Horizontal acquisition of bacterial genes is presently recognized as an important contribution to the adaptation and evolution of eukaryotic genomes. However, the mechanisms underlying expression and consequent selection and fixation of the prokaryotic genes in the new eukaryotic setting are largely unknown. Here we show that genes composing the pathway for the synthesis of the essential vitamin B1 (thiamine) were lost in an ancestor of a yeast lineage, the Wickerhamiella/Starmerella (W/S) clade, known to harbor an unusually large number of genes of alien origin. The thiamine pathway was subsequently reassembled, at least twice, by multiple HGT events from different bacterial donors involving both single genes and entire operons. In the W/S-clade species Starmerella bombicola we obtained direct genetic evidence that all bacterial genes of the thiamine pathway are functional. The reconstructed pathway is composed by yeast and bacterial genes operating coordinately to scavenge thiamine derivatives from the environment. The adaptation of the newly acquired operons to the eukaryotic setting involved a repertoire of mechanisms until now only sparsely documented, namely longer intergenic regions, post-horizontal gene transfer (HGT) gene fusions fostering coordinated expression, gene relocation, and possibly recombination generating mosaic genes. The results provide additional evidence that HGT occurred recurrently in this yeast lineage and was crucial for the reestablishment of lost functions and that similar mechanisms presumably have facilitated a transition from bacterial operon transcription to eukaryotic-style gene expression were proposed, such as gene fusion giving rise to multifunctional proteins (6, 23, 24), increase in intergenic distances between genes to generate room for eukaryotic promoters, and independent transcription producing mRNAs with poly(A) tails have been demonstrated (22). In the best documented study, which concerns a bacterial siderophore biosynthesis operon acquired by yeasts belonging to the Wickerhamiella/Starmerella (W/S) clade, the bacterial genes acquired as an operon were shown to be functional (22).

Thiamine, commonly known as vitamin B1, is essential for all living organisms because its active form, thiamine pyrophosphate (TPP), is an indispensable cofactor of enzymes participating in amino acid and carbohydrate metabolism (26–30). Some organisms that are unable to synthesize thiamine de novo are nevertheless capable of using a salvage pathway to rescue the pyrimidine (hydroxymethylpyrimidine or HMP) and thiazole (hydroxyethylthiazole or HET) precursors and similar compounds that result from natural thiamine degradation in the surrounding environment (31, 32).

In the present study we describe a composite thiamine salvage pathway made up of yeast and bacterial genes found in several species of the yeast W/S clade. Our recent work (5) revealed that the W/S-yeast clade harbors an unusual large number of HGT events in yeasts, mostly as single genes, a finding independently confirmed in a large study in which ~300 yeast genomes were examined (9). Here we show that in W/S-clade species most of

### Significance

**Food is the only source of the essential vitamin B1 for humans, but many microorganisms such as yeast and bacteria can synthesize it themselves. Here we report on a group of yeasts that have lost part of the vitamin B1 biosynthetic pathway in the past but have managed to rebuild it by capturing multiple genes from bacteria through horizontal gene transfer (HGT). We show a mosaic pathway composed of yeast and bacterial genes working coordinately to accomplish the synthesis of an essential nutrient. This involved adaptation of the bacterial genes to the very different expression rules in their new environment using several different mechanisms. Our results endorse HGT as an important mechanism for evolutionary adaption in eukaryotes.**

Author contributions: C.G. and P.G. designed research; C.G. and P.G. performed research; P.G. contributed new reagents/analytic tools; C.G. analyzed data; and C.G. and P.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. Published under the PNAS license.

Data deposition: All the alignment files and complete phylogenies and can be accessed in figshare (DOI: 10.6084/m9.figshare.9800636).

To whom correspondence may be addressed. Email: pmz@fct.unl.pt.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1909844116/-/DCSupplemental.

First published October 14, 2019.
the genes in the thiamine salvage pathway were originally acquired from bacteria as part of an operon and we use genetic dissection to link each of the transferred genes with an observable phenotype in the yeast setting. Moreover, we present evidence for the occurrence of 2 independent horizontal operon transfer (HOT) events in distinct subclades within the W/S lineage, that is, the independent horizontal acquisition of single functionally related genes, endow the yeast host with the ability to salvage a wide range of thiamine precursors.

Results
Thiamine Genes in the W/S-Yeast Clade. Many organisms lack the genes encoding enzymes of the 2 upper branches of the pathway used to generate the pyrimidine and thiazole precursors for de novo biosynthesis of thiamine, while retaining only the genes composing the salvage pathway for thiamine or its precursors (31–33). Thi5 and Thi4 are the most important enzymes in each of the 2 pathway branches required for de novo thiamine synthesis (34) (Fig. 1A). THI5 is absent in all W/S clade species and THI4 is missing in about half of the species in the clade. They maintain however the indispensable genes encoding a thiamine transporter, Dur31 (35), and the kinase Thi80, responsible for TPP synthesis (Fig. 1A and B). Species that lack THI5, or both THI5 and THI4, often maintain the salvage pathway genes THI6 and THI20 that enable them to scavenge thiamine derivatives from the environment (36, 37). In W/S-clade species, significant tBLASTx hits (e-value < e−10) were retrieved from available genomes for the THI6 and THI20 genes (Fig. 1A and B). These 2 genes are contiguous in the genome of all W/S-clade species, but not in the genomes of most Saccharomyces yeasts (SI Appendix, Fig. S1). THI20 seems to be encoded by 2 separate genes, while for THI6 a single gene was predicted in most species. A subsequent BLASTp search using the 3 predicted proteins from St. bombicola (2 corresponding to different domains of THI20 and 1 corresponding to THI6) retrieved bacterial proteins as top hits (Fig. 1B and Dataset S1) and no yeast proteins among the first 1,000 hits. The same result was obtained for all W/S-clade THI6 or THI20 homologs, suggesting that these genes, when present, always have bacterial origin in the W/S clade. The proteins exhibiting the Thi6 and Thi20 enzymatic activities are both encoded by 2 genes each in most bacteria (thiE and thiM, and tenA and thiD, respectively), as depicted in Fig. 1A. It was not possible to ascertain the origin of THI6 and THI20 in Wickerhamiella hasegawae, because preliminary BLASTx searches using the C-terminal and N-terminal domains of the Thi6 and Thi20 from this species showed a mosaic-like pattern for both genes, where part of the protein exhibited high sequence similarity with bacterial proteins while the other domain presented homology with fungal proteins (Fig. 1B and Dataset S2).

Independent Acquisitions of Bacterial Thiamine Operons by W/S-Clade Species. The fact that THI genes are found adjacent to each other in the genomes of W/S-clade yeasts led us to examine the possibility that they were acquired in a single event as an operon. The thiamine operon is organized differently in distinct bacteria (38), which is thought to result from reshuffling during evolution and is observed in many other instances (39–41). Hence, gene order and content, as phylogenetic analysis, can be indicative of the plausibility of HOT occurrences as opposed to

![Fig. 1. (A) Thiamine biosynthetic pathway in bacteria and yeast. Genes involved in thiamine biosynthesis in B. subtilis are represented in blue while their counterparts in S. cerevisiae are represented in red. Yellow arrows represent the salvage pathways that can be present in both yeast and bacteria. The red stars represent thiamine degradation. (B) Presence and absence of the main genes involved in the de novo and salvage thiamine biosynthesis in the W/S clade. Gray squares denote missing genes; blue squares, genes of bacterial origin; and red squares, genes of fungal origin. The origin of the W. hasegawae THI6 and THI20 genes was unclear. Phylogenetic relationships between species are depicted based on a ML phylogeny constructed as described in Materials and Methods. Branches with 100% support are indicated by black dots while bootstrap values >75% are shown next to the respective branches.]
The exceptions were vertical bars represent ends of scaffolds. (Dataset S1), gene content and order of the thiamine clusters identified in the W/S clade were compared to operons found in extant representatives of these 3 bacterial lineages. For most W/S-clade species, the organization of the thiamine cluster resembled that of the thiamine operon in the Bacteroidetes (Fig. 2A and B). Hence, cluster organization in these 2 species resembled instead the Burkholderiales (Betaproteobacteria; Fig. 2A). In independent phylogenies the TenA and ThiD moieties cluster with the cognate proteins in other species of the W/S clade of 2 independent HOT events, as well as the independent origin of the bacterial operons found in the St. bombicola and W. galacta subclades and of tenA in W. galacta. Hence, taken together, our data support the identification in the W/S clade of 2 independent HOT events, as well as the independent acquisition of tenA by W. galacta.

Horizontally Acquired THI Genes Participate in a Thiamine Salvage Pathway. All W/S-clade species examined are predicted to be auxotrophic for thiamine due to the absence of THI genes (Fig. 1B). However, they may be able to rescue 4-amino-2-methyl-5-(phosphohexamethylene)pyrimidine (HMP) from the environment if their salvage pathway of bacterial origin composed of thiEM and thiD (Fig. 1A) is functional. In accordance with this, wild-type

Fig. 2. Thiamine operon organization. (A) Organization of THI genes in the genomes of representative W/S-clade species. Thiamine metabolism-related genes are represented by different colors. Syntenic genes between species are represented in black while nonsyntenic genes are represented in white. Black vertical bars represent ends of scaffolds. (B) Thiamine operon organization in putative donor lineages. A representative species belonging to each order/subclade (Fig. 3 and SI Appendix, Fig. S2), the distinct genomic location of the tenA gene is likely the result of a postacquisition rearrangement. In Wickerhamiella domercqiae, a fusion between the tenA and thiD genes was observed (Fig. 2A). In independent phylogenies the TenA and ThiD moieties cluster with the cognate proteins in other species of the W/S clade of 2 independent HOT events, as well as the independent origin of the bacterial operons found in the St. bombicola and W. galacta subclades and of tenA in W. galacta. Hence, taken together, our data support the identification in the W/S clade of 2 independent HOT events, as well as the independent acquisition of tenA by W. galacta.
(WT) St. bombicola was shown to grow on medium lacking thiamine but supplemented with HMP (Fig. 5A). Next, St. bombicola deletion mutants (tenAΔ, thiDΔ, thiEMΔ, and tenAΔthiDΔthiEMΔ) were constructed and their ability to grow was assessed in medium supplemented with different thiamine precursors.

As expected from their predicted role in the putative salvage pathway, deletion of thiD and/or thiEM rendered the strains unable to use HMP (Fig. 5A). Deletion of tenA did not have an effect on growth on HMP-supplemented medium, in line with this protein having an amino-hydrolase activity capable of producing HMP from 4-amino-5-aminomethyl-2-methylpyrimidine (aminoHMP) (43, 44), and therefore acting upstream of HMP synthesis (Fig. 1A). To assess this possibility, growth in the presence of aminoHMP was also evaluated. As shown in Fig. 5A, the tenAΔ mutant grows notably worse than the WT strain in medium containing aminoHMP as a source of thiamine (Fig. 5A and B), although some residual growth of the mutant was observed (Fig. 5A). Growth on aminoHMP as sole source of thiamine was also evaluated in liquid media (SI Appendix, Fig. S4), where the WT attained much higher cell densities than the tenAΔ mutant. As expected, no differences in growth were observed in thiamine- or HMP-supplemented media in this mutant (SI Appendix, Fig. S4). Together, these results show that aminoHMP is a substrate for TenA, but that an additional source of amino-hydrolase activity, possibly unspecific, seems to be present, supporting residual growth of the tenAΔ mutant in aminoHMP particularly in solid medium.

Assimilation tests were also used to evaluate the functionality of the thiamine salvage pathway in 2 additional W/S-clade species, W. domercqiae and W. galacta. Both species lack THI4, so that to test assimilation of HMP or aminoHMP, it was also necessary to supplement the medium with 5-(2-hydroxyethyl)-4-methylthiazole (HET) (Figs 1A and 5B, Left). As shown in Fig. 5B, Right, when either HET, HMP, or aminoHMP were provided separately, no growth was observed for these species, as expected in the absence of THI4. Also as predicted, a combination of HMP and HET restores growth in both species. Notably, W. galacta was also able to grow when aminoHMP and HET were added to the growth medium showing that although tenA is not part of the thiamine cluster, and was likely acquired from a different bacterial lineage, it is operating as part of the salvage pathway by supplying the pyrimidine moiety of thiamine. Similar results were obtained for W. domercqiae, which could also grow when HET and aminoHMP were simultaneously supplied (Fig. 5B, Right).

In Bacillus subtilis, in addition to TenA that is responsible for the conversion of aminoHMP to HMP, the thiamine salvage pathway involves another protein, YlmB, which is essential for the deformat-lation of N-formyl-4-amino-5-aminomethyl-2-methylpyrimidine (N-formyl-aminoHMP), a compound commonly resulting from thiamine degradation (43, 44). YlmB converts this compound in aminoHMP, thus acting upstream of TenA (Fig. 1A) (43). Interestingly, we found evidence for the presence of a putative YlmB (annotated as an acetylornithine deacetylase)-encoding gene in most W/S-clade species (SI Appendix, Fig. S5). This gene was also horizontally acquired from bacteria, but from a donor belonging to the Acetobacteraceae family (Alphaproteobacteria; SI Appendix, Fig. S5).

Expression of Thiamine Cluster Genes. If the salvage pathway operating in extant W/S-clade species is derived from a bacterial
operon, its expression can be presumed to be currently adapted to eukaryotic canonical transcription. Polycistronic mRNAs (and operons) are rare in eukaryotes (43–47), the few known cases involving processed eukaryotic pre-mRNAs into monocistronic mRNAs by 3′ end formation and polyadenylation and subsequent splicing by a small nuclear ribonucleoprotein (48–50). The siderophore biosynthesis gene cluster described also in W/S-clade yeasts is the best studied case of adaptation of a bacterial operon to the eukaryotic setting (22). In this case, intergenic regions were shown to be lengthier than observed in the putative bacterial donors, which was also observed for THI genes in W/S-clade species (Fig. 24). Expression of the THI genes

**Fig. 4.** Maximum likelihood phylogenies of Thi proteins from W. galacta. Pruned phylogenies showing the closest relatives to W. galacta Thi proteins encoded in the operon are shown. Branches with support higher than 95% (ultrafast bootstrap) are indicated by black dots. The phylogenetic tree for TenA is shown in SI Appendix, Fig. S3.
Next, we examined poly(A)-tailed transcripts along the fused spanning only the within gene amplicons, implying that some of the transcripts encompassing the tails (Fig. 6 which will reveal only mRNAs possessing poly(A) tails. Results the thiEM order of magnitude lower (Fig. 6 A). Assimilation of fragments spanning any 2 independent genes was several expression were observed within gene amplicons while expression qPCR and gene-specific primers (Fig. 6). Nevertheless, the fragment spanning the fusion point between the thiM and thiE moieties in the thiEM gene was slightly less expressed (~4 to 5-fold), which could result from the coexistence of a mRNA spanning the 2 fused genes with shorter mRNAs encompassing only 1 of the genes. To find out to which extent these transcripts conformed with a canonical eukaryotic structure, we repeated the experiment for St. bombicola but using an oligo(dT)20 primer to synthesize the first cDNA strand, followed by qPCR with gene-specific primers, which will reveal only mRNAs possessing poly(A) tails. Results were very similar, emphasizing that most transcripts have poly(A) tails (Fig. 6). However, this time expression of the fragment encompassing the thiM/thiE fusion was similar to the expression within gene amplons, implying that some of the transcripts spanning only the thiM or the thiE moieties may lack poly(A) tails. Next, we examined poly(A)-tailed transcripts along the fused THI genes in 2 additional species, W. domercqiae and W. galacta (Fig. 6). Unlike the observations for St. bombicola, the results indicate that in some cases, in addition to transcripts potentially encompassing the complete fused genes encoding multidomain proteins, shorter poly(A)-tailed mRNA molecules are also produced. For example, in W. galacta, transcripts containing the fusion site between the thiM and thiE moieties are evidently less abundant than transcripts containing the other regions probed within the triple fusion between the thiM, thiE, and thiD genes, suggesting that poly(A) transcripts comprising only the thiM gene on the one hand and the thiE/thiD fusion gene on the other hand, are also generated in significant amounts.

Discussion
De novo synthesis of thiamine is impaired in W/S-clade species because they lack the THI5 and in some cases also the THI4 genes, responsible respectively for synthesis of the pyrimidine (HMP) and thiazole (HET) precursors of thiamine. In addition, yeast orthologs of downstream components of the pathway, THI6 and THI20, are also missing in this lineage. However, we showed there unequivocally for one W/S-clade species using genetic evidence and for other species by examining their ability to salvage various thiamine precursors, that this does not render these organisms dependent solely on import of thiamine from the outside. Instead, clusters of THI genes originating from bacterial operons, complemented by other individual genes, form functional salvage pathways capable of rescuing thiamine degradation products from the environment thereby providing cells with sufficient TPP synthesis capacity to support growth.

Our findings suggest an evolutionary model (Fig. 7) in which the THI6 and THI20 genes, that are rarely found clustered in yeast genomes, were lost in an ancestor of the W/S clade, in addition to THI5, resulting initially in obligate reliance on external sources of thiamine. THI4 was also lost in several species, which could be related to its mode of action as a suicide enzyme (51). We found within the W/S clade several different subsequent outcomes of the loss of thiamine prototrophy. Three of the species examined belonging to 1 of the 2 sister lineages within the W/S clade represented in this study (Wickerhamiella cacticola, Wickerhamiella parangosa, and Wickerhamiella occidentalis) still lack any gene resembling THI6 and THI20, while the 2 remaining species (W. hasegawa and W. galacta) examined in this subclade present evidence of horizontal acquisition of xenologs, in the case of W. galacta from the Burkholderiales (Betaproteobacteria). In 9 of the 10 species forming the second subclade, THI6 and THI20 homologs also seem to be of bacterial origin, but from a different phylum, the Bacteroidetes. According to our evidence, this transfer event took place in the common ancestor of the 9 species, which is consubstantiated by the fact that the phylogenies of THI genes recapitulate the species phylogeny.

The activities of the THI20 and THI6 gene products are encoded by 4 genes of which at least 3 are organized in an operon.
in both putative donor lineages, the Bacteroidetes and the Burkholderiales (Betaproteobacteria). The identity of the donor lineages suggested by the phylogeny is further supported by the fact that gene content and gene order in W/S-clade THI clusters is identical to the operons found in extant representatives of the 2 donor lineages. This in turn strongly suggests that an entire operon was transferred in a single event in both cases. However, inference of the donor lineage originating the THI cluster in St. bombicola and neighboring species was based on phylogenies obtained for the TenA and ThiD proteins only, because it was not possible to obtain a reliable phylogenetic signal for either the ThiE/M fused proteins or the separate ThiE and ThiM moieties, leaving the possibility open that these genes were transferred in an independent event from the tenA and thiD genes and possibly from a different donor. It seems more likely, however, that such a putative second event did not occur entirely independently of the first since sequence similarity may have promoted recombination involving the thiE/thiM genes acquired initially from the Bacteroidetes as part of the operon transfer and homologous genes from a distinct unidentified bacterial lineage, as observed before (52–55). Alternatively, recombination events in the bacterial donor lineage prior to the acquisition by yeasts might also have resulted in the conflicting phylogenetic signal observed for ThiEM. Even more elusive is the origin of the THI6 and THI20 homologs from W. hasegawae, which seem to be mosaics of uncertain origin exhibiting regions of homology with both bacterial and fungal proteins (Dataset S2). Interestingly, no evidence for mosaicism was reported for the siderophore biosynthesis gene cluster in W/S-clade yeasts (22).

In addition to HOTs related to thiamine metabolism we also detected complementary single acquisitions of functionally related genes from bacteria that further extended the range of thiamine precursors that could be salvaged by W/S-clade species (Fig. 7). The best example of this in our data is W. galacta that acquired a tenA-lacking operon from the Burkholderiales (Betaproteobacteria), but nevertheless seems to have been able to acquire a tenA gene from the Actinobacteria. Other W/S-clade species also seem to have acquired a putative ylmB gene from the Acetobacteraceae (Alphaproteobacteria) in addition to a tenA-containing operon from the Bacteroidetes. These 2 genes encode enzymes that mediate the utilization of aminoHMP (tenA) and N-formyl-aminoHMP (ylmB) as sources of HMP. These compounds are originated by thiamine degradation which can occur naturally (43) or by the action of microorganisms (56). The formation of HMP as a degradation product of thiamine has been reported to occur also under different pH and temperature conditions in laboratory experiments (57, 58). Consequently, the products of thiamine degradation (such as HMP, HET,
Fig. 7. Main events in the evolution of THI genes in the W/S clade. Schematic phylogenetic relationships between W/S species are depicted based on the ML phylogeny presented in Fig. 18. Loss of native THI6 and THI20 genes in the most recent common ancestor (MRCA) of the W/S clade is indicated by a red cross. Putative horizontal gene transfer events are represented by boxes above/below which the putative donor lineage is indicated. The number of putative HGT events and fusion events are indicated in yellow and blue circles, respectively; numbers in gray represent cases where the number of HGT events could not be asserted with certainty. Operons are represented by a straight black line linking the genes, while the uncertainty surrounding the origin of the thiE and thiM in the St. bombicola subclade is denoted by question marks. Horizontal acquisition of the putative N-formyl-4-amino-5-aminomethyl-2-methylpyrimidine deformylase (YlmB) is also shown. The origin of the W. hasegawae genes is elusive.

<table>
<thead>
<tr>
<th>Key:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Events</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>\</td>
</tr>
<tr>
<td>//</td>
</tr>
<tr>
<td>Gene relocation</td>
</tr>
<tr>
<td>○</td>
</tr>
<tr>
<td>⬝</td>
</tr>
</tbody>
</table>

Main events in the evolution of THI genes in the W/S clade. Schematic phylogenetic relationships between W/S species are depicted based on the ML phylogeny presented in Fig. 18. Loss of native THI6 and THI20 genes in the most recent common ancestor (MRCA) of the W/S clade is indicated by a red cross. Putative horizontal gene transfer events are represented by boxes above/below which the putative donor lineage is indicated. The number of putative HGT events and fusion events are indicated in yellow and blue circles, respectively; numbers in gray represent cases where the number of HGT events could not be asserted with certainty. Operons are represented by a straight black line linking the genes, while the uncertainty surrounding the origin of the thiE and thiM in the St. bombicola subclade is denoted by question marks. Horizontal acquisition of the putative N-formyl-4-amino-5-aminomethyl-2-methylpyrimidine deformylase (YlmB) is also shown. The origin of the W. hasegawae genes is elusive.

AminoHMP, etc.) may be more abundant than thiamine itself in the environment (56). Moreover, some thiamine breakdown products, such as N-formyl-aminoHMP or aminoHMP, can be toxic (59, 60) so that TenA and YlmB may also have a detoxification role.

Horizontally transferred genes are probably quickly lost if they are not selected for, and for protein coding genes this means that they have to be expressed. Eukaryotic canonical gene expression imposes very different prerequisites from those found in bacterial operons, namely monocystronic mRNAs with appropriate 3′ and 5′ modifications and markedly different promoters. These adaptations were recently studied for the first time for the siderophore biosynthesis gene cluster described in W/S-clade yeasts (22). The occurrence of multiple HOT events in this clade probably reflects the previously reported large number of HGT events in general (5, 9, 22, 61). Common to both events in the W/S clade is the observed increase in intergenic spacing possibly to accommodate de novo evolved promoters. However, adaptation of the W/S-clade siderophore biosynthesis operon did not involve postacquisition gene fusions, of which at least 3 independent examples were observed in the W/S-siderophore cluster (Fig. 7). We interpreted this as an effective means to achieve coordinated expression of the bacterial genes. Gene fusions were also observed in HOT events in protists (6, 23, 24) implying that this may be a general and frequent mechanism employed to transpose coordinated prokaryotic expression to the eukaryotic context. The absence of gene fusions in the long siderophore gene cluster may be due to restrictions imposed by proper functioning of the enzymes which may be incompatible with fusion.

Our data may suggest that individual expression of the genes, possibly driven by spurious promoters, preceded the fusions, because in the pool of poly(A)-tailed mRNAs, shorter transcripts seem to coexist with the full-length mRNAs spanning the complete fusion genes. Overlapping mRNAs were also detected for the siderophore biosynthesis gene cluster from W. versatilis (22).

Taken together, our observations revealed examples of mechanisms facilitating expression of xenologous genes observed so far sporadically in HOT occurrences in various eukaryotic microbes, namely multiple instances of post-HGT gene fusions, increased spacing between genes, dispersion of genes to different genomic locations, and likely instances of recombination leading to the formation of mosaic genes. These observations support the idea that the repertoire of mechanisms for adaptation of gene expression are consistently and frequently used across the eukaryotic domain to functionalize genes horizontally acquired from bacteria. Moreover, they show that successive layers of horizontal transfer events were involved in fine tuning a previously acquired metabolic pathway.

Materials and Methods

Strains. Yeast strains were obtained from the Portuguese Yeast Culture Collection, Caparica, Portugal (PYCC) except for W. versatilis ICM 5958, which was kindly provided by the Japan Collection of Microorganisms, Tsukuba, Japan (ICM), and W. galacta NRRL Y-17645, which was obtained from Agricultural Research Service Culture Collection, Peoria, IL (ARS-NRRL). All strains...
were maintained in YMA medium [1% (wt/vol) glucose, 0.3% (wt/vol) malt extract (wt/vol), 0.3% yeast extract, 0.5% (wt/vol) peptone and 2% (wt/vol) agar].

Identification of Genes Involved in TPP Biosynthesis. Genes related to TPP de novo biosynthesis (TH4, TH5, TH6, TH7, TH20, and TH30) were searched in the genomes of the W/S clade using Saccharomyces cerevisiae S288C homologs as queries (NC_001139.9, NC_001138.5, NC_001148.4, NC_001144.5, NC_001147.6, and NC_001147.6, respectively). The respective best hits were subsequently blasted against the National Center for Biotechnology Information (NCBI) nonredundant (nr) database. Whenever the best hit in NCBI corresponded to the identity of the query gene, it was assumed that the gene was present.

Phylogenetic Analyses. The same dataset represented in Fig. 18 was constructed using the same dataset as in Gonçalves et al. (5) (see SI Appendix, Table S1 for the complete list of taxa used in the phylogeny) based on a previously described methodology (62) with the addition of other W/S-clade species whose genomes were recently published in the context of the Y1000+ project. Briefly, RPA1, RPA2, RPB1, RPB2, RCP1, and RCP2 protein sequences for each species were used to construct the maximum likelihood (ML) tree using RAxML (63) v7.2.8 using the PROT GAMMA I LG model of amino acid substitution and 1,000 rapid bootstraps. Branch support values (>75%) are displayed. This tree is in agreement with the recently published phylogenies from Kominik et al. (22).

Independent phylogenies were constructed for the N-terminal (ThiM) and C-terminal (ThiE) domains of the ThiEM protein. For the putative proteins from bacterial origin TenA, ThiD, ThiE, and ThiM a preliminary BLASTP search against the nr NCBI database was performed and it was confirmed that the top 1,000 hits only included bacterial proteins except for the ThiE portion for which some nonbacterial (fungi, plants) sequences were also recovered (Dataset S1). Preliminary phylogenies using the top 5,000 hits were constructed with their closest relatives. For that, the proteins from (-bb 1,000) (68) for branch support determination.

The protein sequences for other W/S-clade species with clade species whose genomes were recently published in the context of the Y1000+ project were used to construct the maximum likelihood (ML) tree with RAxML (63) v7.2.8 using the PROTGAMMA I LG model of substitution and 1,000 rapid bootstraps. Branch support values (>75%) are displayed. This tree is in agreement with the recently published phylogenies from Kominik et al. (22).

Construction of St. bombicola Deletion Mutants. Standard molecular biology techniques were performed essentially as described in ref. 70 using Escherichia coli DH5α as host. St. bombicola PYCC 5882 was used in all procedures involving this species. Knockout cassettes (SI Appendix, Table S4) were constructed as described in Gonçalves et al. (5) using hygromycin as the selective marker. Transformation of St. bombicola was performed as described in ref. 7. After transformation, mutants were selected on YPD plates containing 650 μg/ml of hygromycin (InvivoGen). Correct integration of the disruption cassettes was verified by appropriate PCR reactions and by sequencing.

Different transformants from 2 independent gene disruption transformations were used for the phenotypic assays.

Construction of St. bombicola WT and mutants and other W/S-clade species were tested for growth in the presence of HMP, aminoHMP, and HET. Strains were first grown for 24 h on YMA [1% (wt/vol) glucose, 0.3% (wt/vol) malt extract (wt/vol), 0.3% yeast extract, 0.5% (wt/vol) peptone, and 2% (wt/vol) agar] medium and transferred to YNB without amino acids and without thiamine (Formedium, Norfolk, UK) plates to exhaust intracellular thiamine pools. Cells were washed with sterile water and resuspended in water to a final OD600 of 0.5. Cell suspensions were subsequently serially diluted 10-fold, spotted onto YNB plates without amino acids and without thiamine (Formedium), and supplemented with CSM (complete supplement mixture, MP Biomedicals), 0.2 μM of thiamine, 0.02 μM of HMP, 0.02 μM of aminoHMP, and 0.2 μM of HET combined with 0.2 μM of HMP or with 0.02 μM of aminoHMP (for strains that lack both THI4 and THI5). Plates were incubated at 25 °C for 5 d.

ACKNOWLEDGMENTS. We thank PYCC for providing the strains used in this work; undergraduate students João Sousa, Ana Sousa, and Andreia Barro for performing some of the preliminary growth assays; and members of the YeastGenomicsLab, in particular Patrícia Brito, for fruitful discussions. We would also like to acknowledge the Y1000+ Project which was the source of genome data used in this work concerning 8 W/S-clade species. This work was supported by the UCIBIO-Unidade de Ciências Biomoleculares Aplicadas, which was financed by national funds from Fundação para a Ciência e Tecnologia, Ministério da Ciência, Tecnologia e Ensino Superior (FCT/MCTES; https://www.fct.pt) UID/Multi/04378/2019 and grants Fruity Yeast–LISBOA-01-0145-FEDER-029529/PTDC/BCIA-MIC/29529/2017 (to P.G.) and SFHR/BID/89489/2012 (to C.G.), both from FCT/MCTES.