

¹Supplementary Material

²Supporting Materials for *Unlocking a signal of introgression from codons in Lachancea kluyveri using a mutation-selection model* by Landerer et al..

³**Table 1** Synonymous mutation codon preference based on our estimates of ΔM . Shown are the
⁴most likely codon in low expression genes for each amino acid in: *E. gossypii*, in the endogenous
⁵and exogenous genes of *L. kluyveri*, and in the combined *L. kluyveri* genome without accounting
for the two cellular environments.

Amino Acid	<i>E. gossypii</i>	Endogenous	Exogenous	Combined
Ala A	GCG	GCA	GCG	GCG
Cys C	TGC	TGT	TGC	TGC
Asp D	GAC	GAT	GAC	GAC
Glu E	GAG	GAA	GAG	GAG
Phe F	TTC	TTT	TTT	TTT
Gly G	GGC	GGT	GGC	GGC
His H	CAC	CAT	CAC	CAC
Ile I	ATC	ATT	ATC	ATA
Lys K	AAG	AAA	AAG	AAA
Leu L	CTG	TTG	CTG	CTG
Asn N	AAC	AAT	AAC	AAT
Pro P	CCG	CCA	CCG	CCG
Gln Q	CAG	CAA	CAG	CAG
Arg R	CGC	AGA	AGG	CGG
Ser ₄ S	TCG	TCT	TCG	TCG
Thr T	ACG	ACA	ACG	ACG
Val V	GTG	GTT	GTG	GTG
Tyr Y	TAC	TAT	TAC	TAC
Ser ₂ Z	AGC	AGT	AGC	AGC

Table 2 Synonymous selection codon preference based on our estimates of $\Delta\eta$. Shown are the most likely codon in high expression genes for each amino acid in: *E. gossypii*, in the endogenous and exogenous genes of *L. kluyveri*, and in the combined *L. kluyveri* genome without accounting for the two cellular environments.

Amino Acid	<i>E. gossypii</i>	Endogenous	Exogenous	Combined
Ala A	GCT	GCT	GCT	GCT
Cys C	TGT	TGT	TGT	TGT
Asp D	GAT	GAC	GAT	GAT
Glu E	GAA	GAA	GAA	GAA
Phe F	TTT	TTC	TTC	TTC
Gly G	GGA	GGT	GGT	GGT
His H	CAT	CAC	CAT	CAT
Ile I	ATA	ATC	ATT	ATT
Lys K	AAA	AAG	AAA	AAG
Leu L	TTA	TTG	TTG	TTG
Asn N	AAT	AAC	AAT	AAC
Pro P	CCA	CCA	CCT	CCA
Gln Q	CAA	CAA	CAA	CAA
Arg R	AGA	AGA	AGA	AGA
Ser ₄ S	TCA	TCC	TCT	TCT
Thr T	ACT	ACC	ACT	ACT
Val V	GTT	GTC	GTT	GTT
Tyr Y	TAT	TAC	TAT	TAC
Ser ₂ Z	AGT	AGT	AGT	AGT

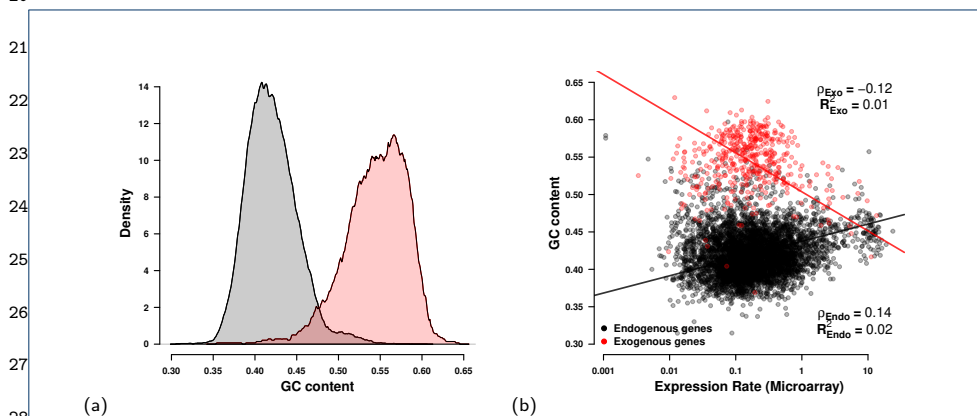


Figure 1 Endogenous and exogenous genes have distinct GC content. (a) Distribution of GC content content in the endogenous and exogenous genes. (b) Correlation of endogenous and exogenous GC content with measured gene expression. While the endogenous GC content shows a slight positive correlation with gene expression ($\rho = 0.14$, $p = 1.2 \times 10^{-21}$), the exogenous GC content is negatively correlated with gene expression ($\rho = -0.12$, $p = 0.014$).

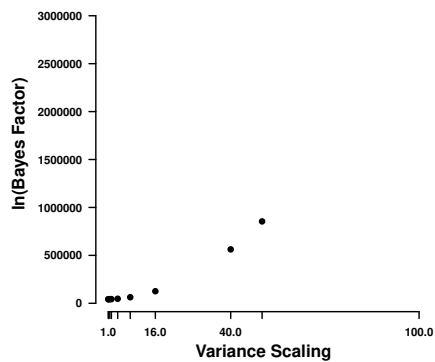
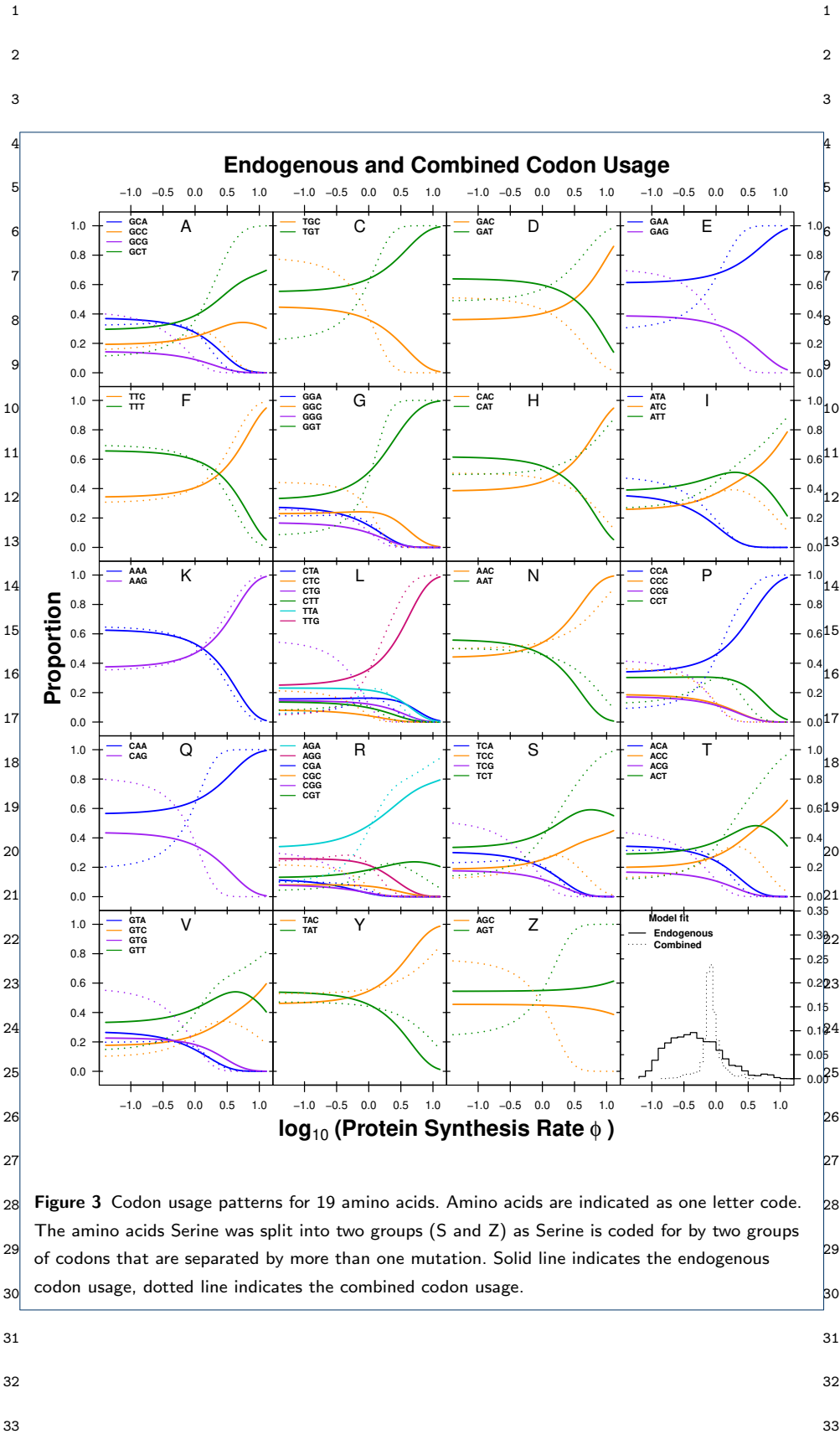
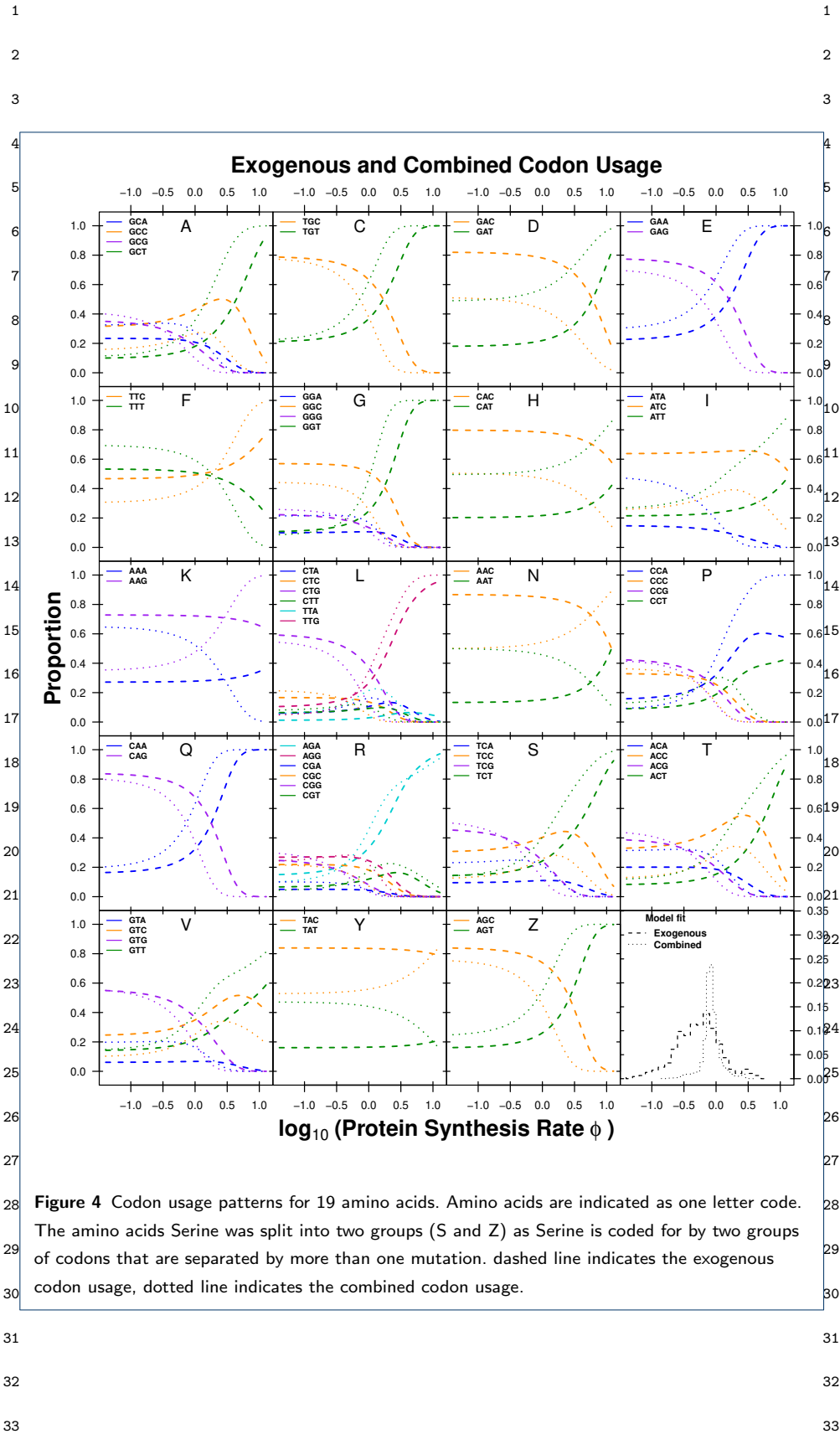


Figure 2 Influence of the variance scaling of the importance distribution on the estimated Bayes factor.





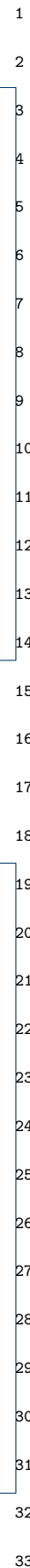
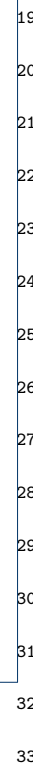


Figure 6 Correlation coefficients of ΔM and $\Delta \eta$ of the endogenous genes with 332 examined budding yeast lineages. Dots indicate the correlation of ΔM and $\Delta \eta$ of the lineages with the exogenous parameter estimates. Blue triangles indicate the *Lachancea* and red diamonds indicate *Eremothecium* lineages. All regressions were performed using a type II regression [?].



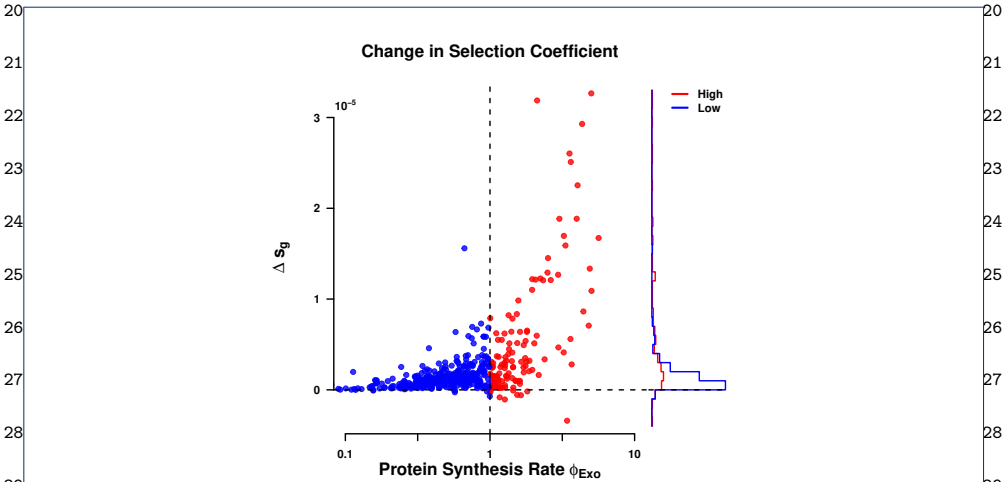
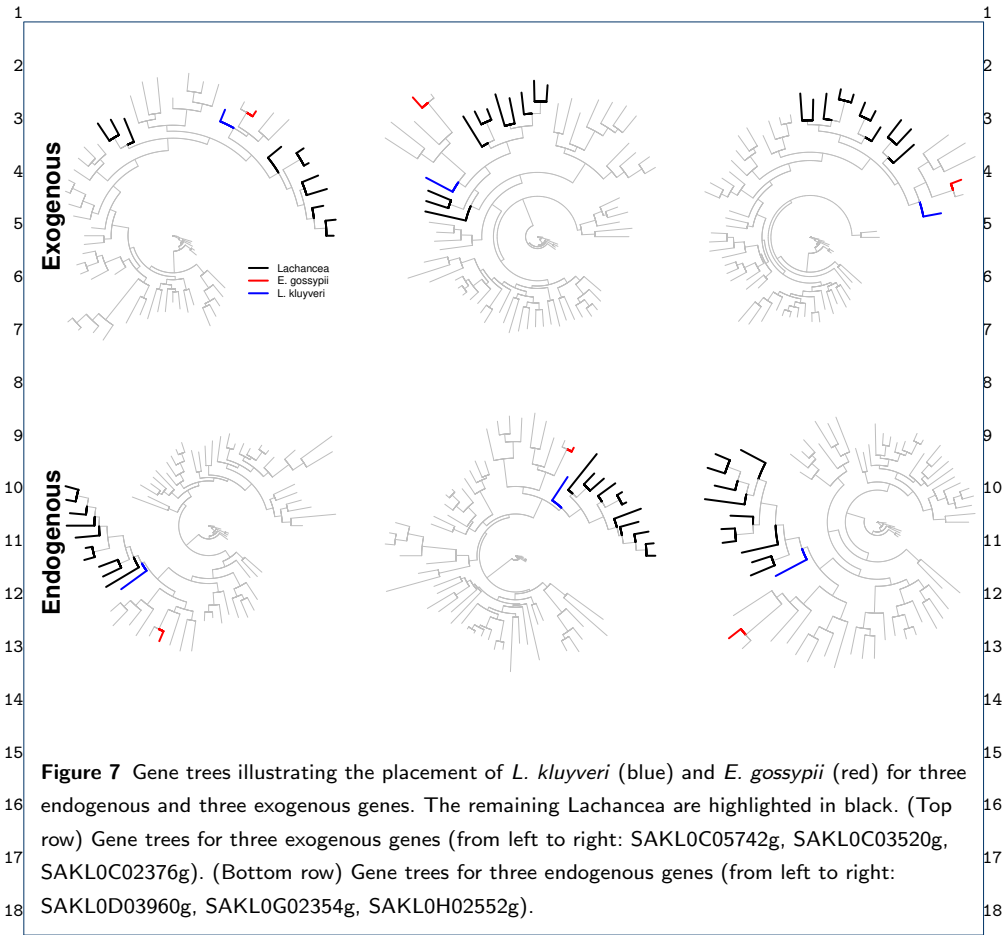


Figure 8 Total amount of adaptation estimated to have occurred between time of introgression and currently observed per gene. Vertical dashed line indicates split between high and low expression genes at $\phi = 1$. Horizontal dashed line indicates no change in selection against mismatched codon usage.

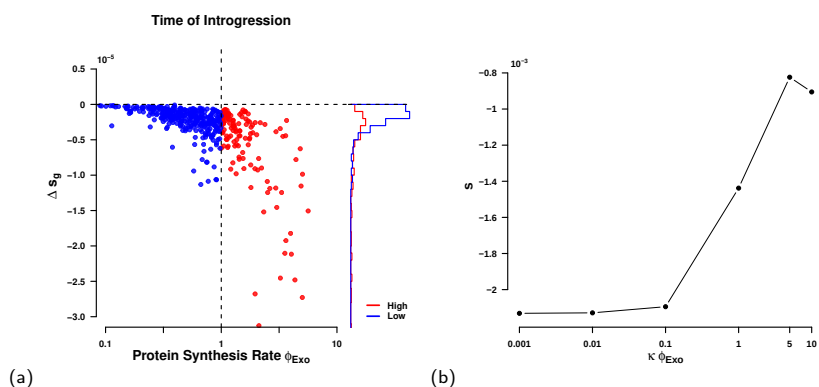


Figure 9 Selection against mismatched codon usage (a) without scaling of ϕ per gene. Vertical dashed line indicates split between high and low expression genes at $\phi = 1$. Horizontal dashed line indicates neutrality. (b) Change of total selection against mismatched codon usage with scaling term κ between *E. gossypii* and *L. kluyveri*