SUPPLEMENTAL MATERIALS

Transposon insertional mutagenesis in *Saccharomyces uvarum* reveals *trans*-acting effects influencing species-dependent essential genes

Sanchez *et. al*

TABLE OF CONTENTS

FIGURES

Supplemental Fig. S7. Orthologous gene comparisons between species.

Supplemental Fig. S8. Network analysis of conserved essential genes across diverged species.

Supplemental Fig. S9. Confirmatory tetrad analysis for predicted *S. uvarum*-specific essential genes.

Supplemental Fig. S10. Confirmed tetrad analysis for predicted *S. cerevisiae*-specific essential genes.

Supplemental Fig. S11. Confirmatory tetrad analysis in *S. uvarum* for all conserved essential and non-essential genes.

Supplemental Fig. S12. Incorrect gene annotation of *DRE2.*

Supplemental Fig. S13. Differential gene expression alone cannot explain difference in gene dispensability.

TABLES

Supplemental Table S1. List of yeast strains used in this study.

Supplemental Table S2. List of plasmids used in this study.

Supplemental Table S3. List of primers used in this study.

Supplemental Table S4. Summary of sequence read coverage for each library.

Supplemental Table S5. List of *S. uvarum* coding coordinates containing insertion counts from each library.

Supplemental Table S6. GO analysis on essential genes.

Supplemental Table S7. GO analysis on non-essential genes.

Supplemental Table S8. List of orthologous coding coordinates containing insertion counts from each library.

Supplemental Table S9. List of genes that are predicted to differ in dispensability between species.

Supplemental Table S10. List of GO enrichment of molecular function terms for each species-specific gene.

Supplemental Table S11. Summary of library coverage.

Supplemental Table S12. Summary of validations.

CODE

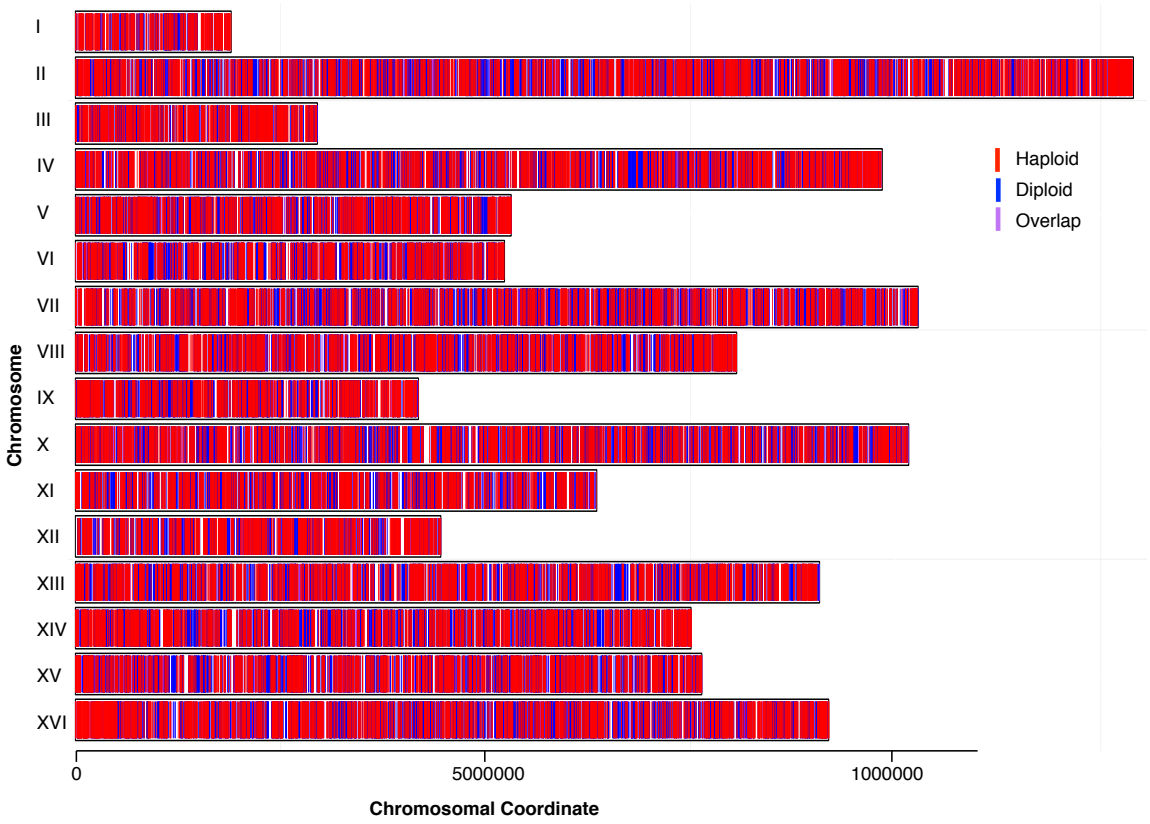
Supplemental_Code.rb. Script used to call insertion positions

Generating Tn7 insertional libraries in *S. uvarum*

The plasmid library containing ligated *S. uvarum* genomic DNA with Tn7 insertion sites contains ~50,000 unique genomic insertion sites that were integrated into a diploid and a haploid *MAT***a** strain (Figure **1B)**. We isolated mutants with the ClonNat resistance marker on plates and created a final diploid pool of ~500,000 (10X coverage of plasmid library) transformants (**Figure 1B)**. The haploid pool was obtained from the Caudy lab containing ~300,000 transformants. We added an additional 200,000 transformants to the haploid pool to increase the total coverage.

Distribution of insertion sites across the *S. uvarum* genome

We first sequenced the plasmid pool of *S. uvarum* mutants to determine the genome coverage of genes containing insertion sites in the original plasmid library. We sequenced the library by extracting plasmid DNA and enriching for fragments of DNA containing the Tn7 sequence. Primers were designed to target the Tn7 sequence and the Illumina linker sequences that were ligated to randomly sheared plasmid DNA (**Figure 1C)**. The PCR amplicons of the Tn7 library were sequenced, trimmed, mapped and processed through an in-house Ruby script to determine the position of the insertion site (Material and Methods). Insertion sites with fewer than 10 reads were filtered out. Detailed information about overall sequencing coverage is listed in **Supplemental Table S4**. We used this same method to determine the insertion sites in the haploid and diploid library by extracting genomic DNA from the pooled libraries. Haploid, diploid and overlapping insertion sites are evenly distributed throughout the *S. uvarum* genome, as illustrated in **Supplemental Fig. S1.**

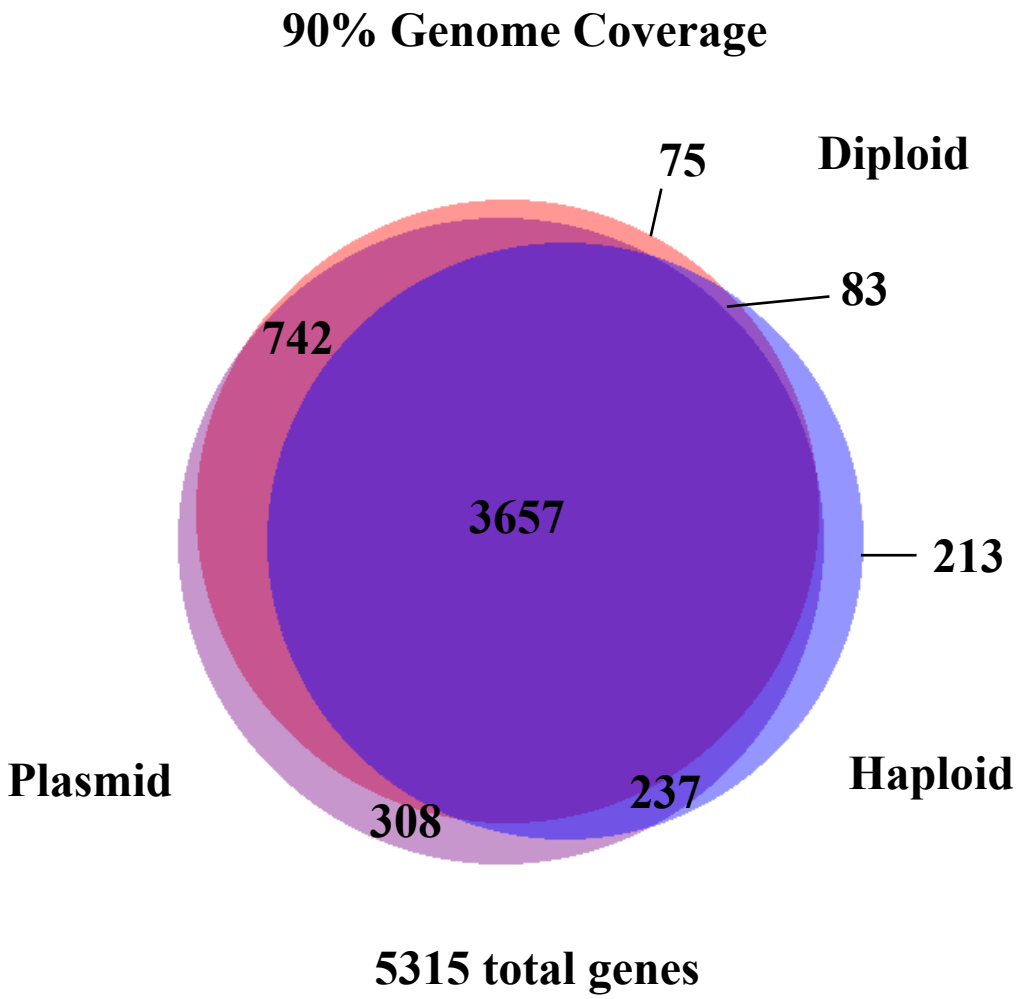


Supplemental Fig. S1. Distribution of haploid and diploid Tn7 insertions across the *S. uvarum* genome. Chromosomal map of sequenced transposon insertions wherein each chromosome is represented by a rectangle outlined in black and numbered one through sixteen from top to bottom. Haploid-only insertion sites are colored red, diploid-only inserts are blue. Insertion sites that were identified in both libraries are indicated in purple.

Once the insertion sites were determined in all three libraries, we counted the number of insertion sites in each annotated open reading frame using an in-house Python script (Materials and Methods). **Supplemental Table S11** summarizes the number of insertion sites and the number of genes that contain insertion sites within the plasmid, haploid and diploid libraries. All annotated *S. uvarum* genes containing the number of insertion sites from each library are fully listed in **Supplemental Table S5.** Of the 5,908 annotated genes, a total of 5,315 (90%) genes harbor insertion sites that were identified in at least one library. Comparisons between shared genes and unique genes with insertion sites are illustrated in **Supplemental Fig. S2.** The number of genes with insertion sites shared amongst all three libraries was 3,657 (69%) of the 5,315 genes summed across the libraries. There are subsets of genes that are library specific or shared between two libraries, likely due to differences in overall transformation coverage per library. There is a subset of 742 genes, however, that is shared only between the diploid and plasmid libraries (14%). Overall, only 3,933 genes harboring insertion sites were determined in the haploid pool, suggesting that at least some of the genes falling out of the haploid pool are likely to be essential based on their dispensability restrictions. We went on to test this assumption using the known essential gene set in *S. cerevisiae.*

Supplemental Table S11. Summary of library coverage.

Library Type	Number of inserts > 10 reads	Number of inserts in ORFs	Number of genes with insert	% Genome covered	Number of orthologs with inserts	% Orthologs with inserts
Plasmid	54,351	33,394 (61.4%)	4,944	83%	4,630	85%
Diploid	46,326	27,121 (58.5%)	4,557	77%	4,283	79%
Haploid	42,904	22,988 (54.5%)	4,190	71%	3,933	72%



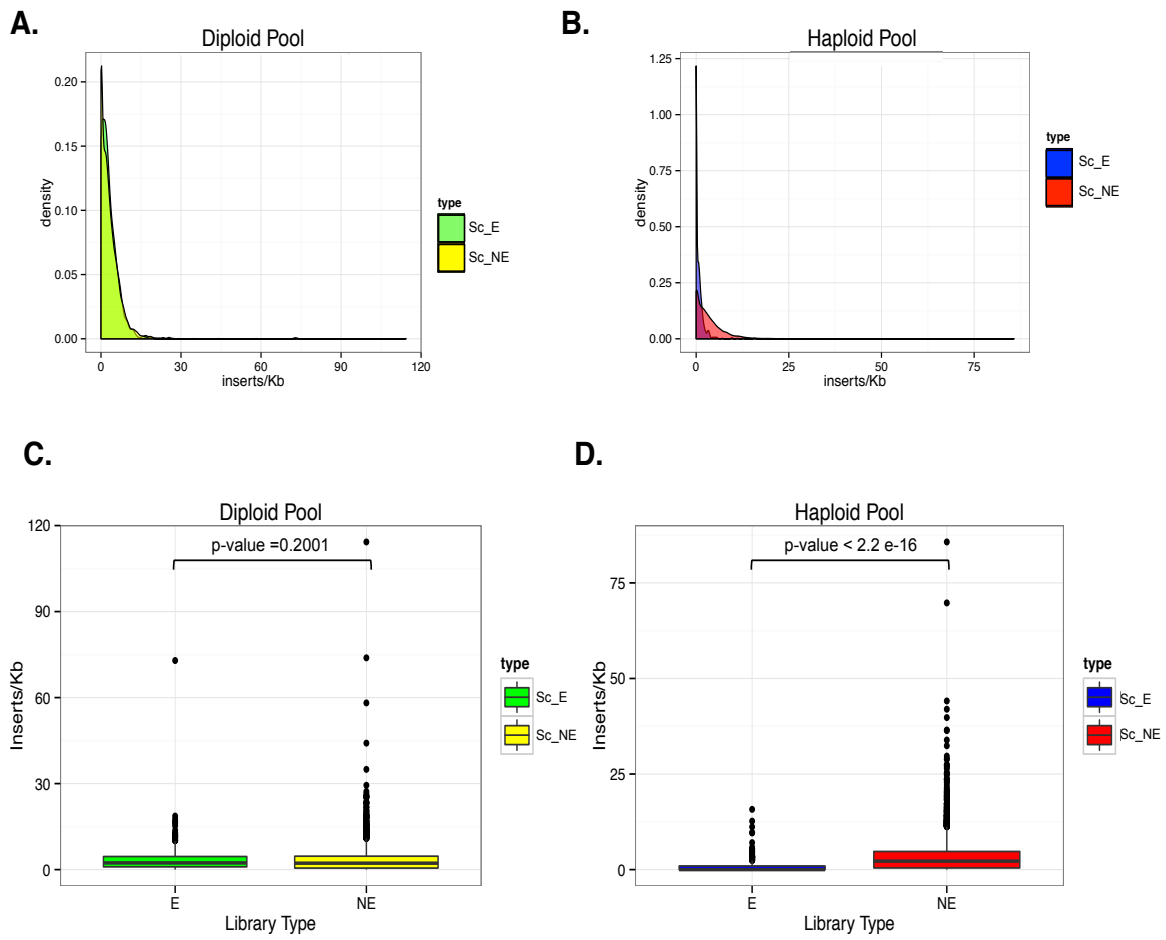
Supplemental Fig. 2. Proportional Venn diagram summarizing the number of insert-containing genes identified in each library. Summary of genes containing at least one insert sampled by 10 or more sequencing reads. The plasmid library is represented in purple, the diploid pool in orange and the haploid pool in blue. Non-overlapping regions represent genes that are library specific.

Known *S. cerevisiae* essential genes contain fewer inserts than known non-essential genes

Although the large number of dropouts suggest that we have identified essential genes, it is possible that these dropouts just reflect an incomplete library. To test if genes in the haploid library without (or with minimal) insertion sites were likely to be essential, we compared the haploid and diploid libraries to determine if differences in the number of insertions between known essential and non-essential genes exist. Since the essentiality of most genes is expected to be conserved between *S. cerevisiae* and *S. uvarum,* we used the known essential set in *S. cerevisiae* to test if orthologous genes in the haploid library contain fewer insertion sites. We normalized the number of inserts within each gene by the length of the gene (inserts/kb) and plotted the distribution of normalized inserts within known essential genes and non-essential genes in both diploid and haploid libraries (**Supplemental Fig. S3A & B)**.

The distribution between known *S. cerevisiae* essential and non-essential genes is similar in the diploid pool, with no significant differences between gene types (essential average inserts/kb=3.8 vs non-essential average inserts/kb = 4.2) (**Supplemental Fig. S3C**, Wilcoxon test p-value = 0.2001). However, the distributions in known *S. cerevisiae* essential and non-essential genes in the haploid library are significantly different, with known essential genes averaging fewer normalized inserts per kb (known essential average inserts/kb=0.88 vs non-essential average inserts/kb = 4) (**Supplemental Fig. S3D**, Wilcoxon test p-value $<2.2 \times 10^{-16}$). This result suggests that the known conserved essential genes can be predicted from the number of inserts in the haploid library. We note that known *S. cerevisiae* essential genes in the haploid library harboring several insertion sites are detected as well. We predict that these genes are candidate *S.*

cerevisiae-specific essential genes, and may not be essential in *S. uvarum*. We explore these genes more fully in the following sections.



Supplemental Fig. S3. Comparison of insertion distributions between haploid and diploid libraries amongst known *S. cerevisiae* essential and non-essential genes. A) Density plot displaying the distribution of the normalized number of insertion sites in known *S. cerevisiae* essential and non-essential genes within the *S. uvarum* diploid and B) haploid libraries. C) Box plot of normalized insertions per kb in known *S. cerevisiae* essential and non-essential genes in the diploid library. No significant difference was detected between known essential and non-essential genes (Wilcoxon test p = 0.2001). Essential average inserts/kb = 3.8 (SD = 3.98) vs. non-essential average inserts/kb = 4.2 (SD = 4.2). D) Box plot of normalized insertions per kb in known *S. cerevisiae* essential and non-essential genes in the haploid library displaying a significant difference between known essential and non-essential genes (Wilcoxon test $p < 2.2 \times 10^{-16}$). Essential average inserts/kb = 0.88 (SD = 1.28) vs. non-essential average inserts/kb = 4 (SD = 4.38).

Predicting *S. uvarum* essential and non-essential genes using an insertion ratio metric

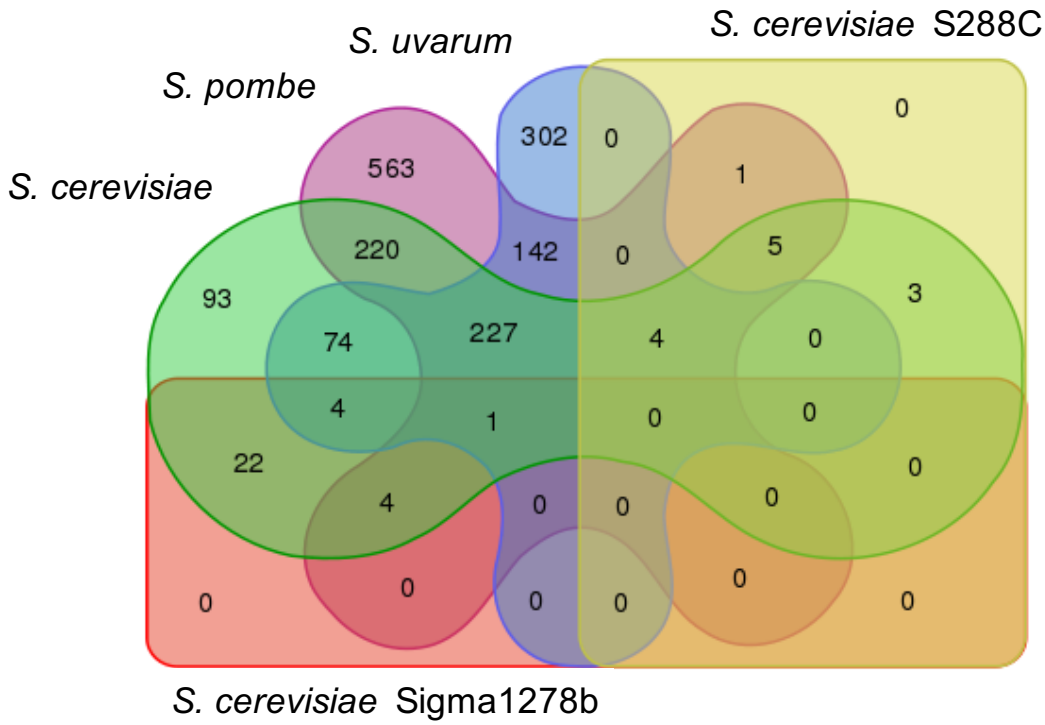
Once we determined that essential genes contain significantly fewer insertion sites than non-essential genes in the haploid library, we created a metric for determining a cut off value to categorize predicted *S. uvarum* essential and non-essential genes. Due to the nature of the library, insertional events at different positions across a gene may result in a partial loss of function, meaning that even essential genes may still tolerate some insertions. Therefore, we relied on comparisons between the diploid and haploid libraries to make inferences about gene essentiality. Specifically, we calculated an insertion ratio using the number of inserts per gene in the haploid library divided by the number of inserts in the diploid library, which inherently normalizes for the length of the gene (Materials and Methods). Using the insertion ratio as a metric, we tested if significant differences exist between *S. uvarum* genes whose orthologs are known to be essential and non-essential in *S. cerevisiae,* as well as *S. uvarum* intergenic regions. Intergenic regions between convergent orientated genes are expected to not be essential, thus, the distribution of intergenic regions is expected to be similar to that of non-essential genes and represents our null distribution.

We note the significant difference that also exists between non-essential genes and intergenic regions and attribute this difference to the possible genes that are differentially essential between species in this category in comparison to intergenic regions in *S. uvarum.* We note these differences are of a lesser magnitude than those that exist between known essential genes. Once we established the intergenic region as our null distribution, we ranked the insertion ratio value for each gene against the

intergenic distribution and determined the proportion of intergenic regions whose insertion ratio was greater than the insertion ratio of that gene. Using this raking metric, we set a cut-off value of 0.25 to categorize all annotated *S. uvarum* genes into essential and non-essential categories. We chose this cut off metric to be more inclusive than restrictive to be more likely to identify essential genes that are species specific, despite the likelihood of increasing our false positive rate. A list of all genes with their predicted classification can be found in **Supplemental Table S7.**

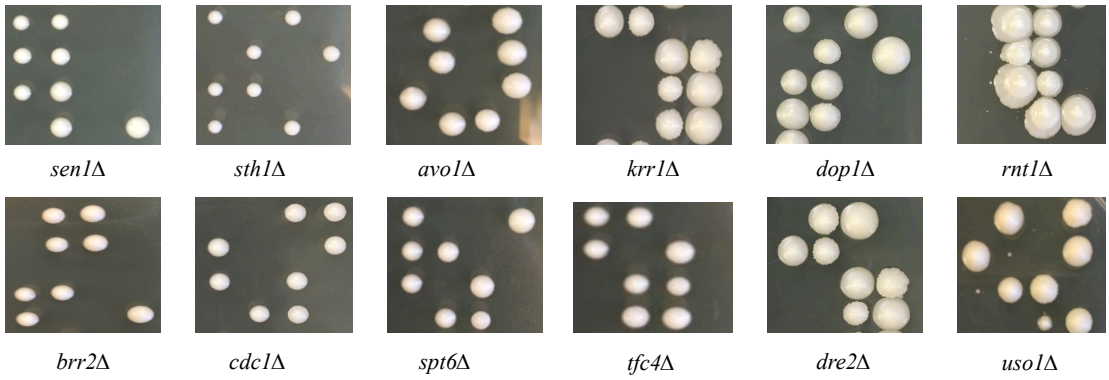
Supplemental Table S12. Validation summary.

Predicted gene type	Number of genes tested	Number of genes confirmed	Number of false positives	% Correct
ScE_SuE	13	12	1	93%
ScNE_SuNE	3	3	0	100%
ScNE_SuE	28	10	19	32%
ScE_SuNE	24	12	12	50%

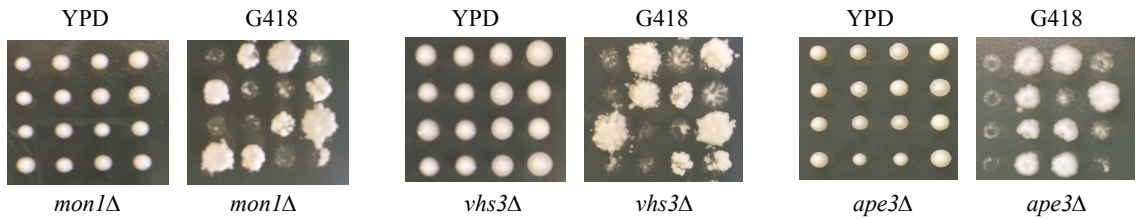


Supplemental Fig. S4. Conservation comparison between *S. cerevisiae*, *S. pombe* and predicted *S. uvarum* essential genes.

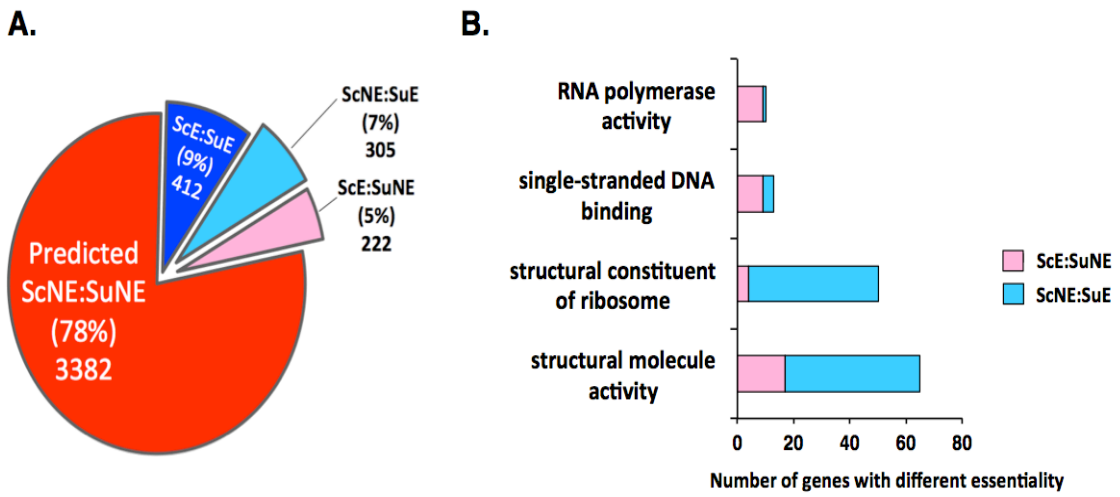
Conserved Essential Genes: ScE:SuE



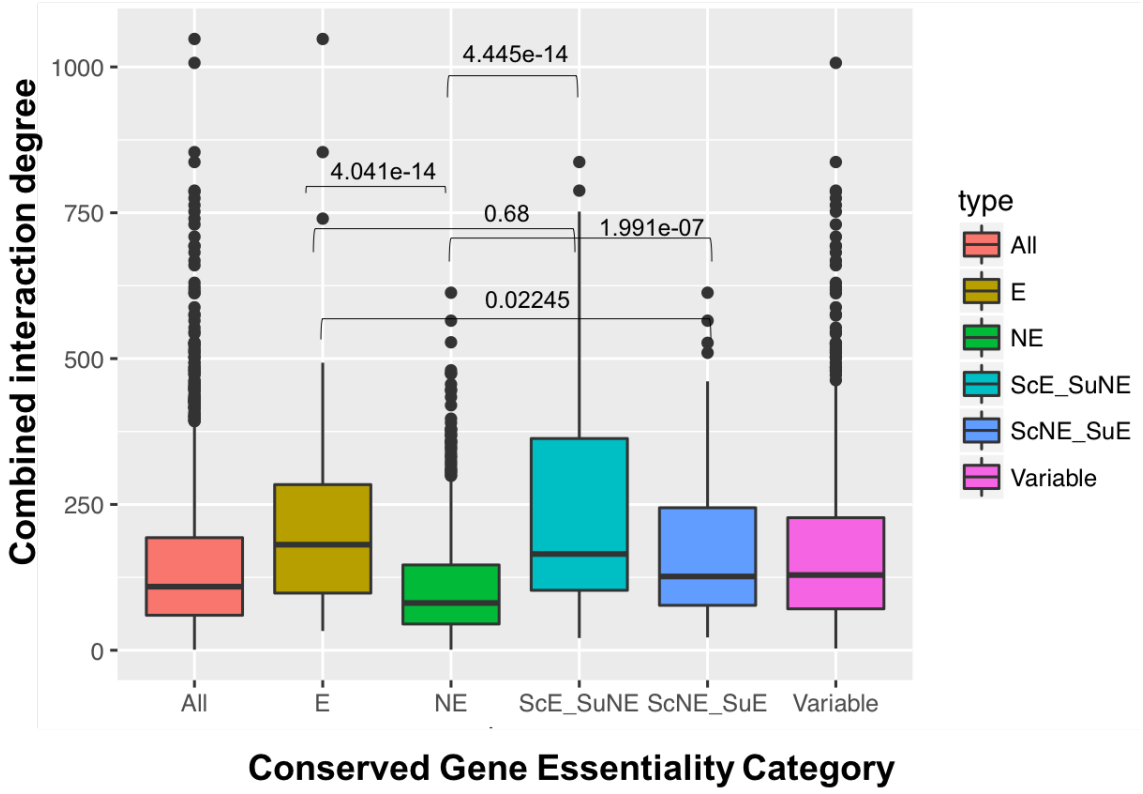
Supplemental Fig. S5. Confirmatory tetrad analysis for conserved, predicted essential genes.



Supplemental Fig. S6. Confirmatory tetrad analysis for conserved, predicted non-essential genes.

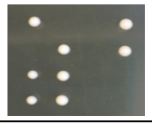
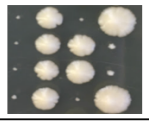
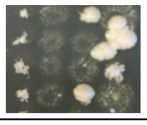
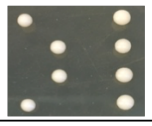
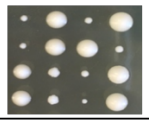
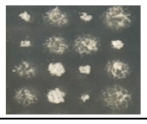
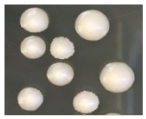
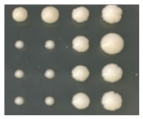
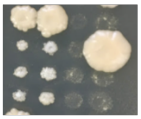
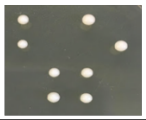
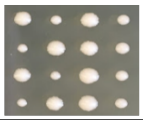
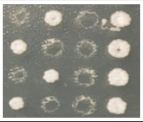
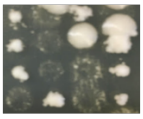
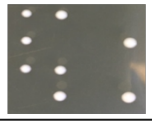
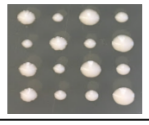
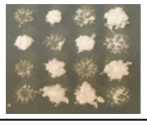
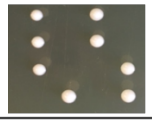
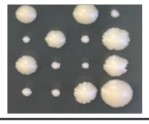
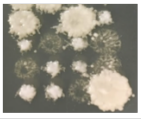
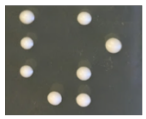
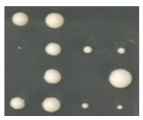
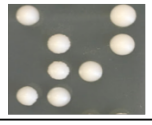
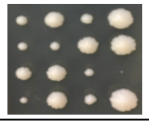
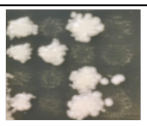
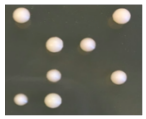
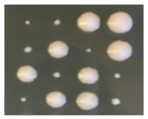
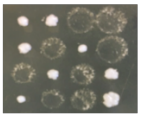


Supplemental Fig. S7. Orthologous gene comparisons between species. A) Essentiality comparisons between 4,321 ortholog pairs between *S. cerevisiae* and *S. uvarum*. A total of 85% orthologs display conserved essentiality, with 12% showing differences in gene dispensability (NE = non-essential, E = essential). **B)** Functional enrichment of orthologs with differential essentiality. Gene Ontology (GO) enrichment was performed on genes that differ in essentiality and a subset of biological functions are represented (**Supplemental Table S10** for complete list). Each color indicates the proportion of total annotated genes categorized for each function, split by the number of genes represented by essential genes in each species. The number of predicted *S. cerevisiae* essential genes in each functional category is indicated in pink and predicted *S. uvarum* essential genes in light blue, where one out of six genes tested for the functional category of the structural constituent of the ribosome was validated (*RSM22).*



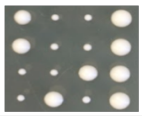
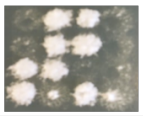
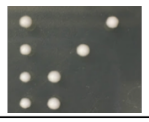
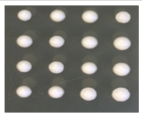
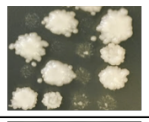
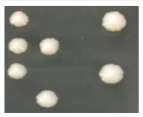
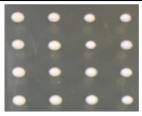
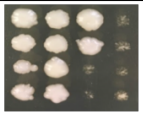
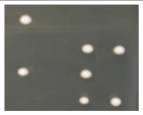
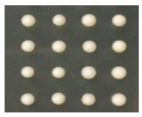
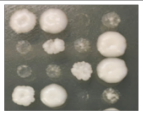
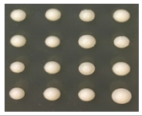
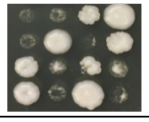
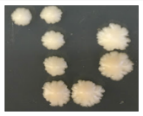
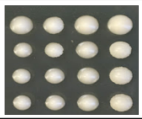
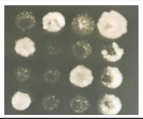
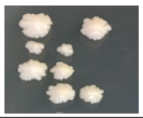
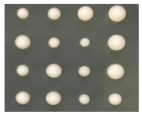
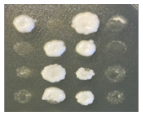
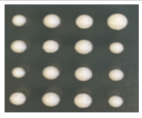
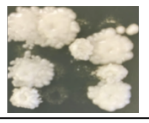
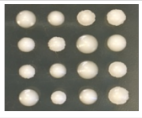
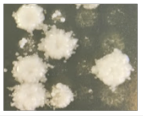
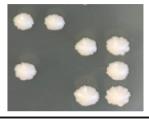
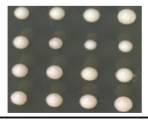
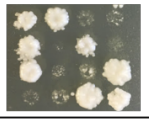
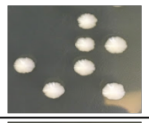
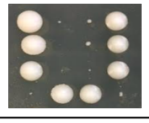
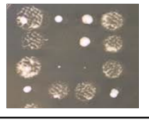
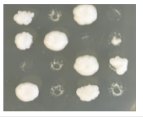
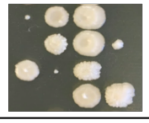
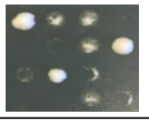
Supplemental Fig. S8. Network analysis of conserved essential genes across diverged species.

S. uvarum Specific Essential Genes:

Gene	*S. uvarum* Tetrads YPD	*S. cerevisiae* Tetrads YPD	*S. cerevisiae* Tetrads G418
kap104Δ			
ssq1Δ			
tup1Δ			
aro7Δ			
mdm10Δ			
sac3Δ			
ccm1Δ			
vma5Δ			
aft1Δ			
rsm22Δ			

Supplemental Fig. S9. Confirmatory tetrad analysis for predicted *S. uvarum*-specific essential genes.

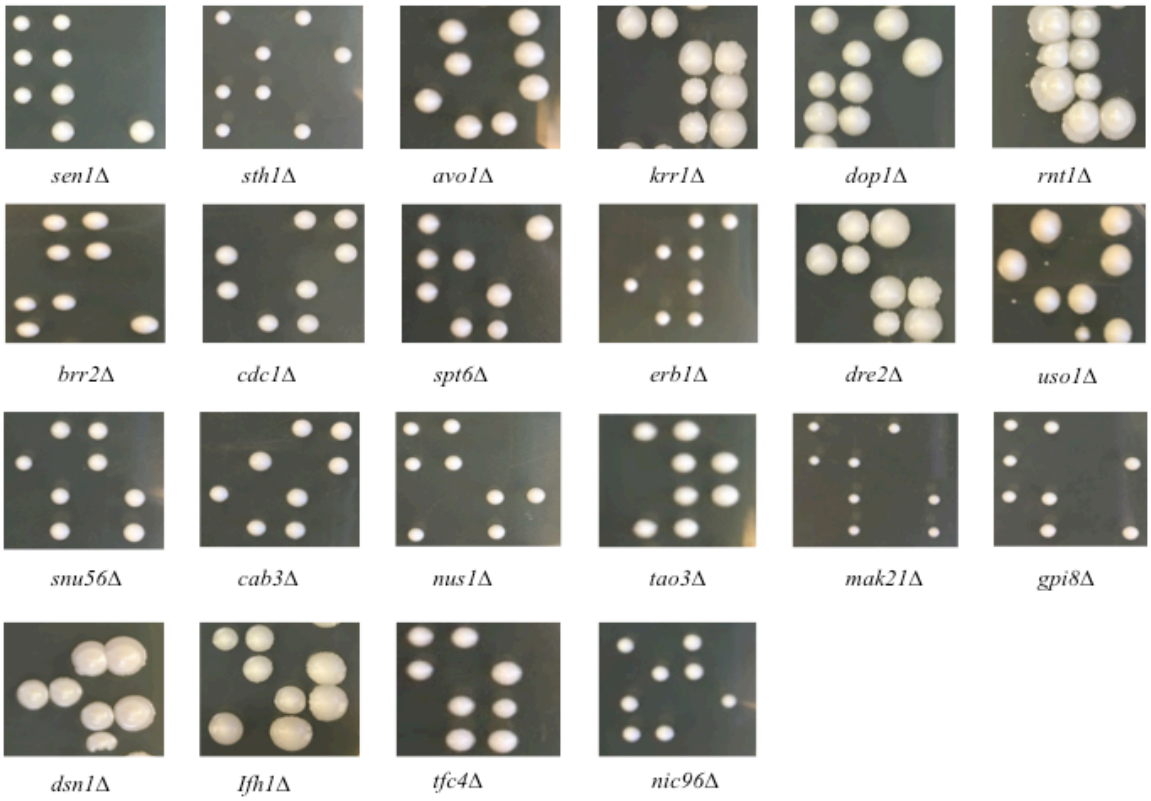
S. cerevisiae Specific Essential Genes:

Gene	*S. cerevisiae* Tetrads YPD	*S. uvarum* Tetrads YPD	*S. uvarum* Tetrads G418
cdc25Δ			
sec24Δ			
mcm10Δ			
alr1Δ ch15			
alr1Δ ch7	NA		
inn1Δ			
vtc4Δ			
shr3Δ			
tfc3Δ			
myo2Δ			
net1Δ			
mdm1Δ			
lcd1Δ			

1

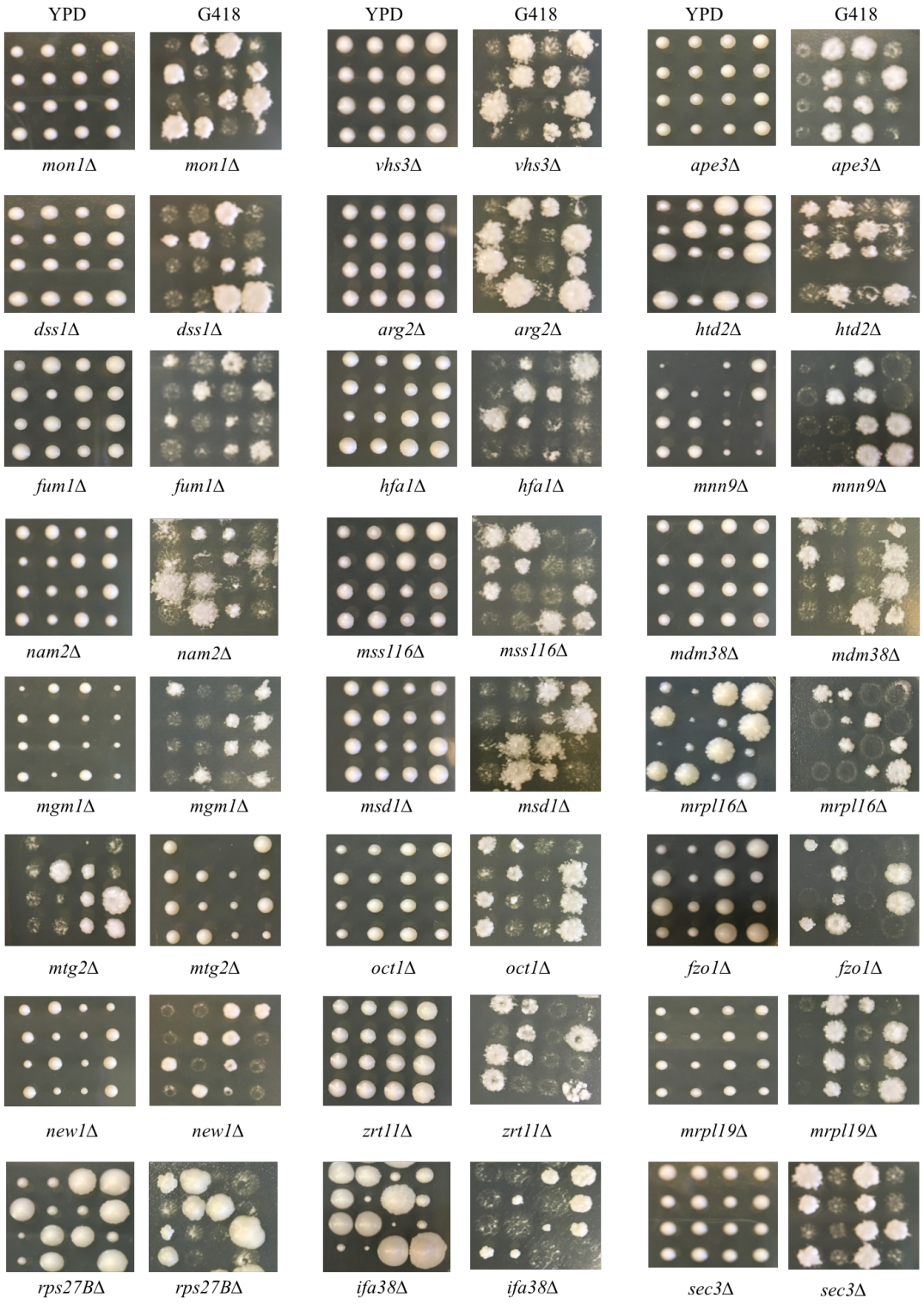
Supplemental Fig. S10. Confirmed tetrad analysis for predicted *S. cerevisiae*-specific essential genes.

Conserved Essential Genes: ScE:SuE

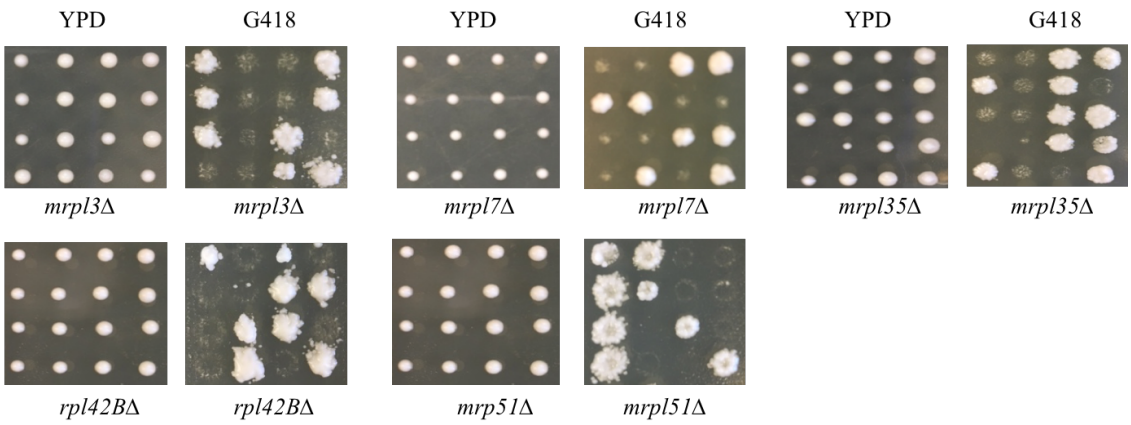


(Figure continued on next 2 pages.)

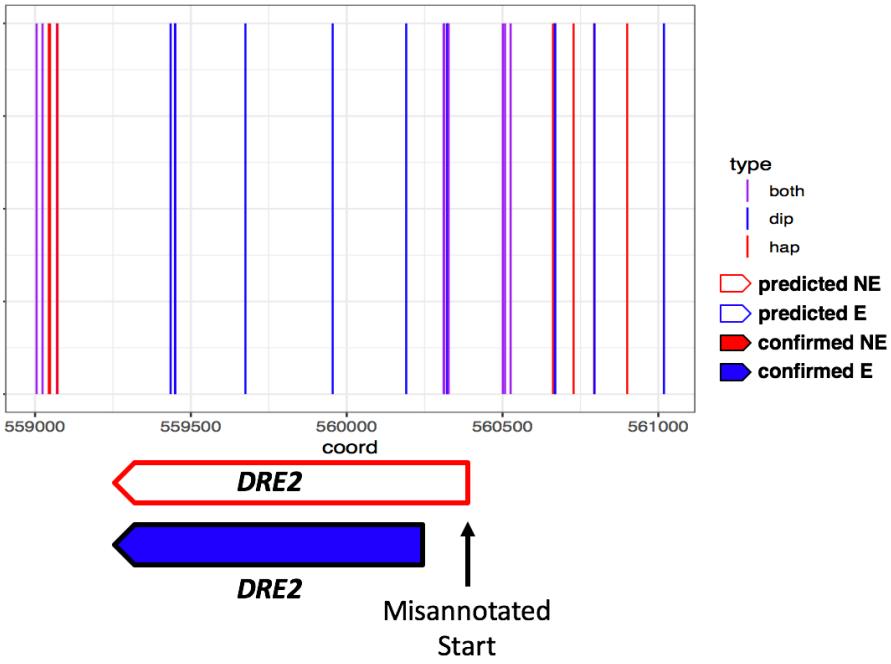
Conserved Nonessential Genes: ScNE:SuNE



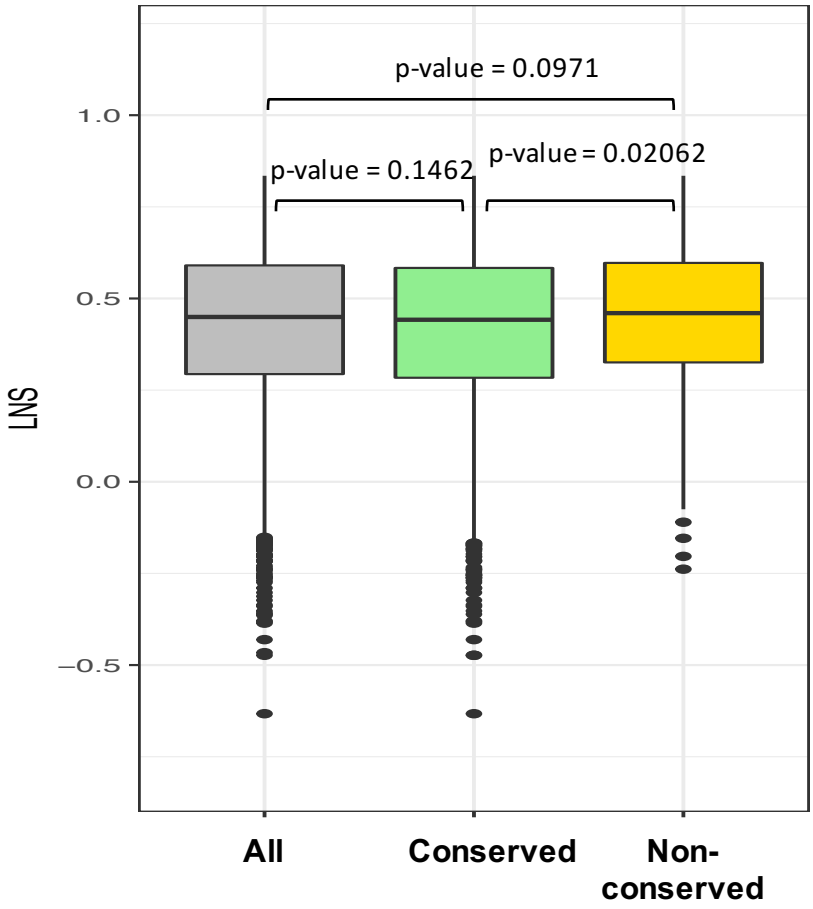
1



Supplemental Fig. S11. Confirmatory tetrad analysis in *S. uvarum* for all conserved essential and non-essential genes.



Supplemental Fig. S12. Incorrect gene annotation of *DRE2*.



Supplemental Fig. S13. Differential gene expression alone cannot explain difference in gene dispensability.