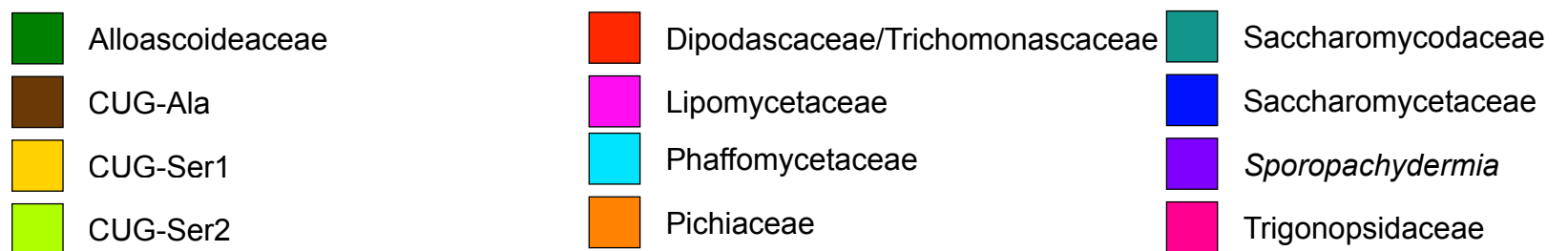
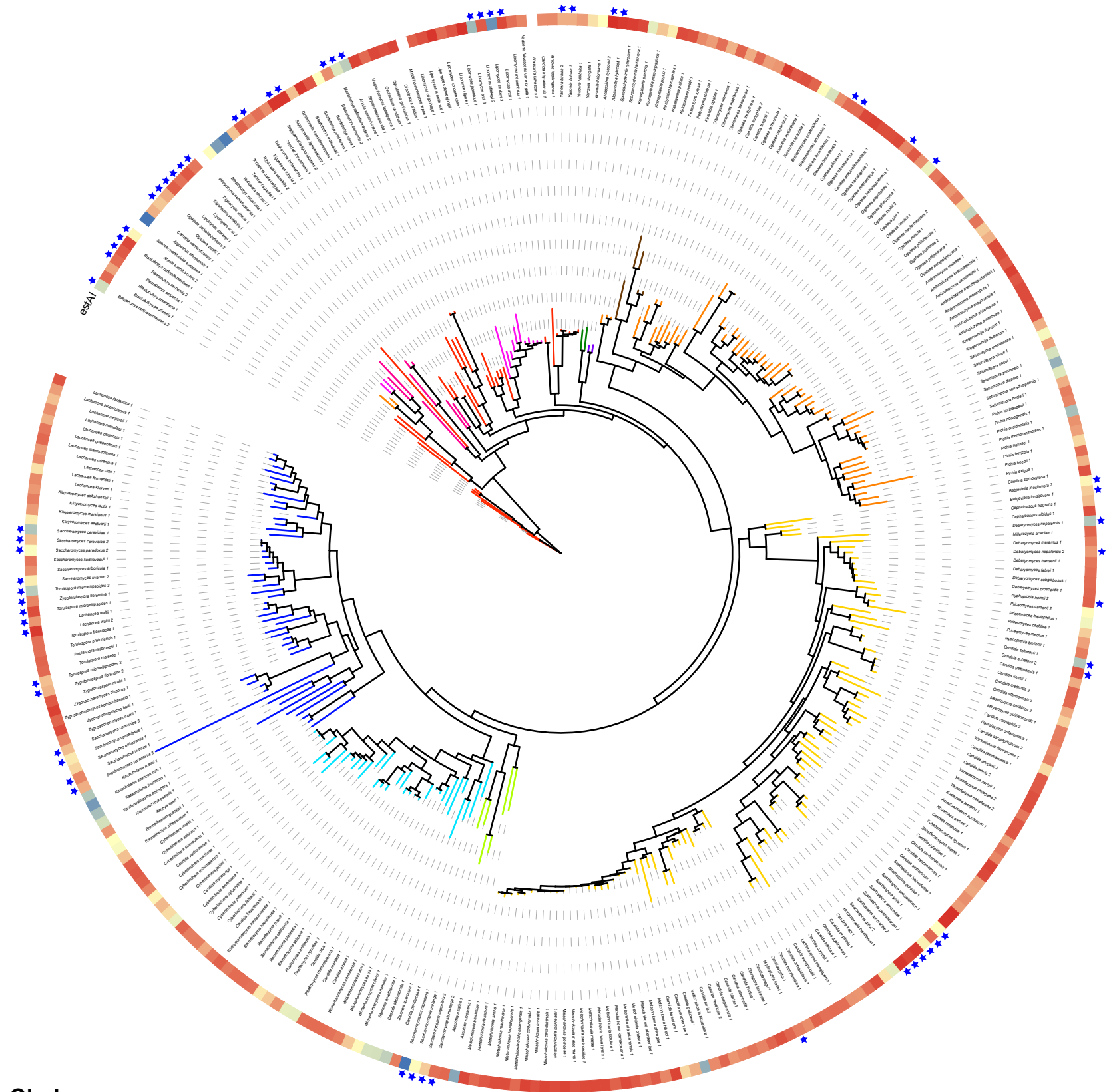
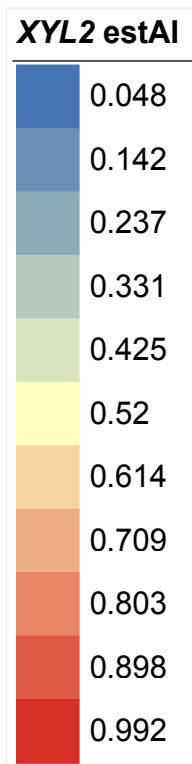


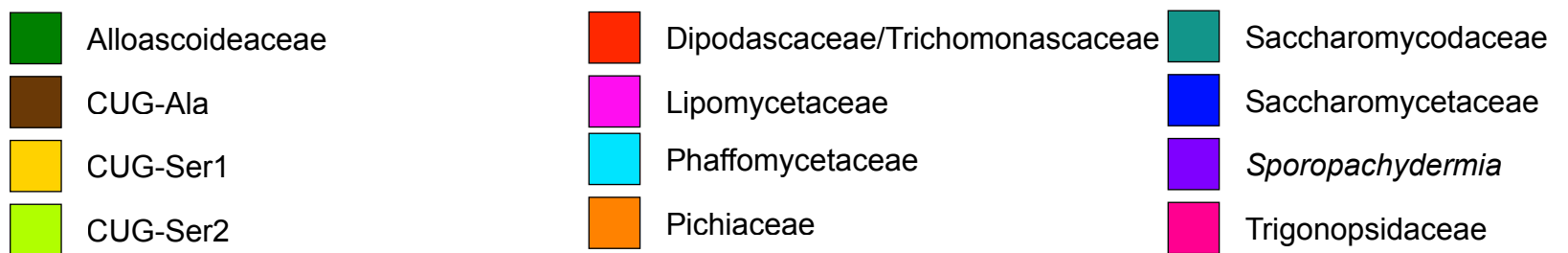
### Clades



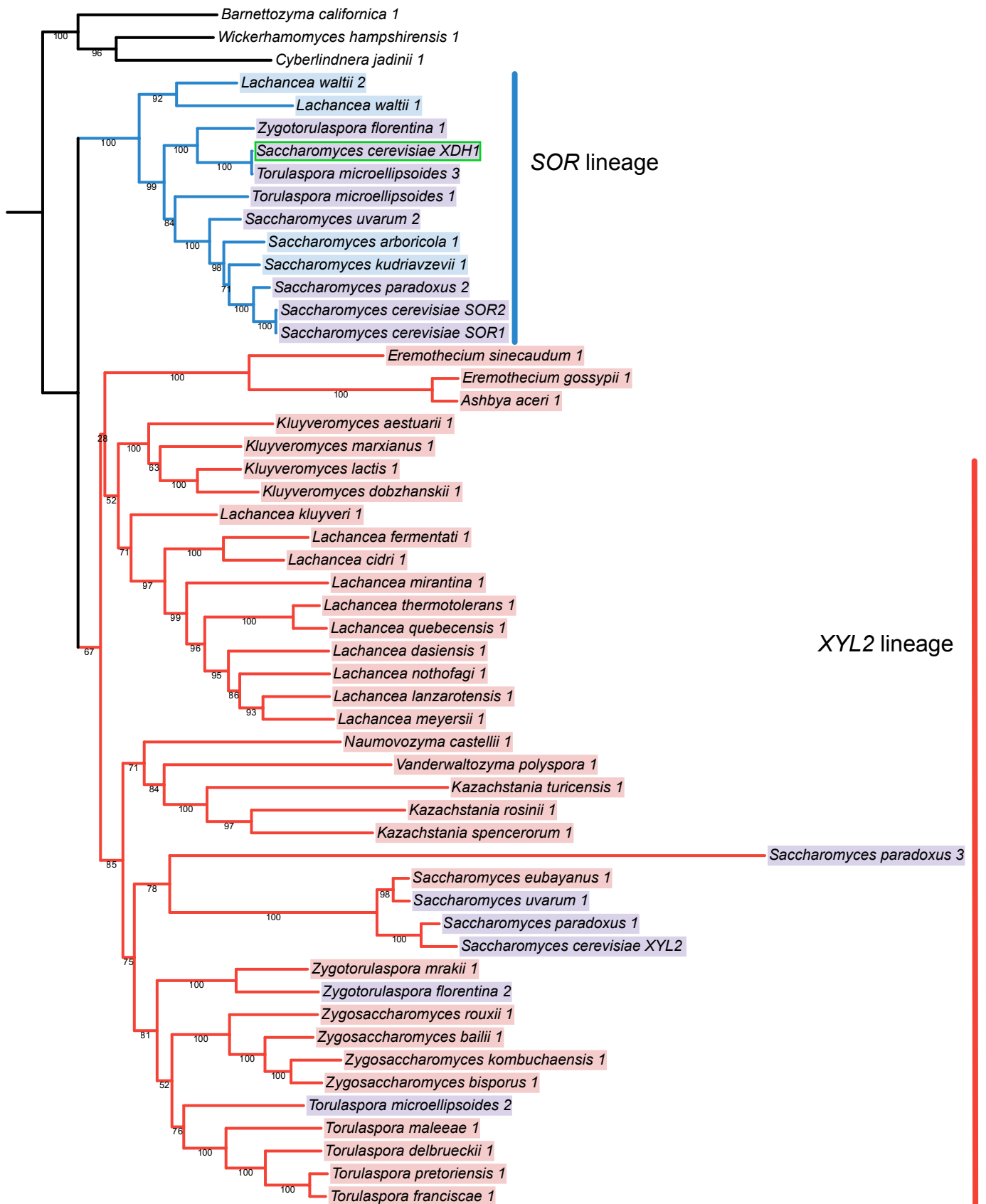
**S1. Maximum likelihood tree of Xyl1 protein sequences.** Maximum likelihood phylogeny of Xyl1 protein sequences was built using IQTree with an Auto substitution model based on 1,000 bootstrap replications. Major yeast clades are depicted by branch color using tree topology from Shen et al. (1). Color gradient indicates codon optimization indices (estAI values) with color ranging from blue (lowest estAI value) to red (highest estAI value). Stars indicate paralogs belonging to multi-copy lineages. Note that these analyses show that *S. cerevisiae* GRE3 is a XYL1 homolog.



**Clades**

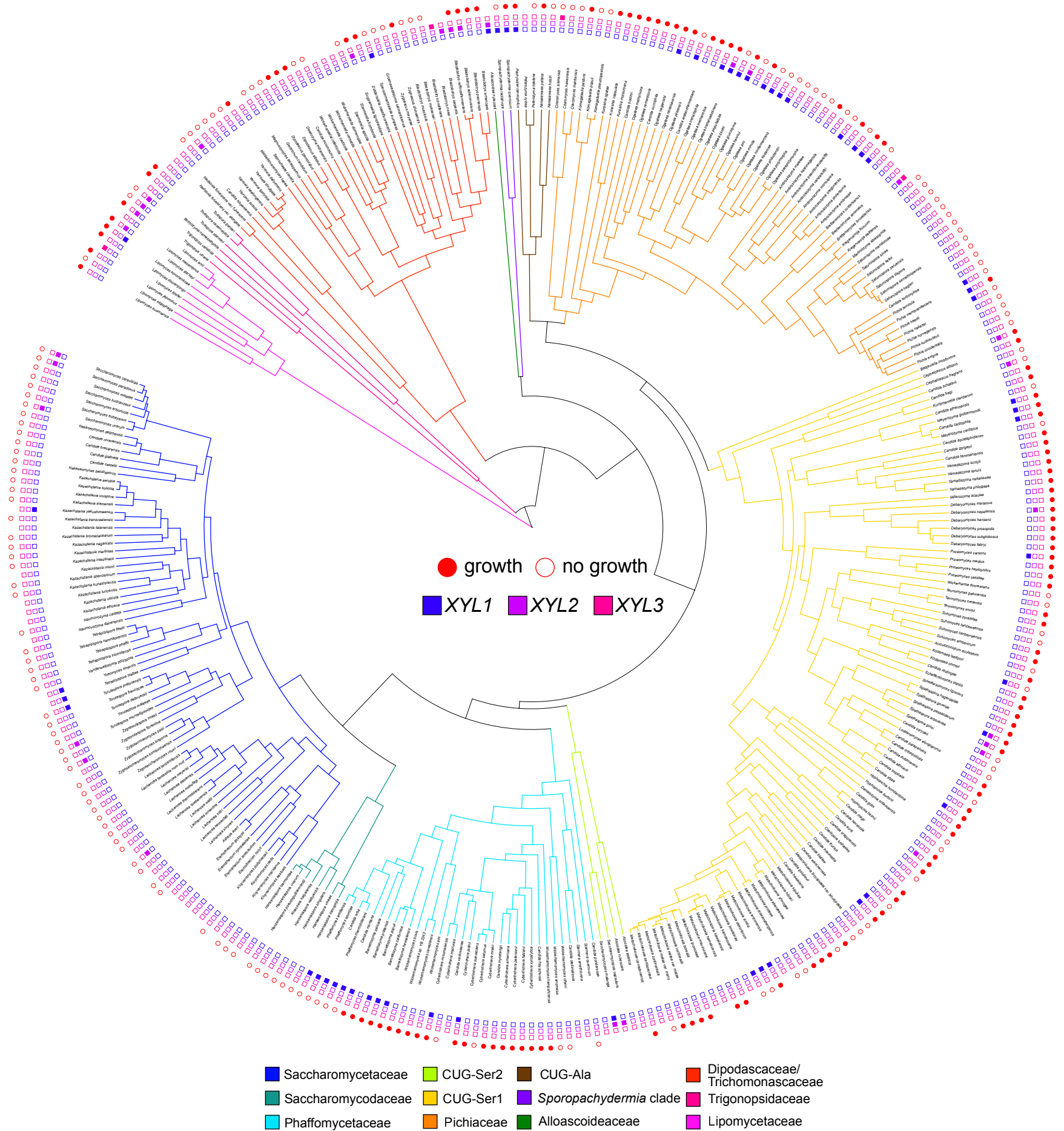


**S2. Maximum likelihood tree of Xyl2 protein sequences.** Maximum likelihood phylogeny of Xyl2 protein sequences was built using IQTree with an Auto substitution model based on 1,000 bootstrap replications. Major yeast clades are depicted by branch color using tree topology from Shen et al. (1). Color gradient indicates codon optimization indices (estAI values) with color ranging from blue (lowest estAI value) to red (highest estAI value). Stars indicate paralogs belonging to multi-copy lineages. Notably, the stars depict the lineage-specific duplication of *SOR1/SOR2* in *S. cerevisiae*. The nearly identical *S. cerevisiae* paralogs *SOR1* and *SOR2* fall within the Saccharomycetaceae *XYL2* clade, suggesting that *SOR1/SOR2* are both paralogs of *S. cerevisiae* *XYL2*. Associated *S. cerevisiae* gene nomenclature: *Saccharomyces cerevisiae* 1 - *SOR2*, *Saccharomyces cerevisiae* 2 - *SOR1*, *Saccharomyces cerevisiae* 3 - *XYL2*.



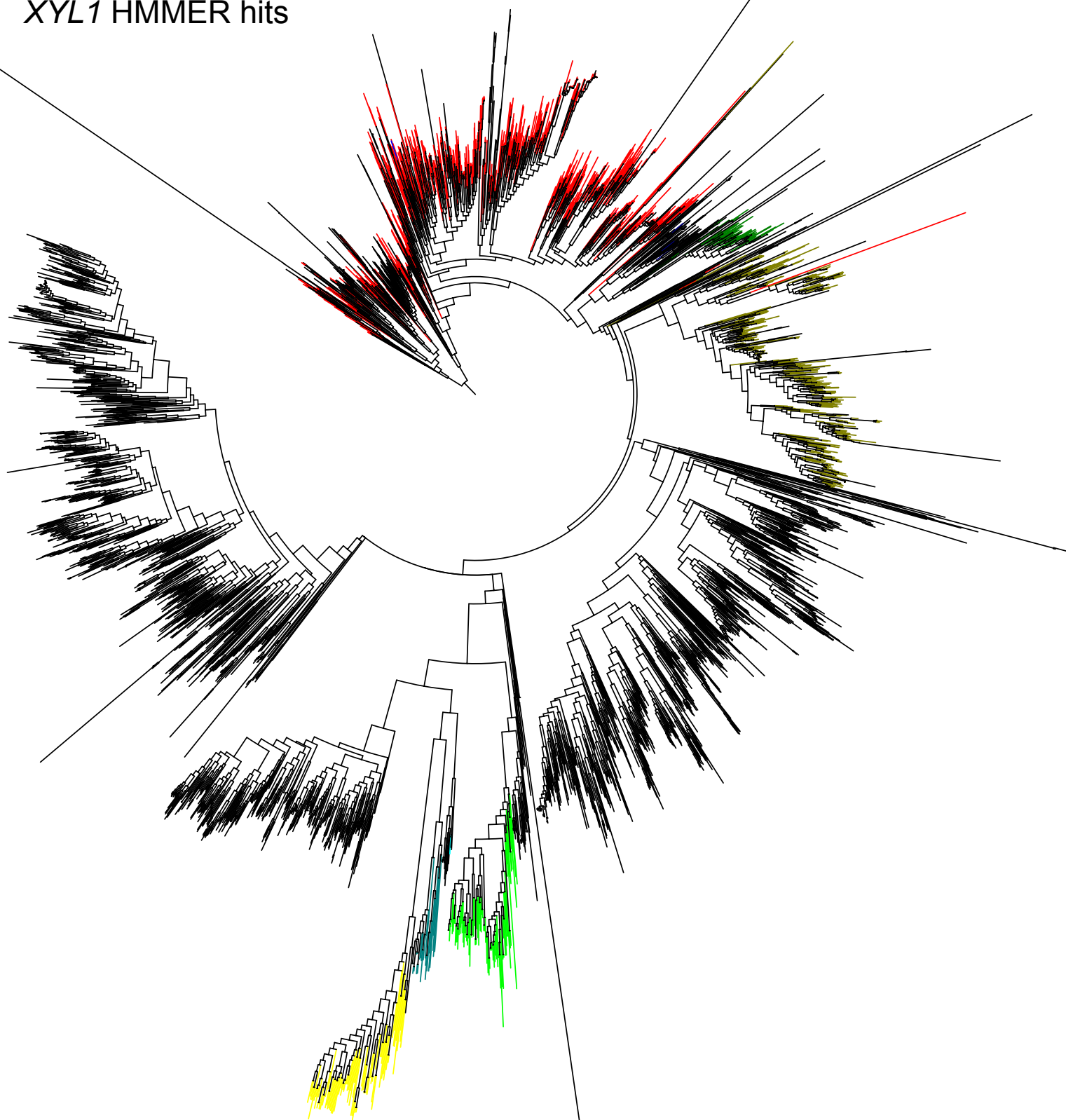
**S3. Maximum likelihood tree of Saccharomycetaceae Xyl2 protein sequences.** Independent maximum likelihood tree of Xyl2 protein sequences in the family Saccharomycetaceae with the addition of *S. cerevisiae XDH1* (2) (green box) generated using IQ tree with an LG+I+G4 substitution model and node support based on 1,000 bootstrap replications. The duplication of XYL2 prior to the yeast whole genome duplication resulted in two major lineages sorting within the Saccharomycetaceae, the SOR lineage (blue) and the XYL2 lineage (red). Species that retained sequences from both lineages are indicated in purple. XDH1 (green box), previously identified as responsible for weak xylose utilization in some wine *S. cerevisiae* strains, was included and found to be horizontally transferred from *Torulaspota microellipsoides* (3).
















**Figure S4. XYL gene copy number xylose growth across 332 Saccharomycotina species.** Major yeast clades are depicted by branch color. Filled boxes indicate multi-copy lineages for the respective XYL gene. Species with non-zero growth rates in xylose medium are indicated by a filled red circle, and species unable to assimilate xylose are indicated by an empty red circle. Species without circles were not assayed for growth. Time-calibrated phylogeny from (Shen et al. 2018).

## XYL1 HMMER hits

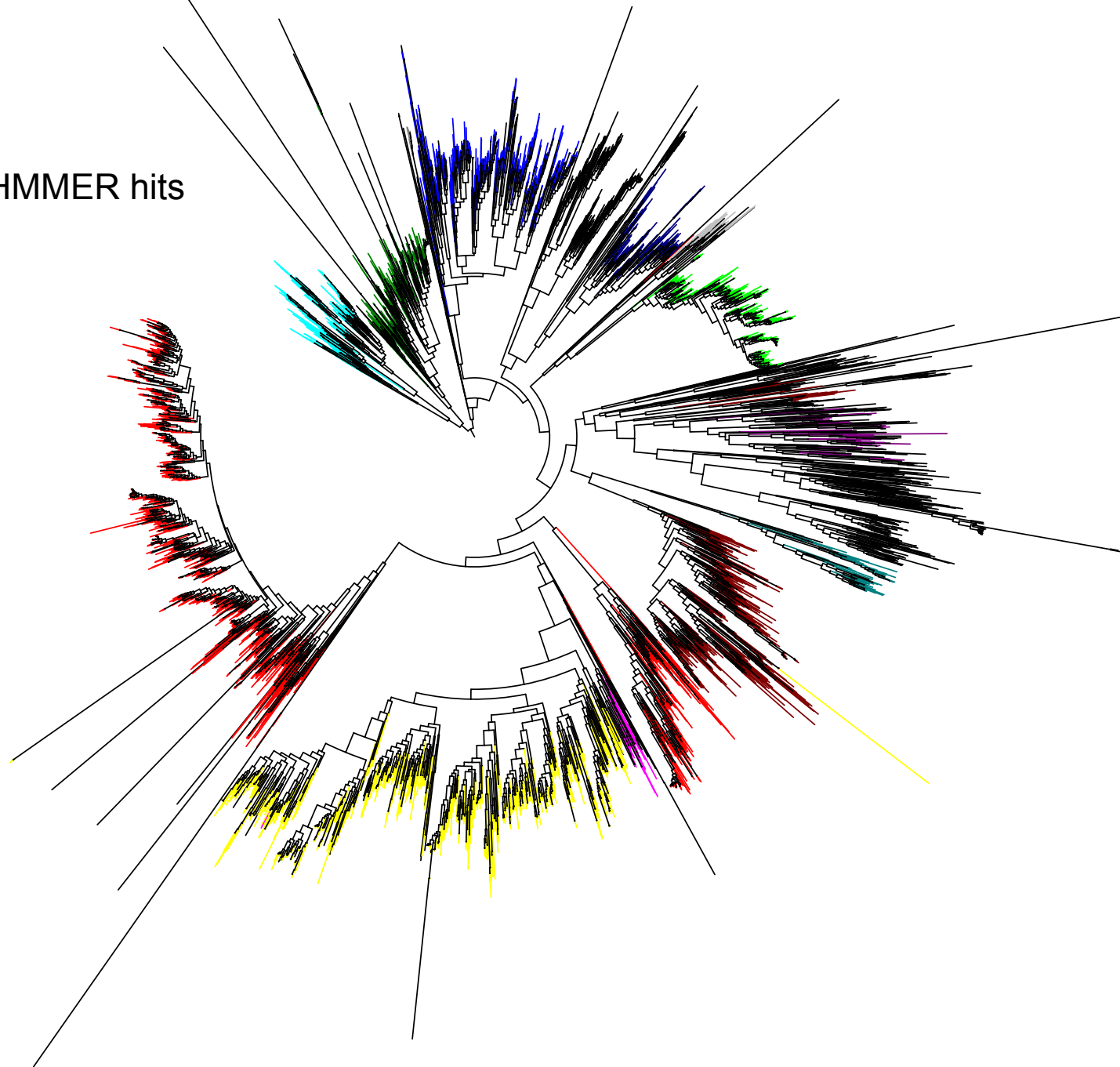


### Kegg Orthology















 glycerol 2-dehydrogenase (NADP+)	 potassium voltage-gated channel Shaker-related subfamily A, beta member 1
 D-arabinose 1-dehydrogenase	 aflatoxin B1 aldehyde reductase
 D-xylose reductase	 aldehyde reductase
 pyridoxine 4-dehydrogenase	 L-glyceraldehyde reductase
 alcohol dehydrogenase (NADP+)	 5-oxoprolinase (ATP-hydrolysing)
 No annotation	

**S5. XYL1 homologs identified in sequence similarity search.** HMMER search for XYL1 homologs identified a large gene family of aldo-keto reductases. Genes were annotated by KEGG orthology via BLASTKoala. Maximum likelihood tree of XYL1 KEGG-annotated hits were built with FastTree. The subclade containing xylose reductase homologs was identified for manual sequence curation (dark tan, XYL1 - K17743).

## XYL2 HMMER hits

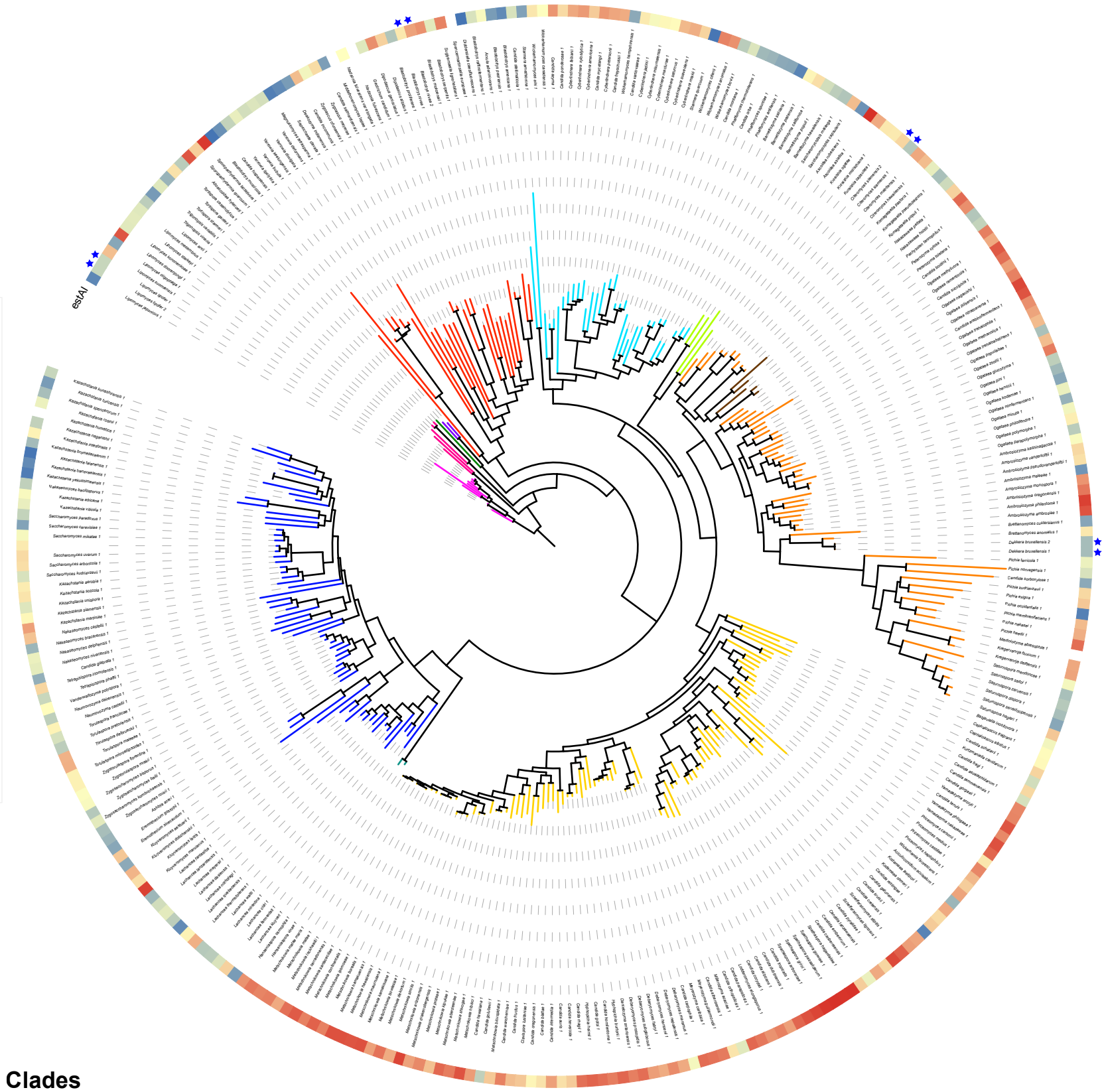
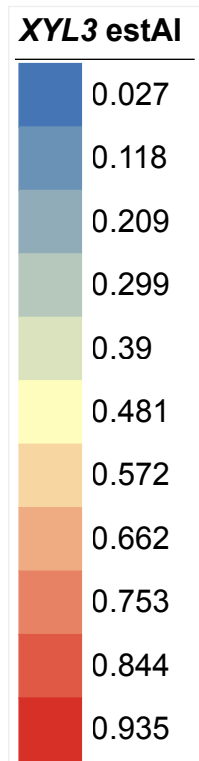


## Kegg Orthology

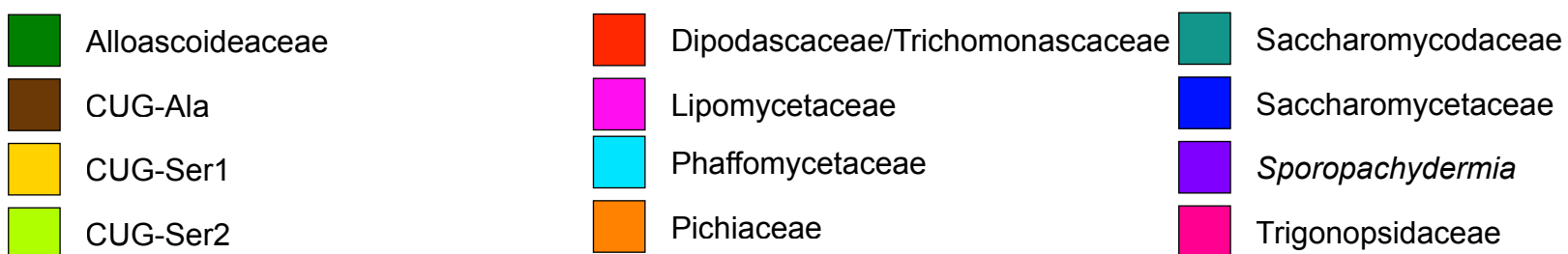
	alcohol dehydrogenase (NADP+)		mitochondrial enoyl-[acyl-carrier protein] reductase / trans-2-enoyl-CoA reductase
	(R,R)-butanediol dehydrogenase / meso-butanediol dehydrogenase / diacetyl reductase		alcohol dehydrogenase, propanol-preferring
	L-iditol 2-dehydrogenase		alcohol dehydrogenase (NADP+)
	aryl-alcohol dehydrogenase		D-arabinitol dehydrogenase (NADP+)
	S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase		alcohol dehydrogenase
	NADPH:quinone reductase		reticulon-4-interacting protein 1, mitochondrial
	D-xylulose reductase		No annotation

**S6. XYL2 homologs identified in sequence similarity search.** HMMER search for XYL2 homologs identified a large gene family of medium-chain dehydrogenases. Genes were annotated by KEGG orthology via BLASTKoala. Maximum likelihood tree of XYL2 KEGG-annotated hits were built with FastTree. The subclade containing xylitol dehydrogenase (xylulose reductase) homologs was identified for manual sequence curation (dark green, XYL2 - K05351).

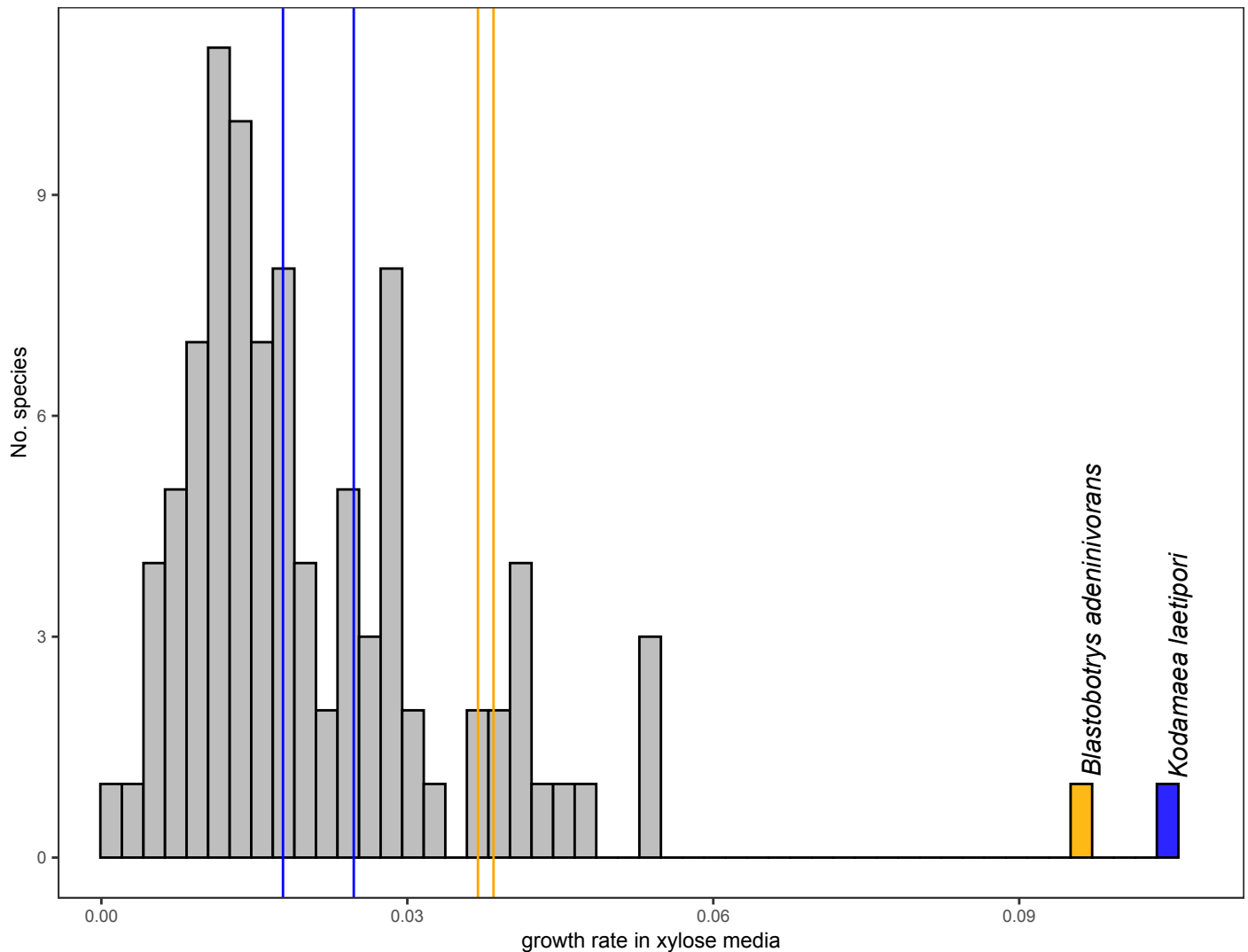




### Clades



**S7. Maximum likelihood tree of Xyl3 protein sequences.** Maximum likelihood phylogeny of Xyl3 protein sequences was built using IQTree with an Auto substitution model based on 1,000 bootstrap replications. Major yeast clades are depicted by branch color using tree topology from Shen et al. (1). Color gradient indicates codon optimization indices (estAI values) with color ranging from blue (lowest estAI value) to red (highest estAI value). Stars indicate paralogs belonging to multi-copy lineages.



**S8. Distribution of xylose growth rates.** Growth rates in medium containing xylose as a sole carbon source were measured for 281 species. *Kodamaea laetipori* (blue) and *Blastobotrys adeninivorans* (orange) both have much higher growth rates than their closest relatives (blue lines - *Candida parapsilosis* and *Candida restingae*, orange lines - *Blastobotrys raffinofermentans* and *Blastobotrys peoriensis*). *Kodamaea laetipori* and *Blastobotrys adeninivorans* were removed as outliers prior to generating phylogenetically independent contrasts for growth rate data.



**Supplemental Table 1.** Pagel's (1994) method was used to fit coevolutionary models to pathway presence and xylose-growth data. The best supported model is interdependence.

<b>Model</b>	<b>Weighted AIC</b>
Independent	$2.6 \times 10^{-5}$
Growth dependent on pathway	0.19
Pathway dependent on growth	0.14
Interdependent	0.67

**Supplemental Table 2.** Pagel's (1994) method was used to fit coevolutionary models to gene copy number for *XYL1* and xylose-growth data. The best supported model is interdependence.

<b>Model</b>	<b>Weighted AIC</b>
Independent	$3.5 \times 10^{-4}$
Growth dependent on <i>XYL1</i> copy number	0.48
<i>XYL1</i> copy number dependent on growth	$2.9 \times 10^{-3}$
Interdependent	0.51

**Supplemental Table 3.** Pagel's (1994) method was used to fit coevolutionary models to gene copy number for *XYL2* and xylose-growth data. The best supported model is interdependence.

<b>Model</b>	<b>Weighted AIC</b>
Independent	0.71
Growth dependent on <i>XYL2</i> copy number	0.17
<i>XYL2</i> copy number dependent on growth	$7.3 \times 10^{-2}$
Interdependent	$5.1 \times 10^{-2}$

**Supplementary Table 4.** Predictors of growth on xylose for all three *XYL* genes calculated with a Bayesian phylogenetic mixed model (MCMCglmm). Codon optimization indices were first scaled to have a mean of 0 and a variance of 1. Significant predictor *XYL3* codon optimization is highlighted. CI=confidence interval. DIC=deviance information criterion. pMCMC is equal to double the probability that the coefficient is above 0 based on the posterior distribution.

<b>Fixed effects</b>	<b>Posterior mean</b>	<b>Lower 95% CI</b>	<b>Upper 95% CI</b>	<b>Sample size</b>	<b>pMCMC</b>
<i>XYL1</i> estAl	24.13	-19.33	66.65	3990	0.226
<i>XYL2</i> estAl	12.65	-26.39	56.58	3776	0.536
<i>XYL3</i> estAl	43.56	-1.18	90.87	3594	0.039
<b>Random effects</b>					
Phylogeny	87326	5821	178832	3408	
<b>Summary statistics</b>					
DIC	6.02				