PNAS
www.pnas.org

Supplementary Information for

Multi-layered horizontal operon transfers from bacteria reconstruct a thiamine salvage pathway in yeasts

Carla Gonçalves and Paula Gonçalves*
*corresponding author
Email: pmz@fct.unl.pt

This PDF file includes:

Supplementary text: Methods for the construction of preliminary phylogenies
Figures S1 to S6
Tables S1 to S4
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3
Phylogenetic analyses related files can be accessed in Figshare (DOI: 10.6084/m9.figshare.9800636)

Methods for the construction of preliminary phylogenies presented in Figure S2 and S5.

For the preliminary phylogenies presented in Figure S2, the top 5.000 BLASTp hits on NCBI (Refseq database) were selected using *St. bombicola* proteins as queries and an *e-value* cut off of $1e^{-10}$. *tenA*, *thiD*, *thiE* and *thiM* (recovered as *thiEM* because these genes are fused in all W/S-clade species) sequences from other W/S-clade species were retrieved by tBLASTx in local genome databases and the proteins were subsequently predicted in *AUGUSTUS* (web interface) using *Saccharomyces cerevisiae* as reference. Sequences with more than 95% similarity were removed with *CD-HIT* v4.6.7 (1) and the remaining sequences were aligned with *MAFFT* v7.222 (2) using an iterative refinement method (L-INS-i). For TenA, ThiD, ThiE and ThiM, poorly aligned sequences were removed with *trimAl* v1.2 (3) using its "gappyout" option. For the alignment comprising the fused ThiEM protein which encodes the ThiE and ThiM domains, columns with gaps in more than 10% of the sequences were removed (-gt 0.9). If this strategy removed more than 40% of the columns in the original alignment, at least 60% were conserved by adding the necessary number of columns in decreasing order of scores (-cons 60). In this way it was ensured that both ThiE and ThiM portions were conserved in the alignment. Phylogenies were constructed in *IQ-TREE* v1.6.10 (4) with 1.000 ultrafast bootstrap (5) replicates and with selection of the best fitting model. In all cases, W/S-clade sequences grouped with bacteria in the preliminary phylogenies and *W. galacta* groups with different bacteria than other W/S-clade species, suggesting an independent origin.

For the preliminary phylogeny of YlmB presented in Figure S5A, the top 5.000 BLASTp hits on NCBI (Refseq database) were selected using *St. bombicola* putative YlmB as query and an e-value cutoff of $1e^{-10}$. Sequences with more than 95% similarity were removed with *CD-HIT* v4.6.7 (1) and the remaining sequences were aligned with *MAFFT* v7.222 (2) using an iterative refinement method (L-INS-i). For the final phylogeny presented in Figure S5B, only the top 750 hits in UniprotKB (reference_proteomes) were selected with the addition of the closest related hits in *Bacillus subtilis* which include the characterized YlmB protein (6). Sequences with more than 98% similarity were removed with *CD-HIT* v4.6.7 (1) and the remaining sequences were aligned with *MAFFT* v7.222 (2) using an iterative refinement method (L-INS-i). Phylogenies were constructed in *IQ-TREE* v1.6.10 (4) with 1.000 ultrafast bootstrap (5) replicates and with the selection of best fitting model.

Supplementary figures

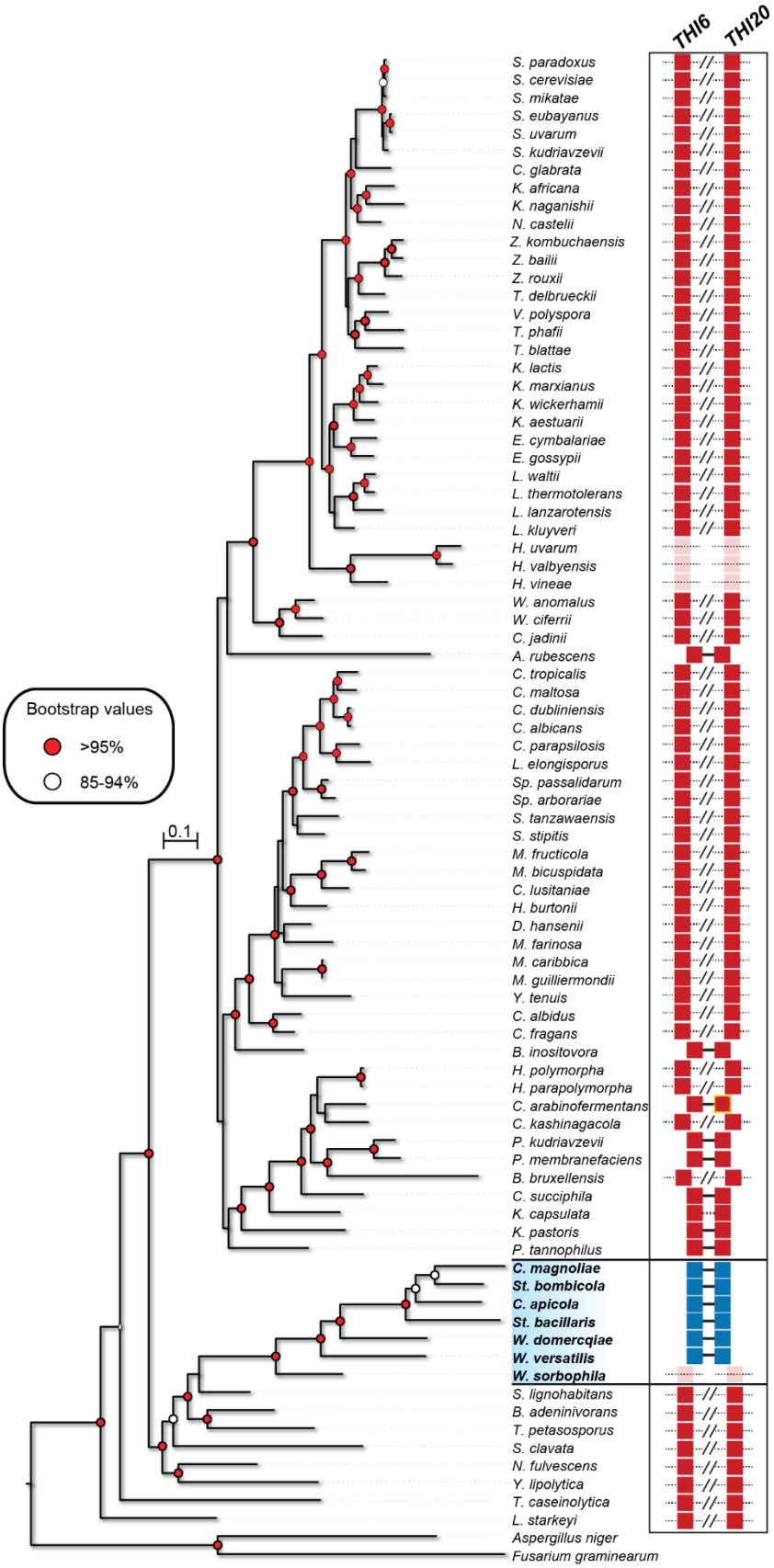

Fig. S1. Distribution of *THI6* and *THI20* homologues in the Saccharomycotina. The maximum likelihood (ML) tree is the one depicted in Gonçalves C. et al., 2018 (7). A tBLASTx in local genome databases (see Table S4, which includes the respective accession numbers) was performed (*E*-value < e^{-3}). A BLASTp in

NCBI was performed with the top hits and whenever the top BLASTp hits corresponded to the identity of the query gene, the gene was considered as being present. Physical linkage was asserted by inspecting if the top hits for *THI6* and *THI20* were contiguously located. The red and blue squares represent homologues of fungal origin (top BLASTp hits corresponded to fungal proteins) and bacterial origin (top BLASTp hits corresponded to bacterial proteins), respectively. Faded squares denote absent genes (e-value > e^{-3}). Physical linkage between *THI6* and *THI20* is represented by a black line linking the two squares while "//" means that the genes are distantly located in the respective genome. For *Candida arabinofermentans* several paralogues of *THI20* were detected and one of the paralogues is in cluster with *THI6* while the others are located in other genomic regions. In *Kuraishia capsulata THI6* and *THI20* are separated by three genes.

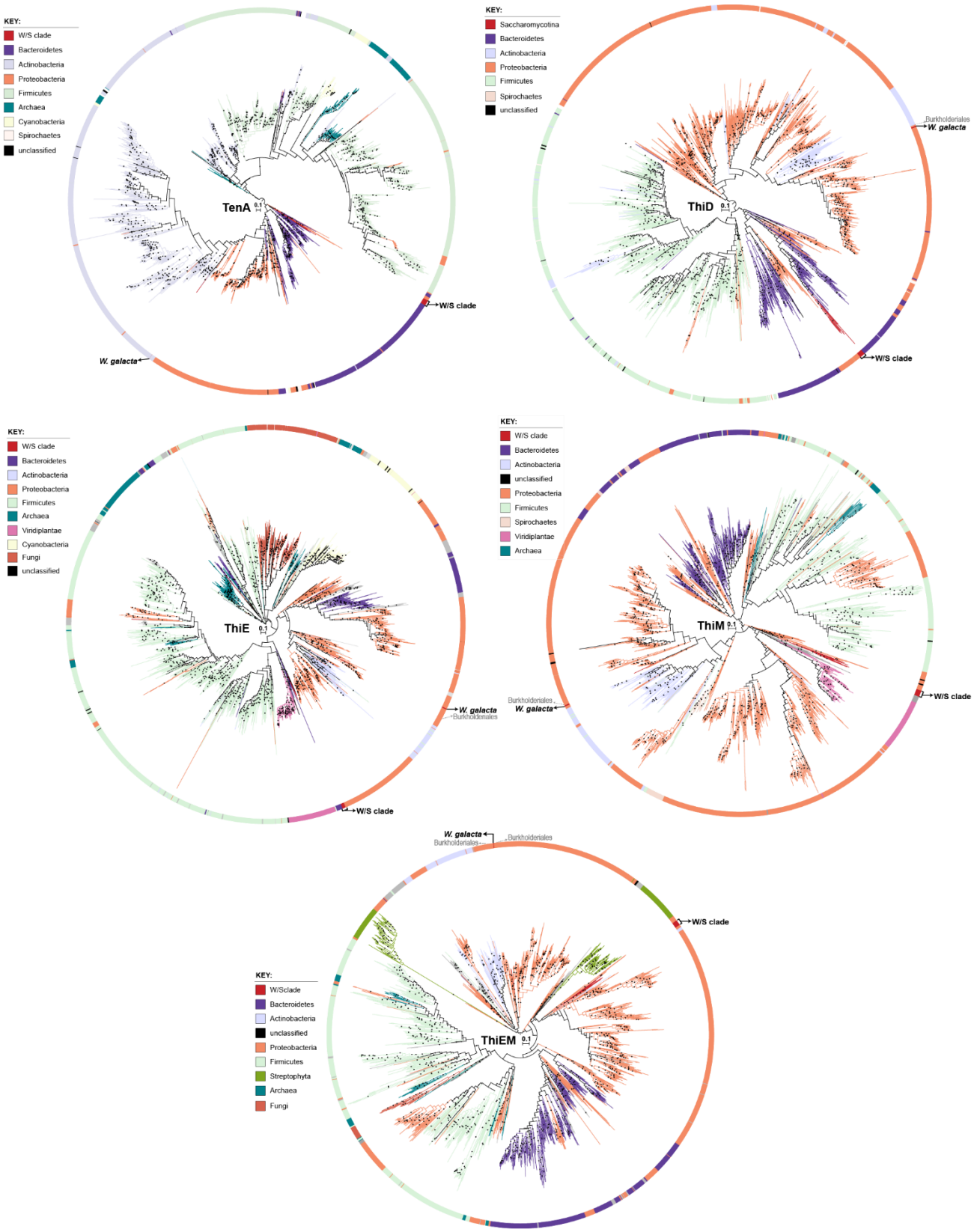

Fig. S2. Preliminary phylogenies of Thi proteins. The top 5.000 hits on NCBI (Refseq database) were selected using *St. bombicola* proteins as queries. Taxonomic classification (Phylum for Bacteria and Kingdom/Domain for other organisms) was automatically assigned using a customized *E*-utilities script provided by Entrez Direct and is presented in the key on the left next to the respective phylogenetic tree.

Lineages with less than 20 representatives are colored in grey. All phylogenies were midpoint rooted and visualized in iTOL v4 (8).

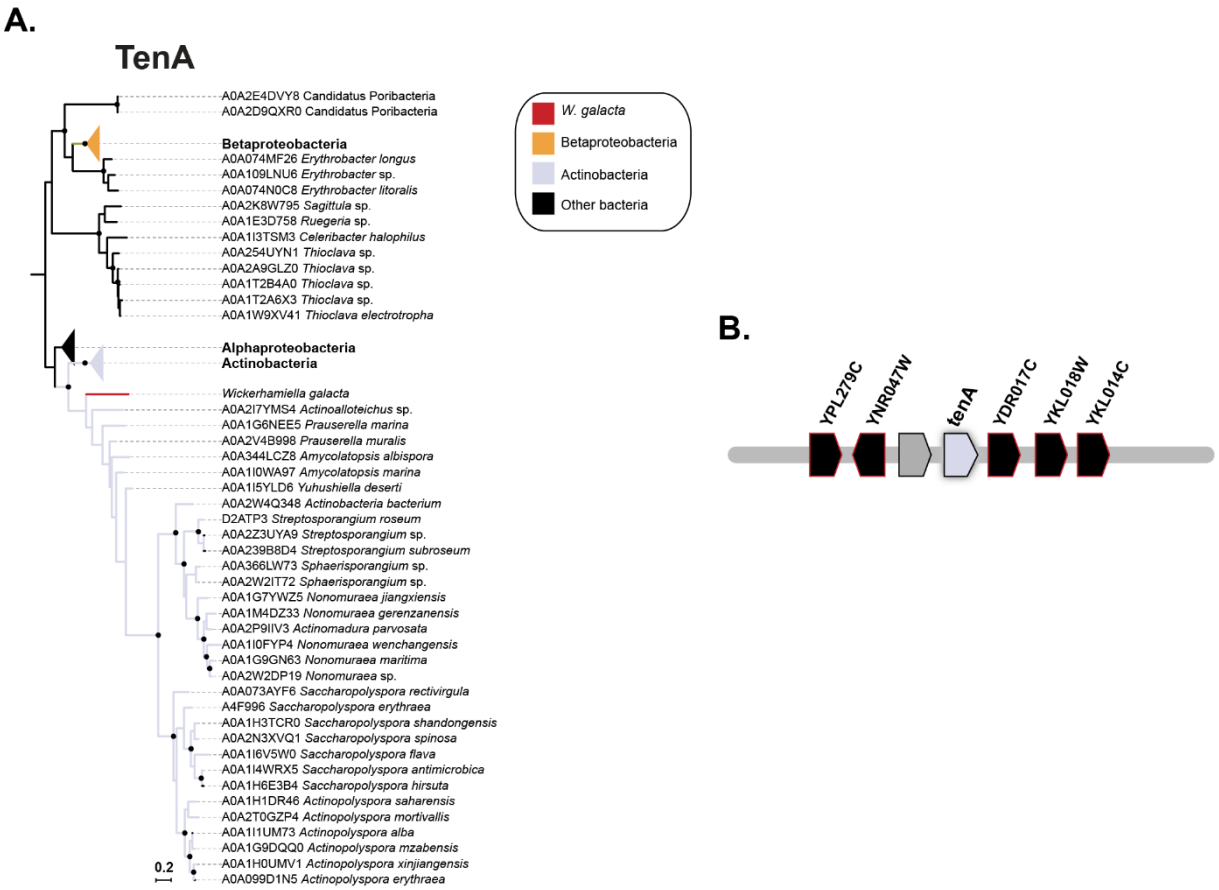

Fig. S3. Maximum likelihood phylogeny of proteins closely related to TenA from *W. galacta*. (A) Phylogeny constructed with the top 750 BLASTp hits in UniprotKB. Branches with support higher than 95% (ultrafast bootstrap) are indicated by black circles. **(B)** Gene organization in the vicinity of *tenA* from *W. galacta*. Genes are represented by arrows denoting relative transcriptional direction. Genes from fungal origin are outlined in red while genes with dubious annotations are represented in grey. The closest related genes in *S. cerevisiae* are shown on top of the arrows.

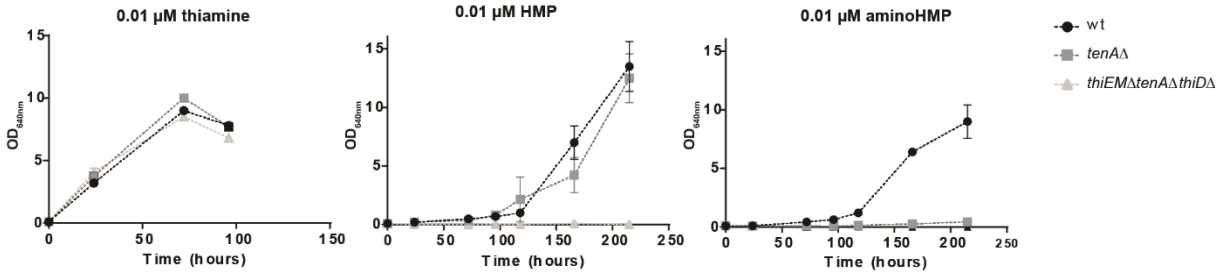

Fig. S4. Growth curves for wt and *THI* deletion mutants (***tenAΔ*** and ***thiEMΔtenAΔthiDΔ***) grown in YNB medium without aminoacids and without thiamine supplemented with 0.01 µM of thiamine, 0.01 µM of HMP or 0.01 µM of aminoHMP, as indicated. Cultures were pre-grown in the respective media for one to four days (depending on the lag-phase) until late exponential phase was reached and were subsequently inoculated in the same medium to an OD_{640nm} of 0.2. Growth was monitored for 220 hours at 30ºC with shaking (180 r.p.m.).

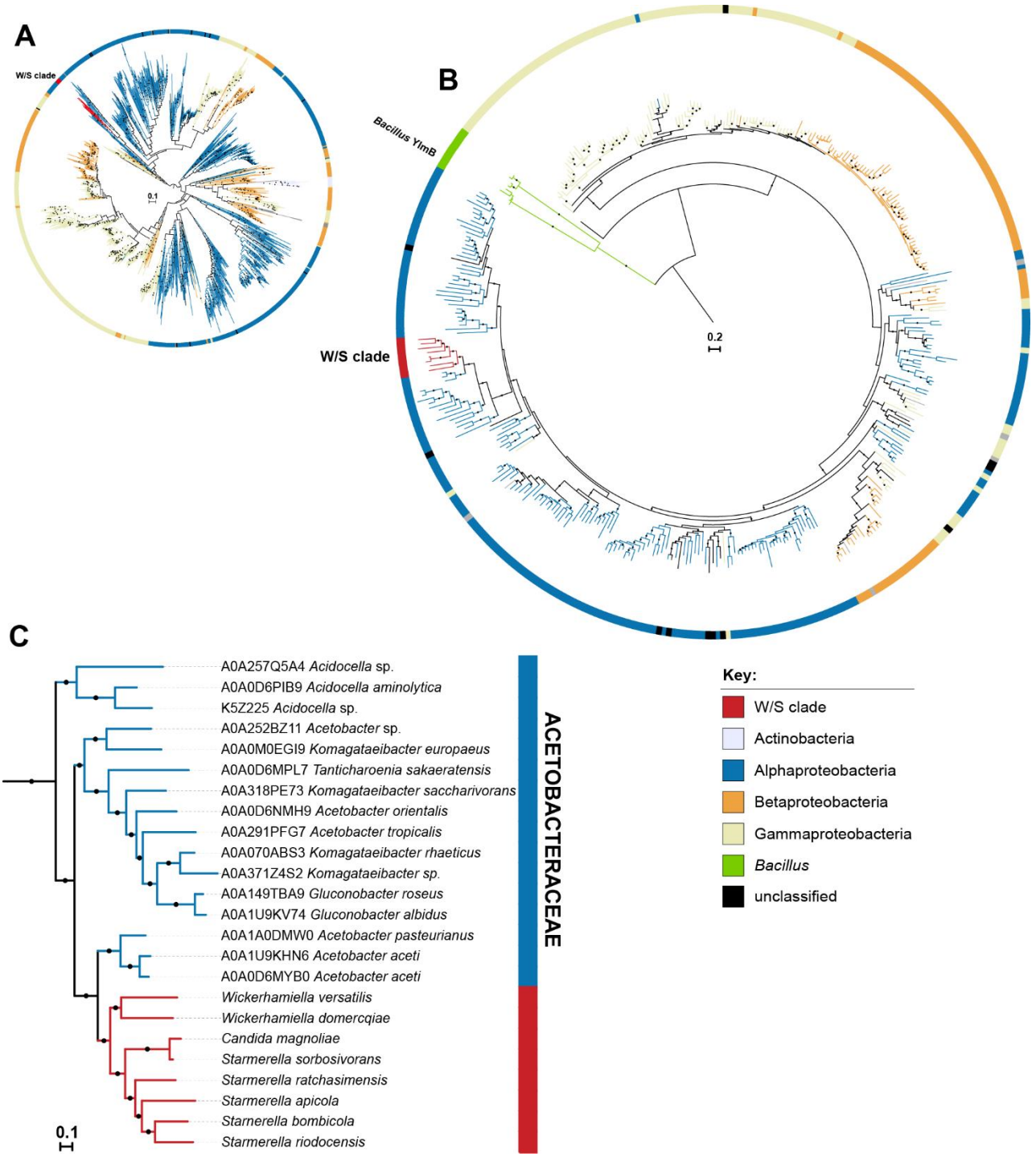

Fig. S5. Maximum likelihood phylogeny of proteins closely related to putative YlmB from W/S-clade species. **(A)** Preliminary phylogeny constructed with the top 5.000 hits to putative YlmB from *St. bombicola*. Complete **(B)** and pruned **(C)** ML phylogeny constructed with the top 750 hits to putative YlmB from *St. bombicola* with the addition of functionally characterized YlmB proteins from *Bacillus subtilis* (indicated in green). Branches with support higher than 95% (ultrafast bootstrap) are indicated by black dots. Phylogenies were midpoint rooted and visualized in iTOL v4 (8) and branches are colored according to the taxonomic classification as indicated in the key.

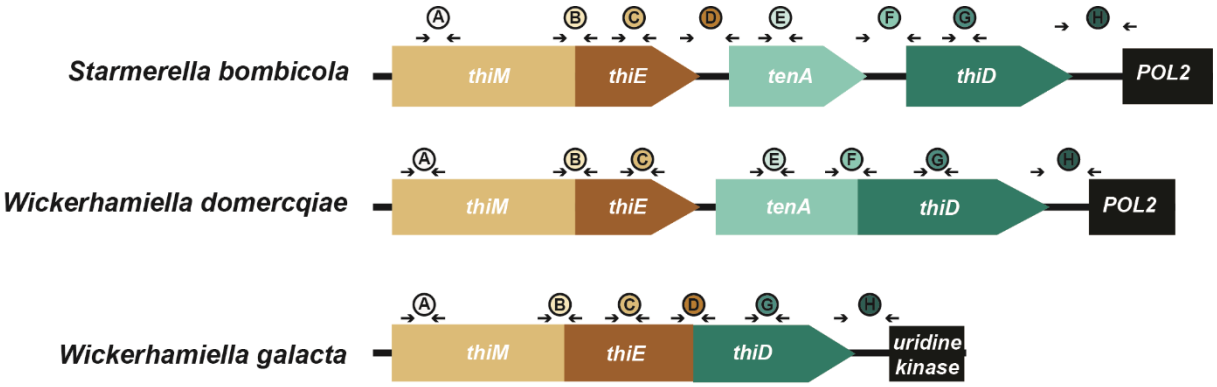

Fig. S6. Primer pairs used for RT-qPCR. Primer sequences and efficiencies are shown in Table S3.

Supplementary Tables

Table S1. Complete list of fungal taxa, abbreviated species names and genome databases and accession numbers of RNA polymerase proteins used to construct the species tree (Figure 1A and Figure S1). Abbreviated species names are given for each species as used in the phylogeny. W/S-clade species are highlighted in blue. Species used as outgroups are highlighted in orange.

Species	Abbreviated names	Strain(s)	Genome database
Aspergillus niger	*A.* niger	CBS 513.88	BROAD/JGI
Fusarium graminearum (Gibberella zeae)	*F. graminearum*	PH-1	SGD/BROAD
Blastobotrys adeninivorans	*B. adeninivorans*		NCBI
Ascoidea rubescens	*A. rubescens*	NRRL Y17699	JGI
Babjeviella inositovora	*A. inositovora*	NRRL Y-12698	JGI
Candida tenuis	*C. tenuis*	NRRL Y-1498	JGI
Candida albicans	*C. albicans*	WO-1	SGD/BROAD
[Candida] apicola	*C. apicola*	NRLL Y-50540	NCBI
Candida arabinofermentans	*C. arabinofermentans*	NRRL YB-2248	JGI
Candida caseinolytica	*C. caseinolytica*	NRRL Y-17796	JGI
Candida dubliniensis	*C. dubliniensis*	CD36	SGD/YGOB
Candida glabrata	*C. glabrata*	CBS 138	SGD/YGOB
[Candida] magnoliae	*C. magnoliae*	PYCC 2903	NCBI
Candida parapsilosis	*C. parapsilosis*	CDC317	BROAD/NCBI
Candida tanzawaensis	*C. tanzawaensis*	NRRL Y-17324	JGI
Candida tropicalis	*C. tropicalis*	MYA-3404	SGD/BROAD
Clavispora lusitaniae	*C. lusitaniae*	ATCC 42720	BROAD/NCBI
Cyberlindnera jadinii	*C. jadinii*	NBRC 0988	NCBI
Debaryomyces hansenii	*D. hansenii*	CBS767	SGD/NCBI
Dekkera bruxellensis	*D. bruxellensis*	AWRI1499	NCBI
Eremothecium cymbalariae	*E. cymbalariae*	DBVPG#7215	NCBI/YGOB
Eremothecium gossypii	*E. gossypii*	ATCC 10895	SGD/NCBI
Hyphopichia burtonii	*H. burtonii*	NRRL Y-1933	JGI
Kazachstania africana	*K. africana*	CBS 2517	YGOB
Kazachstania naganishii	*K. naganishii*	CBS 8797	YGOB
Kluyveromyces aestuarii	*K. aestuarii*	ATCC 18862	NCBI
Kluyveromyces lactis	*K. lactis*	NRRL Y-1140	SGD/YGOB
Kluyveromyces wickerhamii	*K. wickerhamii*	UCD 54-210	NCBI
Komagataella pastoris	*K. pastoris*	GS115 + CBS 7435	SGD/NCBI
Lachancea kluyveri	*L. kluyveri*	NRRL Y-12651	SGD/YGOB
Lachancea thermotolerans	*L. thermotolerans*	CBS 6340	SGD/YGOB
Lachancea waltii	*L. waltii*	NCYC 2644	SGD/YGOB
Lipomyces starkeyi	*L. starkeyi*	NRRL Y-11557	JGI
Lodderomyces elongisporus	*L. elongisporus*	NRRL YB-4239	BROAD/NCBI
Metschnikowia bicuspidata	*M. bicuspidata*	NRRL YB-4993	JGI
Meyerozyma guilliermondii	*M. guilliermondii*	ATCC 6260	BROAD/NCBI
Millerozyma farinosa	*M. farinosa*	CBS 7064	SGD/NCBI

Species	Abbreviated names	Strain(s)	Genome database
Nadsonia fulvescens var. elongata	*N. fulvescens*	DSM 6958	JGI
Naumovozyma castellii	*N. castellii*	CBS 4309	YGOB
Ogataea polymorpha	*O. polymorpha*	NCYC 495 leu1.1	JGI
Ogataea parapolymorpha	*O. parapolymorpha*	DL-1	NCBI
Pachysolen tannophilus	*P. tannophilus*	NRRL Y-2460	JGI/NCBI
Pichia kudriavzevii	*P. kudriavzevii*	M12	NCBI
Pichia membranifaciens	*P. membranifaciens*	NRRL Y-2026	JGI
Saccharomyces cerevisiae	*S. cerevisiae*	S288c	SGD/BROAD
Saccharomyces eubayanus	*S. eubayanus*	FM1318	NCBI
Saccharomyces kudriavzevii	*S. kudriavzevii*	IFO 1802	SSS Website
Saccharomyces mikatae	*S. mikatae*	IFO 1815	SSS Website SGD/BROAD
Saccharomyces paradoxus	*S. paradoxus*	NRRLY-17217	SSS Website NCBI
Saccharomyces uvarum	*S. uvarum*	CBS 7001	SSS Website SGD/BROAD
Saprochaete clavata	*Sa. clavata*	CNRMA 12.647	NCBI
Scheffersomyces stipitis	*S. stipitis*	CBS 6054	JGI/NCBI
Spathaspora passalidarum	*S. passalidarum*	NRRL Y-27907	JGI/NCBI
Starmerella bacillaris	*St. bacillaris*	PYCC 3044	NCBI
Starmerella riodocensis	*St. riodocensis*	NRRL Y-27859	NCBI (Y1000+ project)
Starmerella ratchamatensis	*St. ratchamatensis*	CBS 10611	NCBI (Y1000+ project)
Starmerella sorbosivorans	*St. sorbosivorans*	CBS 8768	NCBI (Y1000+ project)
Starmerella bombicola	*St. bombicola*	PYCC 5882	NCBI
Sugiyamaella lignohabitans	*Su. lignohabitans*	NRRL YB-1473	NCBI
Tetrapisispora blattae	*T. blattae*	CBS 6284	YGOB
Tetrapisispora phaffii	*T. phafii*	CBS 4417	YGOB
Torulaspora delbrueckii	*T. delbrueckii*	CBS 1146	YGOB
Trichomonascus petasosporus	*T. petasosporus*	NRRL YB-2093	JGI
Vanderwaltozyma polyspora	*V. polyspora*	DSM 70294	YGOB/NCBI
Wickerhamiella domercqiae	*W. domercqiae*	PYCC 3067	NCBI
Wickerhamiella cacticola	*W. cacticola*	NRRL Y-27362	NCBI (Y1000+ project)
Wickerhamiella hasegawae	*W. hasegawae*	JCM 12559	NCBI (Y1000+ project)
Wickerhamiella galacta	*W. galacta*	NRRL Y-17645	NCBI (Y1000+ project)
Wickerhamiella occidentalis	*W. occidentalis*	NRRL Y-27364	NCBI (Y1000+ project)
Wickerhamiella pararugosa	*W. pararugosa*	NRRL Y-17089	NCBI (Y1000+ project)
Wickerhamiella versatilis	*W. versatilis*	JCM 5958	RIKEN
Wickerhamomyces anomalus	*W. anomalus*	NRRL Y-366	JGI
Yarrowia lipolytica	*Y. lipolytica*	CLIB122	SGD/NCBI
Zygosaccharomyces bailii	*Z. bailii*	CLIB 213	NCBI
Zygosaccharomyces kombuchaensis	*Z. kombuchaensis*	CBS 8849	Local database
Zygosaccharomyces rouxii	*Z. rouxii*	CBS 732	SGD/JGI

Table S2. Information concerning strains and genome assemblies used in this work.

Species	Genome assembly reference	Information was included in:
Sugiyamaella lignohabitans CBS10342	ASM164002v2	Figure 1B
Wickerhamiella galacta NRRL Y-17645	ASM304524v1	Figure 1B, Figure 2, Figure 4
Wickerhamiella hasegawae JCM 12559	ASM412510v1	Figure 1B
Wickerhamiella pararugosa NRRL Y-17089	ASM412523v1	Figure 1B
Wickerhamiella occidentalis NRRL Y-27364	ASM412509v1	Figure 1B
Wickerhamiella sorbophila DS02	ASM225199v2	Figure 1B
Wickerhamiella versatilis JCM 5958	JCM_5958_assembly_v001	Figure 1B, Figure 2, Figure 3
Wickerhamiella domercqiae PYCC 3067	ASM303370v1	Figure 1B, Figure 2, Figure 3
Starmerella sorbosivorans CBS 8768	ASM412500v1	Figure 1B, Figure 2, Figure 3
Candida magnoliae PYCC 2903	ASM303343v1	Figure 1B, Figure 2, Figure 3
Starmerella bacillaris PYCC 3044	ASM303376v1	Figure 1B, Figure 2, Figure 3
Starmerella apicola NRRL Y-50540	ASM100541v1	Figure 1B, Figure 2, Figure 3
Starmerella ratchamatensis CBS 10611	ASM412497v1	Figure 1B, Figure 2, Figure 3
Starmerella riodocensis NRRL Y-27859	ASM412495v1	Figure 1B, Figure 2, Figure 3
Starmerella bombicola PYCC 5882	ASM303378v1	Figure 1B, Figure 2, Figure 3
Dysgonomonas mossii DSM22836	ASM37640v1	Figure 2
Mangrovibacterium marinum DSM 28823	ASM304662v1	Figure 2
Caballeronia udeis ES_PA-B7	ASM361007v1	Figure 2
Burkholderia novacaledonica LMG 28615T	LMG28615	Figure 2
Peptostreptococcus russellii RT-10B	ASM301205v1	Figure 2
Veillonella dispar DORA_11	GCA_000508585.1	Figure 2

Table S3. Primers used for Real Time PCR. Efficiencies of each primer pair were calculated using five ten-fold DNA dilutions.

Name	Sequence 5'-3'	Purpose	Primer Efficiency (%)
qPCR_thiD_Fw	AGATACTGATGGCACTCTCATCG	primers for qRT-PCR of *thiD* from *St. bombicola*	101
qPCR_thiD_Rv	GTGAACTGGGCCATGACCTC	primers for qRT-PCR of *thiD* from *St. bombicola*	
qPCR_thiE_Fw	TTATGGCGTCGGAGCCTACC	primers for qRT-PCR of *thiM* portion from *St. bombicola*	99,6
qPCR_thiE_Rv	CGAGTAGTGGCTTCAGATGTGTC	primers for qRT-PCR of *thiM* portion from *St. bombicola*	
qPCR_thiEfim_Fw	GAAGCGTGCAAAGCAATTAGTGG	primers for qRT-PCR of *thiE* portion from *St. bombicola*	92,5
qPCR_thiEfim_Rv	CAGCTTTTGGGTCTTCTGCCAG	primers for qRT-PCR of *thiE* portion from *St. bombicola*	
qPCR_tenA_Fw	GCTGAGCAGGAATTGCACGAC	primers for qRT-PCR of *tenA* from *St. bombicola*	94,5
qPCR_tenA_Rv	GCGCATAGGGATGTTCGACG	primers for qRT-PCR of *tenA* from *St. bombicola*	
qPCR_thiM718_Fw	CAGTTGCTGGAGAGCTCGCTG	primers for qRT-PCR of the fragment in between *thiM and thiE* from *St. bombicola*	95,7
qPCR_thiE1024_Rv	GGAACATGGAACTCATCGCATACC	primers for qRT-PCR of the fragment in between *thiM and thiE* from *St. bombicola*	
qPCR_Pol2_Fw	TGACGGCAGAGGCTATCTTGTC	primers for qRT-PCR of *POL2* from *St. bombicola*	94,1
qPCR_Pol2_Rv	CGTTAAACGTAGCGATGACAGTGG	primers for qRT-PCR of *POL2* from *St. bombicola*	
qPCR_thiD_779Fw	GTAGAGGTCATGGCCCAGTTCA	primers for qRT-PCR of the fragment in between *thiD and POL2* from *St. bombicola*	94,5
qPCR_Pol2_82Rv	TGGCATCCAACATGTCCGTCAG	primers for qRT-PCR of the fragment in between *thiD and POL2* from *St. bombicola*	
qPCR_Wicdom_tenA334_Fw	CTAAGGTGTGCTGCAGTTGAGC	primers for qRT-PCR of *tenA* from *W. domercqiae*	94,8
qPCR_Wicdom_tenA545Rv	GTGGCTGCTGCAAGCTCATCG	primers for qRT-PCR of *tenA* from *W. domercqiae*	
qPCR_Wicdom_thiD850Fw	CTGGTGCGTTTGTCCAAGCTC	primers for qRT-PCR of *thiD* from *W. domercqiae*	93,1
qPCR_Wicdom_thiD1089Rv	TGCAAGCGGAATTAAGCGATCACG	primers for qRT-PCR of *thiD* from *W. domercqiae*	
Wicdom_thiM67Fw	GCGGTAGCATCTAACTACGCTG	primers for qRT-PCR of *thiM* portion from *W. domercqiae*	94.3
Wicdom_thiM308Rv	GCCACTAGAGCTCTGCATGTATC	primers for qRT-PCR of *thiM* portion from *W. domercqiae*	
Wicdom_thiE1116Fw	TCTTGGACTCAGCGTGGGATC	primers for qRT-PCR of *thiE* portion from *W. domercqiae*	97.4
Wicdom_thiE1364Rv	GCTACTCCAGCAGCACCTGTG	primers for qRT-PCR of *thiE* portion from *W. domercqiae*	
Wicdom_thiM684Fw	TGGTGTTGCCGGTGAGCTTGC	primers for qRT-PCR of the fragment in between *thiM and thiE* from *W. domercqiae*	108.3
Wicdom_thiE953Rv	CTTGACCTCAACCGCGCGAG	primers for qRT-PCR of the fragment in between *thiM and thiE* from *W. domercqiae*	
Wicdom_tenA518_Fw	CATGCGATGAGCTTGCAGCAGC	primers for qRT-PCR of the fragment in between *tenA and thiD* from *W. domercqiae*	86,7
Wicdom_thiD790_Rv	CGTCTGAGAATACACTCCAAGCG	primers for qRT-PCR of the fragment in between *tenA and thiD* from *W. domercqiae*	
Wicdom_POL2_661Fw	CGTCTCGCAATTGATCTTGATGTGC	primers for qRT-PCR of *POL2* from *W. domercqiae*	90.0
Wicdom_POL2_915Rv	CAATGTCCTGACTTACTAGCTCACG	primers for qRT-PCR of *POL2* from *W. domercqiae*	
Wicdom_tenAthiD_1451Fw	GCTGCGTCGGATCAACAACTGG	primers for qRT-PCR of the fragment in between *thiD and POL2* from *W. domercqiae*	94,4
Wicdom_POL2_36Rv	CGCGATGAAAGACGCGACGC	primers for qRT-PCR of the fragment in between *thiD and POL2* from *W. domercqiae*	
Wgalacta_thiM266Fw	GTCTCCATTCGTGCGGCCATTG	primers for qRT-PCR of *thiM* portion from *W. galacta*	94,9
Wgalacta_ThiM485Rv	CGTCTTCAGAGCTAACAGCGC	primers for qRT-PCR of *thiM* portion from *W. galacta*	
Wgalacta_thiE312Fw	GCTCGATGGTCTGGCAGAGAT	primers for qRT-PCR of *thiE* portion from *W. galacta*	100,5
Wgalacta_thiE503Rv	GCTTCAAGTTGACTACCTGCCG	primers for qRT-PCR of *thiE* portion from *W. galacta*	
Wgalacta_thiD527Fw	GGCATCTACCATCAGACCTTACTAG	primers for qRT-PCR of *thiD* from *W. galacta*	101,6
Wgalacta_thiD700Rv	CCTTGACGGATTGTGCAAGGC	primers for qRT-PCR of *thiD* from *W. galacta*	
Wgalacta_thiM798Fw	GGAGCAGCTGACTTTGAGCG	primers for qRT-PCR of the fragment in between *thiM and thiE* from *W. galacta*	105,6
Wgalacta_thiE81Rv	CGAGTGCCACGTCTACGTCG	primers for qRT-PCR of the fragment in between *thiM and thiE* from *W. galacta*	
Wgalacta_thiE312Fw	GCTCGATGGTCTGGCAGAGAT	primers for qRT-PCR of the fragment in between *thiD and thiE* from *W. galacta*	92,5
Wgalacta_thiD105Rv	CACGCTGAGACCATAAGCTCC	primers for qRT-PCR of the fragment in between *thiD and thiE* from *W. galacta*	
Wgalacta_thiD750Fw	TGTAGGTCATGGAACTGGTCCG	primers for qRT-PCR the fragment between thiD and uridine kinase from *W. galacta*	116.9
Wgalacta_uridineRv	GCCACTCCCTGAACATCCAGC	primers for qRT-PCR the fragment between thiD and uridine kinase from *W. galacta*	

Table S4. Primers used for the disruption of *THI* genes in *St. bombicola* PYCC 5882.

Name	Sequence 5'-3'	Purpose
thiE_upNotI_Fw	TATCATGCGGCCGCGTGATACCTGGCATCAG	Amplification of the upstream and downstream regions of thiEM
thiE_downNcoI_Rv	ATGATACCATGGGGATTCCAGCACCTCC	
thiE_transf_Fw	TGAGTGGATTCTCAGGATC	Amplification of final construction for thiEM deletion
thiE_transf_Rv	GGCAATGGAAAGTGCTG	
tenAthiD_upstreamSphI_Fw	CGGGCACGATGCATTC	Amplification of the upstream and downstream regions of thiD
tenAthiD_downstream_KpnI_Rv	ATGATAGGTACCCACAGGTTGTGCACGC	
thiD_transf_Fw	TAGCTCTCATTGCTGCC	Amplification of final construction for thiD deletion
thiD_transf_Rv	AATTAGATCGCGACGCGT	
tenAKO_KpnI_Rv	GACTTAGGTACCGAGTGCCATCAGTATCTTC	Amplification of the upstream and downstream regions of tenA
tenA_603_SmaI_Fw	GACTTACCCGGGGATGACCGAGGCATTTGT	
tenA_transf_Fw	GGTTACAGATCCAGAGC	Amplification of final construction for tenA deletion
tenA_transf_Rv	GAGTGCCATCAGTATCTTC	
thiE_upNotI_Fw	TATCATGCGGCCGCGTGATACCTGGCATCAG	Amplification of the upstream and downstream regions of the entire operon
tenAthiD_downNotI_Rv	ATGATAGCGGCCGCTCACGAGCATATGGAGTC	
thiE_transf_Fw	TGAGTGGATTCTCAGGATC	Amplification of final construction for the deletion of the entire operon
THIope_transf_Rv	CTTCGCTACCGATATCG	

References

1. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)* **22**, 1658-1659 (2006).
2. K. Katoh, D. M. Standley, MAFFT: iterative refinement and additional methods. *Methods in molecular biology (Clifton, N.J.)* **1079**, 131-146 (2014).
3. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)* **25**, 1972-1973 (2009).
4. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular biology and evolution* **32**, 268-274 (2015).
5. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular biology and evolution* **35**, 518-522 (2017).
6. A. H. Jenkins, G. Schyns, S. Potot, G. Sun, T. P. Begley, A new thiamin salvage pathway. *Nature chemical biology* **3**, 492-497 (2007).
7. C. Goncalves *et al.*, Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. *eLife* **7** (2018).
8. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-w259 (2019).