Supplementary Table 1 Spearman rank correlations between d_N, d_S (estimated by PAML) and recombination rate for both broad- and fine-scale recombination rates.

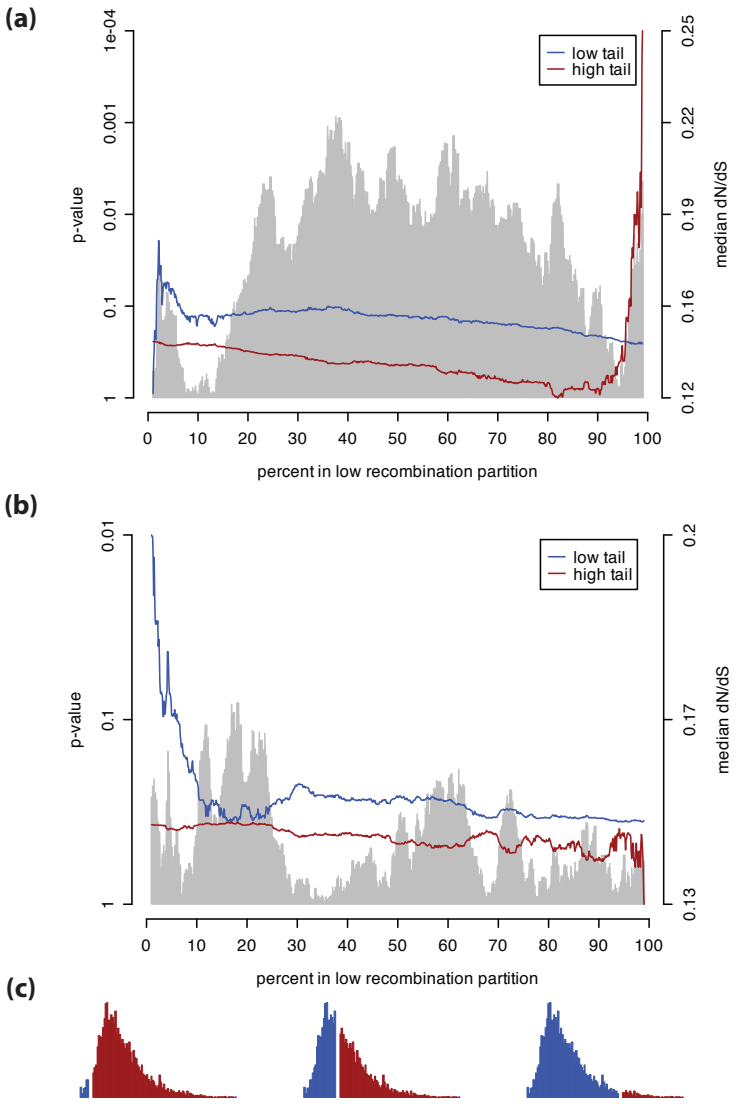
	correlation	**p-value**
d_N and broad-scale recombination rate	0.053	1.4×10^{-7}
d_S and broad-scale recombination rate	0.211	$< 10^{-16}$
d_N and fine-scale recombination rate	0.059	5.9×10^{-9}
d_S and fine-scale recombination rate	0.153	$< 10^{-16}$

Supplementary Table 2 Summary of linear models and Spearman correlation tests comparing broad- and fine-scale recombination rates, after stratification into d_N/d_S bins. Three sets of linear models and correlations were tested for genes in three different d_N/d_S bins. 10% of the genes all have the minimum d_N/d_S value (0.0001); these were placed in the lowest bin. The central 85% of genes and the top 5% of genes by d_N/d_S comprise the second and third bins. For the linear models, the fine-scale recombination rate is estimated using broad-scale recombination rate as the predictor and requiring the intercept to be 0 (results are similar when not requiring the intercept to be zero). Spearman's ρ and the associated p-value are given for correlation tests of the correlation between fine- and broad-scale recombination rates.

	linear models			**Spearman correlations**	
d_N/d_S **bin**	**slope**	**std. error**	**p-value**	ρ	**p-value**
low 10%	0.79	0.031	$< 10^{-16}$	0.35	$< 10^{-16}$
medium 85%	0.84	0.011	$< 10^{-16}$	0.35	$< 10^{-16}$
high 5%	0.83	0.050	$< 10^{-16}$	0.37	$< 10^{-16}$

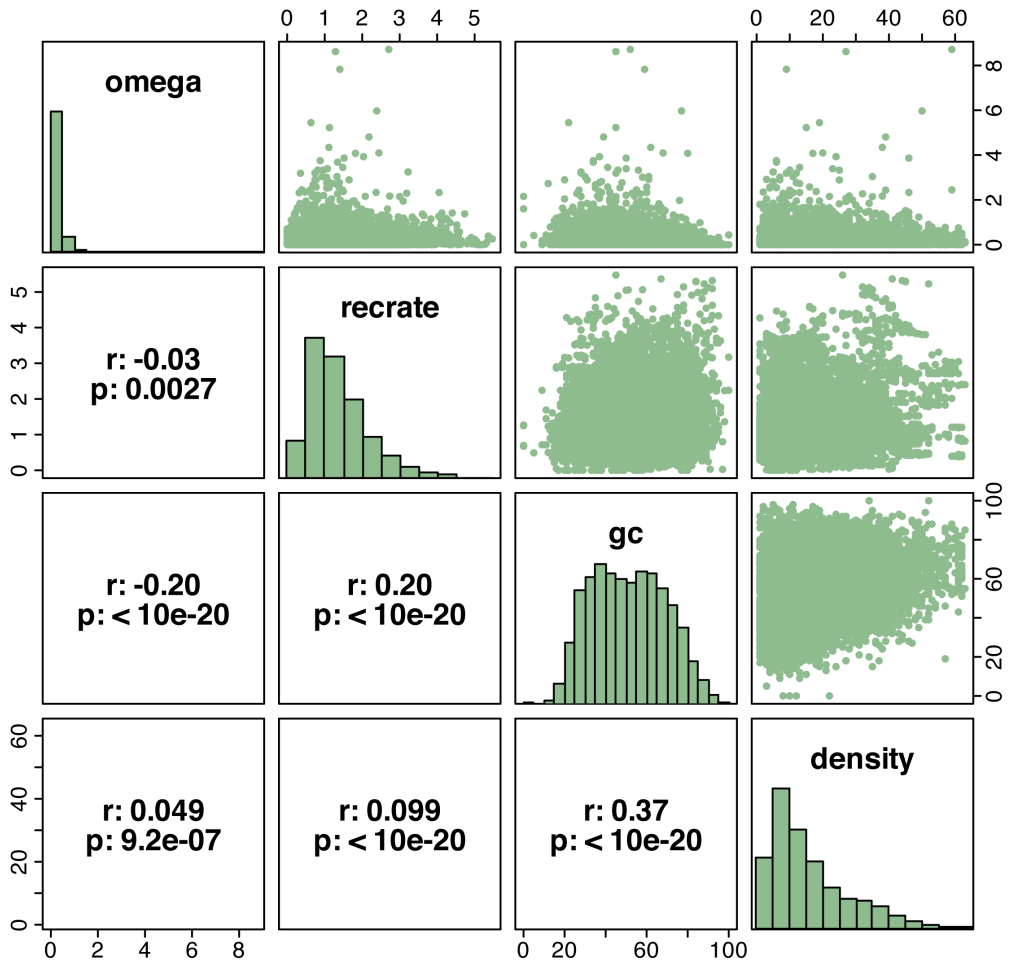
Supplementary Table 3 Summary of two linear models testing HRI due to an effect of recombination on non-synonymous polymorphism levels. The two columns under the heading "broad-scale" show the coefficients and p-values for the linear model that incorporates broad-scale recombination rate as one of the predictors, and similarly the two columns under the heading "fine-scale" are for the model incorporating fine-scale recombination rates. Both models include as predictors: recombination rate, the number of synonymous segregating sites (p_S), linear and quadratic effects of GC content, and gene density. The response variable in both models is the number of non-synonymous segregating sites. The models can be summarized as: $p_N \sim \mathrm{GC} + GC^2 + p_S + genedensity + recombinationrate$.

	broad-scale		**fine-scale**	
model term	**coefficient**	**p-value**	**coefficient**	**p-value**
intercept	0.47	0.008	0.56	0.002
GC	0.013	0.06	0.011	0.12
GC^2	-0.00021	0.001	-0.00019	0.004
p_S	0.42	$< 10^{-16}$	0.42	$< 10^{-16}$
gene density	0.0012	0.5	0.0018	0.3
recombination rate	0.066	0.02	0.026	0.05

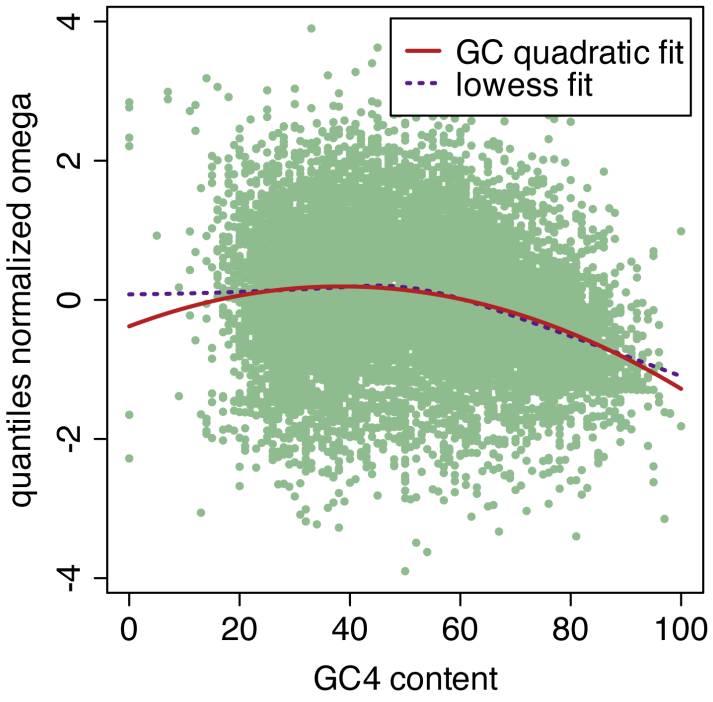


Supplementary Figure 1: Summary of Wilcoxon rank sum tests of whether partitions of the data into high and low recombination-rate bins have significantly different median d_N/d_S values. (a) Broad-scale recombination rates. (b) Fine-scale recombination rates. (c) Examples of three partitions of the broad-scale recombination rates into high and low recombination-rate bins. (a-b) The horizontal axis indicates the percent of the dataset, sorted by recombination rate, that is placed in the low recombination rate bin. Tests were performed for each one-percent increment of genes included in the low tail. The gray bars indicate the p-value of each two-tailed, two sample, Wilcoxson rank sum test, as indicated by the left-hand axis. The colored lines show the mean d_N/d_S ratio of genes in the low (blue) and high (red) recombination-rate bins, which is quantified on the right-hand axis.

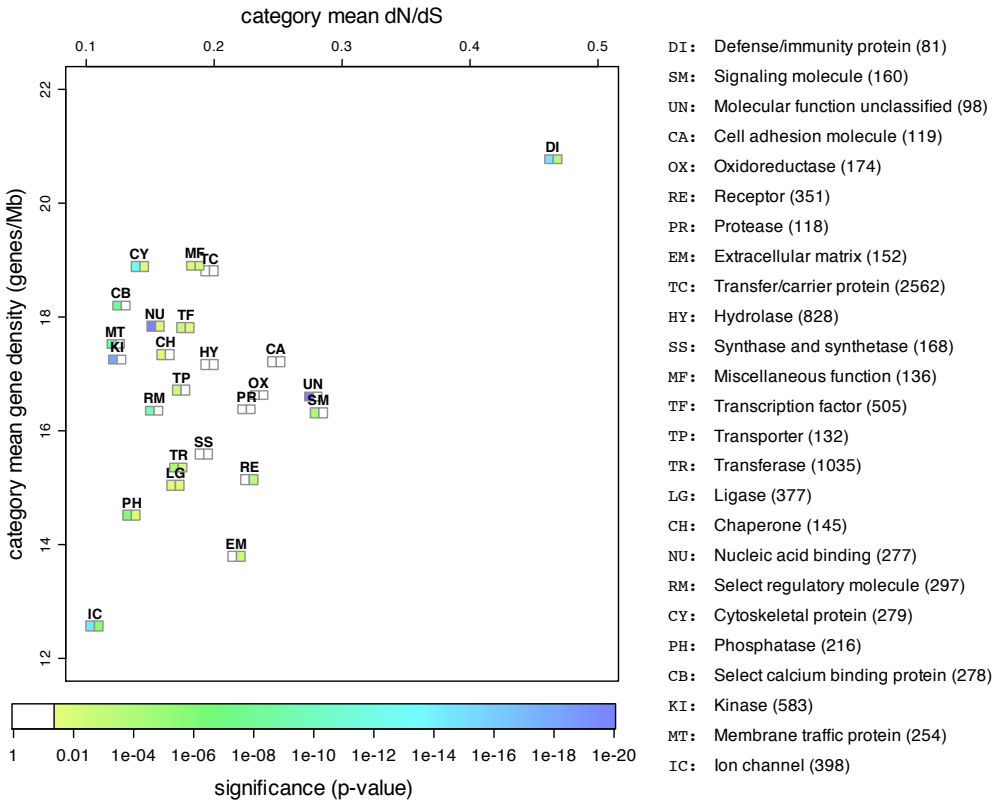
Supplementary Figure 2: Plots of broad-scale recombination rates for each chromosome. The recombination rate is plotted on the vertical axis as a function of physical position. The horizontal and vertical axes are on the same scale across chromosomal plots. Blue boxes mark regions spanning 4 cM, centered on the centromeres. Hash marks above each plot indicate the locations of genes in our data set, with genes falling in the lowest 2% of recombination rates genome wide indicated in red.



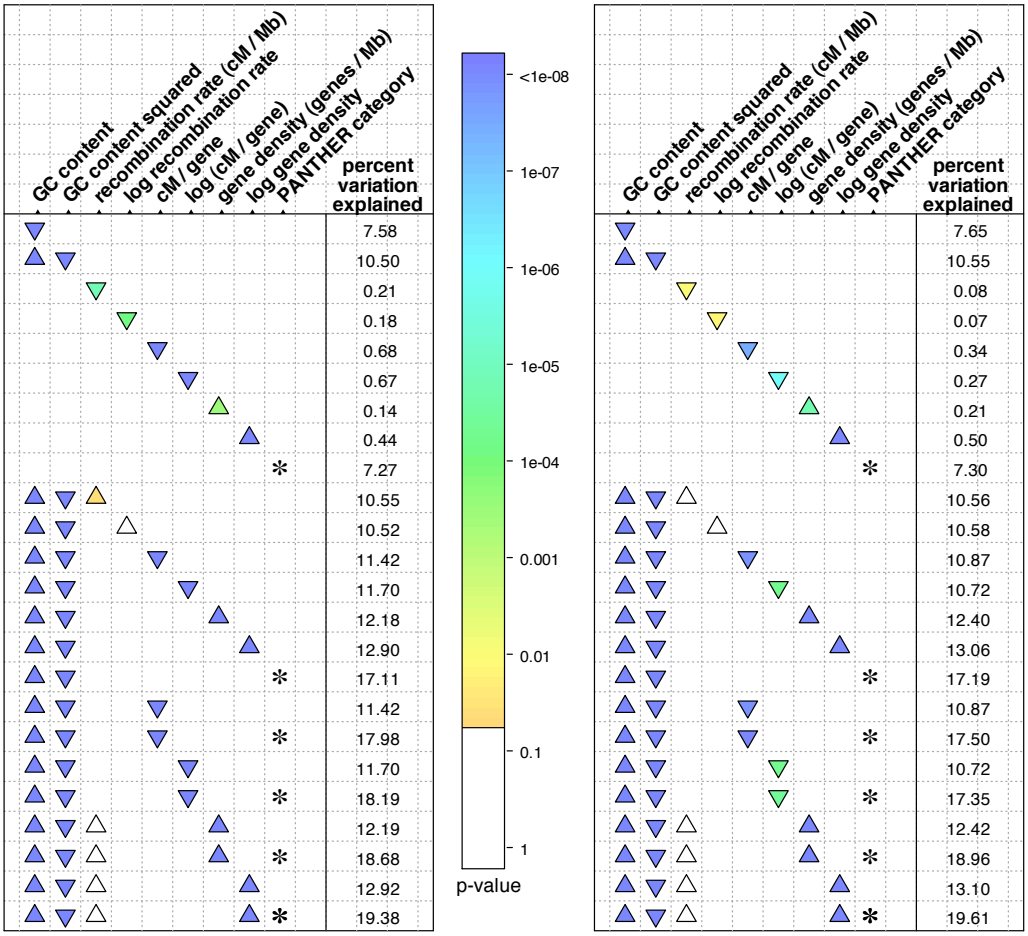
Supplementary Figure 3: Scatter plots, histograms and Spearman rank correlation estimates for pairwise comparisons of d_N/d_S, broad-scale recombination rate, percent GC content, and gene density. The diagonal contains histograms of each variable. Pairwise scatter plots are shown above the diagonal for each combination of variables. Below the diagonal are the Spearman rank correlation coefficients (r) and associated p-values (p) for each pairwise correlation test.



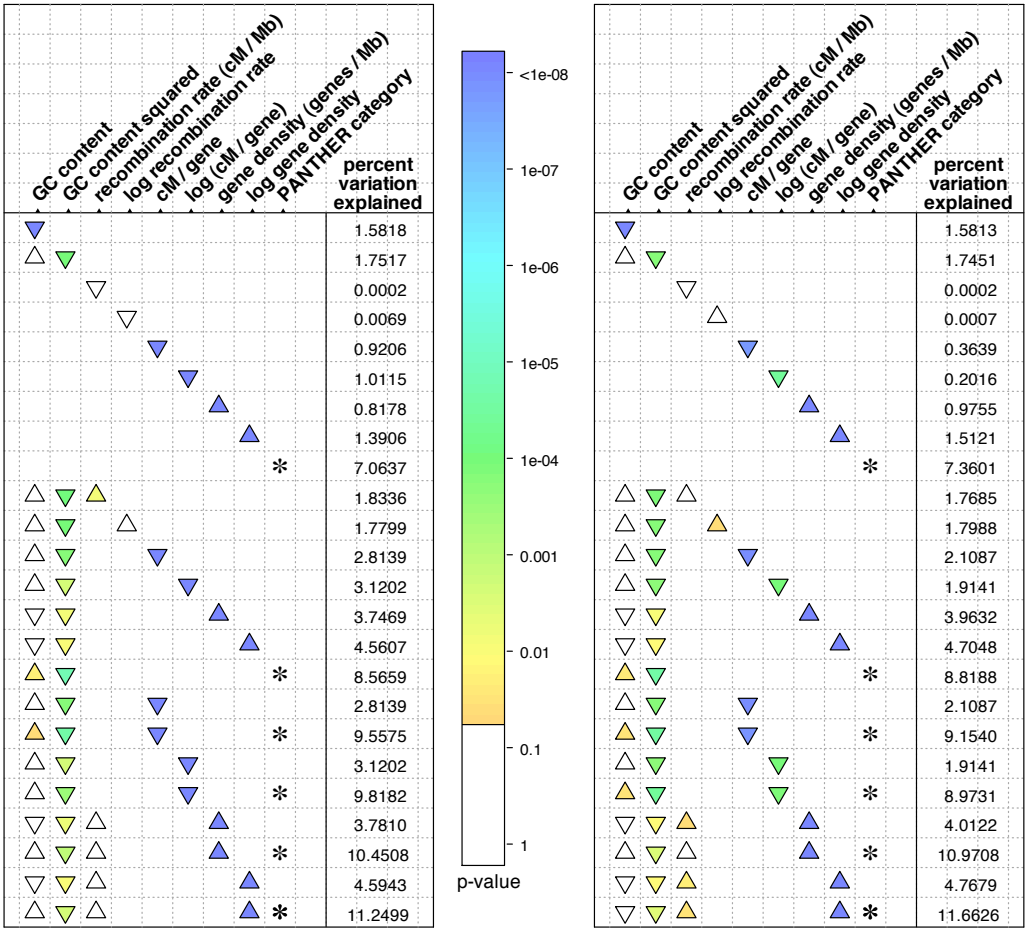
Supplementary Figure 4: Highly non-linear relationship between GC content and normalized d_N/d_S (estimated by PAML). The horizontal axis shows GC4 content, which is GC content calculated at codon third positions when the codons encoding the amino acid are four-fold degenerate; the vertical axis shows quantile-normalized d_N/d_S (see Methods). The solid, red line shows the fit of a linear model with first order and second order predictors of GC content, while the dotted, purple line shows a lowess fit for comparison.



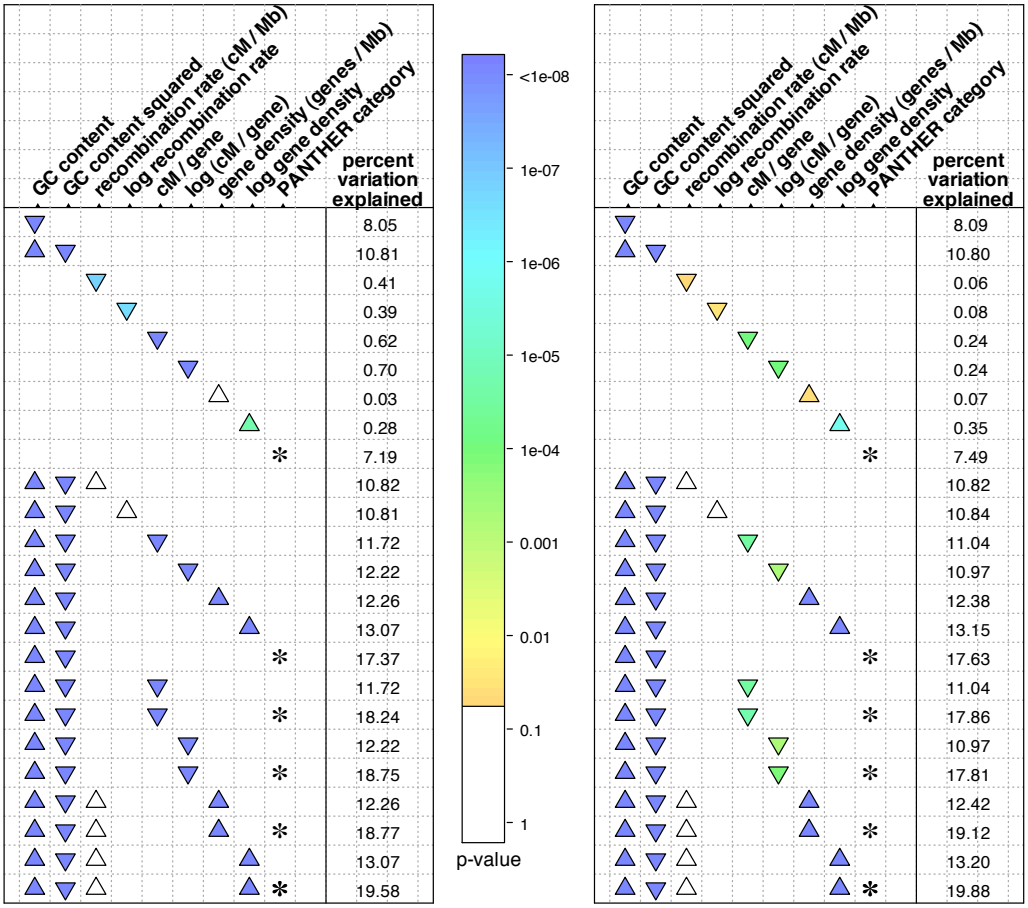
Supplementary Figure 5: Scatter plot of d_N/d_S and gene density (genes/Mb) means for each category of PANTHER molecular functions. This figure suggests that a few (i.e. defense/immunity and ion channel) functional categories are contributing heavily to the correlation between d_N/d_S and gene density. Two-letter codes indicate the PANTHER molecular function category for which mean d_N/d_S and mean density were calculated. PANTHER molecular function categories are listed on the right with the two-letter abbreviation and the number of genes in the indicated category shown in parentheses. Only categories with more than 80 genes are shown. The location on the scatter plot of the two-letter category code indicates the d_N/d_S (horizontal axis) and gene density (vertical axis) calculated for genes in the indicated category. Boxes below each two-letter category abbreviation indicate the significance of a two-sided, two-sample unpaired Student's t-test of whether genes with the indicated molecular function have a significantly different mean d_N/d_S (left colored box) or significantly different mean gene density (right colored box) than genes without the indicated functional annotation among their annotations. Significance levels are color-coded by p-value, as indicated by the bar below the scatter plot. Boxes colored white are not significant at the 0.05 level.



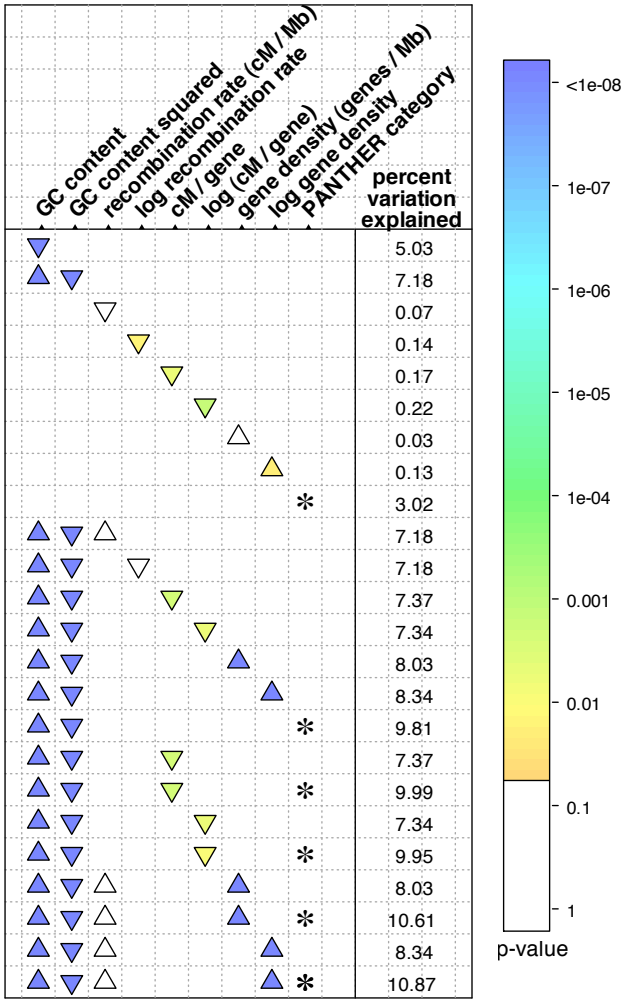
Supplementary Figure 6: More extensive set of linear models, with normalized d_N/d_S as a response variable. The left and right panels show results for broad- and fine-scale recombination rates, respectively. Each row describes a linear model used to explain variation in quantile-normalized d_N/d_S. The columns list the predictor variables, and the presence of absence of a triangle indicates if a particular variable was used in the model. The color of the triangles displays the p-value, coded by the colored bar in the center; it is based on a Student's t-test, which tests whether the coefficient for that term differs significantly from zero. White triangles represent non-significant coefficients at the 0.05 level. The orientation of the triangles indicates the sign of the coefficient, with upward- and downward-pointing triangles indicating positive and negative coefficients, respectively. Models including PANTHER category as a factor are indicated with an asterisk. The column of numbers on the right is the percentage of the variation in quantile-normalized d_N/d_S that is explained by the model i.e., 100*(r-squared). GC content refers to GC4 content (see Methods). Although the gene density and GC content calculations in the two data sets are the same for all genes, the genes included differ slightly based on whether broad- or fine-scale rates could be calculated; thus, even when the only difference between models is the scale over which recombination rates are estimates, the results can differ slightly.



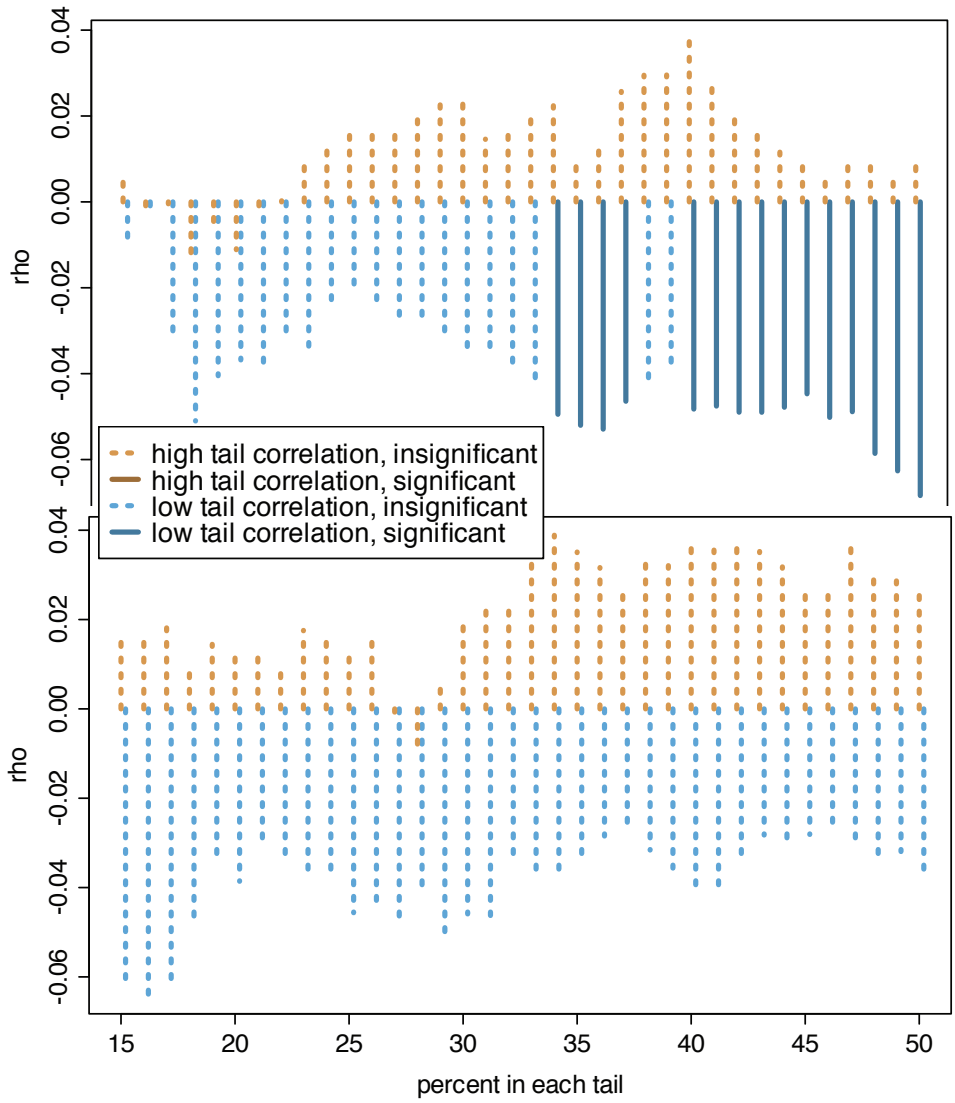
Supplementary Figure 7: More extensive set of linear models, with normalized d_N/d_S based on Nei-Gojobori distances as a response variable. The left and right panels show results for broad- and fine-scale recombination rates, respectively. Each row describes a linear model used to explain variation in quantile-normalized d_N/d_S. The columns list the predictor variables, and the the presence of absence of a triangle indicates if a particular variable was used in the model. The color of the triangles displays the p-value, coded by the colored bar in the center; it is based on a Student's t-test, which tests whether the coefficient for that term differs significantly from zero. White triangles represent non-significant coefficients at the 0.05 level. The orientation of the triangles indicates the sign of the coefficient, with upward- and downward-pointing triangles indicating positive and negative coefficients, respectively. Models including PANTHER category as a factor are indicated with an asterisk. The column of numbers on the right is the percentage of the variation in quantile-normalized d_N/d_S that is explained by the model i.e., 100*(r-squared). GC content refers to GC4 content (see Methods). Although the gene density and GC content calculations in the two data sets are the same for all genes, the genes included differ slightly based on whether broad- or fine-scale rates could be calculated; thus, even when the only difference between models is the scale over which recombination rates are estimates, the results can differ slightly.



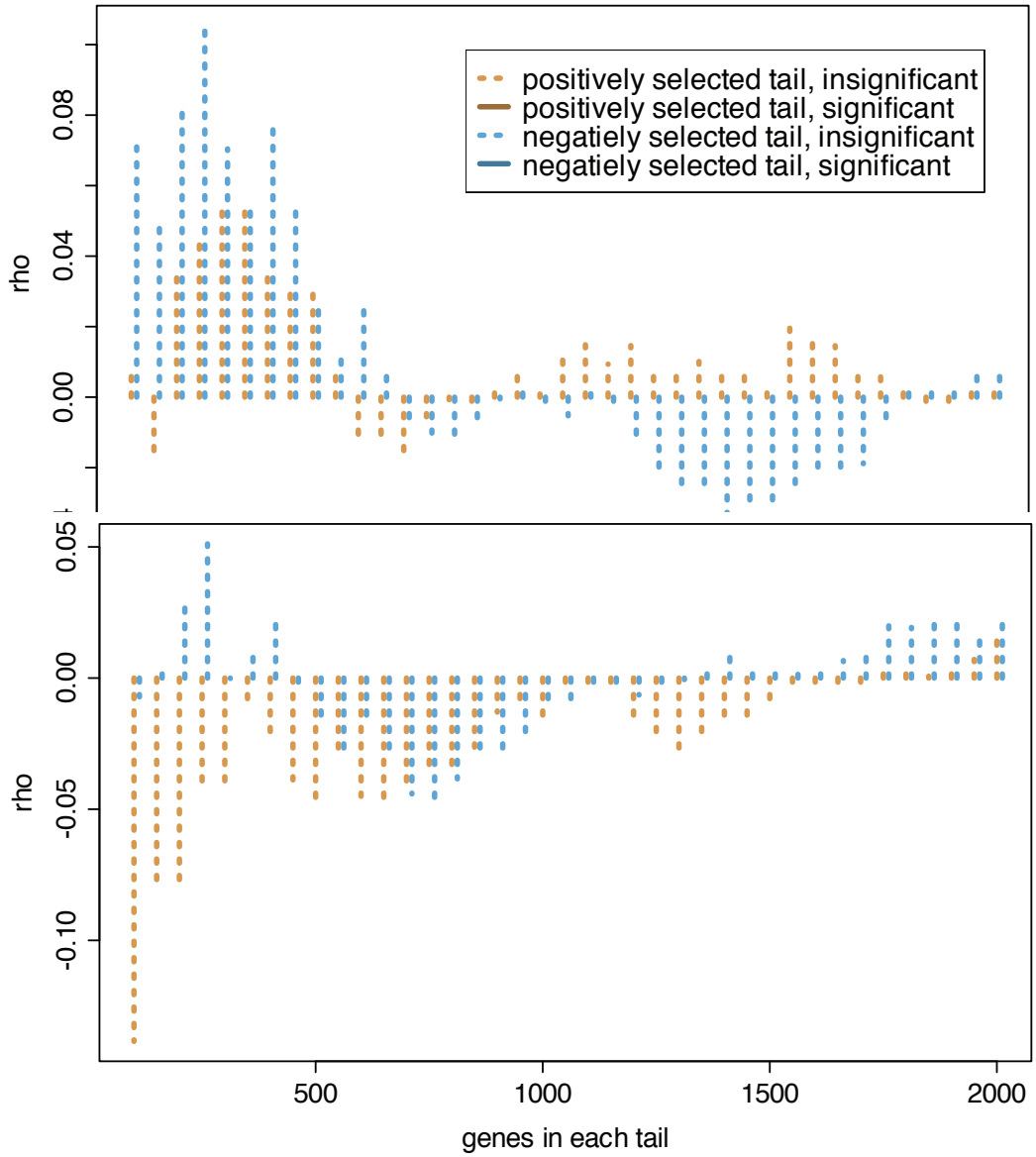
Supplementary Figure 8: Set of linear models excluding genes on human chromosomes 2, 7, 14, 15, 20, 21, 22, on which fusions or fissions occurred somewhere on the phylogeny. The response variable is the phylogeny-wide d_N/d_S. The left and right panels show results for broad- and fine-scale recombination rates, respectively. Each row describes a linear model used to explain variation in quantile-normalized d_N/d_S. The columns list the predictor variables, and the presence of absence of a triangle indicates if a particular variable was used in the model. The color of the triangles displays the p-value, coded by the colored bar in the center; it is based on a Student's t-test, which tests whether the coefficient for that term differs significantly from zero. White triangles represent non-significant coefficients at the 0.05 level. The orientation of the triangles indicates the sign of the coefficient, with upward- and downward-pointing triangles indicating positive and negative coefficients, respectively. Models including PANTHER category as a factor are indicated with an asterisk. The column of numbers on the right is the percentage of the variation in quantile-normalized d_N/d_S that is explained by the model i.e., 100*(r-squared). GC content refers to GC4 content (see Methods). Although the gene density and GC content calculations in the two data sets are the same for all genes, the genes included differ slightly based on whether broad- or fine-scale rates could be calculated; thus, even when the only difference between models is the scale over which recombination rates are estimates, the results can differ slightly.



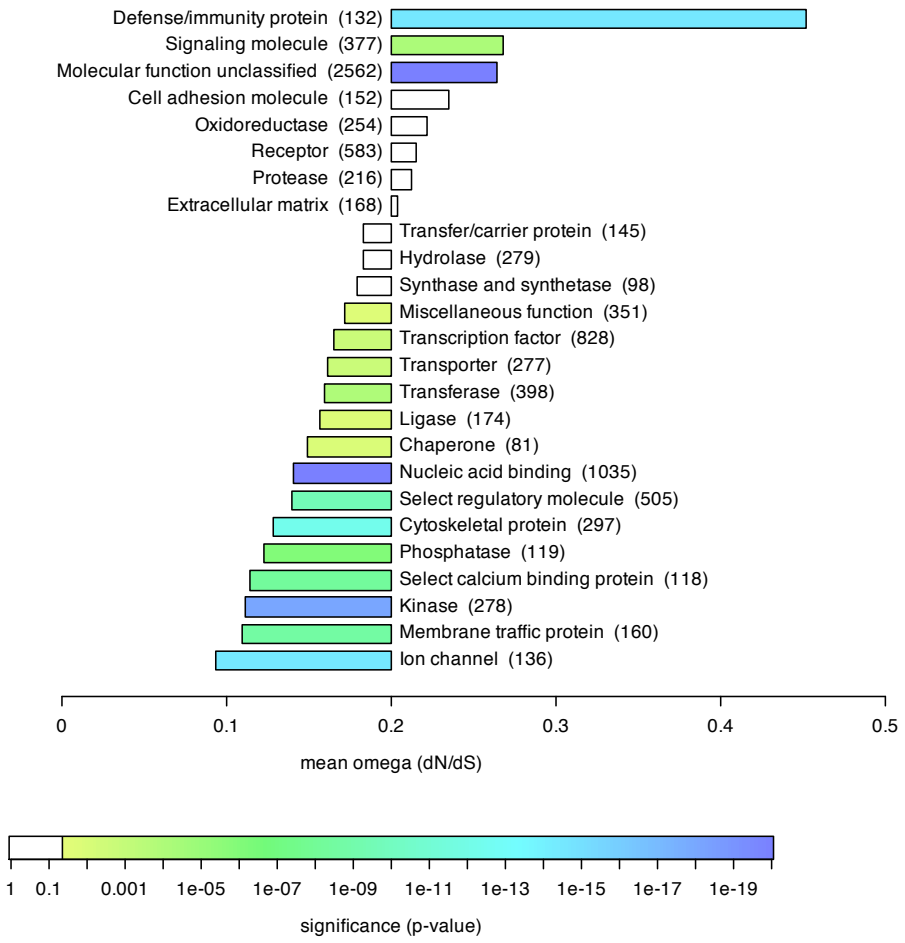
Supplementary Figure 9: Set of linear models with quantile-normalized human-lineage-specific d_N/d_S as the response variable. Data for fine-scale recombination rates are shown. Each row describes a linear model used to explain variation in quantile-normalized d_N/d_S. The columns list the predictor variables, and the presence of absence of a triangle indicates if a particular variable was used in the model. The color of the triangles displays the p-value, coded by the colored bar in the center; it is based on a Student's t-test, which tests whether the coefficient for that term differs significantly from zero. White triangles represent non-significant coefficients at the 0.05 level. The orientation of the triangles indicates the sign of the coefficient, with upward- and downward-pointing triangles indicating positive and negative coefficients, respectively. Models including PANTHER category as a factor are indicated with an asterisk. The column of numbers on the right is the percentage of the variation in quantile-normalized d_N/d_S that is explained by the model i.e., 100*(r-squared). GC content refers to GC4 content (see Methods). Although the gene density and GC content calculations in the two data sets are the same for all genes, the genes included differ slightly based on whether broad- or fine-scale rates could be calculated; thus, even when the only difference between models is the scale over which recombination rates are estimates, the results can differ slightly.



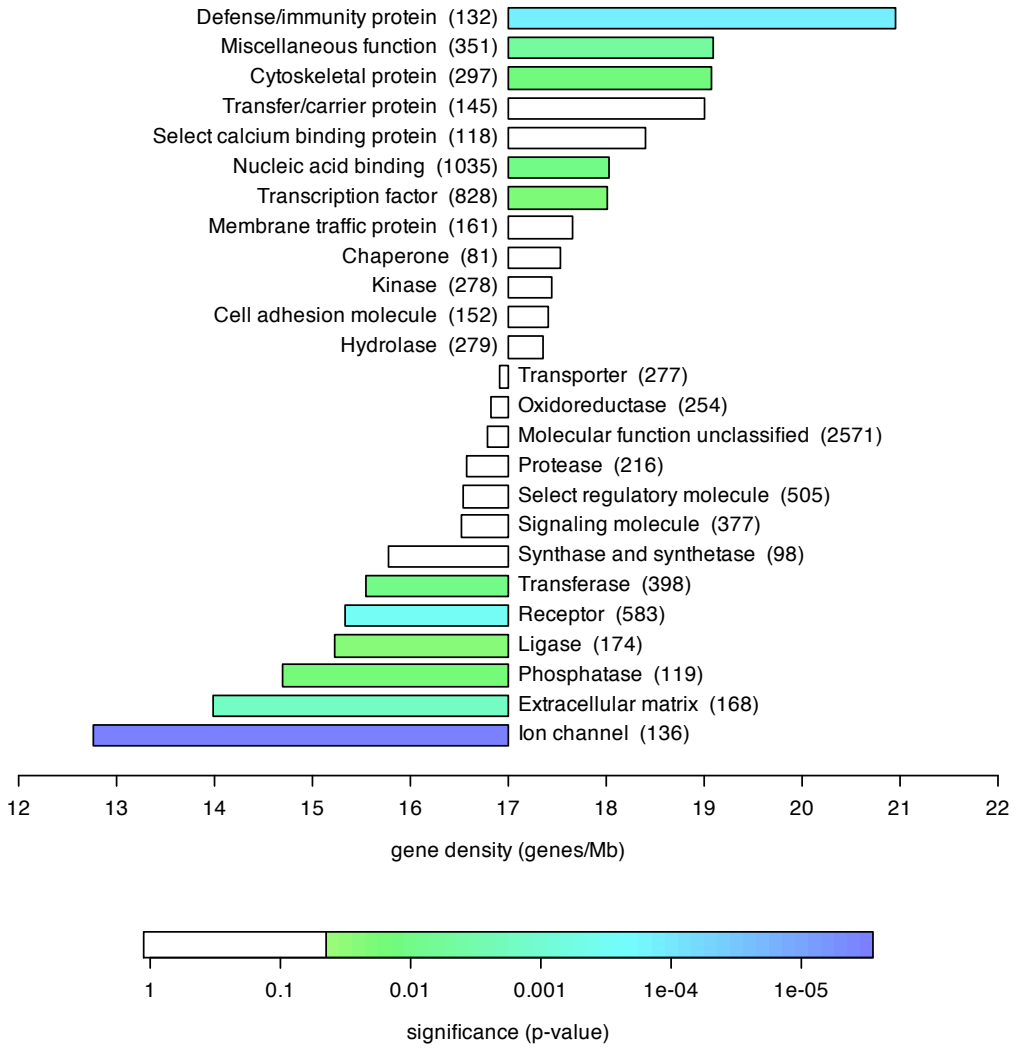
Supplementary Figure 10: Summary of Spearman rank correlation tests of the correlation between d_N/d_S and the recombination rate for both broad-scale (top panel) and fine-scale (bottom panel) rates. Tests were performed on sets of genes that lie in either the high or low tail of the distribution of d_N/d_S estimates, excluding genes with no observed non-synonymous substitutions (see Methods). The horizontal axis indicates the percent of genes included in either the low tail or the high tail. For tests of genes in the low tail, tests were conducted in increments of one percent of the genes, starting at the genes with the low 15% of d_N/d_S scores, followed by genes with the low 16% of d_N/d_S scores, etc. Tests of genes in the high tail were conducted similarly. The height of each colored line indicates the Spearman rank correlation coefficient calculated for that test. Tan solid or dotted lines represent tests for correlations among genes in the high tail of d_N/d_S, while blue solid or dotted lines represent tests for correlations among genes in the low tail of d_N/d_S. Dotted lines represent tests that are not significant at the 0.01 level, while solid lines indicates significant tests at this level.



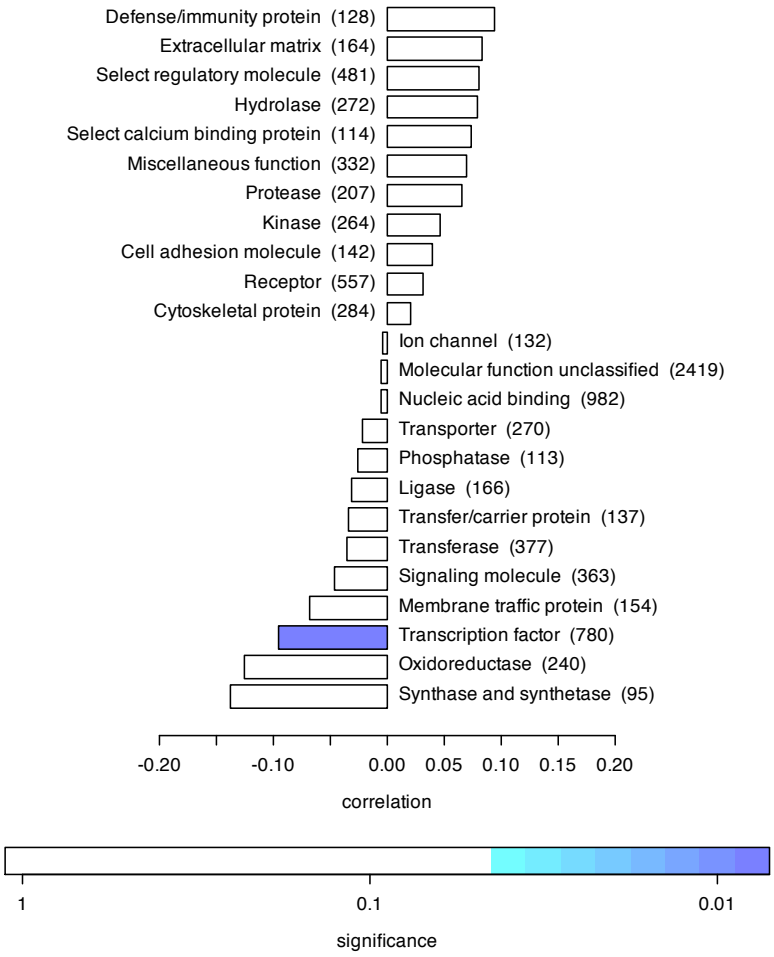
Supplementary Figure 11: Summary of Spearman rank correlation tests of the correlation between d_N/d_S and the recombination rate for both broad-scale (top panel) and fine-scale (bottom panel) rates. Tests were performed on sets of genes that lie in either the high or low tail of the distribution of γ estimates, excluding genes with no observed non-synonymous substitutions (see Methods). The horizontal axis indicates the number of genes included in either the low tail or the high tail. Tests were conducted in increments of 50 genes, starting with 100 genes. The height of each colored line indicates the Spearman rank correlation coefficient calculated for that test. Tan solid or dotted lines represent tests for correlations among genes potentially evolving under positive selection (high γ estimates), while blue solid or dotted lines represent tests for correlations among genes more likely to be evolving under purifying selection (low γ estimates). Dotted lines represent tests that are not significant at the 0.01 level, while solid lines indicates significant tests at this level.



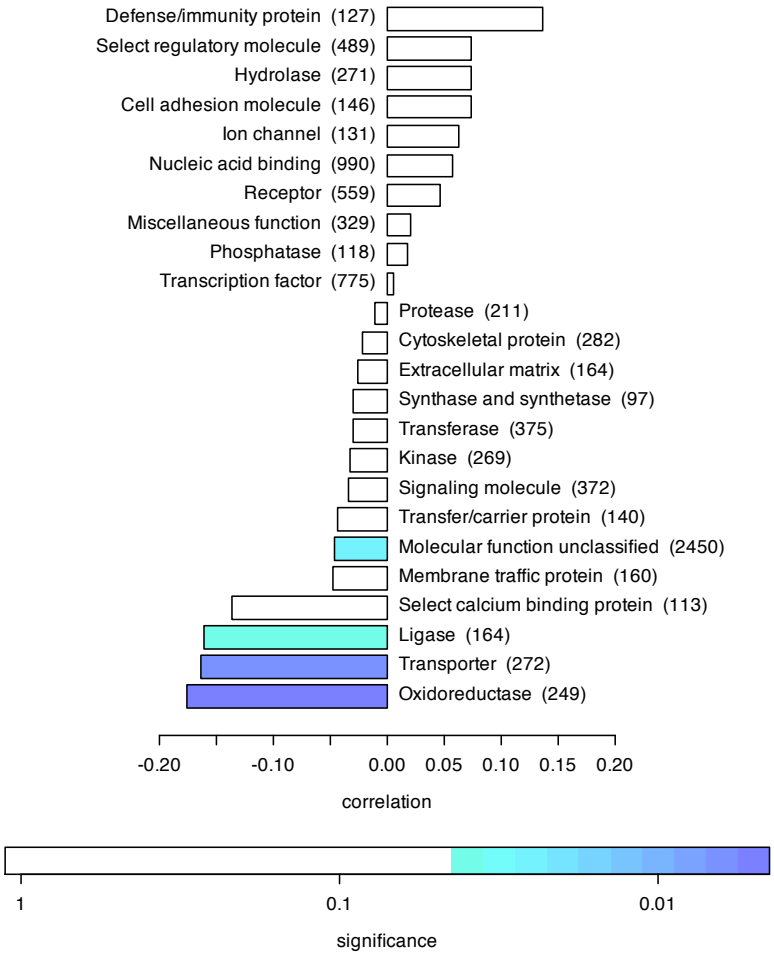
Supplementary Figure 12: Summary of mean d_N/d_S values for sets of genes based on the PANTHER classification of molecular functions. The length of each colored bar indicates the deviation of the category mean from the grand mean over all genes. Bars are ordered by the deviation from the grand mean. Colors indicate the p-value for a two-sample, two-tailed Student's t-test comparing d_N/d_S values for genes annotated as having the indicated molecular function and genes that do not have the indicated annotation among their annotations. White bars indicate non-significant tests at the 0.05 level. Numbers in parentheses indicate the number of genes annotated as having a given molecular function.



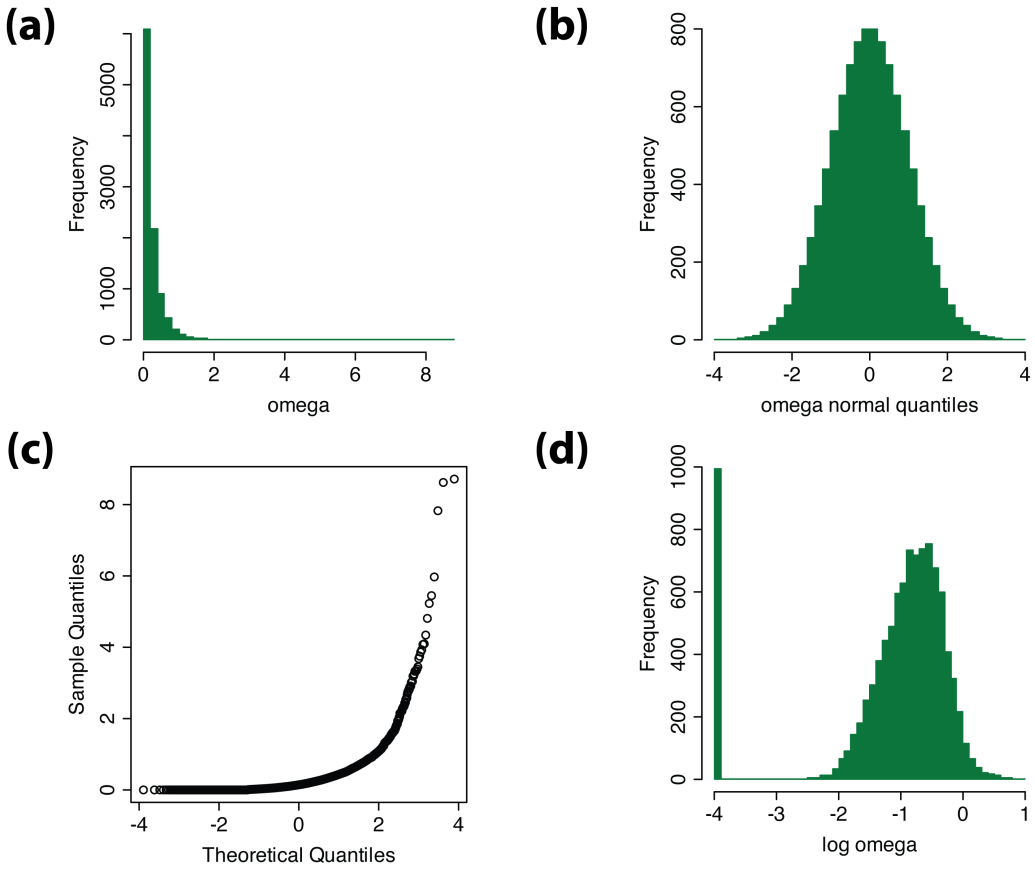
Supplementary Figure 13: Summary of mean density (genes/Mb) values for sets of genes based on the PANTHER classification of molecular functions. The length of each colored bar indicates the deviation of the category mean from the grand mean over all genes. Bars are ordered by the deviation from the grand mean. Colors indicate the p-value for a two-sample, two-tailed Student's t-test comparing density values for genes annotated as having the indicated molecular function and genes that do not have the indicated annotation among their annotations. White bars indicate non-significant tests at the 0.05 level. Numbers in parentheses indicate the number of genes annotated as having a given molecular function.



Supplementary Figure 14: Correlation between d_N/d_S and the broad-scale recombination rate *within* each PANTHER category. The horizontal length of each bar indicates the Spearman rank correlation coefficient between the recombination rate and d_N/d_S for the genes in the indicated PANTHER category. The numbers in parentheses indicates the number of genes in the PANTHER category. Categories are ordered vertically by correlation coefficient. Coloring indicates the significance of the correlation coefficient (uncorrected for multiple tests) i.e., the probability that a correlation coefficient of equal or greater magnitude will be achieved by chance. White bars indicate correlation coefficients that are not significantly different from zero at the 5% level.



Supplementary Figure 15: Correlation between d_N/d_S and the fine-scale recombination rate *within* each PANTHER category. The horizontal length of each bar indicates the Spearman rank correlation coefficient between the recombination rate and d_N/d_S for the genes in the indicated PANTHER category. The numbers in parentheses indicates the number of genes in the PANTHER category. Categories are ordered vertically by correlation coefficient. Coloring indicates the significance of the correlation coefficient (uncorrected for multiple tests) i.e., the probability that a correlation coefficient of equal or greater magnitude will be achieved by chance. White bars indicate correlation coefficients that are not significantly different from zero at the 5% level.



Supplementary Figure 16: Quantile normalization of d_N/d_S. (a) Original distribution of d_N/d_S. (b) Distribution of the quantile-normalized d_N/d_S values. (c) QQ-plot showing the relationship between the empirical and theoretical quantiles. (d) Distribution of log-transformed d_N/d_S values.