% identity between FOT2 and a T. microellipsoides FOT gene that was named FOT2_TOMI. In contrast, Fot1, Fot2, and the newly identified Fot3 in S. cerevisiae have only 97.8%, 93.3%, or 98% identity, respectively, with the second copy of T. microellipsoides FOT, which was named FOTX_TOMI. Fot1 and Fot3 exhibit
distinct composite amino acid patterns (supplementary fig. S3, Supplementary Material online), suggesting that both FOT1 and FOT3 are the result of two independent gene conversions between FOTX_TOMI and FOT2_TOMI, eventually accompanied by the independent loss of downstream regions (fig. 3A).

These results are supported by a phylogenetic tree showing the evolutionary relationships among the FOT genes (fig. 3B). The location of region C (or the conserved portion of this region) differs according to strain (fig. 2). Strains 22A4, L1414, and Lava32_15 contain the four first genes of region C on chromosome XV, similar to EC1118 but on the left arm subtelomeric region instead of the right arm (confirmed by PCR analysis). An analysis of the junctional sequence where region C is inserted into the genomes of the different strains revealed the presence of a poly(T) sequence (23–30 Ts) in the 14 strains possessing this region (fig. 2). Interestingly, we found a poly(T) sequence at the same locus in both S288C (27 Ts) and T. microellipsoides (21 Ts; fig. 4). This finding suggests that region C was inserted in the subtelomeric region of the chromosome XV of a recipient strain by homologous recombination between these poly(T) sequences.

We also identified in four strains (MO1A, 22A4, CBS7957, and CLIB382) a 7.3-kb-long sequence extension of region C immediately upstream of the second gene (POT3-like) that was also present in the T. microellipsoides genome. Two open reading frames (ORFs) encoding a probable fluconazole resistance protein and a hexose transporter were identified in this fragment. In addition, we identified a 7.4-kb contig

---

**Fig. 3.** Hypothetical formation of different FOT genes. (A) FOT3 and FOT1 may have been generated through two independent gene conversion events between FOTX and FOT2. (B) Phylogenetic relationships between protein sequences Fot 2 FotX, different parts of Fot 1 and Fot 3 (Fot 1.1, Fot 1.2, Fot 3.3, Fot 3.4, and Fot 3.5 represented in part A of the figure), and their closest homologs searched using BLASTP in the National Center for Biotechnology Information and GRYC databases. Protein sequence alignment was performed with Muscle and visualized with Mega 5.2.2. Bootstrap values (percentage) based on 1,000 replicates are indicated at the nodes.

**Fig. 4.** Mechanism of insertion of region C by homologous recombination between poly(T) sequences located upstream of the region C in Torulaspora microellipsoides and downstream of YOR393W on Saccharomyces cerevisiae chromosome XV.
encompassing five genes, with 99.9% of similarity to the up-
stream part of the 93-kb region identified in the genome of T.

t. microellipsoides, in CBS7957 and CLIB382 strains. These five
ORFs encode a putative complete ATP-binding protein Arb1-
like, a Cog1 protein involved in protein trafficking, a pyrimi-
dine nucleotidase Sdt1, a killer toxin resistance protein Kre29-
like, and a Golgi GDP-mannose transporter Vrg4-like. The
presence of these proteins suggests that the large genomic
fragment of 158 kb was initially transferred to a recipient
S. cerevisiae strain closely related to CLIB382 with a mosaic
genome (Schacherer et al. 2009; Cromie et al. 2013) or to
CBS7957, which exhibits close evolutionary relationships
with the latter strain (supplementary fig. S4, Supplementary
Material online).

These results indicate that region C has undergone many
chromosomal rearrangements through gene conversion as
well as gene losses following its acquisition from the donor
yeast. In all cases, FOT genes were highly conserved among
wine yeasts, suggesting that their presence confers a strong
adaptive advantage in this environment.

The Effect of FOT Genes on Biomass, Viability, and
Grape Must Fermentation
To investigate whether region C and FOT genes can confer an
advantage upon exposure to the selective conditions of grape
must fermentation, we deleted either the entire region C or
the two tandem-duplicated genes FOT1–2 in the wine yeast
59A, a haploid derivative of the EC1118 strain (Ambroset et al.
2011). During fermentation in a synthetic medium simulating
a grape must, the wild-type and deletion mutants exhibited similar
growth, fermentation rates, and metabolite profiles (data not shown).
We therefore assessed the role of the
FOT genes in a Chardonnay grape must containing more
diversified nitrogen resources (e.g., γ-amino butyric acid, pep-
tides, or proteins) in addition to ammonium and amino acids.
Although the deletion mutants showed superior growth and
fermentation rates during the first hours of fermentation (fig.
5A and B), these advantages depreciated very quickly, result-
ing in an 8% and 12% lower biomass production in ΔC
(183 mg/l, P = 0.04) and Δfot1–2 (270 mg/l, P = 0.07) mutants,
respectively, compared with the wild-type. Furthermore, the
deletion mutants exhibited a higher mortality rate by the end
of fermentation (25% and 18% higher for ΔC and Δfot1–2,
respectively, compared with the wild-type strain) and per-
formed incomplete fermentations, with 6 g/l of residual fruc-
tose for ΔC and 2.5 g/l for Δfot1–2 (fig. 5C and D). Thus, our
data suggest that region C, and particularly the FOT1–2 genes,
plays a key role in maintaining cell viability and fermentation
efficiency during the last step of wine fermentation. Nitrogen
is the main limiting nutrient during wine fermentation. Its
availability is directly related to biomass production, which
governs the fermentation rate (Varela et al. 2004). As a con-
sequence, nitrogen deficiency is the most prevalent cause of
stuck and sluggish fermentations. As FOT genes encode trans-
porters of a broad range of oligopeptides (Damon et al. 2011),
the phenotypic differences between the wild-type and
Δfot1–2 strains may result from a higher uptake by 59A of
nitrogen sources, such as small peptides.

Effect of FOT Genes on the Uptake of Oligopeptides
To evaluate the amount of oligopeptide in grape must, we pu-
rified the oligopeptide fraction less than 1 kDa, which includes
small peptides of two to nine amino acids. This oligopeptide
fraction (< 1 kDa) represents a substantial part of the total
nitrogen available in the grape must (17%), and approximately
10% of the total assimilated nitrogen. We then determined the
total content of nitrogen, free amino acids, and oligopeptides in
the grape must and in the wine fermented by the wild-type and
mutant strains at the beginning of the stationary phase (40 g/l
of CO₂ produced, after the nitrogen source was totally
exhausted). Although all strains consumed the same amount
of free amino acids and ammonium, 59A consumed approxi-
ately 35% more oligopeptide-derived nitrogen than the de-
letion mutants (~10 mg/l, P = 0.027 and 7.7 mg/l, P = 0.037
more than Δfot1–2 and ΔC, respectively). This difference is
in agreement with the measured difference in biomass be-
tween the wild-type strain and the deletion mutants.

An analysis of the amino acid composition of the oligo-
peptide fraction of the grape juice shows that it is mainly
composed of glutamate or glutamine (these two com-
pounds are indistinguishable; table 1). The oligopeptide frac-
tion consumed by 59A was particularly rich in glutamate/
glutamine and to a lesser extent in several other amino acids (fig. 6).

Because glutathione is a tripeptide composed of L-glutamate, L-cysteine, and glycine, which can be found in grape must in amounts ranging from 0 to 100 mg/l (Cheynier et al. 1989; Kritzinger et al. 2013), we also quantified this compound. In S. cerevisiae, glutathione is involved in many cellular processes, such as redox homeostasis and response to oxidative stress, and can also be used as a source of nitrogen and sulfur. We found that most glutathione in grape must supernatant is in the form of oxidized glutathione (GSSG; 16 mg/l vs. less than 2 mg/l reduced glutathione, GSH). The wild-type strain consumed more GSSG than deletion strains (fig. 7A). Furthermore, the wild-type strain grew better on SD medium supplemented with glutathione as the sole nitrogen source than the deletion mutant, which suggests that Fot1–2 is also able to transport glutathione (fig. 7B). Therefore, the consumption of glutathione contributes to the formation of a higher amount of biomass, even if the extra amount of nitrogen from glutathione consumed by 59A does not exceed 2 mg/l (corresponding to assimilation of 20% of the oligopeptides rich in glutamate/glutamine by Fot1–2).

**FOT1–2 Genes Increase Strain Fitness during Wine Fermentation**

To evaluate the adaptive role of FOT genes, we performed competition experiments under controlled wine fermentation conditions. Figure 6 illustrates the amino acid composition of the oligopeptide fraction of the Chardonnay must specifically assimilated by 59A in comparison with ΔC (blue) or Δfot1–2 (red) strains at the beginning of the stationary phase (40 g/l of CO2 production). Data are presented as means ± SD (n = 3).

**Table 1. Amino Acid Composition of the Oligopeptide Fraction (<1 kDa) of the Chardonnay Must.**

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>mg/l</th>
<th>mg N/l</th>
<th>μmol/l</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>0.519 ± 0.05</td>
<td>0.081 ± 0.008</td>
<td>5.83 ± 0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Arg</td>
<td>7.51 ± 2.7</td>
<td>2.41 ± 0.8</td>
<td>43.19 ± 15</td>
<td>1.7</td>
</tr>
<tr>
<td>Asp + Asn</td>
<td>12.31 ± 1.4</td>
<td>1.29 ± 0.14</td>
<td>92.58 ± 10</td>
<td>3.6</td>
</tr>
<tr>
<td>Glu + Gln</td>
<td>292.64 ± 3.9</td>
<td>27.87 ± 0.3</td>
<td>1,990.78 ± 26</td>
<td>77.3</td>
</tr>
<tr>
<td>Gly</td>
<td>12.48 ± 0.6</td>
<td>2.33 ± 0.1</td>
<td>166.44 ± 8</td>
<td>6.5</td>
</tr>
<tr>
<td>His</td>
<td>4.14 ± 1.2</td>
<td>1.12 ± 0.34</td>
<td>26.77 ± 8</td>
<td>1.0</td>
</tr>
<tr>
<td>Ile</td>
<td>1.34 ± 0.1</td>
<td>0.144 ± 0.03</td>
<td>10.30 ± 2</td>
<td>0.4</td>
</tr>
<tr>
<td>Leu</td>
<td>4.09 ± 0.4</td>
<td>0.437 ± 0.05</td>
<td>31.25 ± 3</td>
<td>1.2</td>
</tr>
<tr>
<td>Lys</td>
<td>3.88 ± 0.3</td>
<td>0.744 ± 1.9</td>
<td>26.57 ± 2</td>
<td>1.0</td>
</tr>
<tr>
<td>Phe</td>
<td>1.88 ± 0.4</td>
<td>0.159 ± 0.04</td>
<td>11.42 ± 2.8</td>
<td>0.4</td>
</tr>
<tr>
<td>Ser</td>
<td>6.54 ± 1.1</td>
<td>0.873 ± 0.1</td>
<td>62.36 ± 10</td>
<td>2.4</td>
</tr>
<tr>
<td>Thr</td>
<td>6.23 ± 1.3</td>
<td>0.733 ± 0.1</td>
<td>52.39 ± 11</td>
<td>2.0</td>
</tr>
<tr>
<td>Tyr</td>
<td>6.52 ± 1.4</td>
<td>0.504 ± 0.1</td>
<td>36.05 ± 8</td>
<td>1.4</td>
</tr>
<tr>
<td>Val</td>
<td>2.43 ± 1</td>
<td>0.291 ± 0.1</td>
<td>20.80 ± 8</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**NOTE.—N, assimilable nitrogen.**
conditions. Cocultures of 59 A expressing green fluorescent protein (GFP) and FOT1–2 mutant yeast expressing red fluorescent protein (t-dimer2) (fig. 8A) were achieved during three successive fermentation experiments on grape must. The relative abundance of each strain was monitored over time by flow cytometry (fig. 8B). These competition assays revealed a considerable competitive advantage of the strain with FOT genes in natural must. Despite a transient advantage of the deletion mutant during the early steps of fermentation, which is in line with its higher growth and fermentation rates at this stage, the /C1 fot1–2 strain was quickly dominated during the stationary phase (fig. 8C). At the beginning of the third successive coculture, the FOT1–2 mutant completely disappeared from the coculture (fig. 8C). We verified that these results were not biased by the use of different fluorochromes and that the presence of these fluorochromes did not affect the fitness of the strains. These results demonstrate the strong selective advantage provided by FOT1–2 genes during wine fermentation.

**Discussion**

Saccharomyces cerevisiae wine yeasts have been used for winemaking for several millennia and display outstanding fermentative capacities. During wine fermentation, these yeasts are challenged with various stresses, including high sugar content, low pH, the presence of ethanol and sulfites, and long periods of nutrient starvation, which have prompted their adaptive evolution.

In this study, we identified T. microellipsoides as the donor of the large 65-kb genomic region previously characterized in the genome of S. cerevisiae wine yeasts (Novo et al. 2009). The high identity (99.9 %) of the regions found in these two distant species of the Saccharomycetaceae family indicated a recent transfer event. In addition, the maximum divergence within a conserved 9-kb fragment of the region C containing the first three genes (ARB1, PUT3, and SEO1) between 12 wine strains is very low, about 0.2 nucleotidic difference per kilobase. Considering that vine and yeast domestication occurred during Neolithic era (~9,000 years ago) (McGovern et al. 2004; Fay and Benavides 2005; Legras et al. 2007) and that divergence between wine strains is about 1.0 nucleotidic difference per kilobase (Fay and Benavides 2005; Liti et al. 2009), we can estimate that this HGT event occurred more recently than 2,000 years ago. This region is present in the genomes of many wine strains as well as few mosaic strains, including two evolutionarily related strains that are closest to the donor, as shown by the presence of an additional larger T. microellipsoides block in their genomes. The genome of one of these strains, CLIB382, was shown to have a major contribution from the wine group (Schacherer et al. 2009). A plausible scenario is the transfer of a large genomic fragment (158 kb) from T. microellipsoides to a recipient S. cerevisiae strain, likely introgressed by a rare unstable hybridization
alleles with potentially specialized functions were generated to contribute new or important functions. Interestingly, novel FOT alleles were identified in various strains of Torulaspora microellipsoides, which have a patchy distribution among various strains. Except for losses and gene conversion through an eukaryote region, the superior viability might be linked to the quality of assimilated oligopeptides. In terestingly, the presence of FOT1–2 genes in grape might be explained, at least partly, by the greater assimilation of glutathione. Glutamate may also play an important role in maintaining cell viability and in tackling oxidative stress. The NADP⁺-dependent glutamate dehydrogenase Gdh3 plays a role in resistance to stress-induced apoptosis in the stationary phase (Lee et al. 2012). Indeed, S. cerevisiae gdh3 mutants rapidly lose viability, are depleted in glutathione and accumulate ROS during the stationary phase.

Overall, this study supports a major role for HGT in the evolution of wine yeast genomes. Membrane transporters are one functional category of gene more prone to horizontal transfer and to evolution by duplication (Richards 2011; McDonald et al. 2012). This plasticity may be linked to the important roles that transporters can play in detoxification processes or in providing a competitive advantage in the struggle for nutrients (Richards 2011; Richards and Talbot 2013). In fact, it is important for osmotrophs such as fungi to acquire nutrients rapidly, thereby making public goods inaccessible to others (Richards and Talbot 2013). Several other cases of HGT involving high-affinity transporters have been reported and likely reflect adaptation to these pressures (Slot and Hibbett 2007; Galeote et al. 2010; Coelho et al. 2013).

Overall, the origin of the Torulaspora microellipsoides region C is reminiscent of the Z. bailii origin of region B, which is also widespread in the genome of wine yeasts. Torulaspora microellipsoides and Z. bailii are closely related to a food and beverage spoilage-yeast cluster (Steels et al. 1999). These species are isolated from various fruit juices, soft drinks, and beverages, including wine (Deak 2007; Kurtzman 2011). Interestingly, it has been reported that ASP3, a S. cerevisiae gene induced by nitrogen starvation, was likely acquired by HGT from Wickerhamomyces anomalus, a distant yeast species used in a variety of biotechnological applications (League et al. 2012). ASP3 can hydrolyze D-asparagine (Dunlop et al. 1978) and is also involved in the utilization of dipeptides with C-terminal asparagine residues (Homann et al. 2005). A high diversity in the capacity of S. cerevisiae strains of various origins to utilize di- and tripeptides as nutrient sources has been highlighted previously (Homann et al. 2005). The presence of assimilated glutathione, an essential metabolite that protects cells against oxidative stress and aging. Glutathione protects cells from the deleterious effects of reactive oxygen species (ROS), which may cause oxidative damage to proteins, nucleic acids and lipids as well as other cellular components, and may lead to cell death (Lee et al. 2001). Strains with disruptions in genes involved in glutathione biosynthesis have a reduced tolerance to a wide range of stress conditions and undergo apoptosis at a high rate compared with the wild-type strain (Grant et al. 1998; Madoe et al. 1999). Conversely, exogenous administration of glutathione rescues the higher mortality of S. cerevisiae strains in the presence of oxygen (Nestelbacher et al. 2000). During wine fermentation, ROS may be part of the cellular response to a variety of stressors (Gibson et al. 2008; Landolfo et al. 2008).

In particular, nitrogen depletion induces ROS accumulation during must fermentation, causing cell membrane damage and cell cycle arrest in G0/G1 and inducing autophagy (Mendes-Ferreira et al. 2010). Therefore, the higher viability of strains with FOT1–2 genes in grape might be explained, at least partly, by the greater assimilation of glutathione. Glutamate may also play an important role in maintaining cell viability and in tackling oxidative stress. The NADP⁺-dependent glutamate dehydrogenase Gdh3 plays a role in resistance to stress-induced apoptosis in the stationary phase (Lee et al. 2012).

Overall, this study supports a major role for HGT in the evolution of wine yeast genomes. Membrane transporters are one functional category of gene more prone to horizontal transfer and to evolution by duplication (Richards 2011; McDonald et al. 2012). This plasticity may be linked to the important roles that transporters can play in detoxification processes or in providing a competitive advantage in the struggle for nutrients (Richards 2011; Richards and Talbot 2013). In fact, it is important for osmotrophs such as fungi to acquire nutrients rapidly, thereby making public goods inaccessible to others (Richards and Talbot 2013). Several other cases of HGT involving high-affinity transporters have been reported and likely reflect adaptation to these pressures (Slot and Hibbett 2007; Galeote et al. 2010; Coelho et al. 2013).

Overall, the origin of the Torulaspora microellipsoides region C is reminiscent of the Z. bailii origin of region B, which is also widespread in the genome of wine yeasts. Torulaspora microellipsoides and Z. bailii are closely related to a food and beverage spoilage-yeast cluster (Steels et al. 1999). These species are isolated from various fruit juices, soft drinks, and beverages, including wine (Deak 2007; Kurtzman 2011). Interestingly, it has been reported that ASP3, a S. cerevisiae gene induced by nitrogen starvation, was likely acquired by HGT from Wickerhamomyces anomalus, a distant yeast species used in a variety of biotechnological applications (League et al. 2012). ASP3 can hydrolyze D-asparagine (Dunlop et al. 1978) and is also involved in the utilization of dipeptides with C-terminal asparagine residues (Homann et al. 2005). A high diversity in the capacity of S. cerevisiae strains of various origins to utilize di- and tripeptides as nutrient sources has been highlighted previously (Homann et al. 2005). The presence
of the ASP3 and FOT genes might contribute to this phenotypic diversity.

The finding of recurrent transfers in food (Novo et al. 2009; League et al. 2012; Cheeseman et al. 2014) suggests that man-made environments offer ecological opportunities for sharing metabolic genes between distant yeasts. These HGT processes may help yeast to evolve to survive fluctuations in the external environment by adapting their metabolic networks to meet the challenges of their ecological niche. In line with this hypothesis, we found that FOT genes facilitate the utilization of the nitrogen resources in grape must, which are present in limiting amounts in grape juice. Nitrogen is a key nutrient consumed at the beginning of the fermentation process, and its deficiency can lead both to sluggish and stuck fermentations (Blateyron and Sablayrolles 2001). It is noteworthy that several of the transferred genes gained by the acquisition of the three non-Saccharomyces regions identified in wine yeasts (Novo et al. 2009) have putative functions related to nitrogen metabolism, including an asparaginase, an oxoprolinase, an ammonium transporter, an allantoate transporter, and two transcription factors related to the biosynthesis enzymes involved in lysine and proline utilization. This study strongly supports the view that HGT plays an important role in the adaptation of fungi to their environments and that these events are favored in the ecological niches created by food and beverages, where many yeast species cohabit. As a consequence, a better knowledge of the functional and ecological impacts of HGT also increases our understanding of the unintentional domestication of wine yeasts through elucidating the genetic bases of yeast strain adaptation to specific ecological niches.

Materials and Methods

Yeast Strains and Fermentation Conditions

The S. cerevisiae strain 59A is a haploid derivative of the commercial wine strain EC1118 (Ambroset et al. 2011). The recombinant strains constructed and used in this study are listed in table 2. Fermentations were performed in synthetic medium, which mimics a natural grape must (Bely et al. 1990) and contains 200 g/l glucose/fructose, 425 mg/l assimilable nitrogen contained in a mixture of amino acids and ammonium, 1.12 mg/l oleic acid and 3.75 mg/l ergosterol, pH 3.3. For a natural grape must, we used a Chardonnay (Coursan, France 2012) containing 230 g/l of sugars (100 g/l glucose and 130 g/l of fructose) and 240 mg/l of yeast assimilable nitrogen (free amino acids and ammonium).

Fermentation experiments were performed in 1.2-l fermenters inoculated at a density of 10⁶ cells/ml. Yeast cells were first grown overnight at 28 °C in 10-ml flasks containing YPD (1% Bacto yeast extract, 2% bactopeptone, 2% glucose) with shaking (180 rpm) and were transferred to 20-ml flasks containing the grape must for 24 h. These precultures were used to inoculate fermenters equipped with airlocks to maintain anaerobiosis, and fermentations were performed at 24 °C with constant magnetic stirring (150 rpm). The CO₂ released was estimated by an automatic measurement of the weight loss of the fermenter every 20 min, and the rate of CO₂ production was calculated by polynomial smoothing of the CO₂ released (Sablayrolles et al. 1987).

Strain Construction

Deletion of the 65-kb region C was performed by four successive deletions of 8- to 21-kb fragments. Each fragment was replaced by the loxP–KanMX4–loxP deletion module using the short-flanking homology PCR technique (Schiestl and Gietz 1989). The KanMX4 cassette was PCR amplified from the pUG6 vector using the primers listed in table 3. The PCR fragments were used to transform 59A using the lithium acetate method (Guldener et al. 1996). The deletions were verified by PCR analysis of total DNA isolated from kanamycin-resistant transformants, using primers flanking the deleted regions. After each deletion, we used the Cre-loxP recombination system for marker rescue (Guldener et al. 1996). The deleted strain was transformed with the cre expression plasmid pSH65, which carries the phleomycin marker gene and the cre gene under the control of the inducible GAL1 promoter. Cells were plated on YPM medium containing maltose (2%) to induce expression of the Cre recombinase. The plasmid was then removed by streaking cells on YPD without phleomycin. Deletion of the two tandem duplicated FOT1–2 genes was obtained by replacing the two ORFs with the loxP–KanMX–loxP module using the custom primers listed in table 3.

Because strain 59A forms aggregated cells in liquid culture, which causes problems in cell counting and flow cytometry applications, we deleted the AMN1 gene, which encodes a protein required for daughter cell separation and is involved in cell aggregation (Wang et al. 2003; Yvert et al. 2003). Deletion of AMN1 in 59A, ΔC, and Δfot1–2 strains was performed as described above using primers listed in table 3.

For competition experiments, the wild-type and Δfot1–2 strains were tagged with the fluorochromes GFP and tdimer2, respectively. To this end, a cassette containing

Table 2. Strains Used and Constructed in This Study.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>59A</td>
<td>MATa ho</td>
<td>Ambroset et al. (2011)</td>
</tr>
<tr>
<td>Δamn1</td>
<td>MATa ho AMN1::kanMX4</td>
<td>This study</td>
</tr>
<tr>
<td>Δfot1–2</td>
<td>MATa ho FOT1–2::loxP AMN1::kanMX4</td>
<td>This study</td>
</tr>
<tr>
<td>ΔC</td>
<td>MATa ho Region C::loxP; AMN1::kanMX4</td>
<td>This study</td>
</tr>
<tr>
<td>59A-GFP</td>
<td>MATa ho AMN1::TEF2Pr-GFP-ADH1-NATMX4</td>
<td>This study</td>
</tr>
<tr>
<td>Δfot1–2-tdimer2</td>
<td>MATa ho FOT1–2::loxP AMN1::TEF2Pr-tdimer2-ADH1-NATMX4</td>
<td>This study</td>
</tr>
</tbody>
</table>
the TEF2 promoter, GFP or t-dimer2, the ADH1 terminator, and a gene conferring resistance to clonNat was amplified by PCR using specific primers (Fluo-F and Fluo-R, table 3) and the previously described plasmids pFA6a-TEF2Pr-ADH1-NATMX4 or pFA6a-TEF2Pr-t-dimer2-ADH1-NATMX4 (Breslow et al. 2008) as templates. The amplified fragments were inserted into the AMN1 locus, yielding strains 59A-GFP and Δfot1–2-t-dimer2.

### Cell Population, Viability, and GFP Activity

Population size was determined by counting yeast cells with an electronic particle counter (Multisizer 3 coulter counter; Beckman Coulter, CA). Cell viability and GFP activity were determined by flow cytometry using a C6 cytometer (Accuri, BD Biosciences, Piscataway). Cell viability and GFP activity were analyzed by flow cytometry analysis performed after 10 min of staining. Fluorescence data for cells stained by PI were collected in the FL2 channel. For GFP activity, fluorescence (excitation 488 nm, emission 530 nm) was collected in the FL1 channel.

### Fluorescent Microscopy

Fluorescent cells tagged with GFP and t-dimer2 in coculture were visualized at the beginning and the end of fermentation of Chardonnay must. The microscope (Axio Imager Cam MRM A2, Carl Zeiss, New York) equipped with an excitation source and a range of filters for viewing fluorescent cells was used. For this study, the filter 38 (excitation BP 470/40 and emission BP 525/50) was used for cells tagged with GFP (488 nm excitation and 530 nm emission). The filter 20 (excitation: BP 546/12, emission: BP 575–640) was used for cells tagged with t-dimer2 (excitation 568 nm, 630 nm emission).

### Analytical Methods

Residual fructose was analyzed by high-pressure liquid chromatography (HPLC 1100; Agilent Technologies, CA) on an HPX-87H Aminex column (Bio-Rad Laboratories Inc., CA). Dual detection was performed with a refractometer and a UV detector (Hewlett Packard).

The total nitrogen concentration was determined by distillation and back-titration according to the Kjeldahl mineralization method (Scheiner 1976). The residual ammonium concentration in the medium was measured by spectrophotometry using glutamate dehydrogenase (Enzytec fluid Ammonia; R-Biopharm, Darmstadt, Germany).

The oligopeptide fraction of the must was purified by two successive ultrafiltrations using a hollow fiber cartridge filter, UFP-10-C-4, to recover the peptides under 10 kDa followed by ultrafiltration using a hallow fiber cartridge filter, UFP-1-C-4, to recover oligopeptide fractions under 1 kDa. The oligopeptide purified fraction less than 1 kDa was then hydrolyzed by incubation for 24 h at 110°C in 6N HCl. Upon hydrolysis, some amino acids are degraded. A part of cysteine and methionine are oxidized and degraded into H-cysteine and methionine sulfone. Tryptophan, sensitive to heat in an acid medium, is completely degraded and thus not determined. Similarly, the glutamine is converted to glutamate; measured glutamate corresponds to the concentration of glutamate and glutamine. This is also the case for asparagine which is converted to aspartate. Total amino acids were analyzed with a Biochrom 30 analyzer. Molecules with a high molecular weight were precipitated for 1 h at 4°C and centrifuged (4°C, 10 min, 3,000 × g). The sample was then filtered through a 0.22-µm pore-size Millipore nitrocellulose membrane. Amino acids were separated by liquid chromatography on an ion-exchange column (Ultra-pac-8 lithium form; Amersham Pharmacia Biotech, Piscataway) and revealed by the ninhydrin reaction, followed by absorbance measurements at 570 nm, except for proline, which was detected by measuring the absorbance at 440 nm. Norleucine (0.5 mmol/ml) was added to the samples and used as an internal standard. Total amino acids detected in the hydrolysates of

<table>
<thead>
<tr>
<th>Table 3. Primers Used in This Study.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer</td>
</tr>
<tr>
<td>Δfot1–2-F</td>
</tr>
<tr>
<td>Δfot1–2-R</td>
</tr>
<tr>
<td>ΔC1-F</td>
</tr>
<tr>
<td>ΔC1-R</td>
</tr>
<tr>
<td>ΔC2-F</td>
</tr>
<tr>
<td>ΔC2-R</td>
</tr>
<tr>
<td>ΔC3-F</td>
</tr>
<tr>
<td>ΔC3-R</td>
</tr>
<tr>
<td>Δamn1F</td>
</tr>
<tr>
<td>Δamn1R</td>
</tr>
<tr>
<td>Fluo-F</td>
</tr>
<tr>
<td>Fluo-R</td>
</tr>
</tbody>
</table>

**Note:** Homologous sequence to KanMX4 cassette of pug6 are underlined. Homologous sequence to clonNat cassette of pFA6a-TEF2Pr-GFP-ADH1-NATMX4 or pFA6a-TEF2Pr-t-dimer2-ADH1-NATMX4 are in italic.
the extracts minus the free amino acids detected in the same extract before hydrolysis were classified as bound amino acids—oligopeptides. Glutathione consumption during must fermentation was also analyzed with the Biochrom 30 as described above.

**Genome Sequence and Analysis**

Of the 134 genomes studied, 77 were available in public databases and 57 were newly sequenced in our laboratory (Genowine and Evolya projects).

The 16 (Genowine) and 41 (Evolya) DNA samples were processed to generate libraries of short 400-bp inserts. After passing quality control, the libraries were sequenced using an Illumina HiSeq 2000 platform. Sequencing from both ends generated paired-end reads of 2 × 100 bp, resulting in sequencing depth of 70 to 700× (Genowine project) and 90 to 350× (Evolya project). The entire data sets were deposited in the European Nucleotide Archive (ENA) database under study accession PRJEB6529 (Genowine) and PRJEB7675 (Evolya).

For the PRJEB6529 project, filtered reads were corrected for sequencing errors using Quake v0.3.0. Assembly and scaffolding were carried out using SOAPdenovo v1.05 software, fixing the max_rd_len parameter at 75 and selecting the k-mer lengths resulting in the highest N50 scaffold length. Gaps emerging during the scaffolding process were closed using GapCloser v1.10.

For the PRJEB7675 project, filtering of low-quality regions was performed with trimming if base quality dropped below Q20. Reads shorter than 30 bp were also removed. Assembly and scaffolding were carried out using Velvet v1.2.07 software, selecting the k-mer length resulting in the highest N50 scaffold length. Gaps emerging during the scaffolding process were closed using GapCloser.

Total DNA from the *T. microellipsoides* strain CLIB830T was processed to generate two libraries: 1) A 500-bp insert length paired-end library and 2) a 6-kb insert length mate-paired library. After passing quality controls, the libraries were sequenced using an Illumina HiSeq 2000 platform. Sequencing generated 2 × 100 bp reads per library, resulting in an average sequencing depth of 374×. The entire data set was deposited in the ENA database under study accession PRJEB7632.

Trimomatic v0.32 was used for cleaning and trimming low quality regions and adapters in Illumina data (Bolger et al. 2014). Read sequences were trimmed for the first and last three nucleotides. Filtering of low-quality regions was performed using a sliding window of 20 bp, cutting once the average quality within the window falls below a threshold of 25. Furthermore, reads shorter than 75-bp were further discarded.

In order to remove short inward reads contaminating the mate-pair library, reads were first mapped to the S288C reference genome with BWA v0.6.2 using default parameters (Li and Durbin 2010). Then inwards read pairs were removed using BamTools v2.2.3 “filter command” (Barnett et al. 2011). Picard v1.117 (http://broadinstitute.github.io/picard) “SamToFastq” command was then used to extract resulting reads from the BAM file.

Assembly and scaffolding were carried out using SOAPdenovo2 v2.04, using a kmer value of 61 and the following parameters: paired-end: pair_num_cutoff = 5, map_len = 32; mate-pair: pair_num_cutoff = 10, map_len = 50 (Luo et al. 2012). Closing gap was performed using GapCloser v1.12 using default parameters (Luo et al. 2012).

Reconstruction of the complete nucleotide sequence of the HGT region C of *T. microellipsoides* presents on scaffold2 was performed using targeted PCR and capillary sequencing of the resulting amplicons. This scaffold has been submitted to the ENA (accession number: LN811465).

Similarity search among the yeast genome sequences was performed using BLAST (Altschul et al. 1990).

The nucleotide diversity of region C in EC1118 and *T. microellipsoides* has been calculated from the alignment of EC1118 and *T. microellipsoides* regions C with variscan software for 1-kb sliding windows separated by 200-bp steps.

**Molecular Phylogenetic Analysis Using the Maximum-Likelihood Method**

Protein sequences homologous to the each ORF of the region C protein were identified by a BLASTP search of the National Center for Biotechnology Information and GRYC (Genome Resources for Yeast Chromosomes http://gryc.inra.fr/index.php?page=home, last accessed March 16, 2015) databases. Amino acid sequences were aligned using MUSCLE. The evolutionary history was inferred by using the maximum-likelihood method based on the JTT (Jones, Taylor, and Thornton) matrix-based model (Jones et al. 1992) as proposed by MEGA5.2.2. software (Tamura et al. 2011). For each phylogeny, the tree with the highest log likelihood is shown. For each pairwise comparison, all sites covered are taken into account. The bootstrap value of 1,000 replicates is shown next to each node. Initial trees for the heuristic search were obtained by applying the neighbor-joining method to a matrix of pairwise distances estimated using a JTT model. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site.

**Accession Codes**

Genome sequence data for *Saccharomyces cerevisiae* strains have been deposited in the European Nucleotide Archive (ENA) database under the accession codes PRJEB6529 (Genowine project) and PRJEB7675 (Evolya project). Genome sequence data for *T. microellipsoides* have been deposited in the European Nucleotide Archive (ENA) database under study accession PRJEB7632.

**Supplementary Material**

Supplementary figures S1–S4 and table S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).
Acknowledgments
This work was supported by grants GIS IBiSA-AO 2010-2011, France to S.D., V.G., J.L.L., and F.B.; AIP Bioresources from the National Institute for Agricultural Research, France to S.D., V.G., J.L.L., F.B.; and AGL2012-39937-C02-02 from the Spanish government to E.B. S.M. and A.M. acknowledge support by a doctoral grant and an FPI fellowship from the French and Spanish governments, respectively. They acknowledge Therese Marlin and Pierre Delobel for their help with the glutathione and flux cytometer analyses.

References


