

A Likelihood Ratio Test for Hybridization Under the Multispecies Coalescent

Jing Peng, Sungsik Kong, Laura Kubatko

In the following sections, we show the supplemental figures of detection power and type I error using our LR test, *HyDe* and ABBA-BABA test under JC69 and HKY85+Gamma model with different branch lengths and number of sites in section 2.3.

S1. Comparison of three hybridization tests under HKY85+I model

In this section, we consider 500 CIS datasets simulated under HKY85+I model for each setting. For the speciation times in Figure S1(a)–(c), we assigned the vector $(\tau_1, \tau_2, \tau_3) = (0.25, 0.5, 1.0)$, while in Figures S1(d)–(f), we assigned the vector $(\tau_1, \tau_2, \tau_3) = (0.5, 1.0, 2.0)$. The hybridization parameter γ is chosen to be 0 or to vary from 0.06 to 0.5 by 0.02.

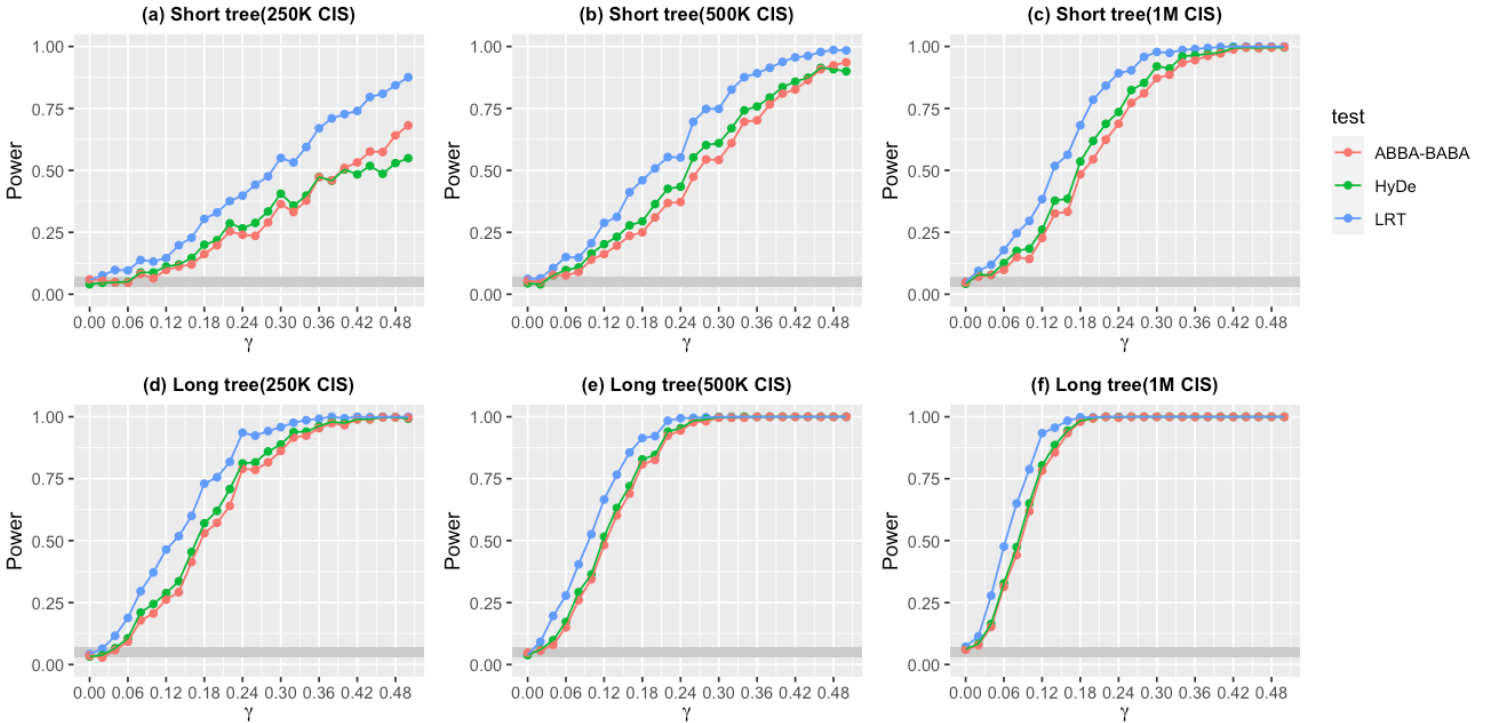


Figure S1: Hybrid detection power under the HKY85+I model. The shaded area is the expected acceptance region (0.031, 0.069) of the empirical type I error rate. (a) – type I error rates are 0.058 (LRT), 0.040 (*HyDe*) and 0.060 (ABBA-BABA). (b) – type I error rates are 0.062 (LRT), 0.044 (*HyDe*) and 0.054 (ABBA-BABA). (c) – type I error rates are 0.050 (LRT), 0.042 (*HyDe*) and 0.048 (ABBA-BABA). (d) – type I error rates are 0.042 (LRT), 0.032 (*HyDe*) and 0.036 (ABBA-BABA). (e) – type I error rates are 0.040 (LRT), 0.038 (*HyDe*) and 0.048 (ABBA-BABA). (f) – type I error rates are 0.072 (LRT), 0.062 (*HyDe*) and 0.060 (ABBA-BABA).

S2. Comparison of three hybridization tests under HKY85+I+ Γ model

In this section, we consider 500 CIS datasets simulated under HKY85+I+ Γ model for each setting. For the speciation times in Figure S2(a)–(c), we assigned the vector $(\tau_1, \tau_2, \tau_3) = (0.25, 0.5, 1.0)$, while in Figures S2(d)–(f), we assigned the vector $(\tau_1, \tau_2, \tau_3) = (0.5, 1.0, 2.0)$. The hybridization parameter γ is chosen to be 0 or to vary from 0.06 to 0.5 by 0.02.

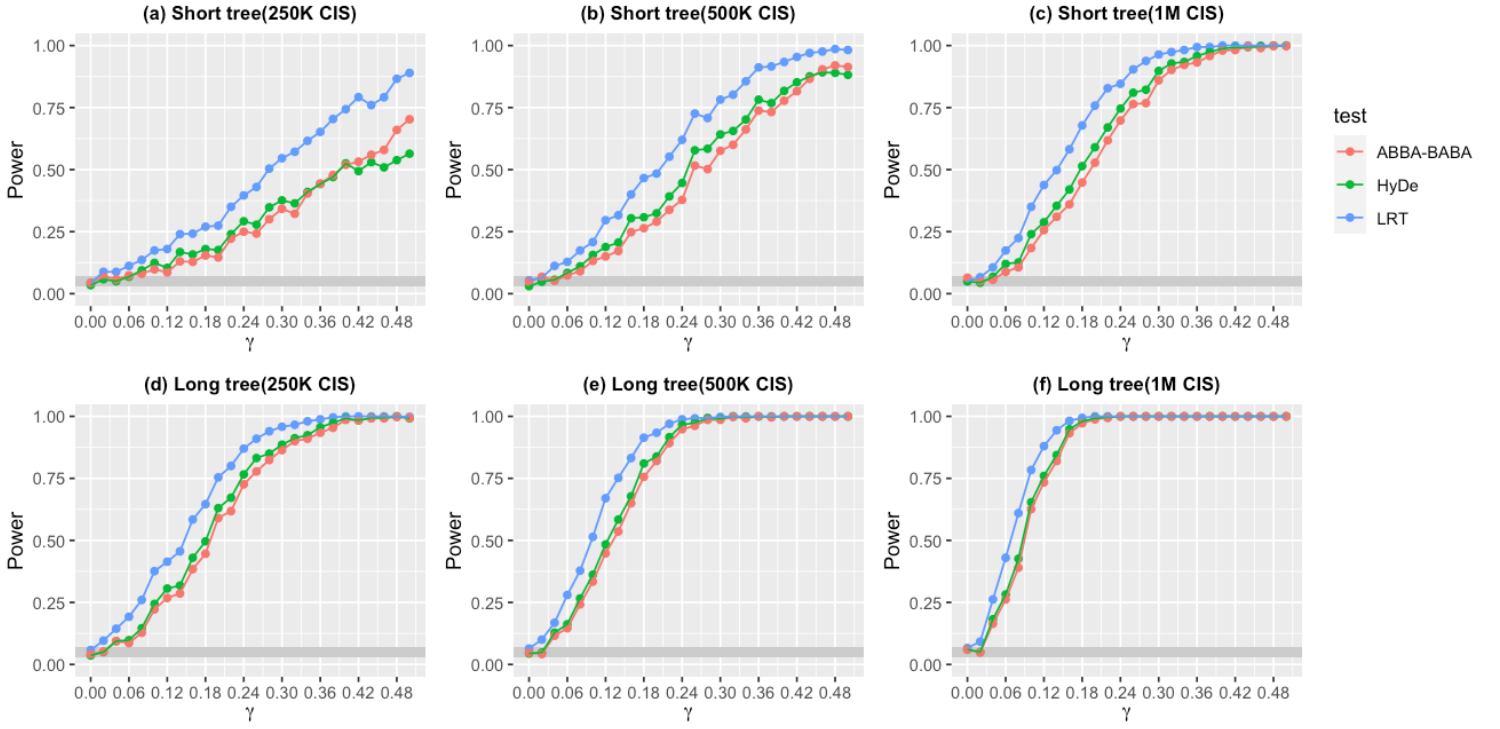


Figure S2: Hybrid detection power under the HKY85+I+ Γ model. The shaded area is the expected acceptance region (0.031,0.069) of the empirical type I error rate. (a) – type I error rates are 0.044 (LRT), 0.034 (*HyDe*) and 0.044 (ABBA-BABA). (b) – type I error rates are 0.054 (LRT), 0.030 (*HyDe*) and 0.048 (ABBA-BABA). (c) – type I error rates are 0.050 (LRT), 0.050 (*HyDe*) and 0.064 (ABBA-BABA). (d) – type I error rates are 0.058 (LRT), 0.036 (*HyDe*) and 0.042 (ABBA-BABA). (e) – type I error rates are 0.064 (LRT), 0.044 (*HyDe*) and 0.048 (ABBA-BABA). (f) – type I error rates are 0.066 (LRT), 0.060 (*HyDe*) and 0.060 (ABBA-BABA).

S3. Comparison of three hybridization tests under JC69 model

S3.1. Short branch tree

In this section, we consider CIS and multilocus datasets simulated under JC69 model. For the speciation times in Figure ??, we assigned the vector $(\tau_1, \tau_2, \tau_3) = (0.25, 0.5, 1.0)$. The hybridization parameter γ is chosen to be 0 or to vary from 0.06 to 0.5 by 0.02.

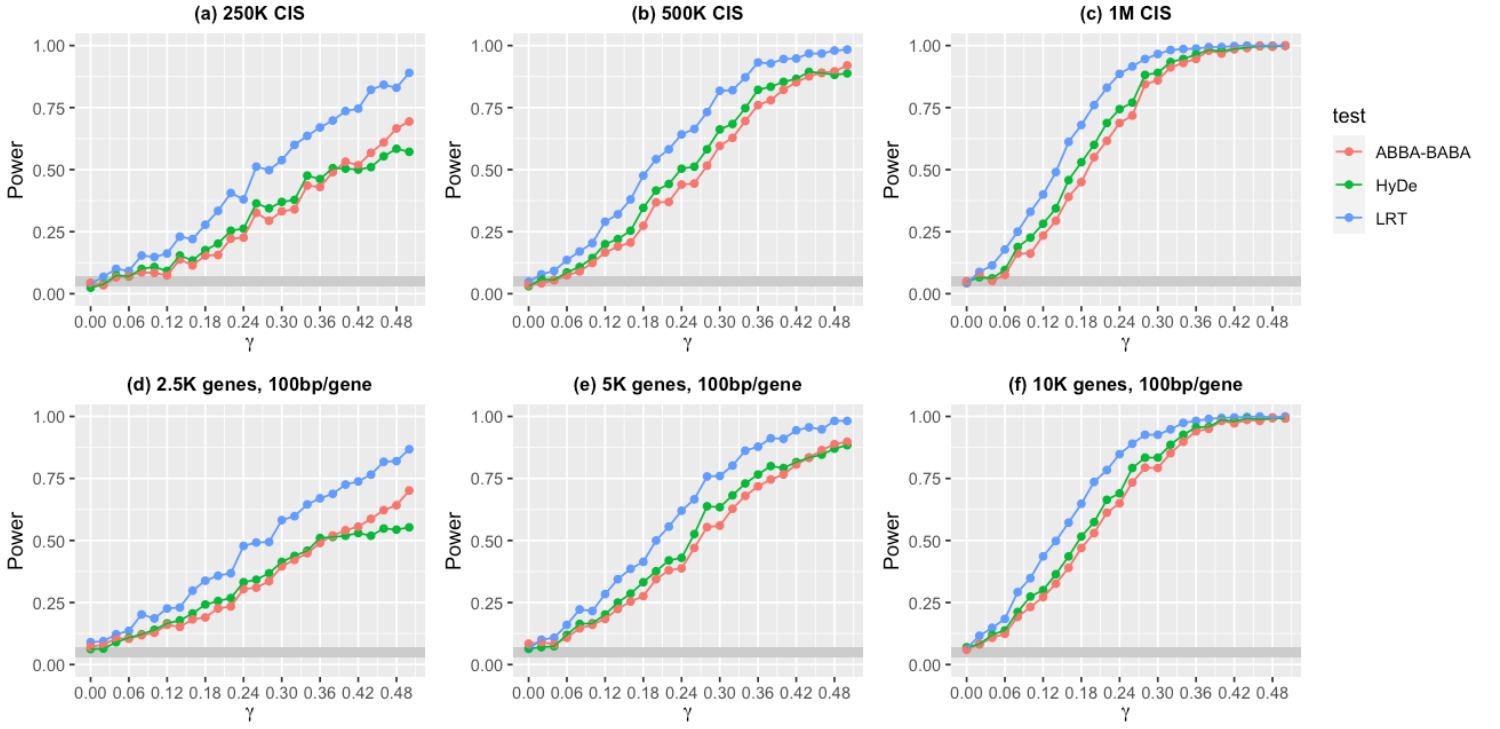


Figure S3: Hybrid detection power under JC69 for the short branch tree. The shaded area is the expected acceptance region (0.031,0.069) of the empirical type I error rate. (a) – 250K CIS data with type I error rates 0.040 (LR), 0.024 (*HyDe*), 0.044 (ABBA-BABA); (b) – 500K CIS data with type I error rates 0.048 (LR), 0.030 (*HyDe*), 0.036 (ABBA-BABA); (c) – 1M CIS data with type I error rates 0.042 (LR), 0.048 (*HyDe*), 0.050 (ABBA-BABA); (d) – multilocus data with 2.5K genes, and type I error rates are 0.090 (LR), 0.062 (*HyDe*), 0.072 (ABBA-BABA); (e) – multilocus data with 5K genes, and type I error rates are 0.064 (LR), 0.064 (*HyDe*), 0.084 (ABBA-BABA); (f) – multilocus data with 10K genes, and type I error rates are 0.064 (LR), 0.070 (*HyDe*), 0.060 (ABBA-BABA).

S3.2. Long branch tree

For the speciation times in Figure ??, we assigned the vector $(\tau_1, \tau_2, \tau_3) = (0.5, 1.0, 2.0)$. The hybridization parameter γ is chosen to be 0 or to vary from 0.06 to 0.5 by 0.02.

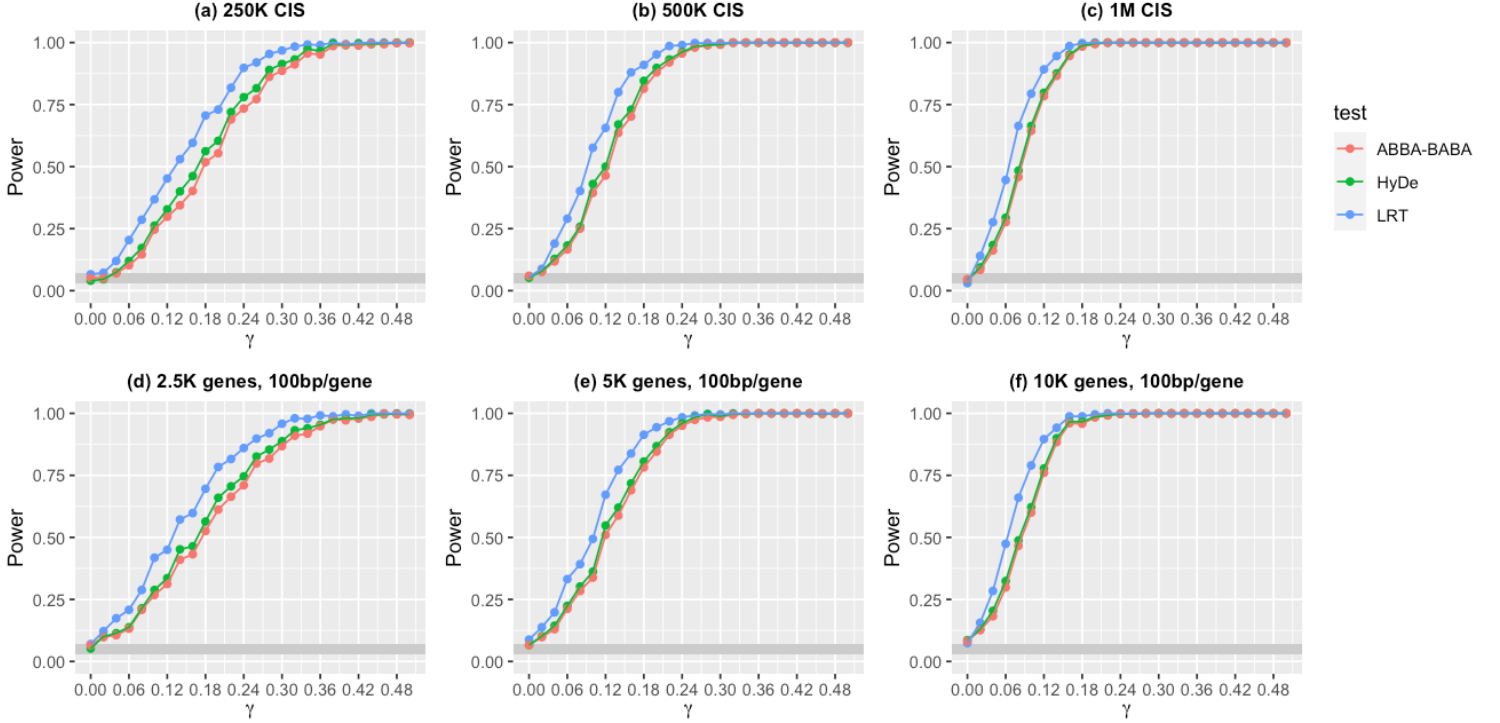


Figure S4: Hybrid detection power under JC69 for the long branch tree. The shaded area is the expected acceptance region (0.031,0.069) of the empirical type I error rate. (a) – 250K CIS data with type I error rates 0.066 (LR), 0.040 (*HyDe*), 0.052 (ABBA-BABA); (b) – 500K CIS data with type I error rates 0.056 (LR), 0.052 (*HyDe*), 0.060 (ABBA-BABA); (c) – 1M CIS data with type I error rates 0.030 (LR), 0.044 (*HyDe*), 0.046 (ABBA-BABA); (d) – multilocus data with 2.5K genes, and type I error rates are 0.070 (LR), 0.052 (*HyDe*), 0.066 (ABBA-BABA); (e) – multilocus data with 5K genes, and type I error rates are 0.088 (LR), 0.066 (*HyDe*), 0.066 (ABBA-BABA); (f) – multilocus data with 10K genes, and type I error rates are 0.074 (LR), 0.086 (*HyDe*), 0.082 (ABBA-BABA).