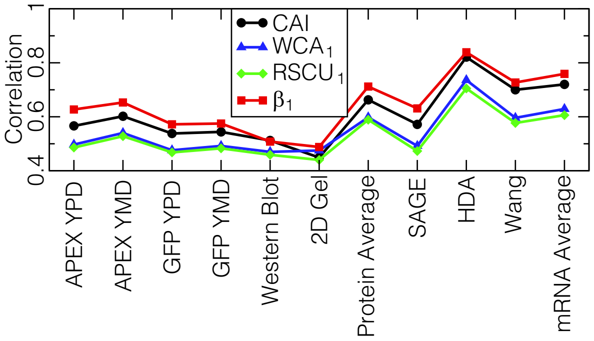

FIG. S1: Pearson correlations between either β_1 from the 4-trend model found by SCUMBLE, the first axis from WCA or CA/RSCU, or the gene's CAI, and various measurements of gene expression levels, for named genes in *S. cerevisiae*. The first seven measurements are of protein levels, while the last four are of mRNA concentrations.

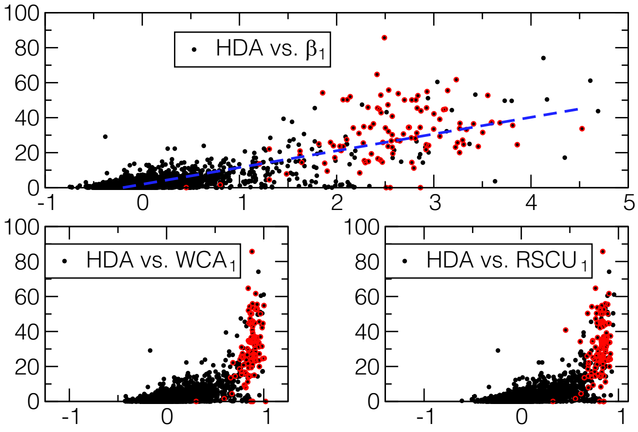

FIG. S2: Single-channel DNA microarray measurements of *S. cerevisiae* mRNA concentrations plotted against the first axis of (a) SCUMBLE, (b) WCA, and (c) CA/RSCU. The expression level seems to be about linearly related to β_1—dashed blue line—while WCA$_1$ and RSCU$_1$ saturate at around 1. Genes for ribosomal proteins are circled in red.

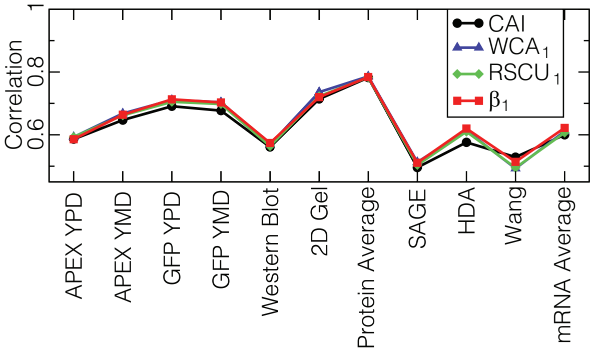

FIG. S3: Spearman correlations between either β_1 from the 4-trend model found by SCUMBLE, the first axis from WCA or CA/RSCU, or the gene's CAI, and various measurements of gene expression levels, for named genes in *S. cerevisiae*. The first seven measurements are of protein levels, while the last four are of mRNA concentrations.

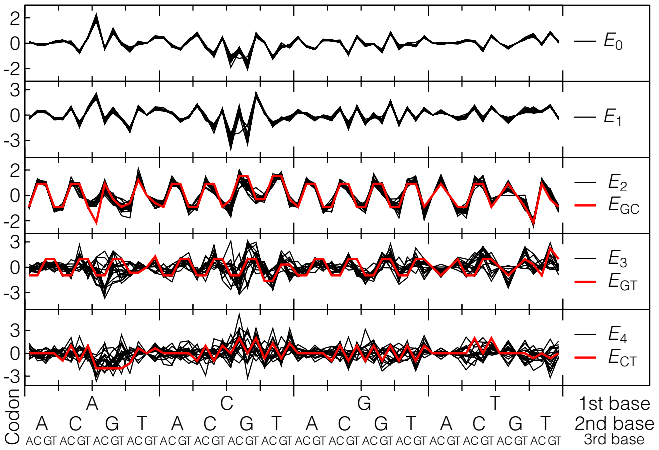

FIG. S4: Preference functions of the 4-trend models identified by applying SCUMBLE to each of the 16 chromosomes in *S. cerevisiae*. The solid red lines show the ideal preferences corresponding to GC, GT and CT bias.

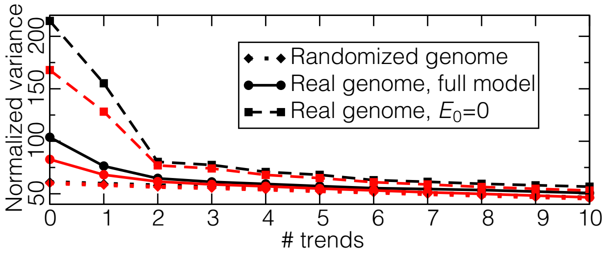

FIG. S5: Average (black) and median (red) normalized variance for named genes in *S. cerevisiae*, for models with up to 10 trends. Full models (circles, solid lines) and models with E_0 set to zero (squares, dashed lines) are compared to results for a randomized genome (diamonds, dotted lines).

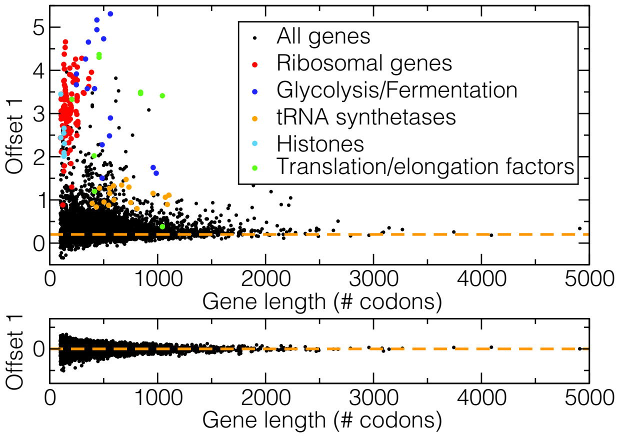

FIG. S6: Top: β_1 from the 4-trend model of *S. cerevisiae* with the constant offset E_0 set to zero, plotted against the number of codons in a gene. The dashed orange line indicates the likely level of bias for a protein with negligible expression level; almost everything below this line can be ascribed to expected statistical fluctuations. Bottom: The spread of β_1 around its average for a randomized genome, where all genes have the same expected codon usage.

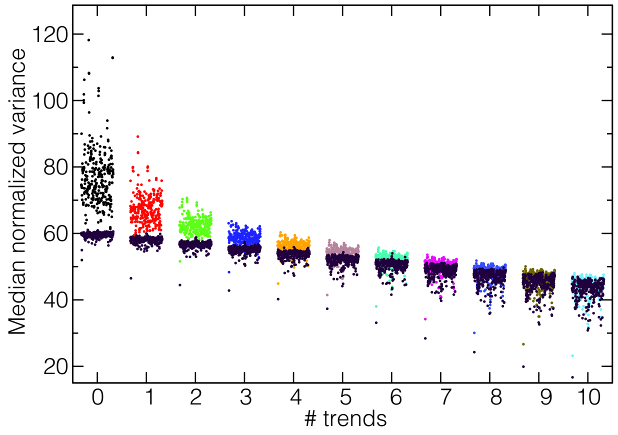

FIG. S7: Median normalized variance for 325 prokaryote genomes, using models with 0–10 trends. The different genomes are slightly offset along the abscissa, in alphabetical order. Median normalized variances for randomized genomes, generated from the various models, are shown in brown (the offsets, but not the preference functions, were re-estimated before calculating the normalized variances).

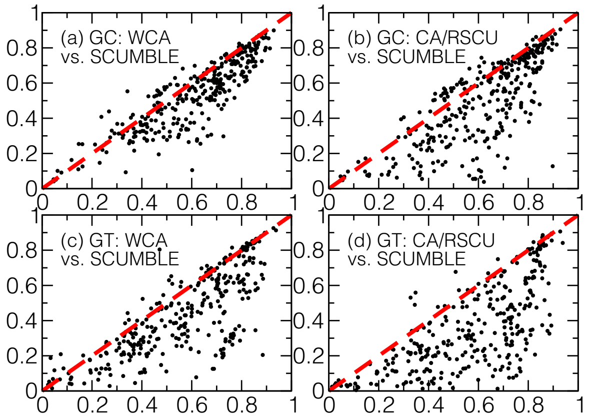

FIG. S8: Maximal square correlations between estimated preference functions/codon weights and the ideal preference function for GC bias [(a) and (b)] or GT bias [(c) and (d)]. The correlations for WCA [(a) and (c)] or CA/RSCU [(b) and (d)] are plotted against the correlations for SCUMBLE for all the prokaryote genomes studied. The codon weights from WCA and CA/RSCU were normalized the same way as SCUMBLE's preference functions before taking correlations. Note that while SCUMBLE yields clearly higher correlation for most prokaryotes, the correlations for WCA and CA/RSCU do tend to be close to one for genomes for which SCUMBLE yields very high correlations, indicating that the ideal preference functions are ideal (or almost ideal) also for WCA and CA/RSCU.