

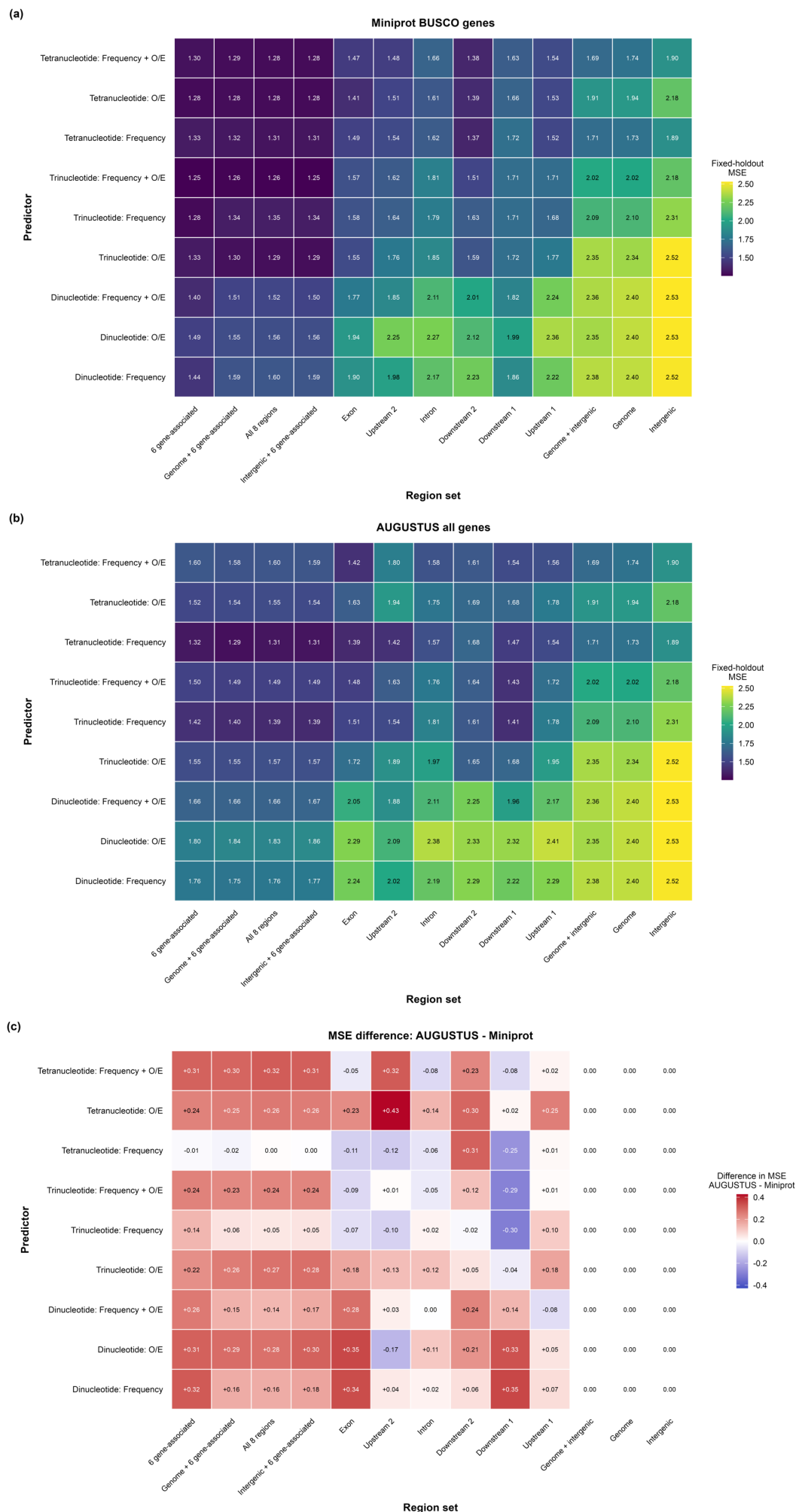


**Figure S1. Comparison of lifespan-prediction performance between Miniprot BUSCO genes and AUGUSTUS all genes.**  
 Legend follows on the next page.

**Figure S1. Comparison of lifespan-prediction performance between Miniprot BUSCO genes and AUGUSTUS all genes.**

Heatmaps show mean squared error (MSE) in the 139 fixed-holdout species for the same 117 predictor configurations evaluated in Fig. 1a. Predictions were averaged across the 1,000 outer models. The gene sets are (a) Miniprot-annotated BUSCO genes and (b) all AUGUSTUS-predicted genes. (c) Difference in fixed-holdout MSE, calculated as AUGUSTUS minus Miniprot. Positive values indicate greater prediction error with AUGUSTUS, whereas negative values indicate greater error with Miniprot. All panels use the predictor-row and region-column ordering from Fig. 1a, which was determined from the outer-test MSE of the Miniprot BUSCO models rather than from fixed-holdout performance.

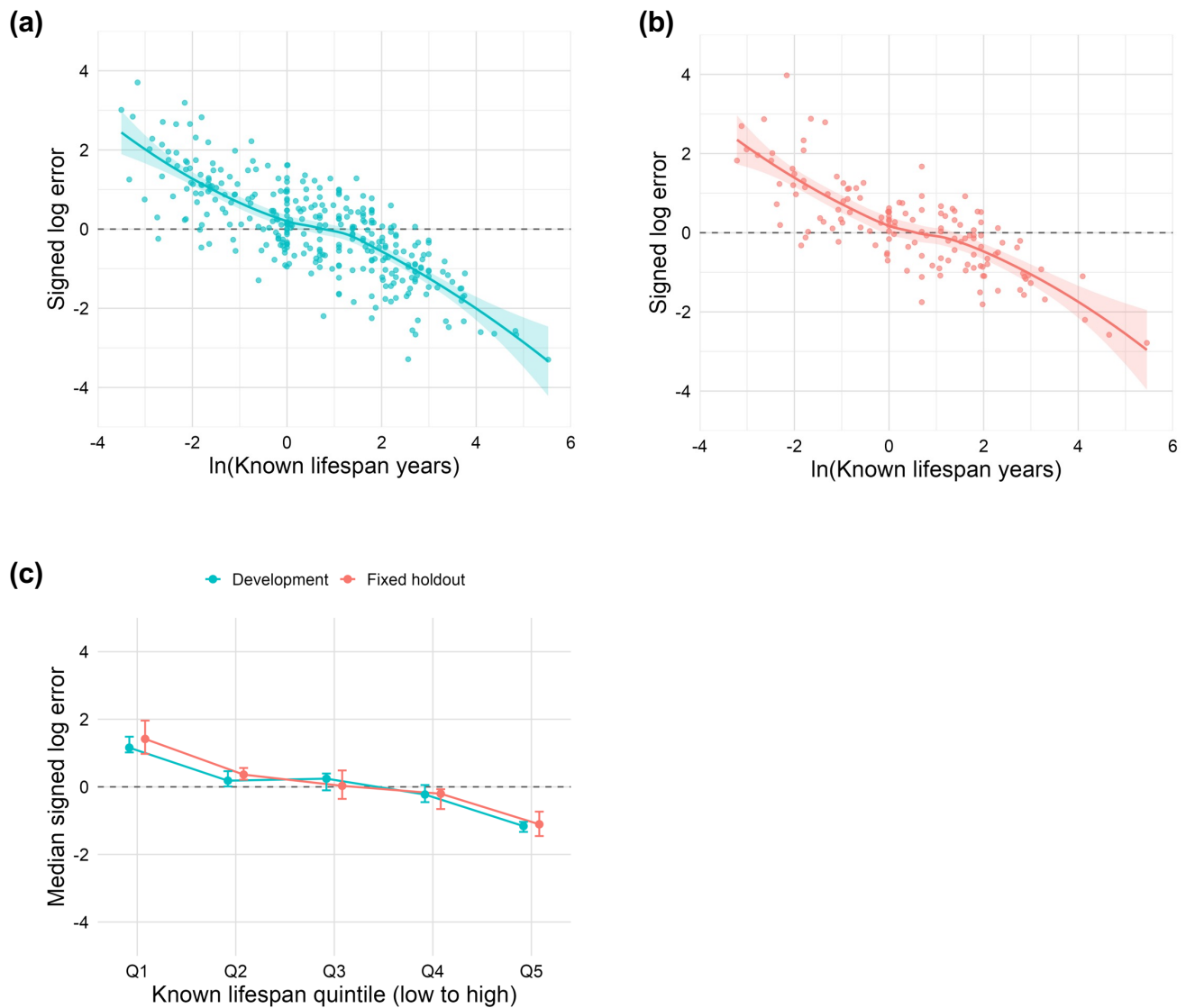


**Figure S2. Relationship between signed prediction error and known lifespan.**

Signed log error was calculated as  $\ln(\text{predicted lifespan}) - \ln(\text{known lifespan})$ , such that positive values indicate overprediction and negative values indicate underprediction. Predictions are the same as those shown in Fig. 2. Dashed horizontal lines indicate zero prediction error. (a) Development species ( $n = 327$ ), for which each prediction was averaged across the outer-resampling iterations in which that species was included in the outer-test subset. (b) Fixed-holdout species ( $n = 139$ ), for which predictions were averaged across 1,000 outer models. In (a) and (b), each point represents one species. Solid curves show locally estimated scatterplot smoothing (LOESS) fits, and shaded bands show 95% confidence intervals for the fitted trends. (c) Median signed log error in five equal-count bins defined from the known-lifespan distribution across all 466 species. Points show medians calculated separately for development and fixed-holdout species, and error bars show percentile bootstrap 95% confidence intervals based on 10,000 resamples.

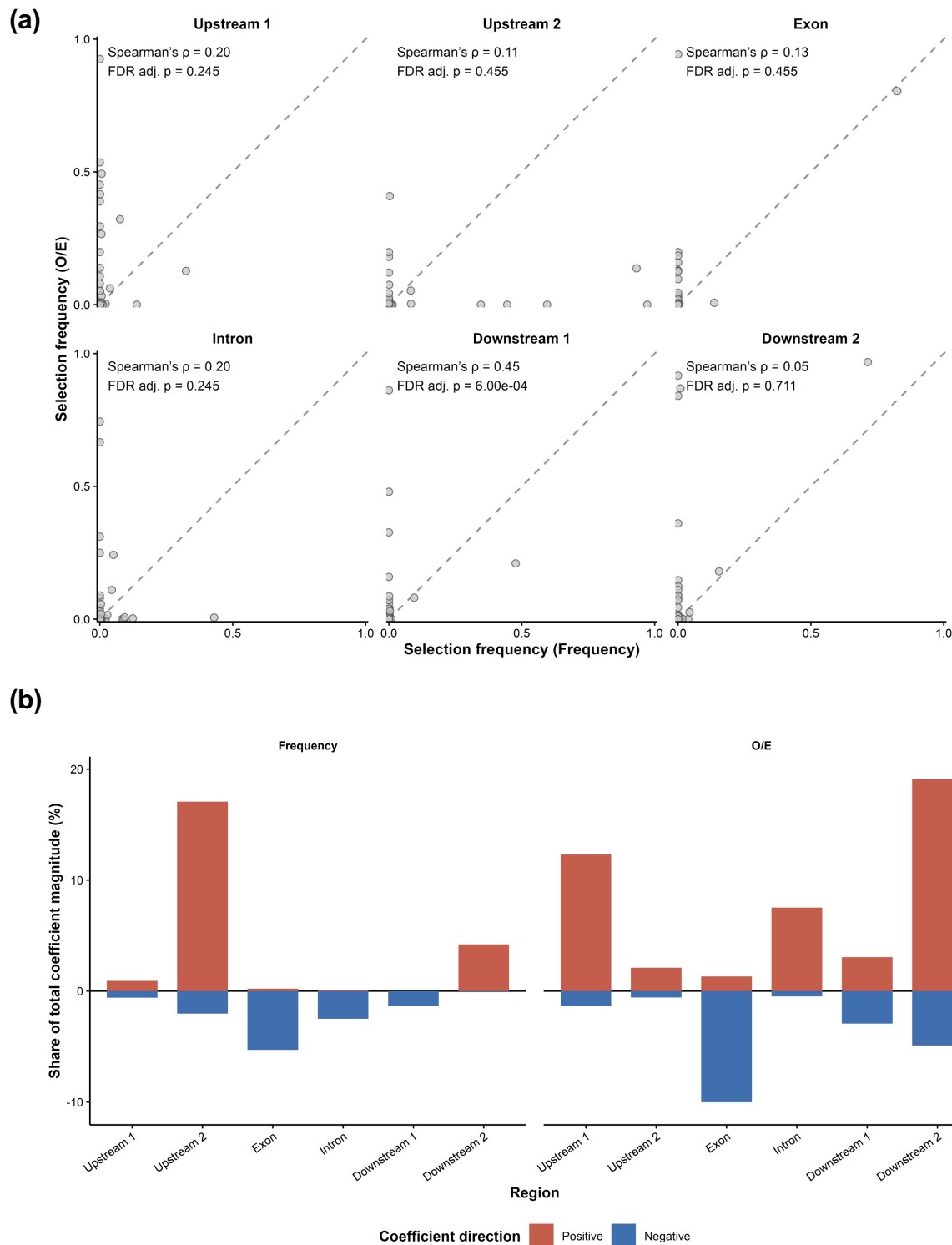


**Figure S3. Selection agreement and signed coefficient contributions of trinucleotide Frequency and O/E features across genomic regions.**

Selection frequencies and coefficients were summarized across 1,000 outer models fitted within the development set, using trinucleotide Frequency and O/E predictors from the six gene-associated regions. Selection frequency was the proportion of models in which a feature had a nonzero coefficient. **(a)** Selection frequencies of matched Frequency and O/E predictors for each of the 64 oriented trinucleotides in each region. The dashed line indicates equal selection frequency. Spearman's  $\rho$  and permutation-based p values adjusted for the false discovery rate (FDR) across the six regions are shown. **(b)** Shares of total coefficient magnitude arising from positive (red) and negative (blue) standardized coefficients in each region and composition measure. Positive and negative coefficient magnitudes were averaged separately across all models, including zero coefficients when a feature was not selected. Positive coefficients increase predicted  $\ln(\text{lifespan})$ , whereas negative coefficients decrease it.

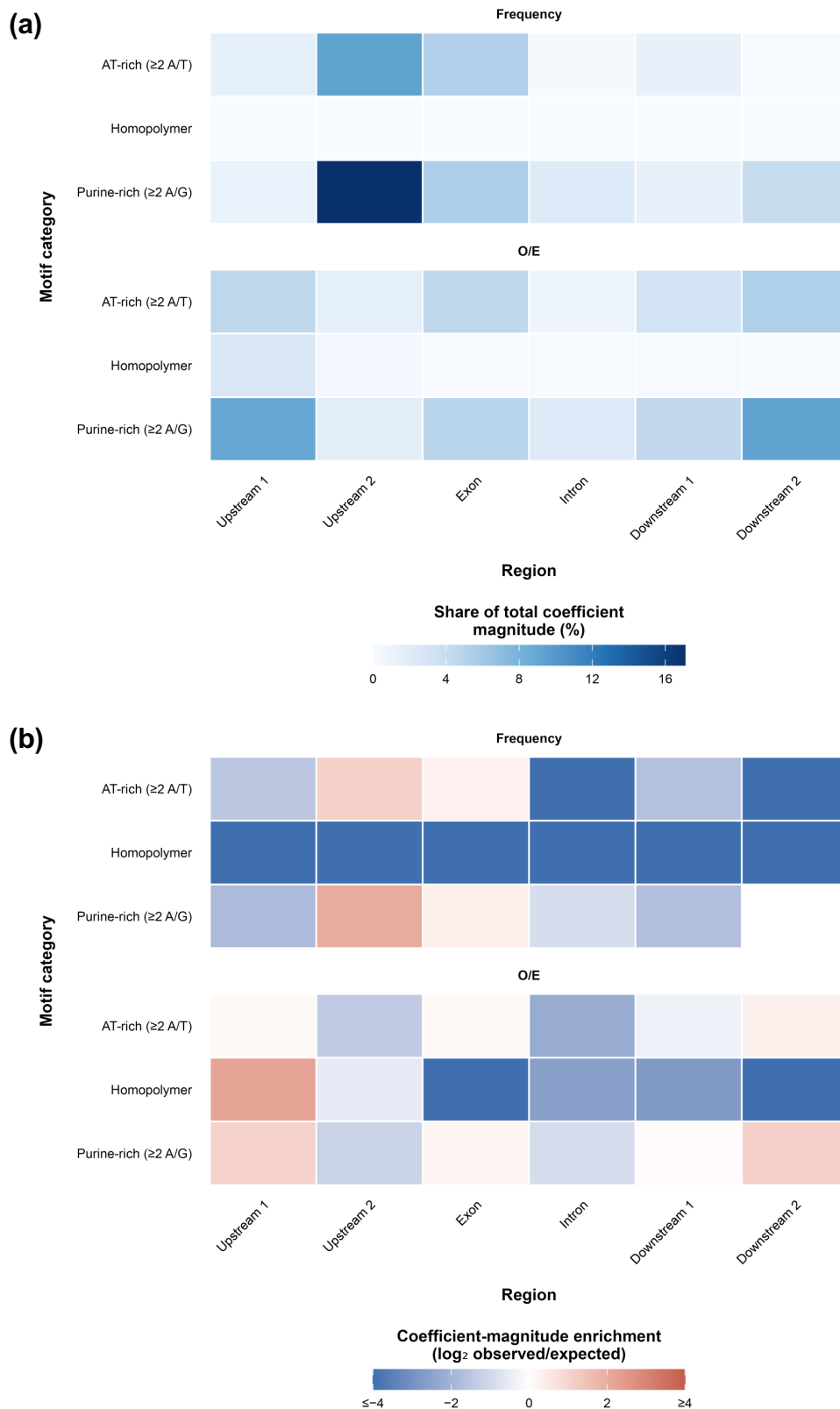
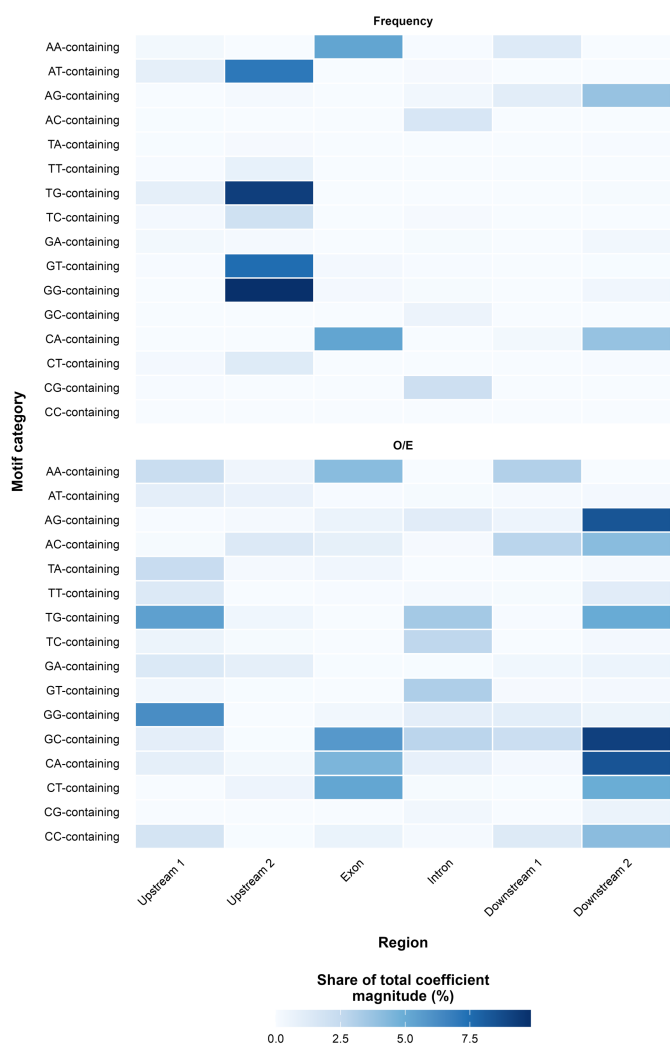


**Figure S4. Distribution of coefficient-magnitude shares and enrichment among trinucleotide motif categories.**

Heatmaps summarize the mean absolute standardized coefficients of the 768 features in the final trinucleotide Frequency + O/E model. (a, c) Shares of total coefficient magnitude across all 768 predictors for each region and composition measure, grouped by sequence motif category (a) or by the dinucleotides contained within each trinucleotide (c). Motif categories comprise AT-rich motifs (at least two A/T bases), homopolymers (three identical bases), and purine-rich motifs (at least two A/G bases). (b, d) Coefficient-magnitude enrichment for the corresponding categories, expressed as the  $\log_2$  ratio of the observed coefficient-magnitude share to the share expected from the number of predictors in the category. Positive values indicate more coefficient magnitude than expected and negative values indicate less. Panels (a) and (c) are descriptive; statistical tests apply to the enrichment shown in (b) and (d). Two-sided empirical p values from the random-set tests were FDR-adjusted across all enrichment tests, and none remained significant. Cells with zero coefficient magnitude are displayed at the lower limit of the color scale.

(c)



(d)

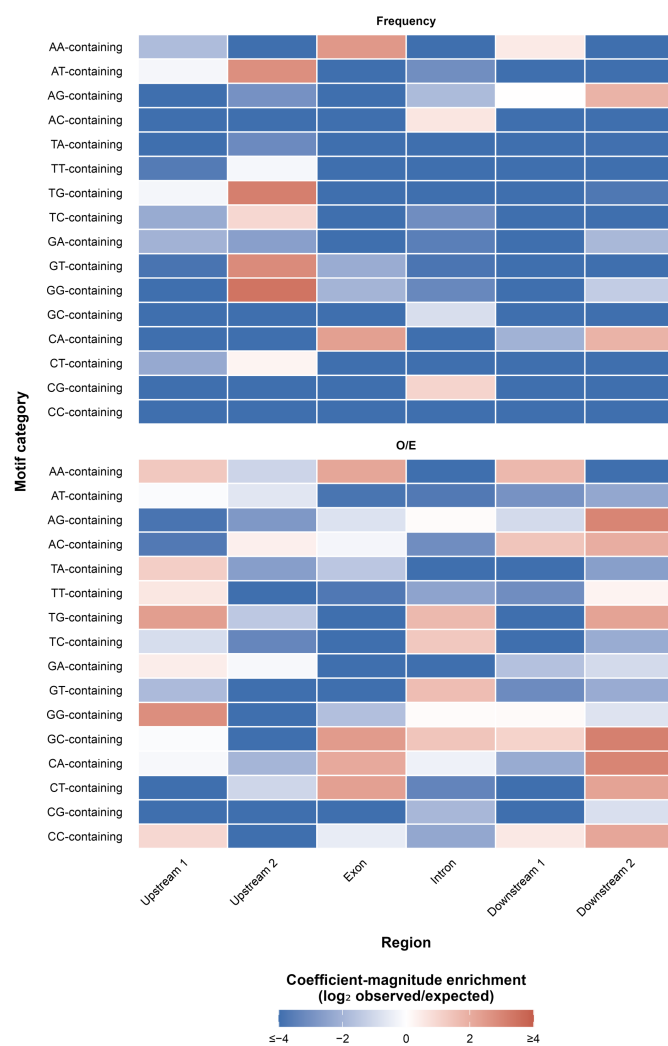


Figure S4. Continued.

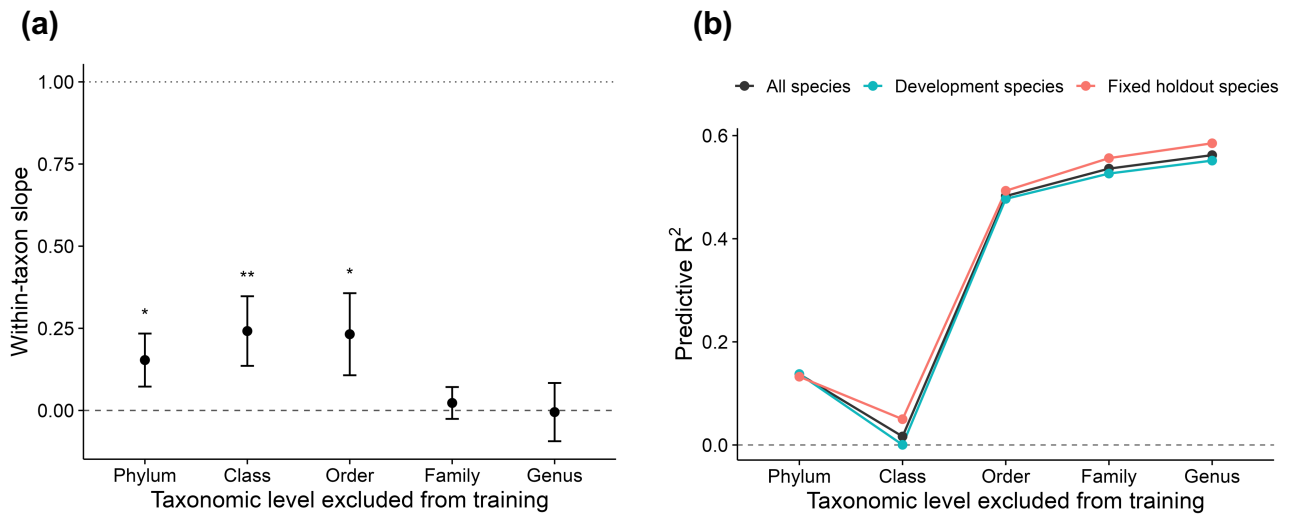


**Figure S5. Leave-one-taxon-out validation of the final trinucleotide lifespan-prediction model.**

Leave-one-taxon-out (LOTO) validation was conducted separately at the phylum, class, order, family, and genus levels. At each level, all development species belonging to one target taxon were excluded, and a model fitted to the remaining development species was used to predict all species in the excluded taxon. This procedure was repeated for every taxon. (a) Common within-taxon slopes between known and predicted  $\ln(\text{lifespan})$ , estimated with taxon-specific intercepts using both development and fixed-holdout species from taxa represented by at least three species and with nonzero known-lifespan range. Points show slope estimates, and error bars show taxon-clustered CR2 95% confidence intervals. The dashed line indicates a slope of zero, and the dotted line indicates the slope of one expected under perfect calibration. Asterisks indicate slopes significantly differing from zero after FDR adjustment across the five ranks (\* $p < 0.05$ ; \*\* $p < 0.01$ ). (b) Predictive coefficient of determination ( $R^2$ ) of the LOTO predictions at each taxonomic level, calculated separately for all species, development species, and fixed-holdout species. An  $R^2$  of zero indicates the same squared error as predicting every species in the displayed subset using that subset's mean known  $\ln(\text{lifespan})$ ; positive values indicate improvement over this reference prediction.