

Supplemental Material for: Minding your CpG's: Context Dependence, Ancestral Misidentification, and Spurious Signatures of Natural Selection

Ryan D. Hernandez, Scott H. Williamson, Carlos D. Bustamante

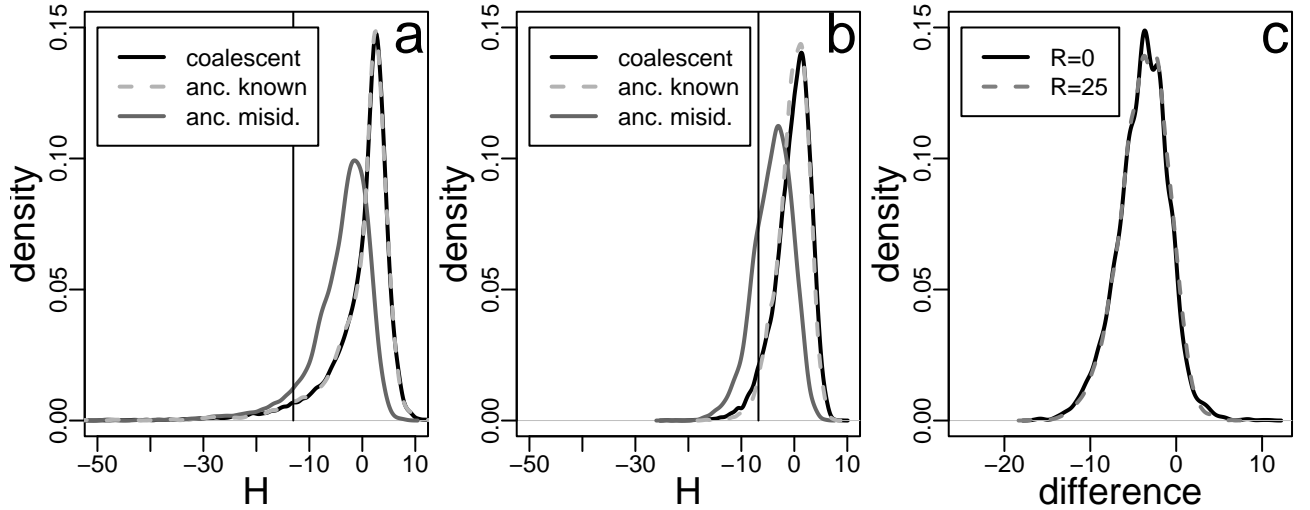


Figure S1: The distribution of the Fay and Wu (2000) H statistic for the average per sequence population scaled recombination rate $R = 0$ (a) and $R = 25$ (b) with the vertical lines indicating the 0.05 critical value obtained from 20,000 coalescent simulations (Hudson, 2002). (c) The distribution of the “difference” between H when the ancestral state of each polymorphism is inferred from the outgroup and H when the ancestral state is known (i.e. saved during the simulation). Each curve represents 5,000 simulations of a sequence of length 2.5kb under the context-dependent mutation model discussed in the text with an average population scaled mutation rate per locus $\theta = 4N_e\mu = 10.75$.

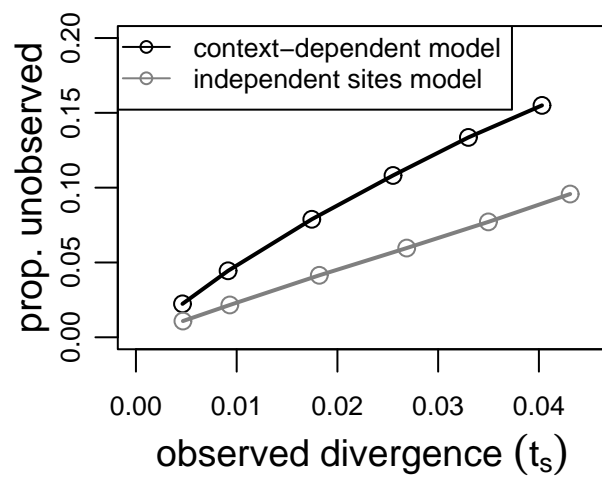


Figure S2: The proportion of substitutions that are unobserved as a function of the average number of substitutions per site (τ_s) for the two mutation models discussed in the text.

Table S1: Comparing the Average Number of Segregating Sites and Fixed Differences When Mutation Rates are Context-Dependent to their Expectations Under the Coalescent.

τ	$L\theta^a$		Segregating sites		Substitutions	
			$\rho^b = 0$	$\rho = 0.01$	$\rho = 0$	$\rho = 0.01$
10	2.15	total ^c	9.403 (4.107)	9.421 (3.423)	20.390 (5.122)	20.303 (4.712)
		obs. ^d	9.364 (4.051)	9.379 (3.405)	18.608 (4.695)	18.503 (4.347)
		exp. ^e	9.624 (4.118)	9.592 (3.658)	20.492 (5.241)	20.445 (4.832)
	4.3	total	18.777 (6.797)	18.819 (5.154)	40.889 (8.150)	40.834 (6.687)
		obs.	18.714 (6.706)	18.756 (5.130)	37.352 (7.464)	37.329 (6.259)
		exp.	19.305 (7.069)	19.262 (5.576)	40.939 (8.317)	41.002 (6.953)
	10.75	total	46.861 (14.825)	46.770 (8.441)	102.258 (16.246)	102.251 (10.888)
		obs.	46.677 (14.506)	46.599 (8.387)	93.626 (14.706)	93.607 (10.162)
		exp.	48.117 (15.252)	48.174 (9.568)	102.431 (16.682)	102.465 (11.234)
50	2.15	total	9.354 (3.980)	9.409 (3.457)	106.379 (10.435)	106.478 (10.455)
		obs.	9.327 (3.934)	9.376 (3.441)	82.453 (8.258)	82.397 (8.452)
		exp.	9.648 (4.144)	9.627 (3.669)	106.382 (10.619)	106.381 (10.494)
	4.3	total	18.676 (6.752)	18.809 (5.137)	212.878 (15.405)	212.775 (14.817)
		obs.	18.605 (6.638)	18.729 (5.108)	165.283 (12.305)	165.111 (11.982)
		exp.	19.234 (6.948)	19.240 (5.533)	213.001 (15.405)	212.937 (14.801)
	10.75	total	46.739 (14.566)	46.932 (8.441)	532.733 (26.454)	531.870 (23.826)
		obs.	46.584 (14.283)	46.766 (8.415)	414.415 (20.920)	413.945 (19.071)
		exp.	48.200 (15.479)	48.077 (9.549)	532.175 (26.683)	532.408 (23.422)

Note: For all simulations, $N_e = 250$ and $n = 50$. Values in parentheses are the simulated standard deviations.

^aAverage population scaled mutation rate per sequence ($2N_e\mu L$).

^bPopulation scaled recombination rate between adjacent sites.

^cIncludes observed and unobserved mutations stored during simulation.

^dIncludes only observed mutations.

^eDetermined from coalescent simulations of the neutral model (Hudson, 2002).