SUPPLEMENTARY FIGURES AND TABLES

Table S1: Predictions of diversity in linked neutral regions for two different DFE realizations, as predicted by Equations 3a and 3b. The expected diversity with no selection is 0.02.

DFE	Length of coding region:	Distance from selected site -->		
		100 bp	**10 kb**	**100 kb**
$f_0 = 0; f_1 = 0, f_2 = 50, f_3 = 50$	**1 kb**	0.0174857	0.019805	0.0199807
	5 kb	0.0150363	0.0192131	0.0199054
	10 kb	0.0141392	0.0187112	0.0198152
$f_0 = 0; f_1 = 0; f_2 = 80; f_3 = 20$	**1 kb**	0.0160713	0.0197472	0.0199892
	5 kb	0.0124679	0.0190189	0.0199474
	10 kb	0.0113456	0.0184565	0.0198979

Table S2: Reduction in neutral and linked neutral diversity calculated analytically as a function of the DFE - as illustrated by considering DFE realizations in which one class is largely over-represented, for different exon lengths (0.5kb, 1kb, 5kb, and 10kb). The expected diversity under neutrality is 0.02.

	$f_0 \geq 80\%$	$f_1 \geq 80\%$	$f_2 \geq 80\%$	$f_3 \geq 80\%$
Neutral Diversity (0.5 kb)	0.01995	0.01982	0.01958	0.01975
Linked neutral diversity (0.5kb)	0.01976	0.01809	0.01795	0.01891
Slope of recovery (0.5kb)	0.00007	0.00054	0.00063	0.00013
Neutral Diversity (1 kb)	0.01992	0.01972	0.01926	0.01951
Linked neutral diversity (1kb)	0.01968	0.01756	0.01674	0.01906
Slope of recovery (1kb)	0.00010	0.00072	0.00100	0.00024
Neutral Diversity (5 kb)	0.01972	0.01934	0.01760	0.01814
Linked neutral diversity (5 kb)	0.01915	0.01634	0.01318	0.01714
Slope of recovery (5kb)	0.00024	0.00109	0.00200	0.00061
Neutral Diversity (10kb)	0.01960	0.01909	0.01673	0.01709
Linked neutral diversity (10kb)	0.01903	0.01615	0.01208	0.01579
Slope of recovery (10kb)	0.00027	0.00112	0.00220	0.00083

Table S3: Statistics ranked by their importance in predicting the DFE classes under equilibrium using the correlation coefficients between the statistics and parameters.

Statistics ranked for f_0	r^2	Statistics ranked for f_1	r^2	Statistics ranked for f_2	r^2	Statistics ranked for f_3	r^2
func_div_sd	0.962	func_hprime_m	0.632	neu_thetaw_m	0.477	func_numSing_m	0.850
func_thetah_sd	0.962	func_rsq_sd	0.511	neu_thetapi_m	0.463	func_numSing_sd	0.748
func_thetah_m	0.960	link_tajimasd_m	0.456	pi_intercept	0.425	func_thetaw_m	0.460
func_div_m	0.958	link_Dprime_m	0.441	neu_thetah_m	0.347	func_Dprime_sd	0.432
func_thetapi_sd	0.896	func_Dprime_sd	0.384	pi_slope	0.340	func_thetaw_sd	0.397
func_rsq_m	0.873	pi_max	0.357	link_thetaw_m	0.320	func_hapdiv_m	0.334
func_Dprime_m	0.866	func_D_sd	0.297	func_tajimasd_m	0.309	func_rsq_sd	0.317
func_thetapi_m	0.855	func_tajimasd_sd	0.280	func_rsq_m	0.278	pi_max	0.297
func_D_m	0.847	func_hprime_sd	0.271	link_thetapi_m	0.252	func_hapdiv_sd	0.295
func_tajimasd_m	0.828	link_thetah_m	0.236	neu_div_m	0.240	link_tajimasd_m	0.267
func_hapdiv_sd	0.731	pi_slope	0.226	func_Dprime_m	0.235	neu_rsq_m	0.264
func_thetaw_sd	0.729	link_thetapi_m	0.219	func_hapdiv_m	0.230	func_thetapi_m	0.258
func_hapdiv_m	0.683	func_numSing_m	0.200	link_div_m	0.229	link_hapdiv_sd	0.252
func_thetaw_m	0.663	func_numSing_sd	0.185	link_thetah_m	0.222	link_rsq_sd	0.234
func_hprime_sd	0.578	link_div_m	0.183	func_hapdiv_sd	0.220	link_D_sd	0.230
func_hprime_m	0.553	pi_intercept	0.165	func_thetapi_m	0.215	link_hapdiv_m	0.218
func_D_sd	0.426	link_hapdiv_sd	0.159	func_thetapi_sd	0.210	link_Dprime_sd	0.217
pi_intercept	0.407	neu_rsq_m	0.143	func_D_m	0.195	pi_slope	0.217
neu_thetaw_m	0.404	func_D_m	0.141	func_thetah_m	0.181	func_thetapi_sd	0.208
func_tajimasd_sd	0.381	link_rsq_sd	0.126	func_D_sd	0.175	link_D_m	0.205
neu_thetapi_m	0.357	link_hapdiv_m	0.125	func_tajimasd_sd	0.171	neu_D_m	0.200
pi_slope	0.345	neu_D_m	0.120	func_thetah_sd	0.168	link_Dprime_m	0.191
func_numSing_sd	0.339	link_Dprime_sd	0.119	func_div_m	0.165	func_div_m	0.187
link_thetapi_m	0.326	link_D_sd	0.117	func_div_sd	0.159	link_hprime_sd	0.183
link_thetaw_m	0.320	func_Dprime_m	0.114	func_thetaw_sd	0.158	link_tajimasd_sd	0.182
link_thetah_m	0.315	link_hprime_sd	0.108	func_thetaw_m	0.148	pi_intercept	0.173
func_numSing_m	0.279	func_tajimasd_m	0.104	func_hprime_sd	0.136	func_div_sd	0.173
link_div_m	0.259	link_thetaw_m	0.103	func_rsq_sd	0.128	link_rsq_m	0.172
func_rsq_sd	0.255	link_div_sd	0.099	neu_tajimasd_m	0.091	link_div_sd	0.164
neu_thetah_m	0.251	link_tajimasd_sd	0.098	link_tajimasd_m	0.081	func_thetah_m	0.163
link_Dprime_m	0.220	neu_D_sd	0.096	link_Dprime_m	0.059	link_thetapi_m	0.156
link_tajimasd_m	0.196	link_D_m	0.090	func_Dprime_sd	0.044	link_div_m	0.155
neu_div_m	0.171	neu_rsq_sd	0.086	neu_Dprime_m	0.043	link_thetah_m	0.154
neu_numSing_m	0.076	link_numSing_sd	0.081	link_numSing_m	0.024	neu_rsq_sd	0.153
neu_Dprime_m	0.053	neu_thetaw_m	0.076	neu_numSing_m	0.023	func_thetah_sd	0.153

neu_tajimasd_m	0.049	link_rsq_m	0.076	func_numSing_sd	0.021	link_thetah_sd	0.144
func_Dprime_sd	0.028	link_thetaw_sd	0.072	neu_hapdiv_m	0.011	neu_D_sd	0.142
link_numSing_m	0.027	link_thetah_sd	0.072	link_Dprime_sd	0.010	link_numSing_sd	0.140
neu_rsq_m	0.009	func_thetaw_m	0.062	neu_thetaw_sd	0.010	link_thetaw_sd	0.134
link_thetapi_sd	0.008	link_thetapi_sd	0.061	neu_thetapi_sd	0.009	link_thetapi_sd	0.127
neu_tajimasd_sd	0.008	func_rsq_m	0.058	link_D_m	0.008	func_tajimasd_sd	0.104
link_thetah_sd	0.007	func_hapdiv_m	0.053	link_D_sd	0.008	link_thetaw_m	0.101
neu_Dprime_sd	0.007	neu_hprime_sd	0.047	neu_numSing_sd	0.007	neu_numSing_m	0.098
neu_D_m	0.007	neu_div_sd	0.043	link_rsq_sd	0.007	func_D_sd	0.094
link_thetaw_sd	0.006	neu_Dprime_sd	0.039	link_tajimasd_sd	0.006	neu_div_sd	0.091
link_rsq_m	0.005	neu_thetapi_m	0.037	link_hapdiv_sd	0.005	neu_hprime_sd	0.078
neu_hprime_sd	0.005	neu_numSing_m	0.035	link_div_sd	0.005	neu_tajimasd_sd	0.075
neu_rsq_sd	0.005	neu_tajimasd_sd	0.034	link_rsq_m	0.005	neu_Dprime_sd	0.073
link_hapdiv_m	0.005	neu_thetah_sd	0.033	func_hprime_m	0.005	neu_thetah_sd	0.065
link_D_m	0.004	func_thetah_sd	0.033	link_hprime_sd	0.005	neu_thetaw_sd	0.053
link_numSing_sd	0.003	func_thetaw_sd	0.030	neu_thetah_sd	0.004	neu_thetapi_sd	0.053
neu_D_sd	0.003	link_hprime_m	0.029	func_numSing_m	0.003	neu_thetaw_m	0.050
link_hprime_m	0.003	func_div_sd	0.028	link_hapdiv_m	0.002	neu_numSing_sd	0.037
neu_div_sd	0.003	neu_thetah_m	0.025	neu_div_sd	0.002	neu_hapdiv_m	0.033
link_D_sd	0.002	func_hapdiv_sd	0.024	neu_rsq_m	0.001	func_rsq_m	0.028
link_rsq_sd	0.002	neu_thetapi_sd	0.024	link_numSing_sd	0.001	neu_tajimasd_m	0.019
neu_hapdiv_m	0.002	func_thetah_m	0.023	pi_numbp50	0.001	func_hprime_sd	0.016
link_tajimasd_sd	0.001	neu_thetaw_sd	0.020	pi_numbp75	0.001	func_hprime_m	0.014
pi_max	0.001	func_div_m	0.020	pi_numbp90	0.001	neu_Dprime_m	0.014
link_hapdiv_sd	0.001	neu_div_m	0.018	link_thetah_sd	0.001	func_Dprime_m	0.012
link_hprime_sd	0.001	neu_hapdiv_m	0.015	neu_rsq_sd	0.001	neu_thetapi_m	0.012
neu_thetapi_sd	0.001	neu_numSing_sd	0.015	link_thetaw_sd	0.000	func_D_m	0.011
link_Dprime_sd	0.001	neu_Dprime_m	0.008	link_thetapi_sd	0.000	link_hprime_m	0.009
neu_hapdiv_sd	0.000	link_numSing_m	0.003	link_hprime_m	0.000	neu_thetah_m	0.005
link_div_sd	0.000	neu_tajimasd_m	0.003	neu_hprime_m	0.000	link_numSing_m	0.004
neu_hprime_m	0.000	func_thetapi_m	0.002	neu_D_m	0.000	neu_div_m	0.004
neu_thetah_sd	0.000	neu_hapdiv_sd	0.002	pi_max	0.000	neu_hapdiv_sd	0.004
neu_numSing_sd	0.000	func_thetapi_sd	0.001	neu_Dprime_sd	0.000	pi_numbp50	0.003
neu_thetaw_sd	0.000	pi_numbp50	0.001	neu_hprime_sd	0.000	pi_numbp75	0.003
pi_numbp50	0.000	pi_numbp75	0.001	neu_D_sd	0.000	pi_numbp90	0.003
pi_numbp75	0.000	pi_numbp90	0.001	neu_hapdiv_sd	0.000	neu_hprime_m	0.001
pi_numbp90	0.000	neu_hprime_m	0.000	neu_tajimasd_sd	0.000	func_tajimasd_m	0.001

Table S4: Statistics ranked by their importance in predicting the DFE classes under equilibrium using a modified algorithm of Joyce and Marjoram (2008) and by averaging the ranking across 10 replicates for each parameter separately.

Statistics ranked for f_0	Avg rank	Statistics ranked for f_1	Avg rank	Statistics ranked for f_2	Avg rank	Statistics ranked for f_3	Avg rank
func_div_m	1.2	func_thetah_m	4.8	func_thetaw_m	2.8	func_thetaw_m	4.3
func_thetah_m	4.7	func_thetaw_m	6.1	func_thetah_m	4.7	func_thetah_m	5.2
func_thetaw_m	8.5	link_thetaw_m	7.9	link_thetaw_m	6.2	neu_thetapi_m	8.2
neu_thetapi_m	13.8	neu_thetapi_m	9.6	neu_thetapi_m	8	link_thetaw_m	10.6
func_hprime_m	14.7	link_thetapi_m	11	func_numSing_m	9.1	func_numSing_m	12.8
func_tajimasd_m	17.6	neu_thetaw_m	14.2	link_thetapi_m	11.4	neu_thetaw_m	12.8
link_hapdiv_m	18.7	func_numSing_m	17.4	link_thetah_m	12.7	link_thetapi_m	14.3
link_thetaw_m	18.7	link_hprime_m	18.8	neu_thetah_m	18.9	func_hprime_m	15.8
func_rsq_m	18.9	pi_max	19.9	func_Dprime_sd	21.5	link_hprime_m	18.3
link_thetapi_m	19.8	pi_numbp90	20.6	func_thetapi_m	21.5	pi_intercept	18.5
link_tajimasd_m	20	func_thetapi_m	20.7	link_div_sd	22.7	func_Dprime_sd	19.5
func_numSing_m	20.3	func_div_m	22	func_div_sd	23.2	pi_max	19.8
neu_rsq_m	20.7	pi_numbp50	23	pi_intercept	23.8	link_thetah_m	20
link_rsq_m	21	link_thetah_m	23.2	pi_numbp90	24.1	pi_numbp90	20.1
func_hapdiv_m	23	func_hprime_m	23.9	pi_slope	24.4	neu_Dprime_sd	21
neu_numSing_m	23	pi_intercept	24.3	neu_Dprime_sd	24.6	pi_numbp75	22.1
func_D_m	25.3	func_Dprime_sd	24.7	link_Dprime_sd	24.8	func_thetapi_m	24.2
link_thetah_m	25.3	func_div_sd	24.8	neu_thetaw_m	24.8	neu_hprime_m	24.5
func_thetapi_m	26.2	pi_slope	26.5	link_hprime_m	25.6	link_hapdiv_sd	24.7
link_hprime_m	26.7	func_tajimasd_m	26.9	func_numSing_sd	26.9	pi_numbp50	24.9
func_Dprime_sd	26.8	neu_hprime_m	27.5	link_D_sd	27.1	neu_D_sd	25
link_D_m	27.3	func_D_sd	27.7	pi_numbp75	27.4	neu_thetah_m	26.3
neu_hprime_m	28	link_div_sd	27.8	func_hprime_m	28.3	func_D_sd	26.4
pi_numbp75	28.4	neu_thetah_m	29.7	pi_max	28.9	func_tajimasd_m	26.5
pi_intercept	28.8	link_D_sd	31.8	link_numSing_sd	29	pi_slope	26.5
func_div_sd	29.1	neu_rsq_m	32	pi_numbp50	29	link_div_sd	28.6
link_numSing_m	29.6	link_tajimasd_m	32.9	neu_D_sd	30.1	neu_div_sd	29.2
neu_D_m	29.7	neu_div_sd	33.6	func_rsq_sd	30.4	func_div_sd	30.1
neu_thetah_m	30	pi_numbp75	34.1	neu_rsq_sd	31.7	func_rsq_sd	31.1
pi_max	30.7	link_rsq_sd	34.5	neu_div_sd	32.2	neu_rsq_sd	32.7
neu_rsq_sd	31.5	neu_D_sd	34.6	func_D_sd	32.7	neu_tajimasd_sd	33.6
neu_thetaw_m	32.5	func_D_m	34.9	func_tajimasd_m	33.6	link_rsq_sd	33.7
func_Dprime_m	36	link_Dprime_sd	35.6	neu_hprime_m	34.6	neu_hapdiv_sd	33.7
func_D_sd	36.5	func_rsq_m	35.9	func_hapdiv_sd	35.2	link_Dprime_sd	34.3
pi_numbp90	38.2	link_hapdiv_m	36	func_thetah_sd	36.1	link_D_sd	37.2

link_Dprime_sd	38.4	func_numSing_sd	36.4	link_hapdiv_sd	36.4	func_numSing_sd	37.6
link_hapdiv_sd	38.6	neu_Dprime_sd	37.5	link_thetapi_sd	36.4	neu_numSing_sd	38.8
func_hapdiv_sd	38.9	link_rsq_m	39.4	neu_hapdiv_sd	36.5	link_tajimasd_m	40.8
pi_numbp50	39.1	neu_hapdiv_sd	40.7	func_hprime_sd	38	neu_thetah_sd	41.2
neu_D_sd	39.2	neu_numSing_sd	41.3	func_tajimasd_sd	38.1	link_thetaw_sd	41.3
link_rsq_sd	40.3	neu_numSing_m	42.1	link_hprime_sd	40.7	func_hapdiv_sd	41.7
link_div_sd	40.6	link_D_m	42.7	func_div_m	41.1	link_numSing_sd	42
func_rsq_sd	41.9	func_Dprime_m	43	neu_tajimasd_sd	41.1	func_thetapi_sd	42.8
pi_slope	43.6	link_thetaw_sd	44.3	link_rsq_sd	41.7	link_hprime_sd	42.8
link_numSing_sd	43.8	link_hapdiv_sd	44.4	neu_numSing_sd	42.1	func_thetaw_sd	44.8
neu_Dprime_sd	43.8	neu_hprime_sd	44.7	link_tajimasd_sd	42.7	func_rsq_m	44.9
neu_hapdiv_sd	44.2	func_tajimasd_sd	44.9	neu_thetaw_sd	43.2	neu_hprime_sd	45
link_tajimasd_sd	44.5	link_hprime_sd	45.1	neu_thetapi_sd	43.6	func_Dprime_m	45.1
func_numSing_sd	44.7	func_hapdiv_sd	45.4	func_D_m	44.6	func_hprime_sd	45.3
link_Dprime_m	44.7	func_rsq_sd	45.5	neu_hprime_sd	44.8	func_tajimasd_sd	46.5
neu_numSing_sd	45	func_thetaw_sd	46.7	link_div_m	46.5	link_tajimasd_sd	46.8
neu_hapdiv_m	45.9	link_tajimasd_sd	46.7	link_Dprime_m	47.7	func_div_m	47.3
neu_tajimasd_m	46.9	neu_rsq_sd	46.7	link_hapdiv_m	48.5	link_thetah_sd	48.2
func_tajimasd_sd	47.2	neu_thetah_sd	47.4	func_Dprime_m	49.7	func_thetah_sd	48.4
link_D_sd	47.9	link_Dprime_m	47.5	neu_div_m	50	func_D_m	48.8
neu_div_sd	48.4	func_thetah_sd	47.8	link_rsq_m	50.1	link_D_m	48.9
func_thetaw_sd	49.2	neu_tajimasd_sd	47.9	neu_Dprime_m	50.1	neu_thetaw_sd	49.6
link_thetaw_sd	49.6	link_numSing_sd	48.2	func_thetaw_sd	50.5	link_div_m	50.1
neu_tajimasd_sd	49.8	func_hprime_sd	49.2	func_thetapi_sd	51.1	neu_D_m	50.4
func_thetapi_sd	51.3	neu_thetaw_sd	49.4	link_thetaw_sd	51.3	neu_div_m	51.1
func_hprime_sd	52.7	neu_D_m	49.5	link_thetah_sd	51.4	link_rsq_m	51.3
neu_Dprime_m	52.8	func_hapdiv_m	49.9	func_hapdiv_m	51.8	neu_rsq_m	54
neu_thetaw_sd	55	link_thetapi_sd	50.1	neu_D_m	52.9	link_thetapi_sd	54.2
neu_thetah_sd	56.3	link_div_m	51.2	neu_thetah_sd	53	neu_thetapi_sd	54.8
neu_hprime_sd	56.7	func_thetapi_sd	51.7	func_rsq_m	53.3	neu_Dprime_m	56.5
neu_div_m	57.1	link_thetah_sd	52	link_tajimasd_m	54.1	link_Dprime_m	57.4
link_hprime_sd	57.5	neu_thetapi_sd	53.8	link_D_m	56.5	neu_tajimasd_m	58.4
link_thetah_sd	58.6	neu_div_m	56.4	neu_rsq_m	58.3	func_hapdiv_m	59.4
neu_thetapi_sd	59.5	neu_tajimasd_m	57.5	neu_numSing_m	63.4	link_numSing_m	61.2
link_div_m	60	link_numSing_m	57.9	neu_tajimasd_m	65.3	neu_numSing_m	62.5
func_thetah_sd	60.6	neu_Dprime_m	59.9	link_numSing_m	65.6	link_hapdiv_m	62.6
link_thetapi_sd	62.5	neu_hapdiv_m	64.3	neu_hapdiv_m	67.9	neu_hapdiv_m	69

Table S5: Ranking of statistics under demographic non-equilibrium. Statistics significantly correlated with parameters of the DFE when statistics from all regions are used and when only functional statistics are used for ranking. Significance was evaluated with $p < 0.05$ with Bonferonni corrections.

Ranking using all statistics

Statistics ranked for f_0	r^2	Statistics ranked for f_1	r^2	Statistics ranked for f_2	r^2	Statistics ranked for f_3	r^2
func_div_m	0.893	func_hprime_m	0.273	func_div_sd	0.125	func_numSing_sd	0.205
func_div_sd	0.850	func_tajimasd_sd	0.121	func_div_m	0.125	func_thetaw_sd	0.200
func_thetah_sd	0.632	func_hprime_sd	0.099	func_thetapi_sd	0.117	func_thetaw_m	0.180
func_thetapi_sd	0.612	func_rsq_sd	0.082	func_thetapi_m	0.114	func_numSing_m	0.180
func_thetah_m	0.585	func_Dprime_sd	0.079	func_thetaw_sd	0.107	func_div_m	0.117
func_thetapi_m	0.556	func_tajimasd_m	0.077	func_thetaw_m	0.098	func_hapdiv_sd	0.114
func_thetaw_sd	0.473	func_div_m	0.065	func_thetah_sd	0.094	func_div_sd	0.108
func_thetaw_m	0.407	func_Dprime_m	0.060	func_hapdiv_sd	0.092	func_hapdiv_m	0.106
func_Dprime_m	0.343	func_div_sd	0.059	func_thetah_m	0.085	func_thetapi_sd	0.104
func_tajimasd_m	0.325	func_numSing_sd	0.056	func_hapdiv_m	0.079	func_thetapi_m	0.102
func_hprime_m	0.286	func_numSing_m	0.056	func_Dprime_m	0.075	func_Dprime_sd	0.077
func_hapdiv_sd	0.284	func_thetah_m	0.053	func_tajimasd_m	0.072	func_thetah_sd	0.074
func_hapdiv_m	0.209	func_thetah_sd	0.048	func_rsq_m	0.045	func_tajimasd_sd	0.062
func_numSing_sd	0.143	func_D_sd	0.034	func_numSing_sd	0.029	func_thetah_m	0.061
func_rsq_m	0.142	func_hapdiv_m	0.020	func_numSing_m	0.020	func_rsq_sd	0.023
func_hprime_sd	0.116	func_thetapi_sd	0.015	func_D_m	0.015	func_hprime_sd	0.010
func_numSing_m	0.102	func_D_m	0.014	func_hprime_sd	0.015	func_D_sd	0.008
func_D_m	0.081	func_hapdiv_sd	0.010	func_rsq_sd	0.010	func_rsq_m	0.005
func_rsq_sd	0.057	func_thetapi_m	0.009	func_D_sd	0.007		
func_D_sd	0.033	func_rsq_m	0.009				
func_tajimasd_sd	0.023	func_thetaw_m	0.008				
		func_thetaw_sd	0.006				

Ranking using statistics calculated from functional regions.

Statistics ranked for f_0	r^2	Statistics ranked for f_1	r^2	Statistics ranked for f_2	r^2	Statistics ranked for f_3	r^2
func_div_m	0.89	func_hprime_m	0.27	func_div_sd	0.12	func_numSing_sd	0.20
func_div_sd	0.85	func_tajimasd_sd	0.12	func_div_m	0.12	func_thetaw_sd	0.20
func_thetah_sd	0.63	func_hprime_sd	0.10	func_thetapi_sd	0.12	func_thetaw_m	0.18
func_thetapi_sd	0.61	func_rsq_sd	0.08	func_thetapi_m	0.11	func_numSing_m	0.18
func_thetah_m	0.59	func_Dprime_sd	0.08	func_thetaw_sd	0.11	func_div_m	0.12
func_thetapi_m	0.56	func_tajimasd_m	0.08	func_thetaw_m	0.10	func_hapdiv_sd	0.11
func_thetaw_sd	0.47	func_div_m	0.06	func_thetah_sd	0.09	func_div_sd	0.11
func_thetaw_m	0.41	func_Dprime_m	0.06	func_hapdiv_sd	0.09	func_hapdiv_m	0.11
func_Dprime_m	0.34	func_div_sd	0.06	func_thetah_m	0.08	func_thetapi_sd	0.10
func_tajimasd_m	0.32	func_numSing_sd	0.06	func_hapdiv_m	0.08	func_thetapi_m	0.10
func_hprime_m	0.29	func_numSing_m	0.06	func_Dprime_m	0.07	func_Dprime_sd	0.08
func_hapdiv_sd	0.28	func_thetah_m	0.05	func_tajimasd_m	0.07	func_thetah_sd	0.07
func_hapdiv_m	0.21	func_thetah_sd	0.05	func_rsq_m	0.04	func_tajimasd_sd	0.06

func_numSing_sd	0.14	func_D_sd	0.03	func_numSing_sd	0.03	func_thetah_m	0.06
func_rsq_m	0.14	func_hapdiv_m	0.02	func_numSing_m	0.02	func_rsq_sd	0.02
func_hprime_sd	0.12	func_thetapi_sd	0.01	func_D_m	0.02	func_hprime_sd	0.01
func_numSing_m	0.10	func_D_m	0.01	func_hprime_sd	0.01	func_D_sd	0.01
func_D_m	0.08	func_hapdiv_sd	0.01	func_rsq_sd	0.01	func_rsq_m	0.01
func_rsq_sd	0.06	func_thetapi_m	0.01	func_D_sd	0.01	func_Dprime_m	0.00
func_D_sd	0.03	func_rsq_m	0.01				
func_tajimasd_sd	0.02	func_thetaw_m	0.01				
		func_thetaw_sd	0.01				

Table S6: Ranking of statistics when distinguishing between demography and purifying selection. Statistics significantly correlated with parameters of demography when statistics from all regions are used, and when only functional statistics are used for ranking. Significance was evaluated with $p < 0.05$ with Bonferonni correction.

Ranking using all statistics				**Ranking using statistics calculated from functional regions.**			
Statistics ranked for N_{anc}	**r^2**	**Statistics ranked for N_{cur}**	**r^2**	**Statistics ranked for N_{anc}**	**r^2**	**Statistics ranked for N_{cur}**	**r^2**
neu_thetah_m	0.99	neu_hapdiv_m	0.92	func_thetapi_m	0.230	func_rsq_sd	0.659
link_thetah_m	0.99	link_hapdiv_m	0.91	func_thetah_m	0.229	func_rsq_m	0.618
neu_thetapi_m	0.93	link_numSing_m	0.84	func_thetapi_sd	0.196	func_numSing_m	0.577
link_thetapi_m	0.93	neu_numSing_m	0.84	func_thetaw_sd	0.192	func_D_sd	0.557
link_thetapi_sd	0.93	neu_rsq_m	0.83	func_thetah_sd	0.179	func_tajimasd_sd	0.487
link_thetah_sd	0.93	link_rsq_m	0.82	func_thetaw_m	0.159	func_numSing_sd	0.482
neu_thetah_sd	0.91	neu_rsq_sd	0.79	func_Dprime_m	0.136	func_Dprime_sd	0.435
neu_thetapi_sd	0.91	link_rsq_sd	0.79	func_tajimasd_m	0.135	func_D_m	0.374
neu_thetaw_sd	0.90	neu_hprime_sd	0.78	func_hapdiv_sd	0.121	func_hprime_sd	0.368
link_thetaw_sd	0.90	link_hprime_sd	0.78	func_hapdiv_m	0.078	func_tajimasd_m	0.275
neu_thetaw_m	0.72	link_hapdiv_sd	0.76	func_hprime_m	0.069	func_Dprime_m	0.216
link_thetaw_m	0.71	neu_hapdiv_sd	0.74	func_numSing_sd	0.042	func_hapdiv_m	0.165
link_div_sd	0.49	link_D_sd	0.68	func_rsq_m	0.041	func_thetaw_m	0.157
neu_div_sd	0.45	func_rsq_sd	0.66	func_Dprime_sd	0.032	func_hapdiv_sd	0.129
neu_Dprime_m	0.45	link_numSing_sd	0.64	func_numSing_m	0.025	func_thetaw_sd	0.070
link_Dprime_m	0.44	neu_numSing_sd	0.64	func_rsq_sd	0.012	func_hprime_m	0.061
link_tajimasd_m	0.43	link_tajimasd_sd	0.64	func_tajimasd_sd	0.009	func_div_m	0.022
neu_tajimasd_m	0.43	neu_D_sd	0.64	func_D_m	0.008	func_div_sd	0.017
neu_Dprime_sd	0.41	func_rsq_m	0.62			func_thetapi_m	0.008
link_Dprime_sd	0.38	neu_tajimasd_sd	0.62				
link_hprime_m	0.35	neu_div_m	0.59				
neu_hprime_m	0.34	link_D_m	0.58				
link_div_m	0.31	func_numSing_m	0.58				
neu_div_m	0.30	link_div_m	0.57				
neu_numSing_sd	0.25	func_D_sd	0.56				
link_numSing_sd	0.25	neu_D_m	0.56				
func_thetapi_m	0.23	func_tajimasd_sd	0.49				
func_thetah_m	0.23	func_numSing_sd	0.48				
func_thetapi_sd	0.20	func_Dprime_sd	0.44				
neu_tajimasd_sd	0.19	link_tajimasd_m	0.43				
func_thetaw_sd	0.19	neu_tajimasd_m	0.43				

func_thetah_sd	0.18	link_Dprime_sd	0.39
link_tajimasd_sd	0.16	neu_Dprime_sd	0.38
func_thetaw_m	0.16	func_D_m	0.37
func_Dprime_m	0.14	func_hprime_sd	0.37
func_tajimasd_m	0.14	link_Dprime_m	0.35
func_hapdiv_sd	0.12	neu_Dprime_m	0.35
neu_numSing_m	0.11	neu_hprime_m	0.34
link_numSing_m	0.11	link_hprime_m	0.34
link_rsq_sd	0.08	func_tajimasd_m	0.27
neu_rsq_sd	0.08	link_thetaw_m	0.22
func_hapdiv_m	0.08	func_Dprime_m	0.22
link_rsq_m	0.08	neu_thetaw_m	0.21
func_hprime_m	0.07	func_hapdiv_m	0.16
neu_rsq_m	0.07	func_thetaw_m	0.16
func_numSing_sd	0.04	neu_div_sd	0.15
func_rsq_m	0.04	link_div_sd	0.14
func_Dprime_sd	0.03	func_hapdiv_sd	0.13
func_numSing_m	0.03	func_thetaw_sd	0.07
func_rsq_sd	0.01	func_hprime_m	0.06
link_D_m	0.01	link_thetaw_sd	0.04
link_hapdiv_sd	0.01	neu_thetaw_sd	0.03
link_hapdiv_m	0.01	link_thetapi_m	0.03
func_tajimasd_sd	0.01	neu_thetapi_m	0.03
func_D_m	0.01	func_div_m	0.02
neu_D_m	0.01	func_div_sd	0.02
		neu_thetapi_sd	0.01
		func_thetapi_m	0.01
		neu_thetah_sd	0.01
		link_thetapi_sd	0.01

Table S7: The mean numbers of fixed differences per site (*i.e.*, polymorphism-adjusted divergence) for different site types in *D. melanogaster*, where different numbers of individuals from the Zambia population were used to identify the set of polymorphic sites.

	Sample size:					
	1	**2**	**5**	**15**	**30**	**76**
exon	0.0238	0.0198	0.0170	0.0160	0.0159	0.0153
coding	0.0228	0.0182	0.0157	0.0146	0.0141	0.0135
4-fold degenerate	0.0497	0.0423	0.0349	0.0316	0.0311	0.0300
0-fold degenerate	0.0182	0.0123	0.0108	0.0102	0.0098	0.0094

Table S8: The increase in divergence values obtained when calculating pairwise divergence (corresponding to a sample size of 1) relative to alternate sample sizes (which exclude polymorphic sites from divergence).

	Sample size:		
	1	**76**	**100**
***D. melanogaster* exon**	1.551	1.000	
***D. melanogaster* 4-fold degenerate**	1.658	1.000	
Simulated exon	1.736		1.000
Simulated neutral	1.634		1.000

Table S9: Inference of the DFE in 94 exons of *D. melanogaster*. Our inference is only comparable to that of Huber *et al*. (2017) for two classes of s – less than and greater than 10^{-4}.

$\|s\|$	**0-10^{-4}**	**$\geq 10^{-4}$**
Huber *et al.* 2017 - Gamma distribution	0.767	0.233
Huber *et al.* 2017 - Lognormal distribution	0.667	0.333
This study	0.781	0.219

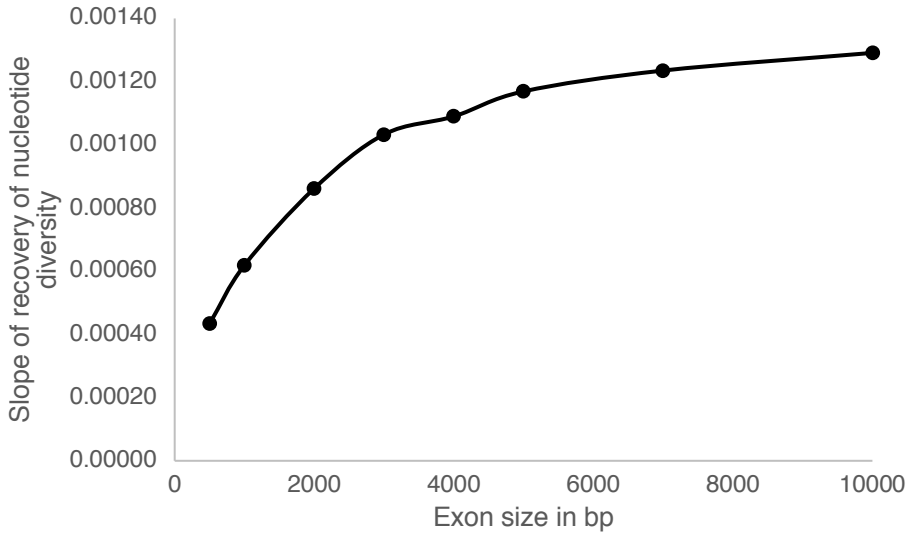


Figure S1: Increase in the slope of recovery of diversity near functional regions of varying sizes as observed via simulations. Larger values of the slope represent a steeper recovery, concordant with larger reduction in diversity observed in the non-coding region.

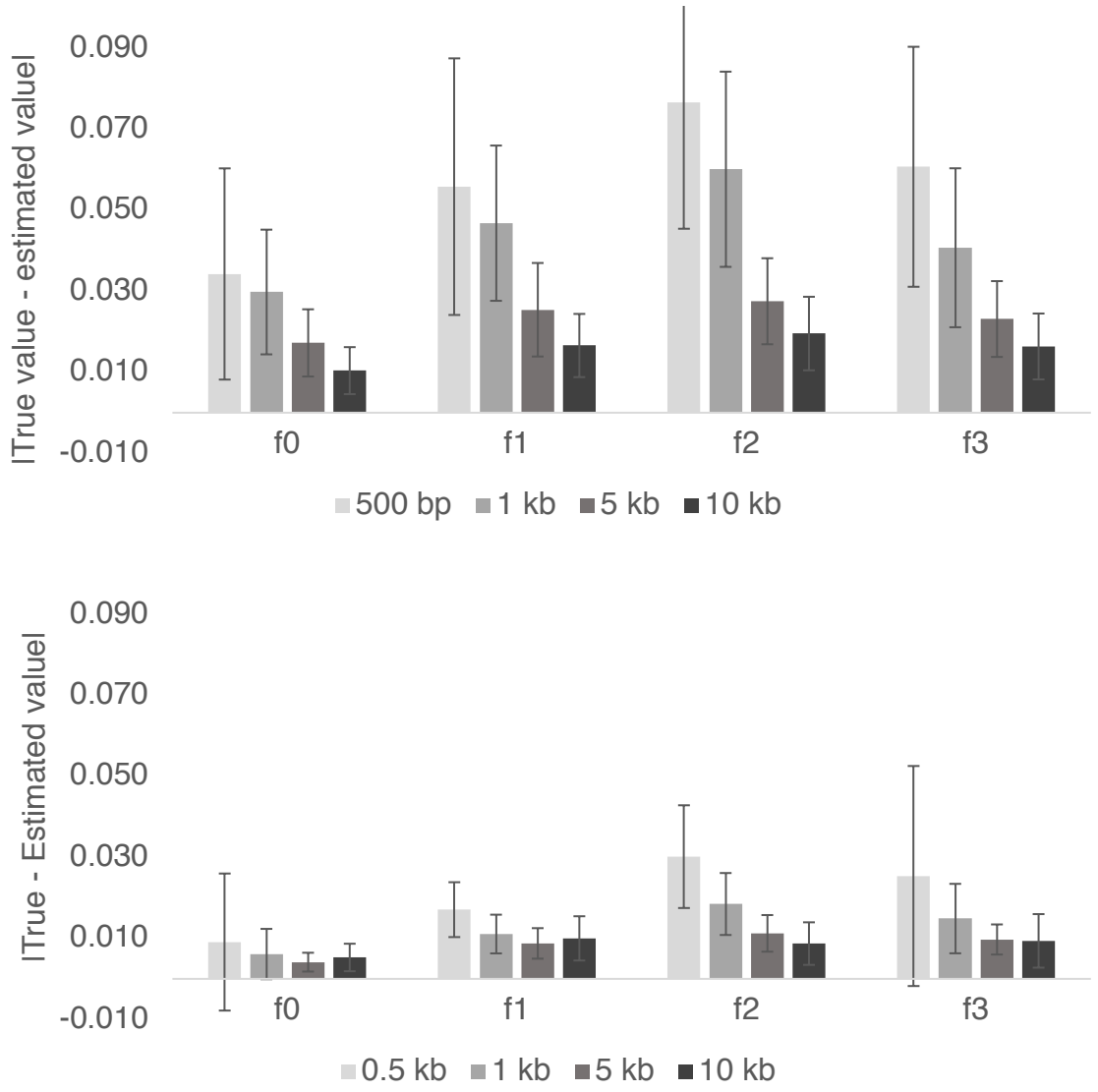


Figure S2: Absolute difference between true and inferred value of parameters characterizing the DFE for lengths of 0.5 kb, 1 kb, 5 kb, and 10 kb of functional regions. The upper panel displays the error in inference when using all statistics, while the lower uses only functional regions.

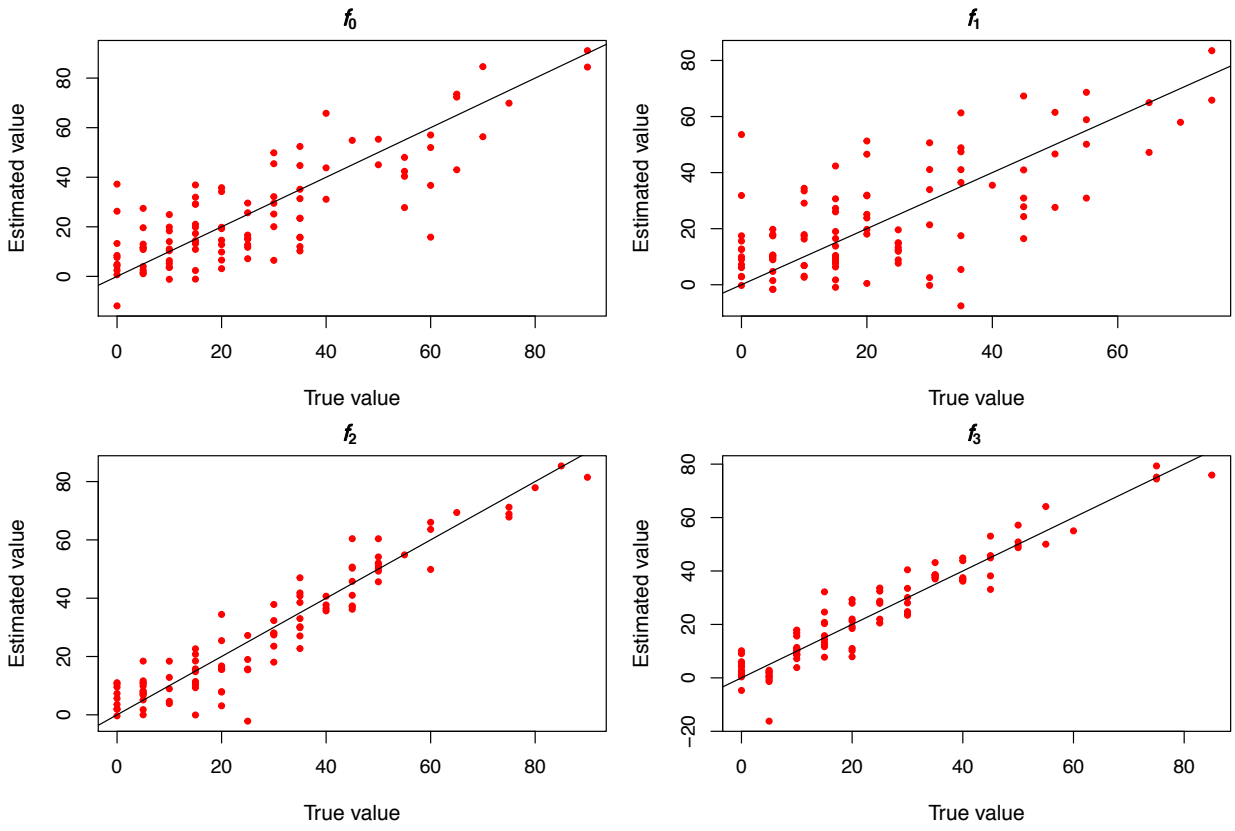


Figure S3: Inference of the DFE under demographic equilibrium using only statistics in the linked neutral regions. The length of the functional region is 10kb.

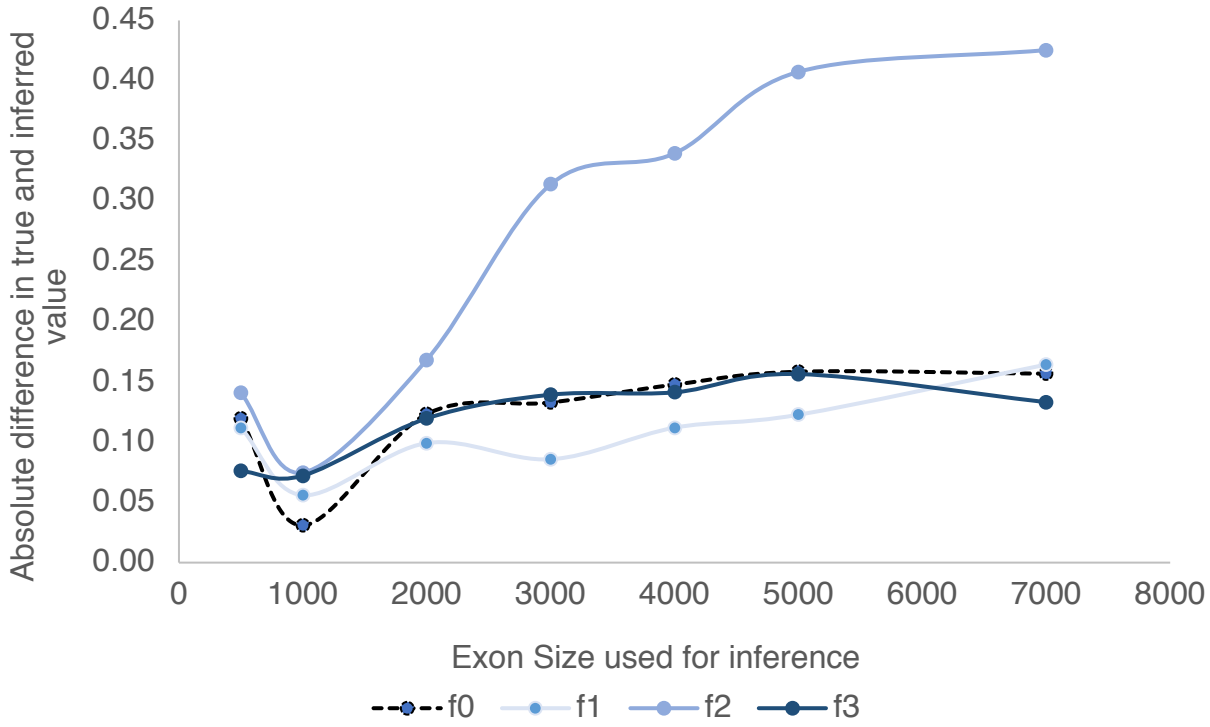


Figure S4: Decrease in accuracy of inference for different DFE classes as the exon size assumed for inference is mis-specified. In this figure, the assumed exon size was 1kb, and the X-axis gives the true exon size.

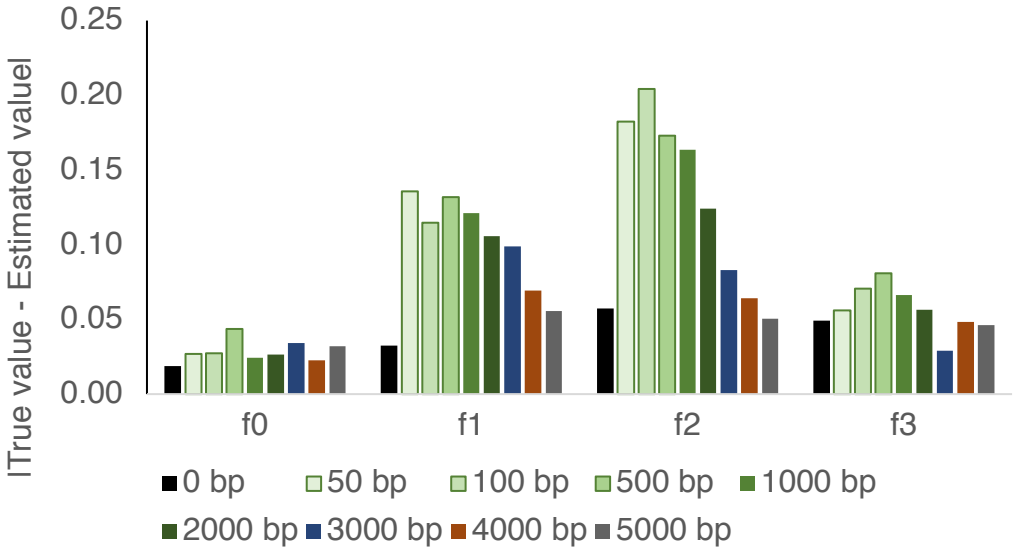


Figure S5: Mis-inference of the DFE in the presence of an additional unaccounted for 1 kb functional region near the target 1kb exon used for inference. The intronic / intergenic distance between the two exons varies from 50-5000 bp, as shown by different colored bars. "0 bp" represents the negative control in which there is no additional 1 kb exon present.

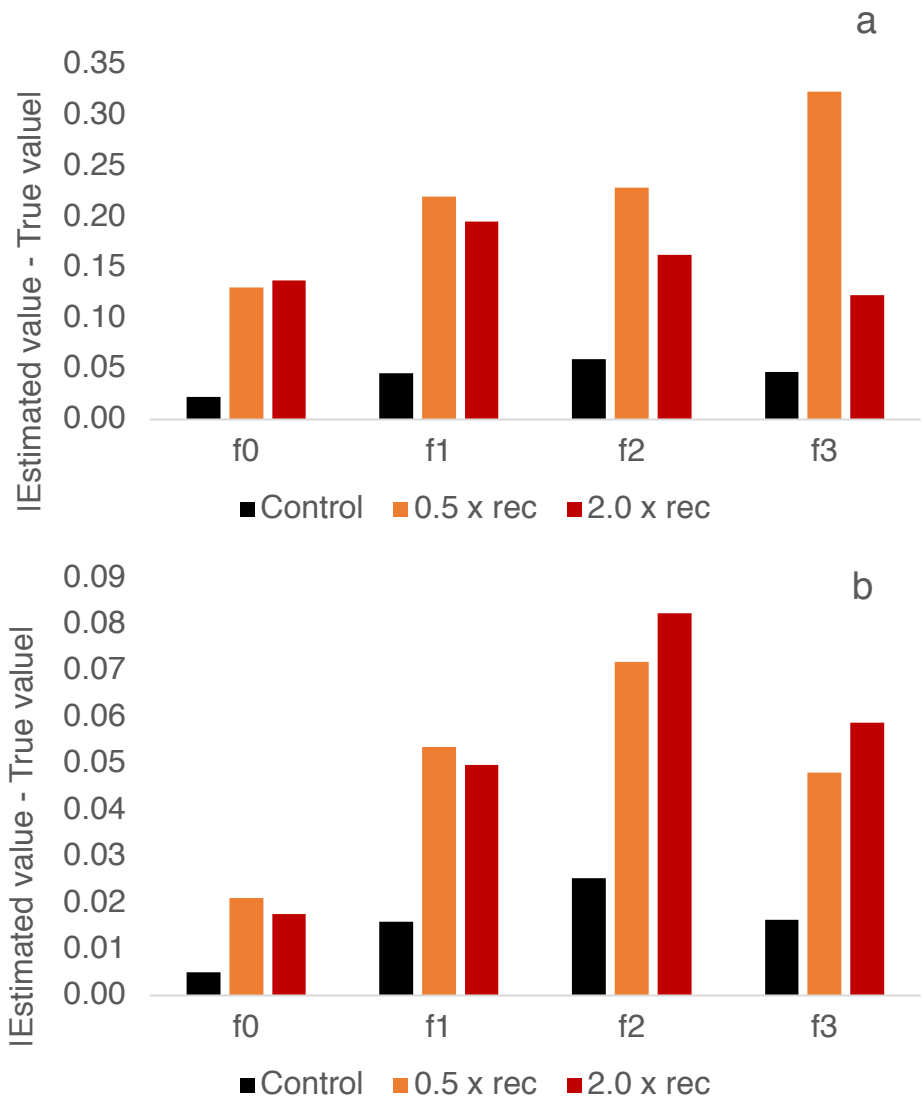


Figure S6: Absolute differences between the true and estimated value of the DFE class, when the true recombination rate is half of that assumed for inference (orange) and when the true value is twice that assumed for inference (red), using a) all statistics and b) statistics only pertaining to the functional region.

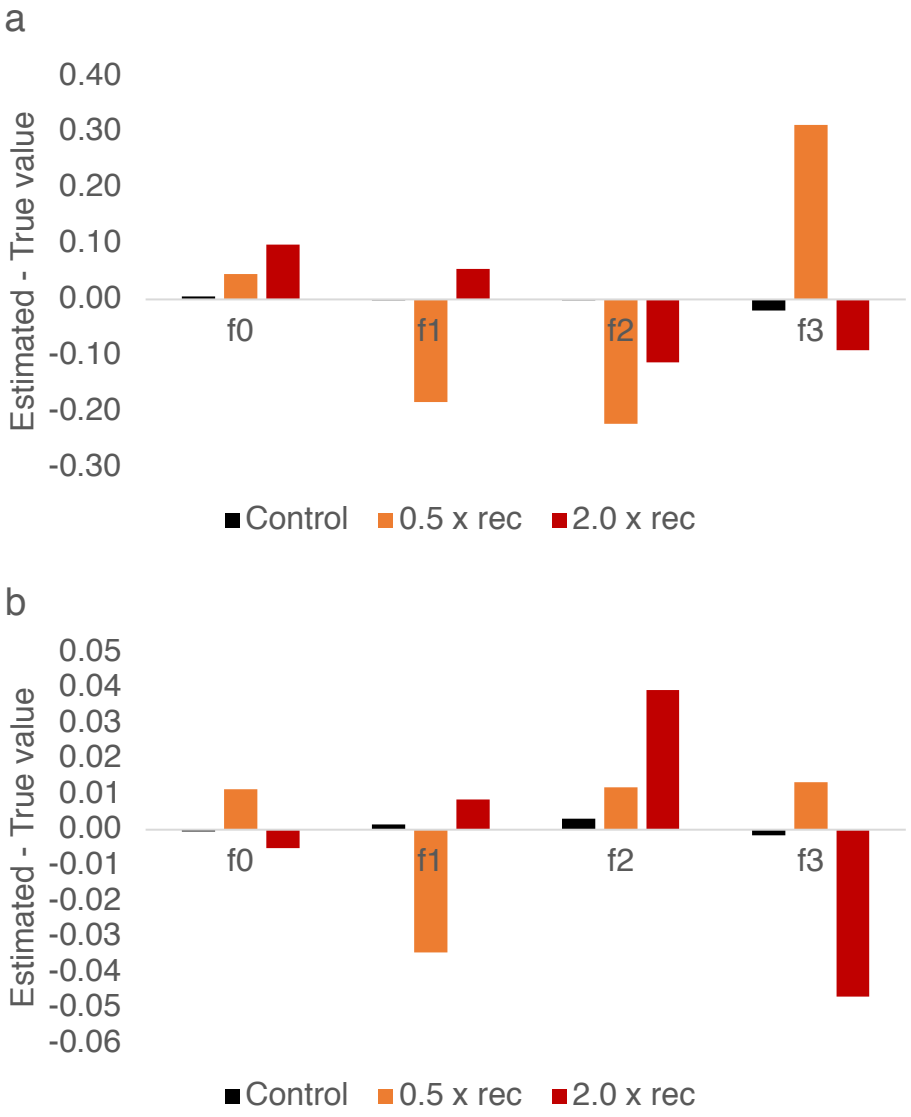


Figure S7: Following Supp Figure 6, the direction of bias in inference of the DFE classes upon mis-specification of the recombination rate is shown, using a) all statistics and b) statistics only pertaining to the functional region.

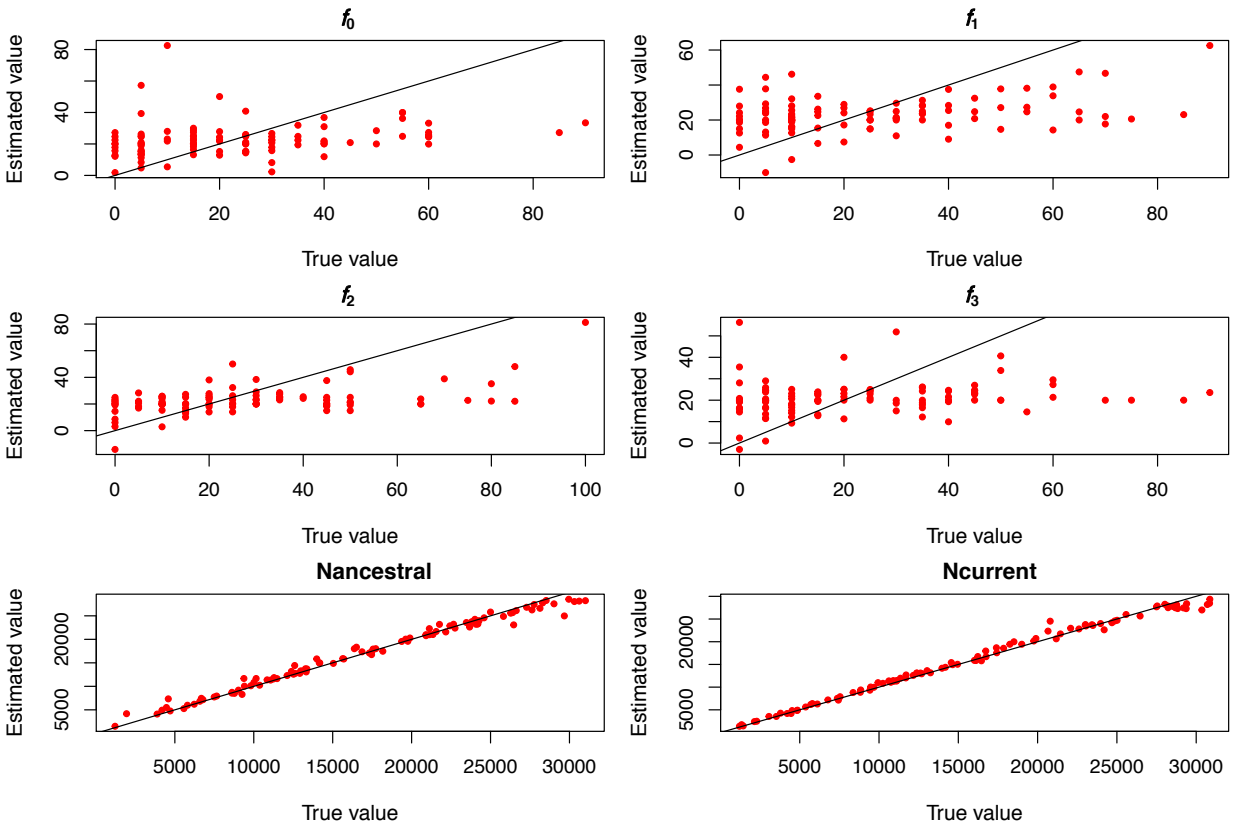


Figure S8: Joint inference of the DFE and demography using statistics only in linked regions (22 summary statistics).

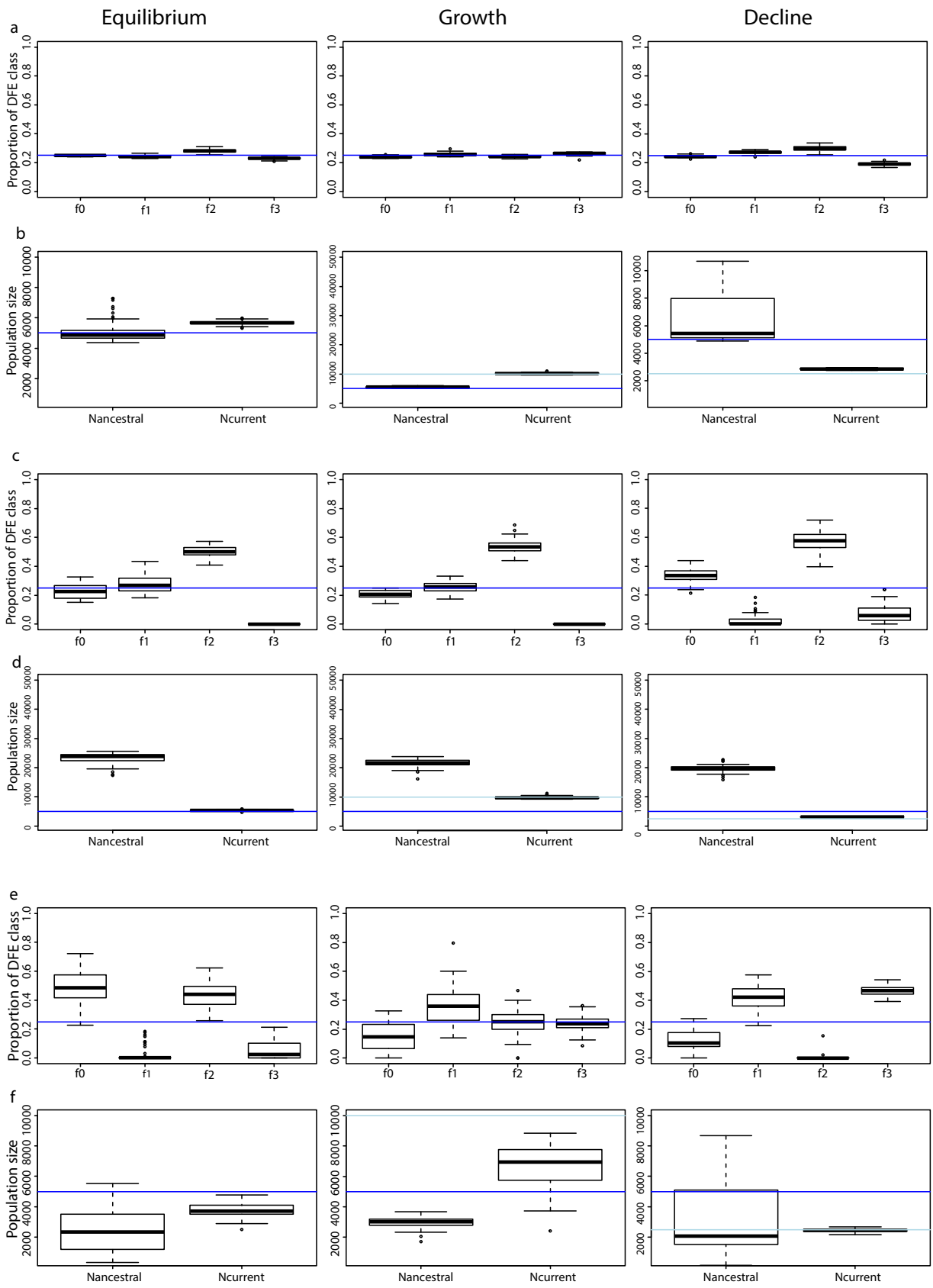

Equilibrium
Growth
Decline
a
b
c
d
e
f
Proportion of DFE class
Population size
f0
f1
f2
f3
Nancestral
Ncurrent

Figure S9: Effect of mis-specification of mutation rate on joint inference of the DFE and demography. (a) Inference of the DFE under equilibrium, 2-fold growth and 2-fold decline, when mutation rate is the same as assumed. Blue line shows the true value. (b) Inference of ancestral and current population sizes under equilibrium, 2-fold growth and 2-fold decline, when mutation rate is the same as assumed. Darker blue and lighter blue lines show the true values of ancestral and current population sizes respectively. (c) Inference of the DFE under equilibrium, 2-fold growth and 2-fold decline when the true mutation rate is twice that assumed. (d) Inference of ancestral and current population sizes under equilibrium, 2-fold growth and 2-fold decline when the true mutation rate is twice that assumed. (e) Inference of the DFE under equilibrium, 2-fold growth and 2-fold decline when the true mutation rate is half of that assumed. (f) Inference of ancestral and current population sizes under equilibrium, 2-fold growth and 2-fold decline when the true mutation rate is half of that assumed.

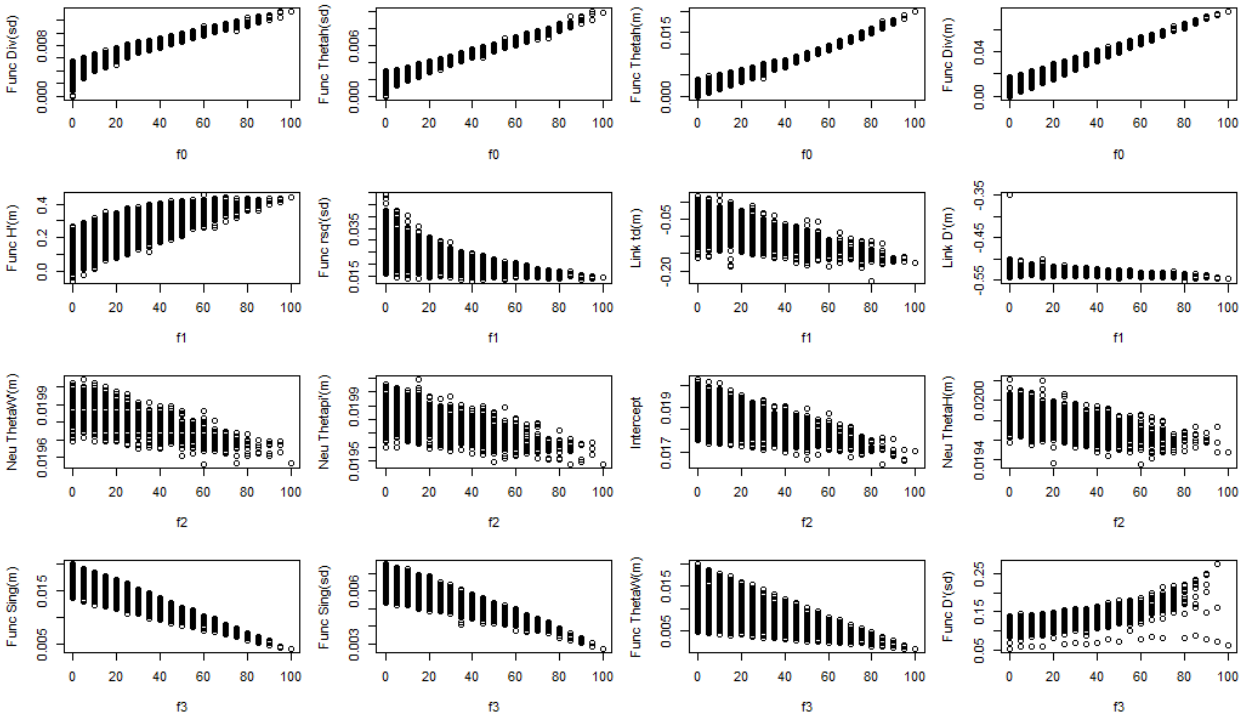


Figure S10: Correlations of the top 4 statistics with parameters characterizing the DFE under demographic equilibrium. “Func” corresponds to the functional region, “Link” to the immediately linked region and “Neu” to the less linked region.

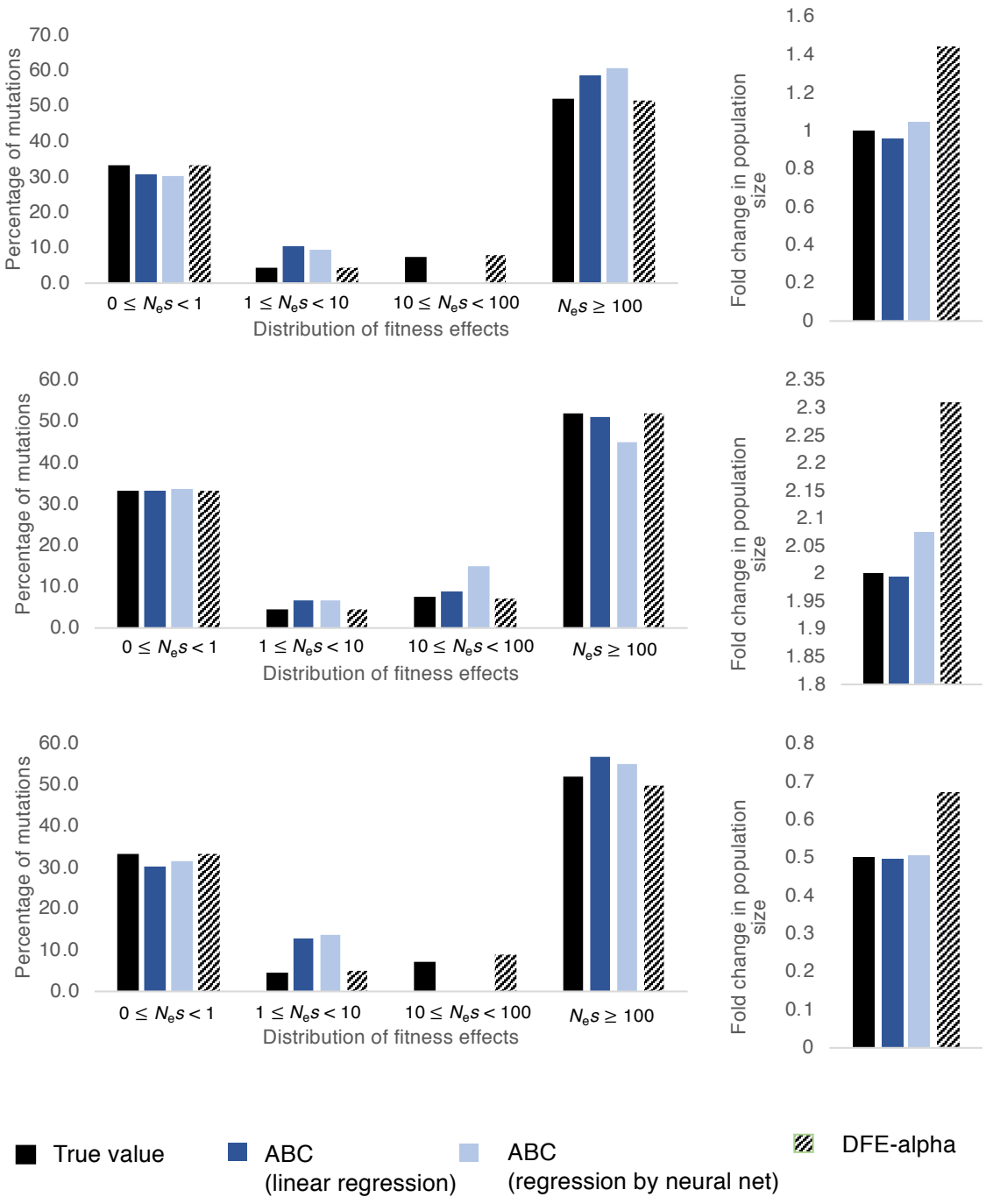


Figure S11: Inference of demography and the DFE by the approach proposed here and DFE-alpha, when the true shape of the DFE is gamma distributed, for equilibrium (top panel), growth (middle panel), and decline (bottom panel). Solid black bars show the true value simulated, dark blue bars show our ABC performance using ridge regression, light blue bars show the ABC performance using linear regression aided by neural net. Patterned bars show the performance of DFE-alpha.

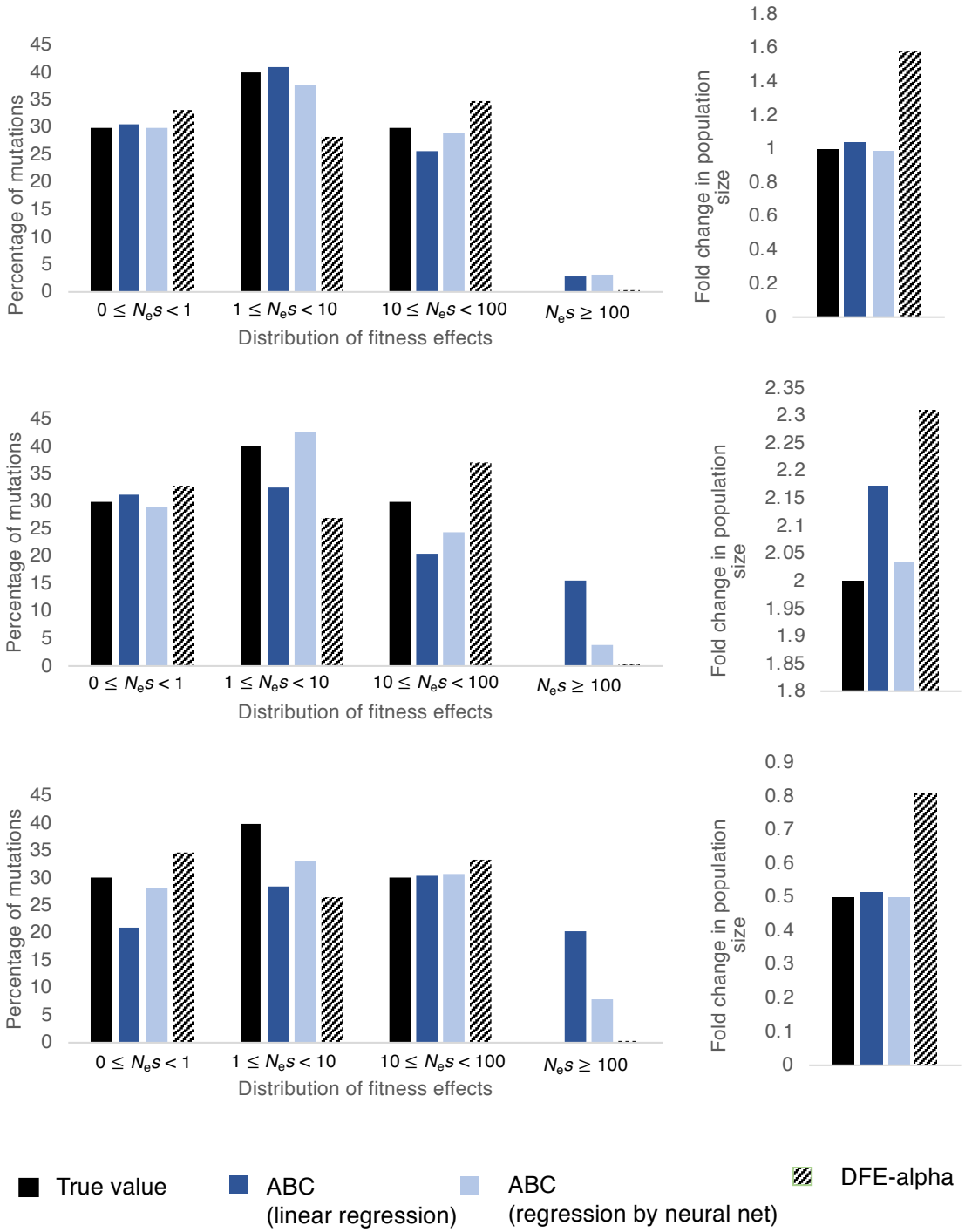


Figure S12: Inference of demography and the DFE when the true shape of the DFE is discrete and skewed towards slightly deleterious class of mutations, for demographic equilibrium (top panel), population size growth (middle panel), and population size decline (bottom panel). Solid black bars show the true value simulated, dark blue bars show the ABC performance using ridge regression, light blue bars show the ABC performance using linear regression aided by neural net. Patterned bars show the performance of DFE-alpha.

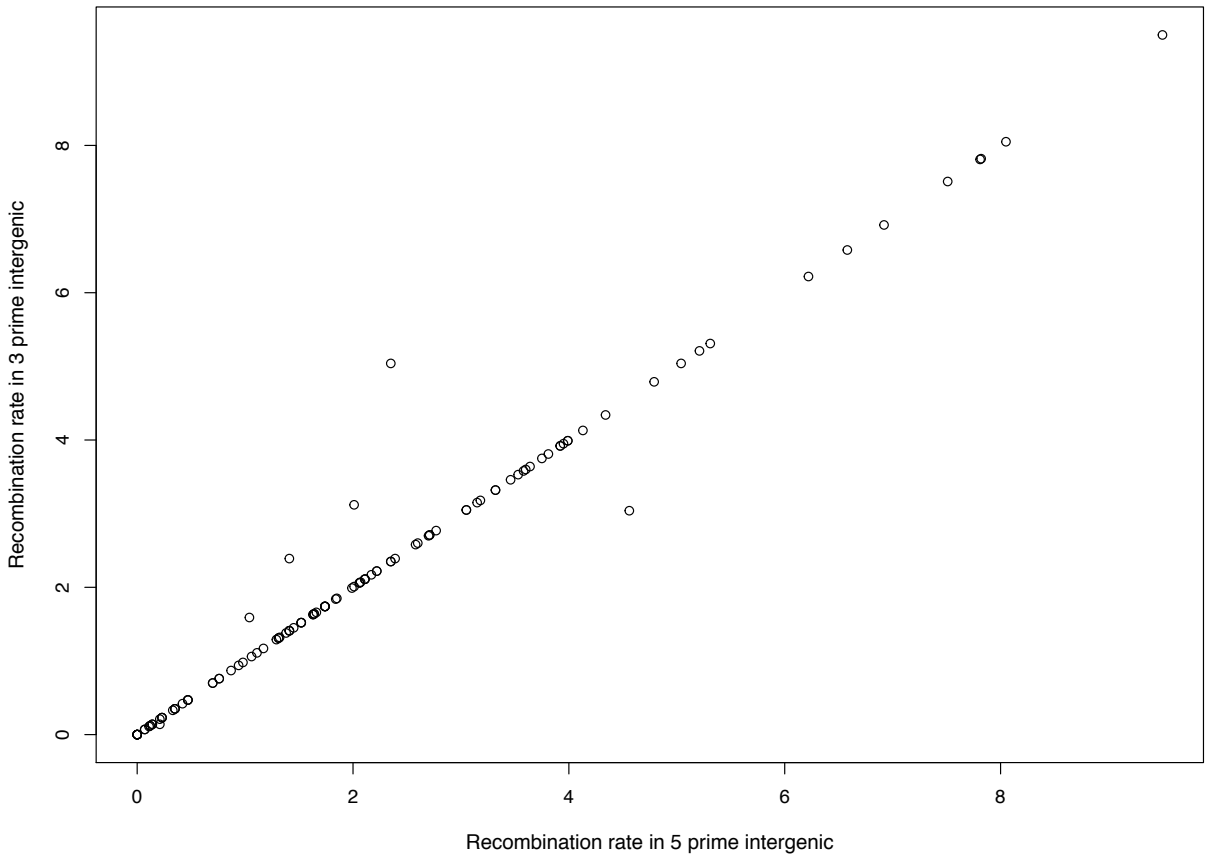


Figure S13: Correlation of recombination rates in 5′ flanking intergenic regions with those in 3′ flanking intergenic regions over all 94 exons chosen for analysis in *D. melanogaster*.

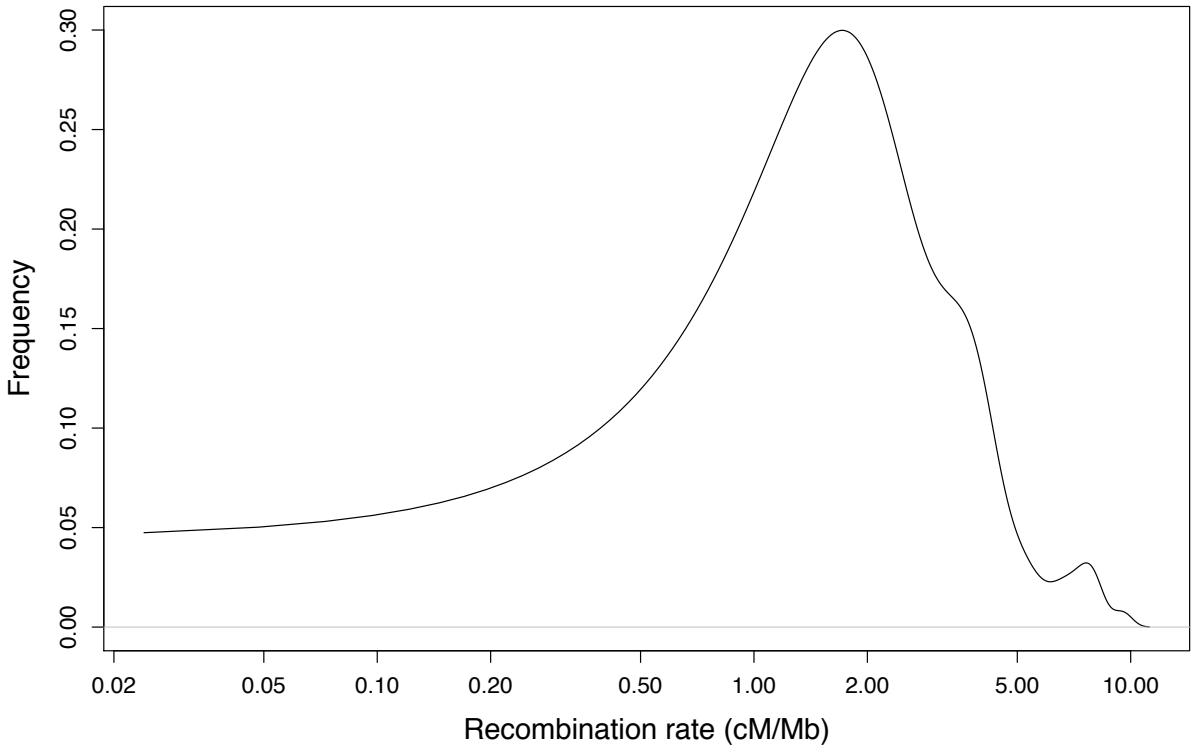


Figure S14: Distribution of the rate of recombination in cM/Mb for all 94 exons selected for analysis in *D. melanogaster*. Note that the x-axis is on a logarithmic scale.

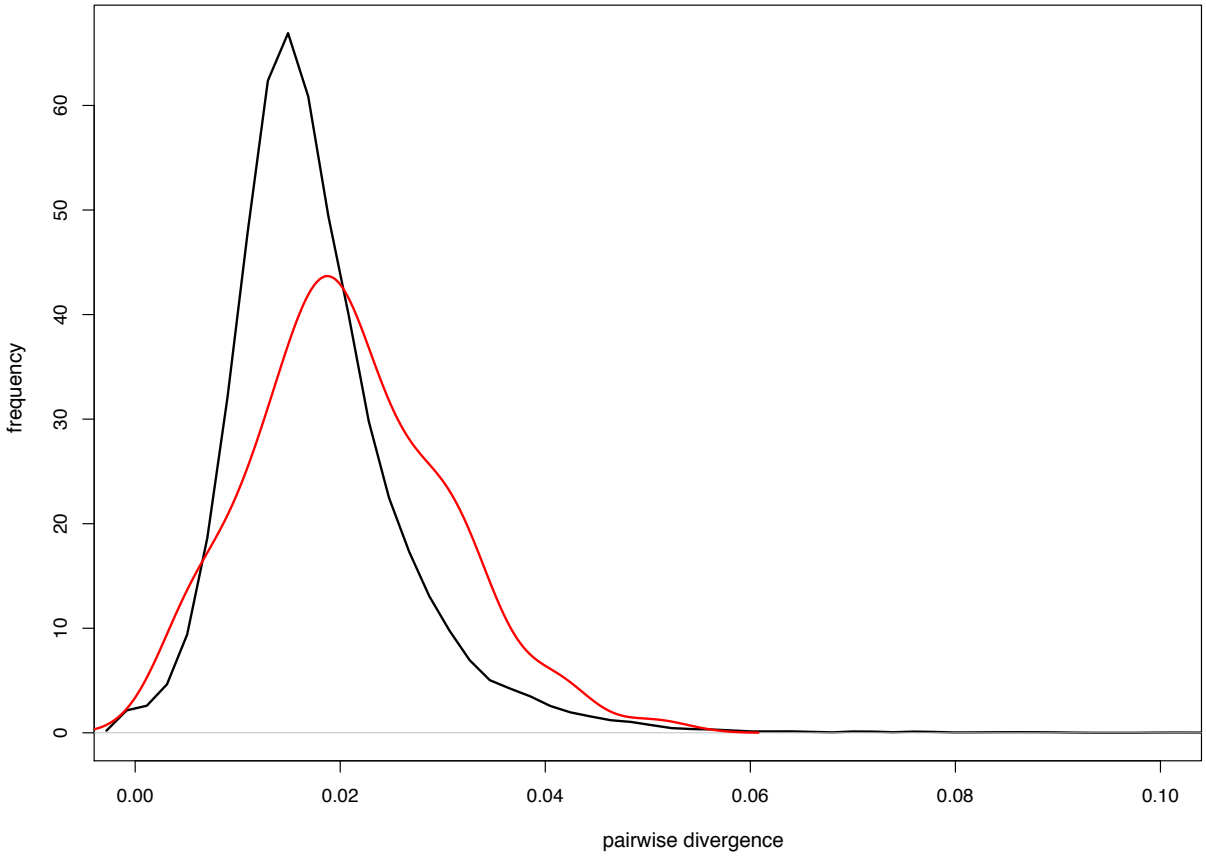
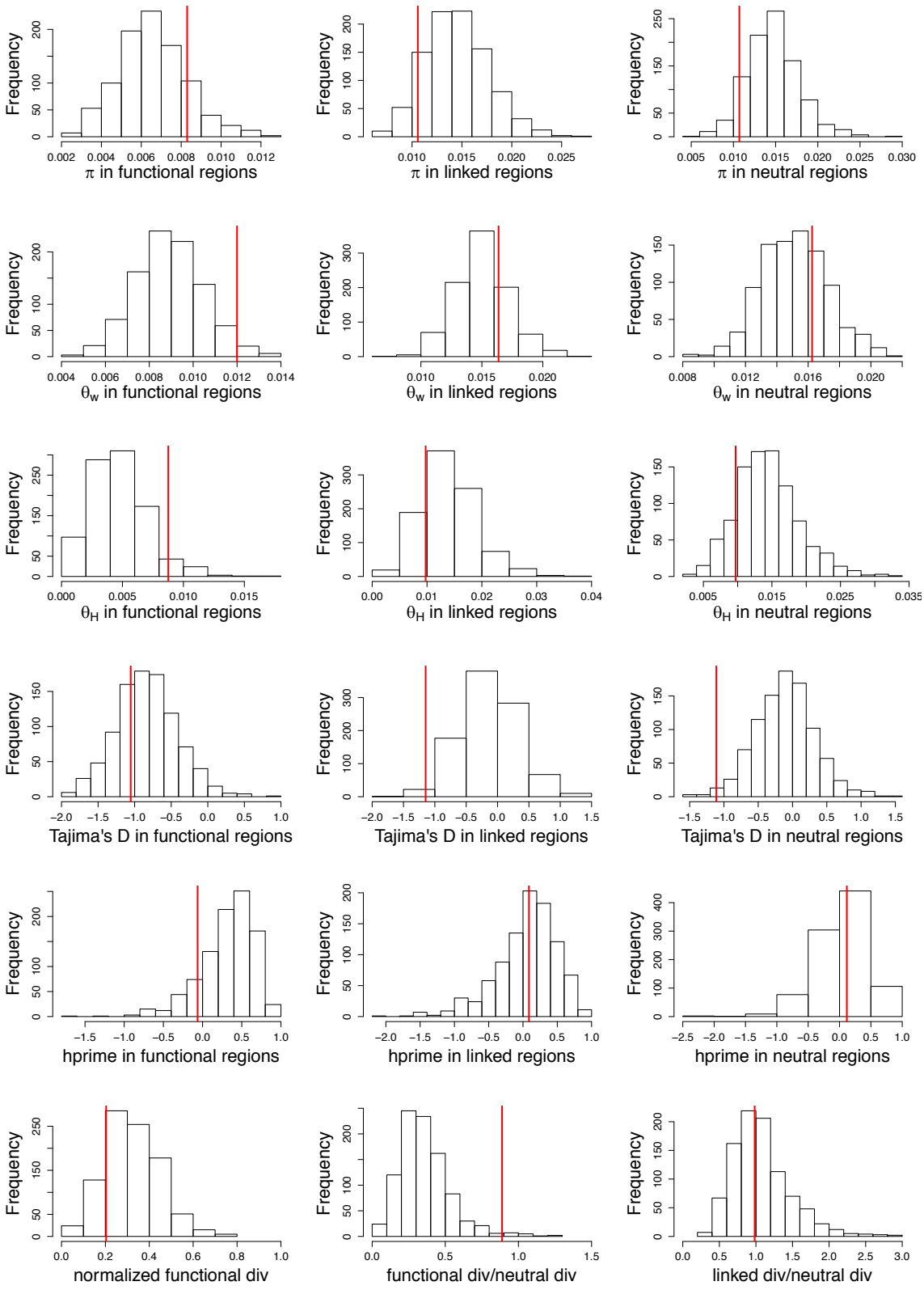


Figure S15: Distribution of divergence per site of single-exon genes that have flanking intergenic regions larger than 4 kb (in red), and for all genes (in black), from *D. melanogaster*.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

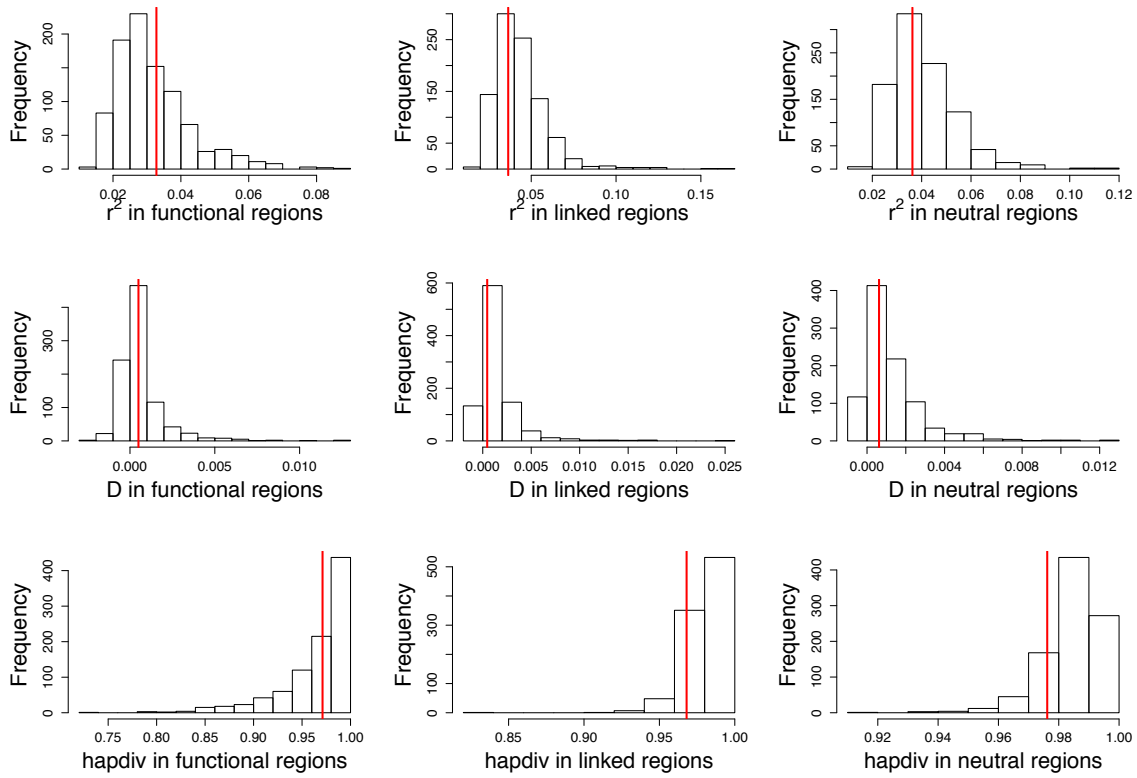
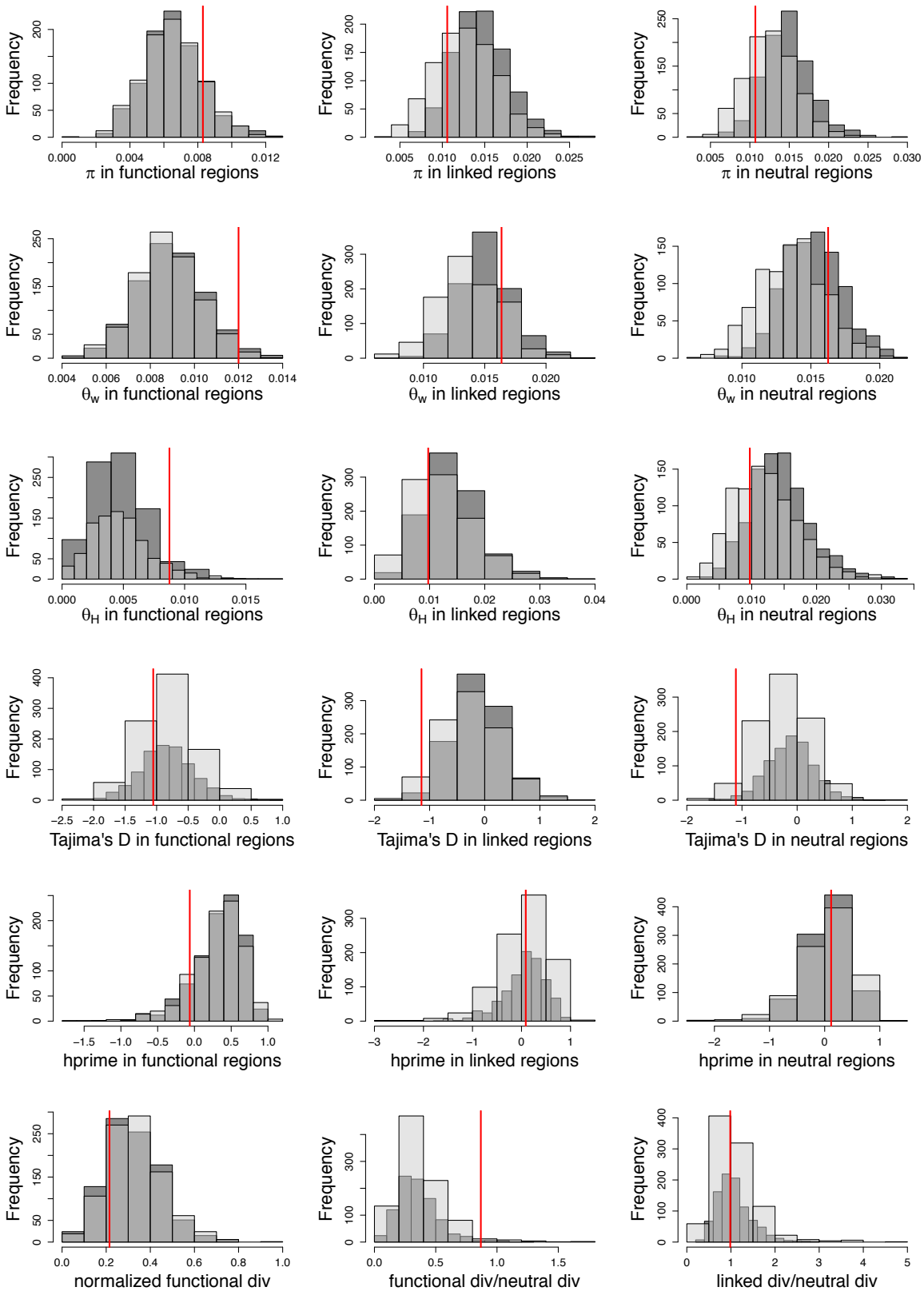


Figure S16: Distributions of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25$, $f_1=0.49$, $f_2=0.04$, $f_3=0.22$, $N_{anc}=$ 1,225,393, $N_{cur} = 1{,}357{,}760$). Red lines indicate the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

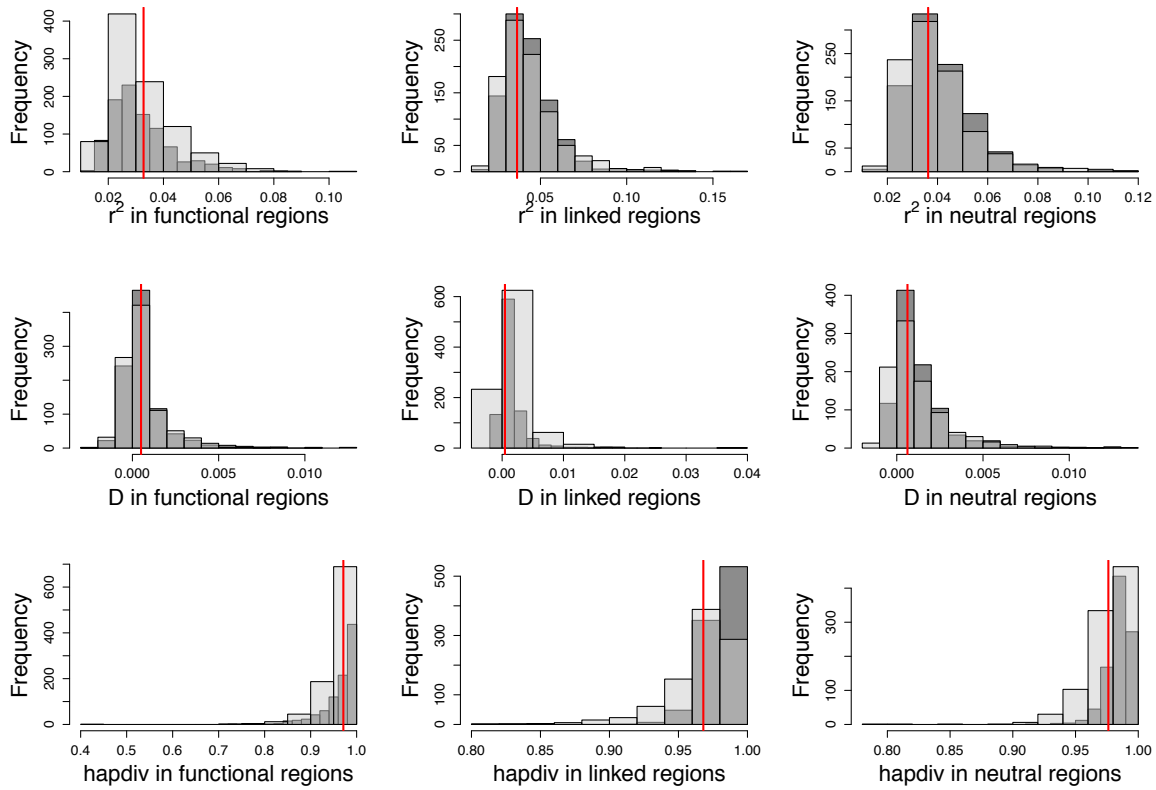
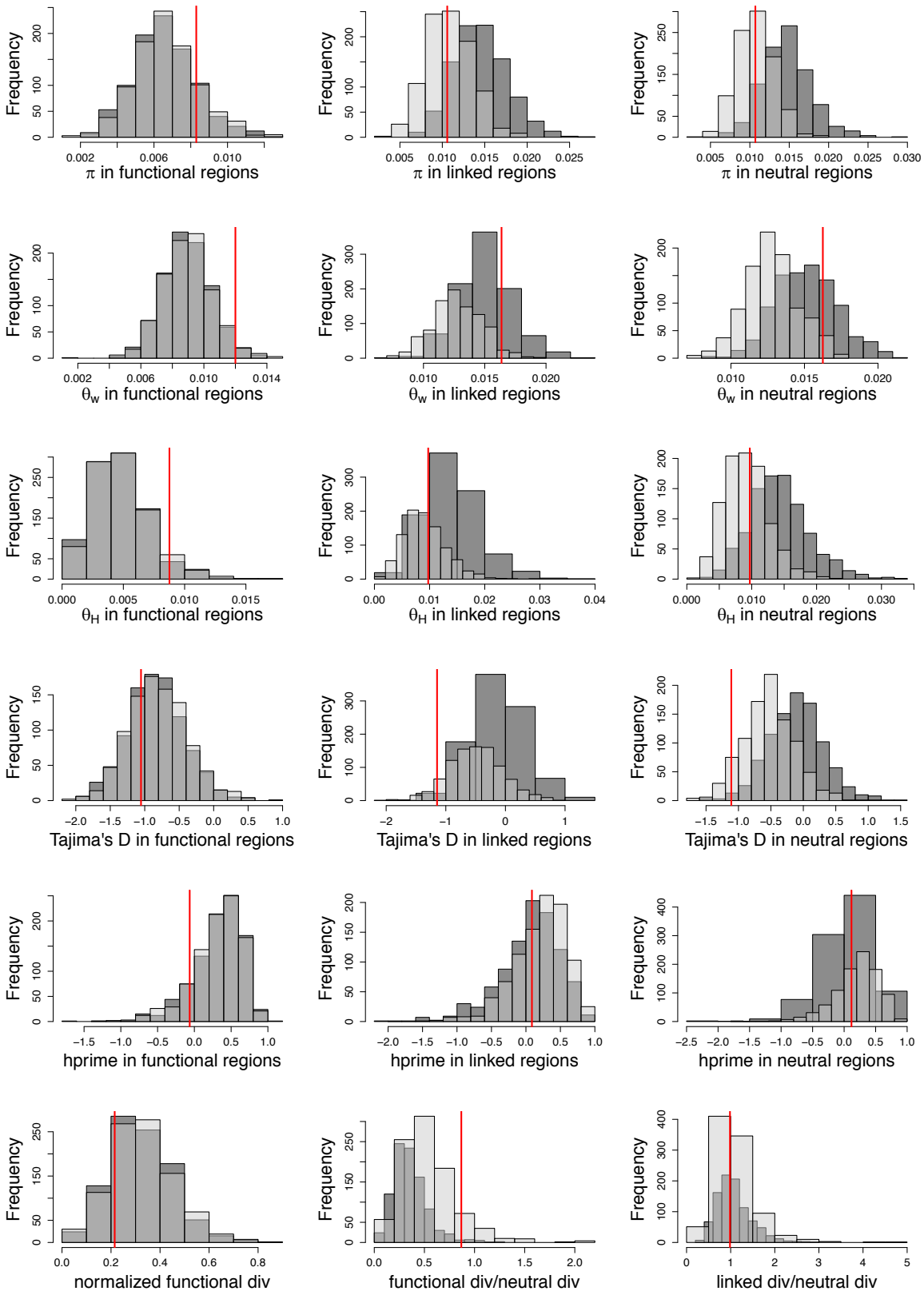


Figure S17: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25$, $f_1=0.49$, $f_2=0.04$, $f_3=0.22$, $N_{anc}=$ 1,225,393, $N_{cur} = 1{,}357{,}760$). In this case, conserved elements that represent 40% of non-coding regions were simulated as experiencing purifying selection with the class of mutations that result in the strongest BGS effects ($-100<2N_e s<-10$). and these sites were masked while calculating statistics. Red line indicates the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no selection on non-coding regions and light grey bars represent simulations with selection on non-coding regions.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

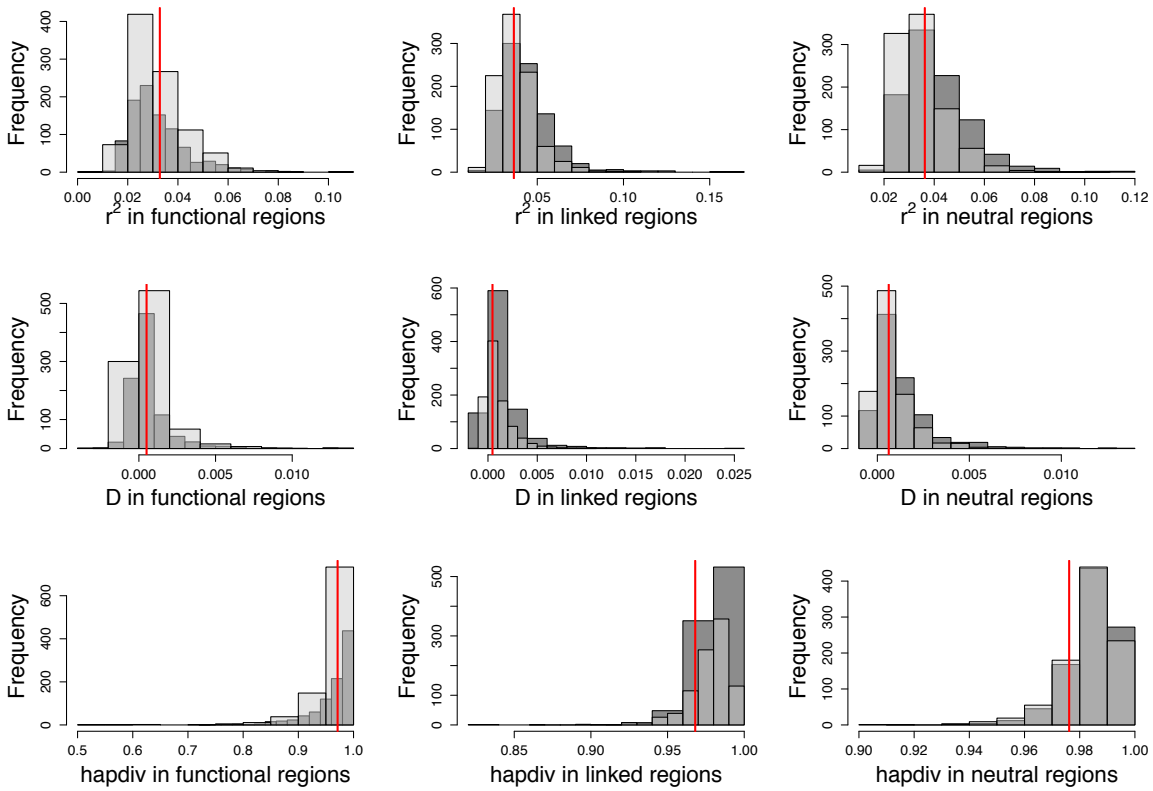
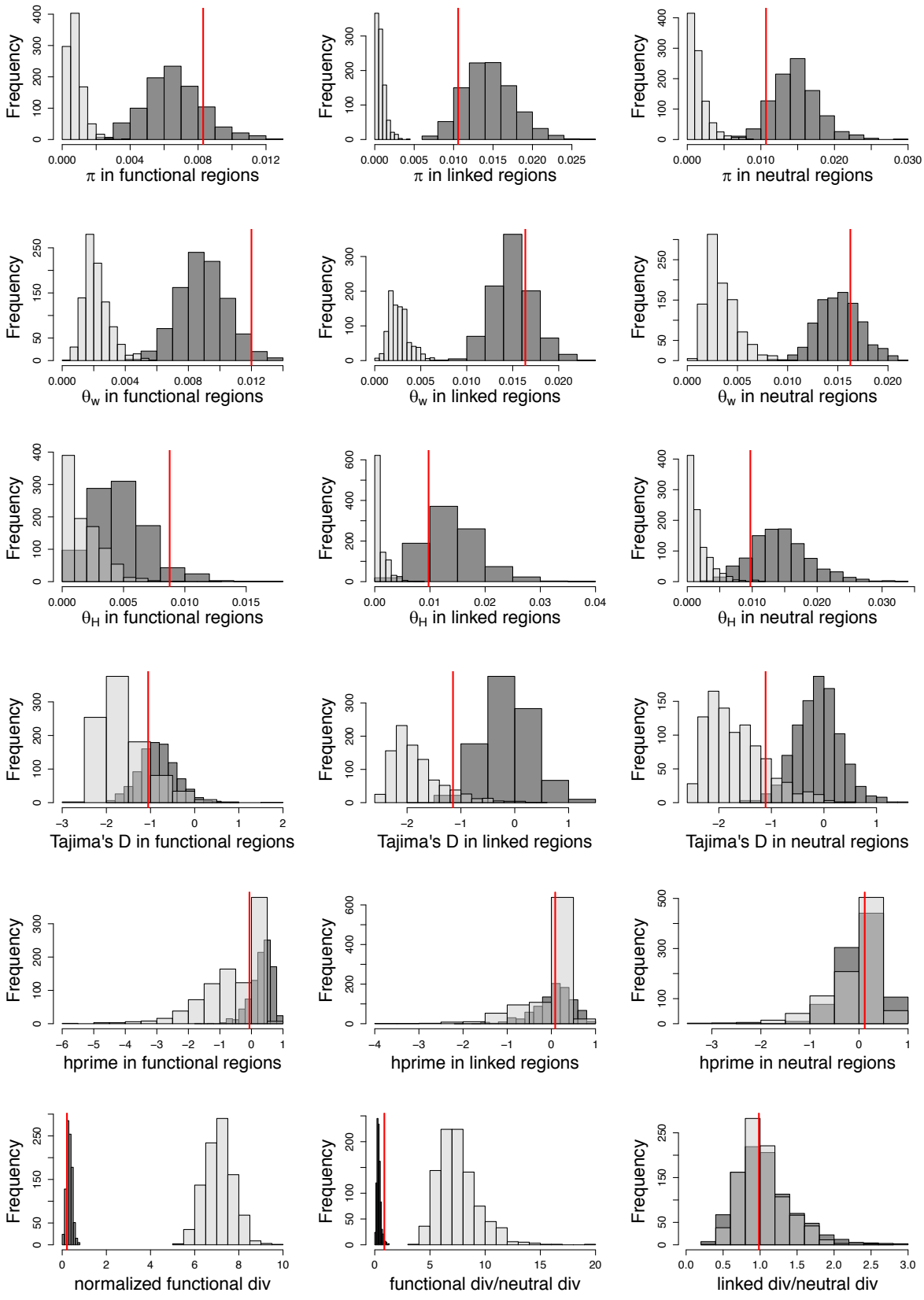


Figure S18: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25$, $f_1=0.49$, $f_2=0.04$, $f_3=0.22$, $N_{anc}=$ 1,225,393, $N_{cur} = 1{,}357{,}760$). In this case, conserved elements that represent 40% of non-coding regions were simulated as experiencing weak purifying selection ($-10<2N_e s<-1$) and these sites were included while calculating statistics. Red line indicates the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no selection on non-coding regions and light grey bars represent simulations with selection on non-coding regions.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

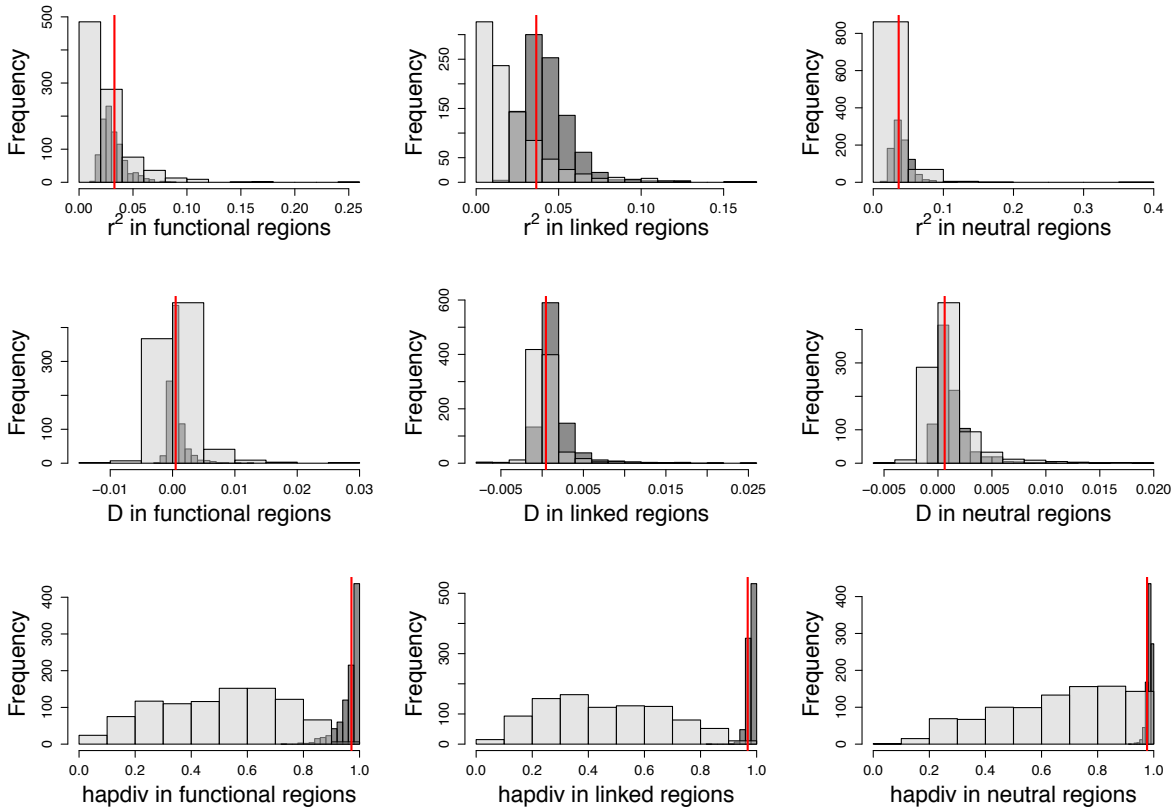
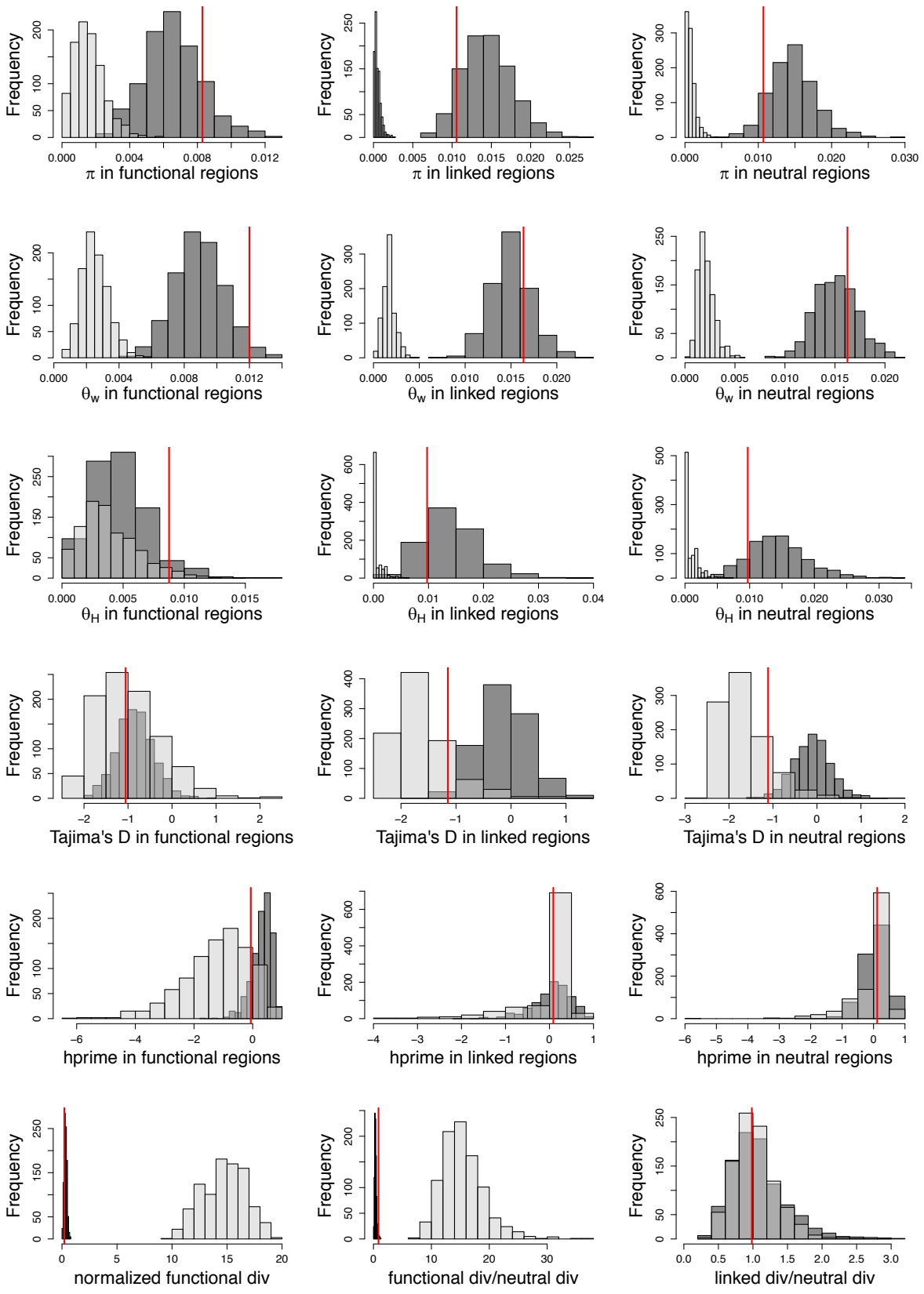


Figure S19: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25$, $f_1=0.49$, $f_2=0.04$, $f_3=0.22$, $N_{anc}=$ 1,225,393, $N_{cur}=$ 1,357,760). Functional regions were simulated as experiencing rare (1%) and strong positive selection ($2N_{anc}s = 1000$). Red lines indicate the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no positive selection and light grey bars represent simulations with positive selection in functional regions.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

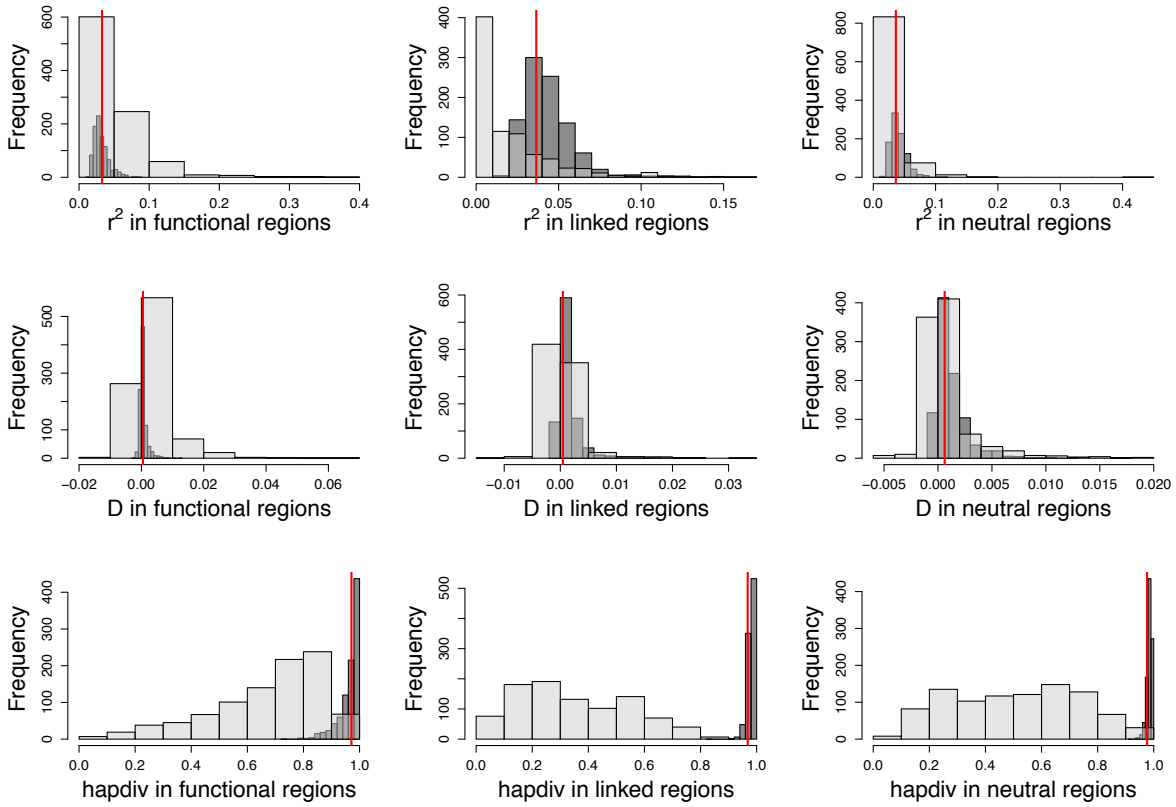
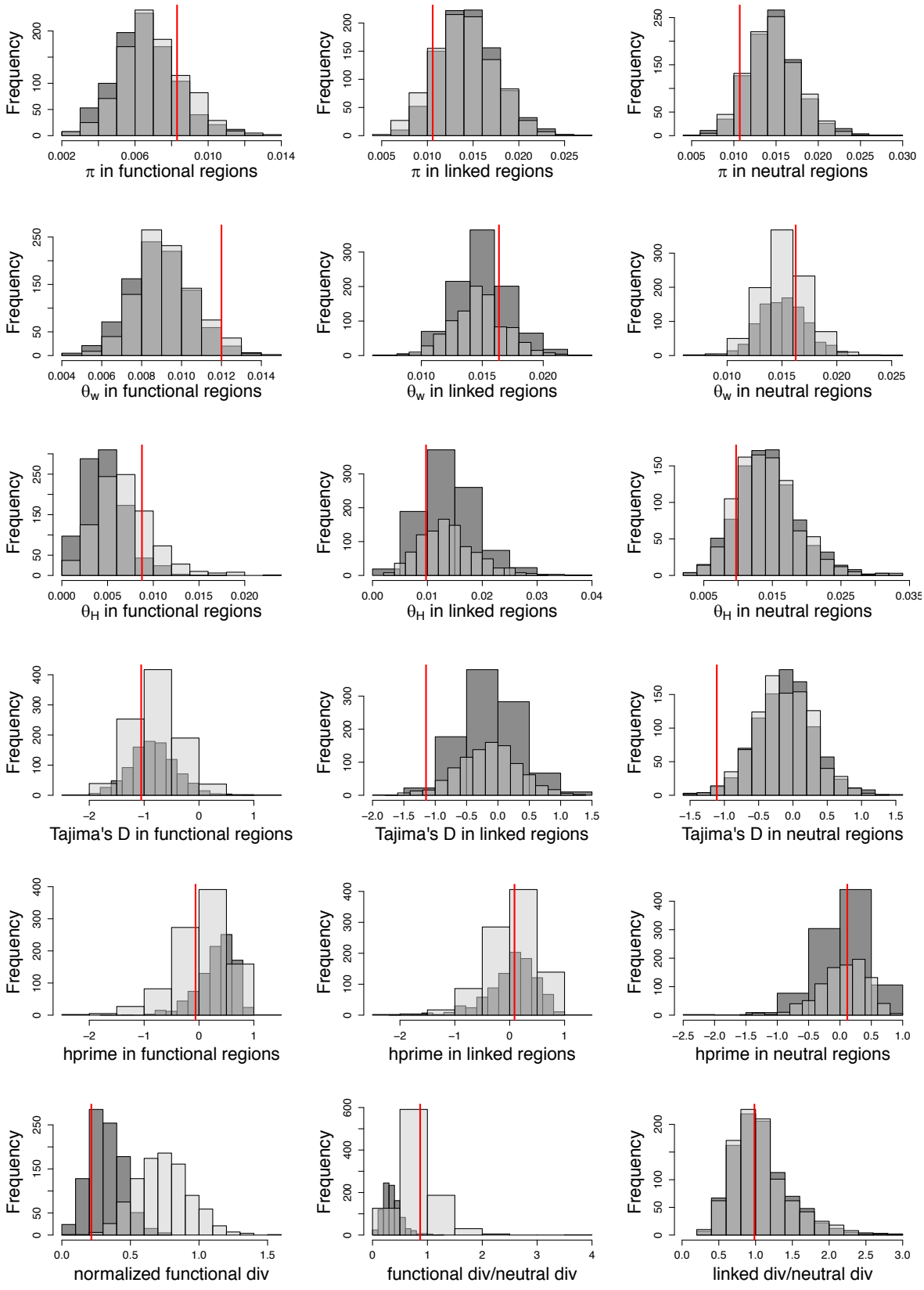


Figure S20: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25$, f_1=0.49, f_2=0.04, f_3=0.22, N_{anc}= 1,225,393, N_{cur}= 1,357,760). Functional regions were simulated to experience common (5%) and strong positive selection ($2N_{anc}s = 1000$). Red line indicates the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no positive selection and light grey bars represent simulations with positive selection in functional regions.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

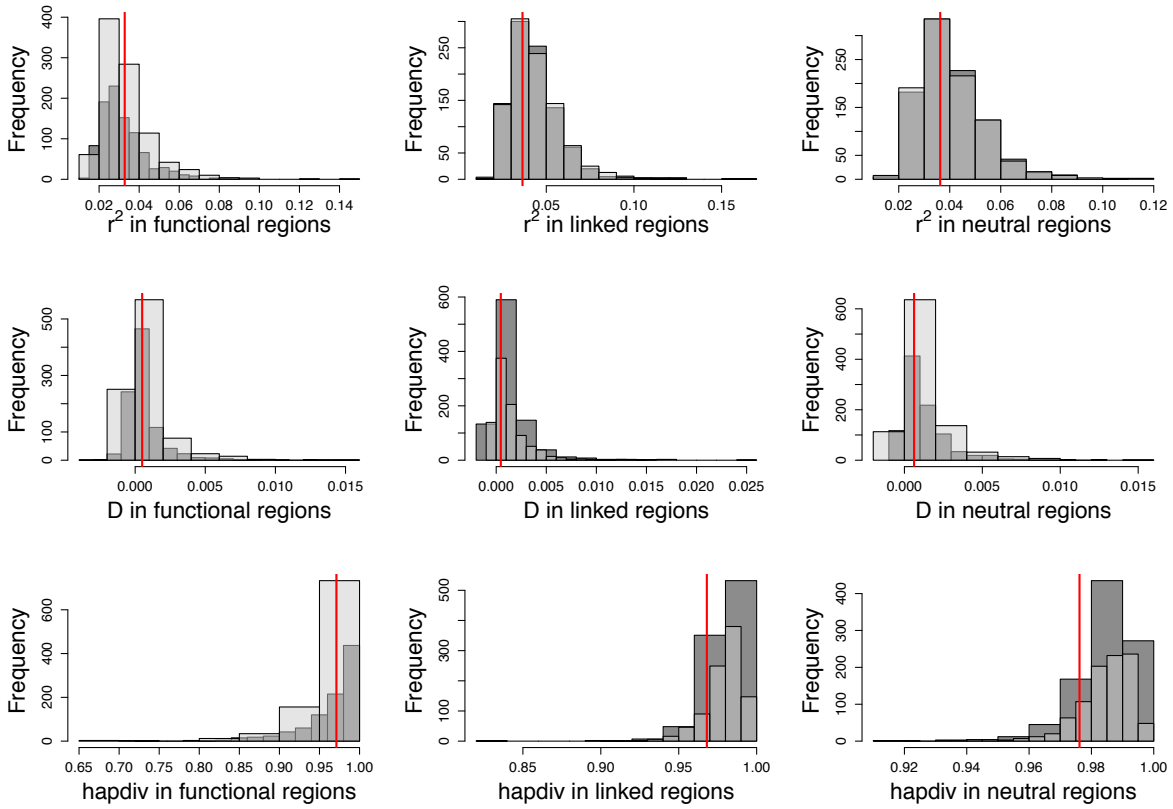
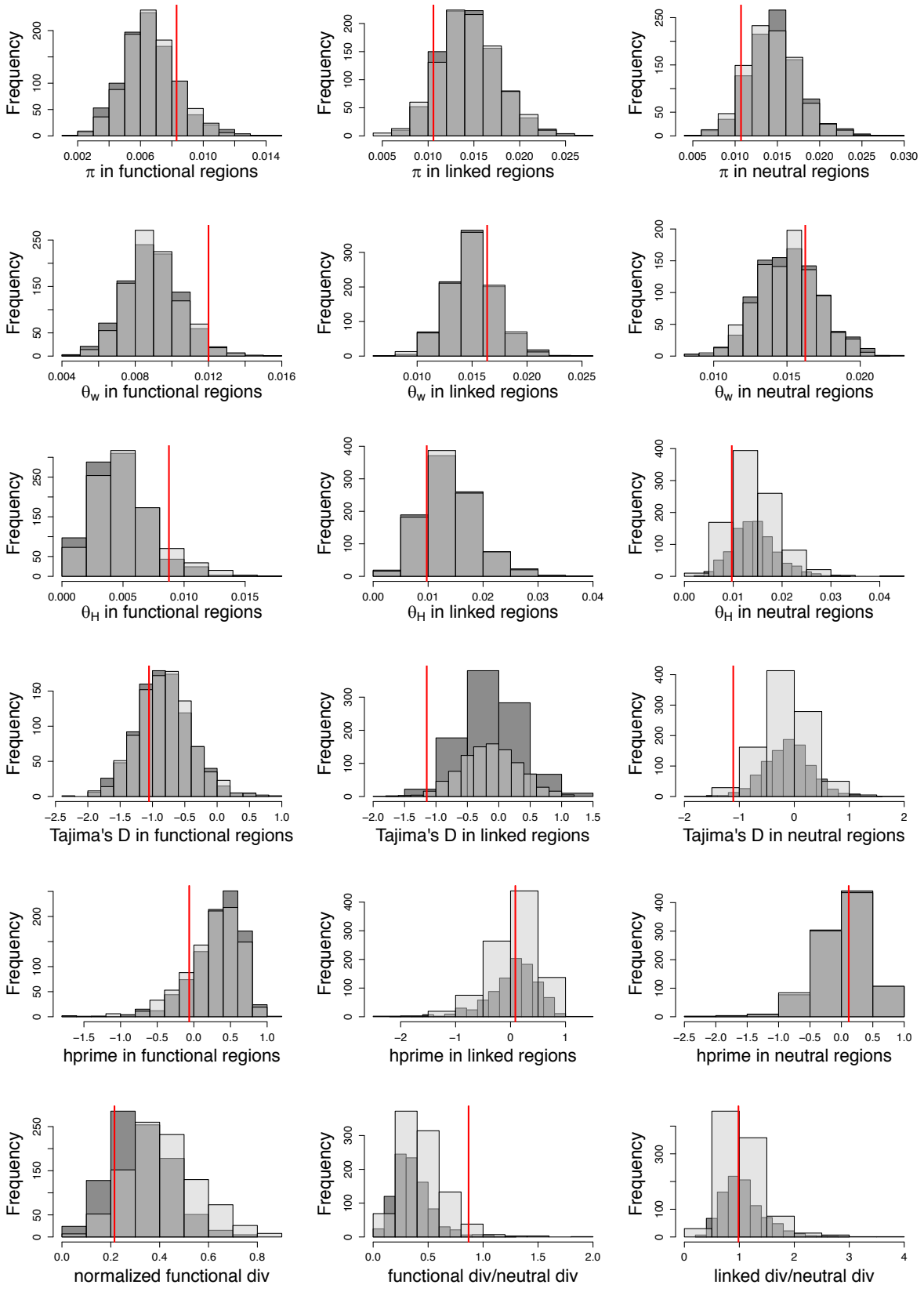


Figure S21: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25$, $f_1=0.49$, $f_2=0.04$, $f_3=0.22$, $N_{anc}=$ 1,225,393, $N_{cur} = 1{,}357{,}760$). Functional regions were simulated to experience common (5%) and weak positive selection ($2N_{anc}s = 10$). Red lines indicate the value observed in 76 individuals of *Drosophila melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no positive selection and light grey bars represent simulations with positive selection in functional regions.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

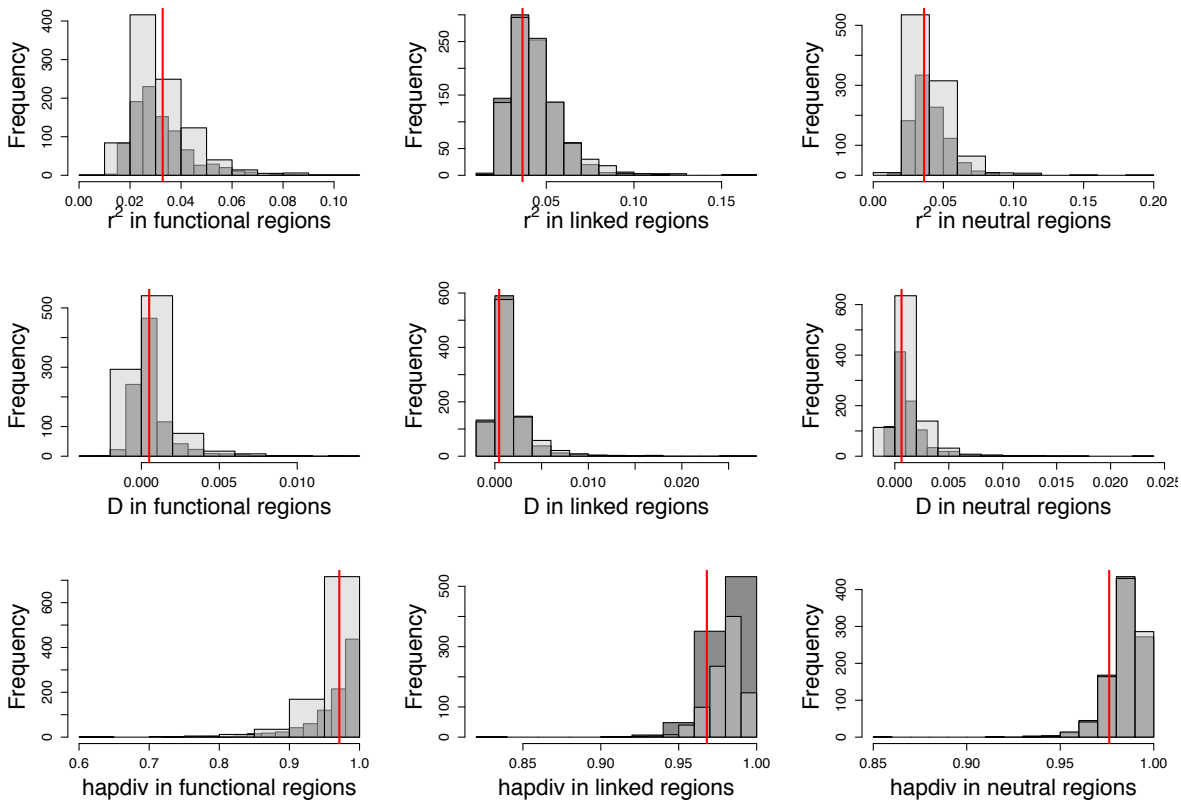
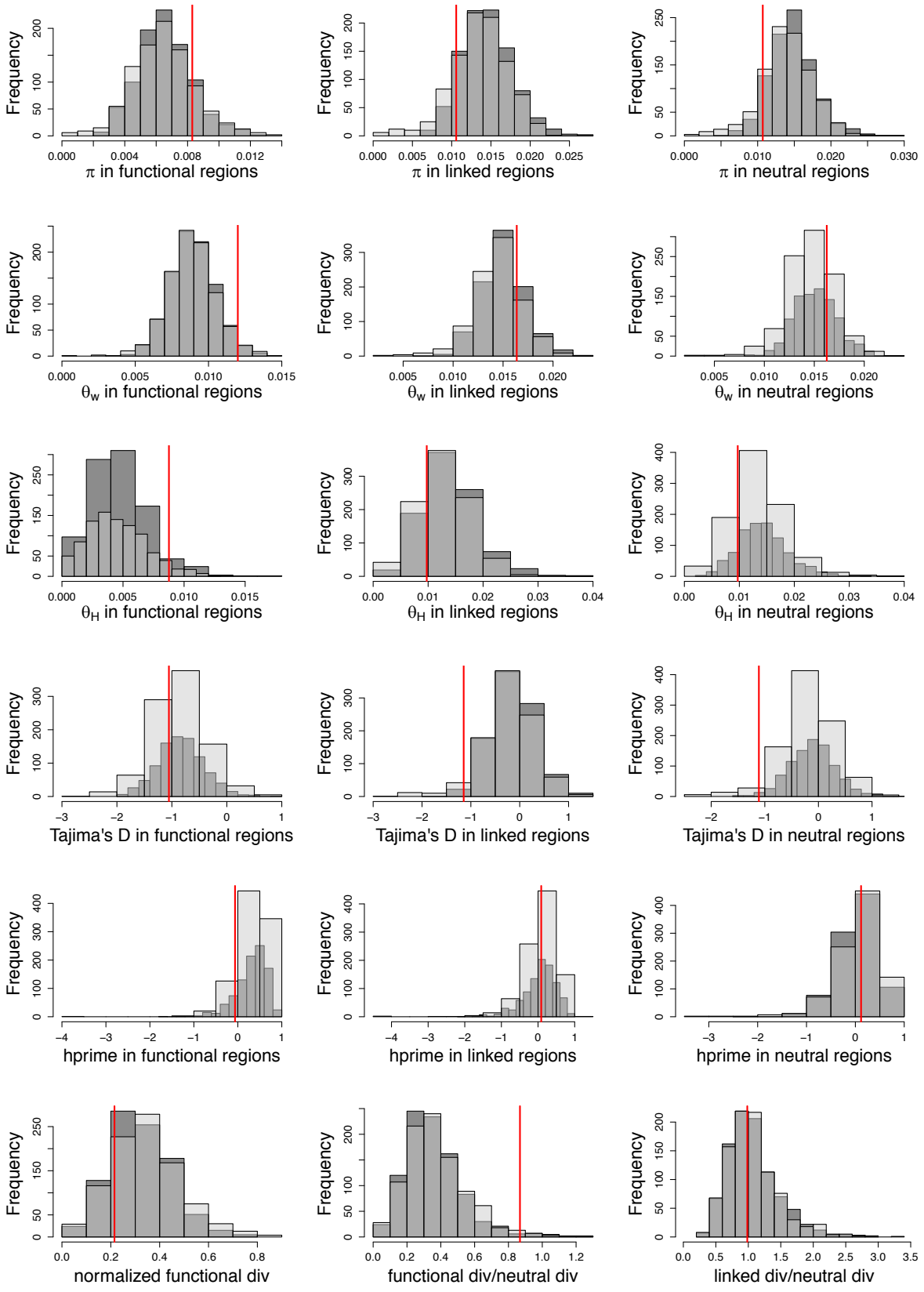


Figure S22: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25, f_1=0.49, f_2=0.04, f_3=0.22$, $N_{anc}=$ 1,225,393, $N_{cur} = 1{,}357{,}760$). Functional regions were simulated as experiencing rare (1%) and weak positive selection ($2N_{anc}s = 10$). Red lines indicate the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no positive selection and light grey bars represent simulations with positive selection in functional regions.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

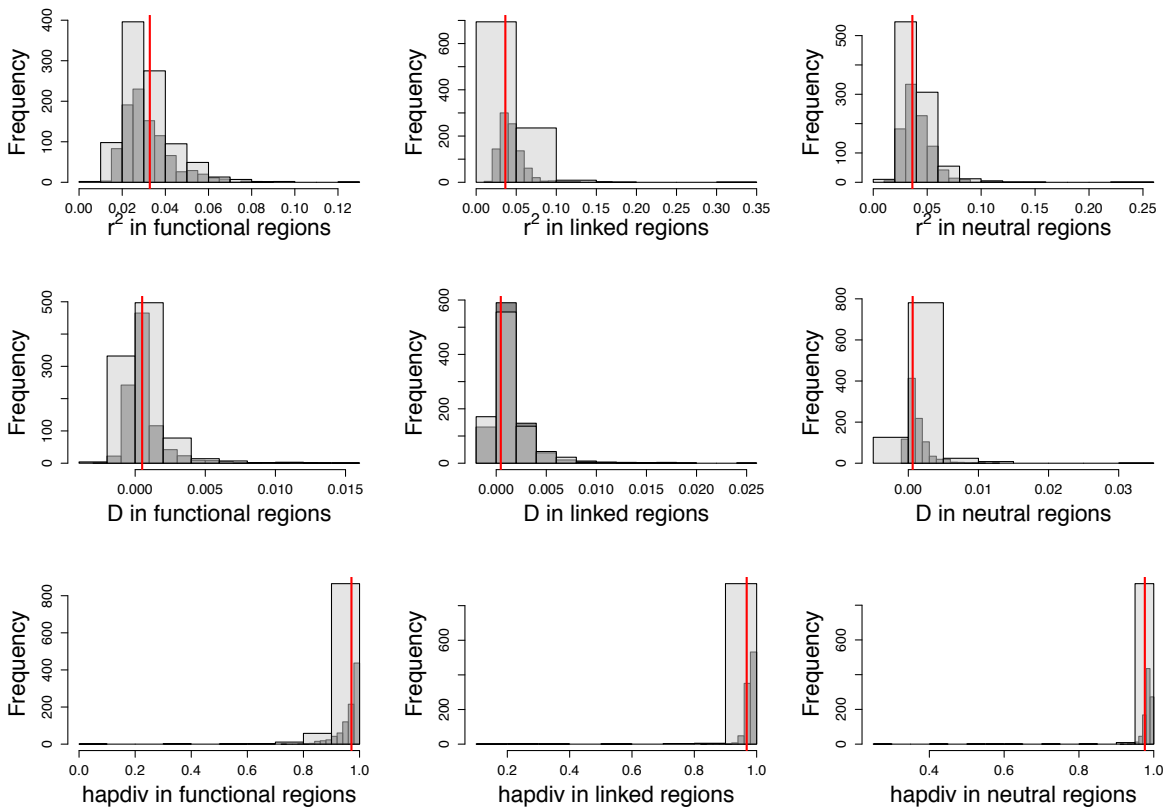
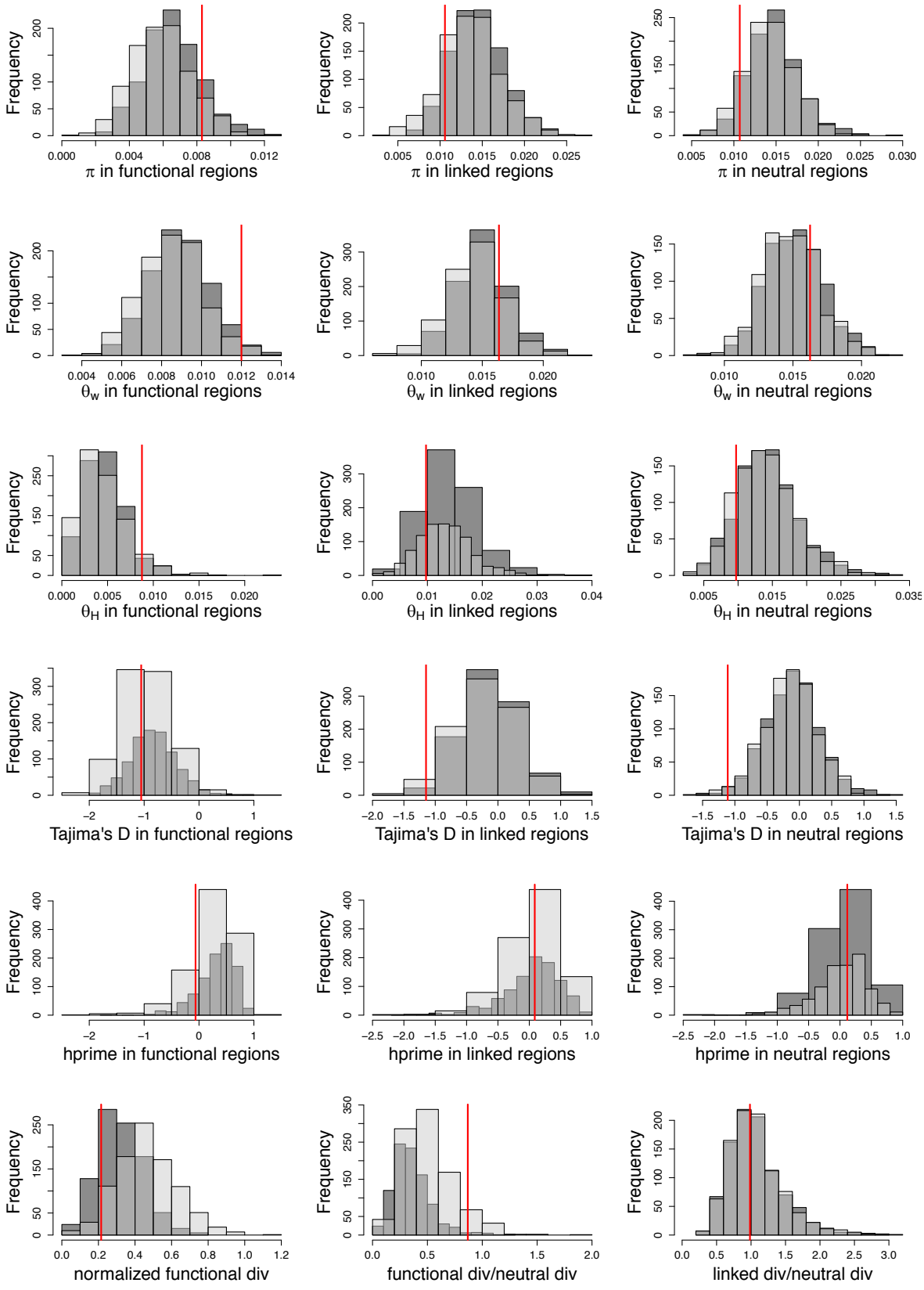


Figure S23: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e.*, $f_0 = 0.25, f_1=0.49, f_2=0.04, f_3=0.22$, $N_{anc}=$ 1,225,393, $N_{cur} = 1{,}357{,}760$). Functional regions were simulated to experience rare (1.28×10^{-4} %) and strong positive selection ($2N_{anc}s = 10000$) as in Lange and Pool (2018). Red lines indicate the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no positive selection and light grey bars represent simulations with positive selection in functional regions.


Frequency
π in functional regions
π in linked regions
π in neutral regions
θw in functional regions
θw in linked regions
θw in neutral regions
θH in functional regions
θH in linked regions
θH in neutral regions
Tajima's D in functional regions
Tajima's D in linked regions
Tajima's D in neutral regions
hprime in functional regions
hprime in linked regions
hprime in neutral regions
normalized functional div
functional div/neutral div
linked div/neutral div

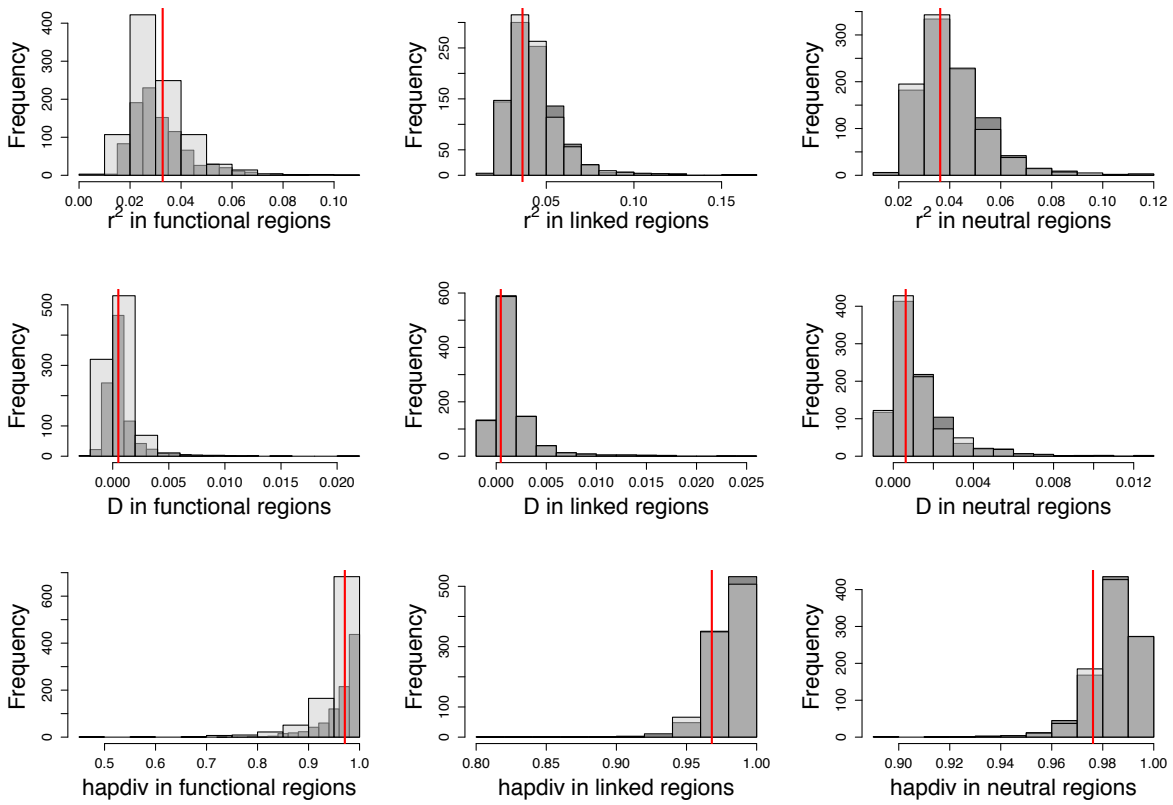


Figure S24: Distribution of summary statistics calculated from 94 exons simulated with 100 replicates each using our inferred model (*i.e* f_0 = 0.25, f_1=0.49, f_2=0.04, f_3=0.22, N_{anc}= 1,225,393, N_{cur} = 1,357,760). Functional regions were simulated to experience rare (0.2%) and weak positive selection ($2N_{anc}s$ = 60) as in Lange and Pool (2018). Red lines indicate the value observed in 76 individuals of *D. melanogaster* from Zambia, after excluding sites with phastCons score ≥ 0.8. Dark grey bars represent no positive selection and light grey bars represent simulations with positive selection in functional regions.