Supporting Information

Noh et al. 10.1073/pnas.1720324115

Table S1. Molecular evolution of chimera-biased genes

	Chimera-biased genes			
Metric	All	Up-regulated	Down-regulated	Genomic Background
N	78	19	59	11,241
Dn	269	52	217	37,595
Ds	1,130	196	934	129,871
Pn	62	14	48	11,554
Ps	159	27	132	15,273
α	0.15 (−0.39; 0.86)	0.25 (−0.13; 0.91)	0.10 (−2.09; 1.00)	−0.72 (−0.74; −0.48)
ω_a	0.001 (−0.02; 0.06)	0.02 (−0.01; 0.06)	0.007 (−0.15; 0.07)	−0.05 (−0.05; −0.03)
f	0.06 (0.01; 0.09)	0.05 (0.01; 0.07)	0.10 (0.00; 0.24)	0.14 (0.12; 0.14)
f_{st}	−0.18 (−0.25; −0.12)	−0.20 (−0.31; −0.08)	−0.18 (−0.26; −0.11)	−0.20 (−0.21; −0.19)
D	−0.61 (−1.11; −0.09)	−0.65 (−1.08; −0.19)	−0.47 (−1.15; 0.41)	−0.50 (−1.05; 0.12)
H	−0.27 (−0.70; 0.09)	−0.36 (−0.79; 0.03)	0.08 (−0.28; 0.37)	0.02 (−0.39; 0.37)

N is the number of genes; Dn and Ds are the counts of nonsynonymous and synonymous divergent sites, respectively, summed across loci. Pn and Ps are the counts of nonsynonymous and synonymous polymorphic sites, respectively, summed across loci, including singletons and without any adjustment to the minor allele frequency. From the DFE-α program, α is the proportion of nonsynonymous substitutions that are adaptive; ω_a is the rate of adaptive substitution; and f is the fraction of nonsynonymous mutations. For these, we give the point estimate and 95% confidence interval in parentheses determined by bootstrapping 1,000 times across loci. For f_{st}, Tajima's D, and Fay and Wu's H, we give the mean followed by the 95% confidence interval determined by bootstrapping the mean 10,000 times across loci. For details, see *Methods*.

Table S2. Polymorphism in genes expressed more in prespore and prestalk cells

Gene set	N	Site class	bp	π
Without Vegetative Expression				
Prespore	113	Non	107,861	3.1×10^{-4} (1.2×10^{-5}; 5.7×10^{-4})
		Syn	27,868	1.2×10^{-3} (4.1×10^{-4}; 2.2×10^{-3})
Prestalk	145	Non	90,571	2.4×10^{-4} (6.7×10^{-5}; 5.3×10^{-4})
		Syn	24,368	1.3×10^{-3} (3.3×10^{-4}; 2.8×10^{-3})
Prespore + Prestalk	258	Non	198,432	2.7×10^{-4} (8.9×10^{-5}; 5.4×10^{-4})
		Syn	52,236	1.2×10^{-3} (3.7×10^{-4}; 2.5×10^{-3})
All				
Prespore	879	Non	1,336,265	2.2×10^{-4} (8.9×10^{-5}; 4.3×10^{-4})
		Syn	347,860	1.0×10^{-3} (5.3×10^{-4}; 1.7×10^{-3})
Prestalk	992	Non	1,260,452	2.1×10^{-4} (6.5×10^{-5}; 5.5×10^{-4})
		Syn	343,576	1.2×10^{-3} (6.2×10^{-4}; 2.0×10^{-3})
Prespore + Prestalk	1,871	Non	2,596,717	2.1×10^{-4} (7.2×10^{-5}; 5.2×10^{-4})
		Syn	691,436	1.1×10^{-3} (5.6×10^{-4}; 1.9×10^{-3})
Genomic background	11,241	Non	15,102,488	1.9×10^{-4} (5.9×10^{-5}; 4.5×10^{-4})
		Syn	3,964,090	9.1×10^{-4} (4.2×10^{-4}; 1.6×10^{-3})

Two datasets are used: prespore/prestalk genes with no expression in vegetation cells and prespore/prestalk genes including those with vegetative expression (All). N refers to the numbers of genes in each dataset; Site Class refers to whether a polymorphic site is Non(synonymous) or Syn(onymous); bp is the total numbers of base pairs summed across loci; π is the Jukes-Cantor corrected pairwise nucleotide diversity, shown as the mean with the 95% confidence interval in parentheses as obtained by bootstrapping the mean 10,000 times. For details, see *Methods*.

Table S3. Prestalk π_N: Prespore π_N bootstrap test results

	Prestalk π_N: Prespore π_N	
Statistic	Without vegetative expression	All
Point estimate	0.771	0.914
Mean (95% CI)	0.799 (0.422; 1.380)	0.919; (0.704; 1.170)
Maximum value	2.236	1.478
Kin selection prediction $Y = 1.03$ bootstrap test result	$P = 0.33$	$P = 0.38$
Kin selection prediction $Y = 1.17$ bootstrap test result	$P = 0.16$	$P = 0.055$
Direct selection prediction $Y = 4$ bootstrap test result	$P << 0.001$	$P << 0.001$

Prestalk π_N: Prespore π_N is the ratio of nonsynonymous pairwise diversity for the prestalk genes to that of the prespore genes. Estimates are shown both for genes without vegetative expression and for all prestalk and prespore genes (as in Table S2). We show the mean of 10,000 bootstrapped ratios, the 95% confidence interval in parentheses, and the maximum ratio obtained by bootstrapping. Two-tailed P values for bootstrap tests against three predicted values, Y, come from doubling the percentage of values in the shorter tail of the bootstrapped distribution cut off by Y. For details, see *Methods*.

Other Supporting Information Files

Dataset S1 (XLSX)