Supporting Information

Hunt et al. 10.1073/pnas.1104825108

SI Text

Correlates of Protein Evolutionary Rate (dN/dS) in *S. invicta*. Analysis of potential evolutionary rate determinants (1–3) in *S. invicta* revealed that gene expression level was negatively correlated with dN/dS (Fig. S2*F*), as is the case in diverse taxa (4). In addition, coding sequence length was positively correlated with dN/dS (1) and intron number was negatively correlated with dN/dS (5), as observed in other taxa (1, 5). Interestingly, normalized CpG dinucleotide content, a measure of CpG depletion that is negatively correlated with DNA methylation in *S. invicta* (6), was positively correlated with dN/dS (Fig. S2*F*) and was largely decoupled from guanine-cytosine (GC) content and expression level (Table S4). Thus, methylated genes may be under greater functional constraint in *S. invicta* compared with unmethylated genes (3, 7). As is the case with the honey bee (8), GC content and codon use were tightly linked in *S. invicta* (third codon position synonymous GC content vs. Fop Spearman rank correlation = −0.996; Table S4).

Relationship Between Caste Bias and Protein Evolution in the Context of Other Forms of Conditional Gene Expression in *S. invicta*. We analyzed data describing tissue specificity of gene expression in *A. mellifera* (9) to determine whether the association between relaxed selection and caste-biased gene expression is driven by differences in pleiotropic constraints associated with tissue specificity (10). Our analyses revealed that the association between rates of protein evolution and caste-biased gene expression in *A. mellifera* was not weakened when controlling for tissue specificity in six tissue types (Spearman's partial correlation between *A. mellifera* dN from Fig. 4*A* and caste bias, when controlling for tissue specificity = 0.120, P = 3.94e-5; n = 1,164; Fig. S3). Furthermore, the association between rate of protein evolution and caste-biased gene expression in *S. invicta* remained significant when controlling for sex bias and developmental bias (Spearman's partial correlation between *S. invicta* dN from Fig. 4*A* and adult caste bias, when controlling for adult sex bias and developmental bias = 0.113, P = 0.0002; n = 1,050). This result is consistent with a scenario in which the observed correlation between caste bias and the rate of protein evolution is not explained by other forms of conditional gene expression.

dS and Caste in *S. invicta*. In the case of sex differences and developmental differences, dN was the primary driver of signal for increased dN/dS relative to unbiased genes (Fig. S5). Although dN is also associated with caste differences, dS was significantly lower for caste-biased genes relative to unbiased genes in both the pupal and adult stages (Fig. S5). Thus, caste-biased genes may have lower mutation rates or be subject to increased selection on synonymous codon use relative to unbiased genes.

1. Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL (2005) Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Mol Biol Evol* 22:1345–1354.
2. Pál C, Papp B, Lercher MJ (2006) An integrated view of protein evolution. *Nat Rev Genet* 7:337–348.
3. Hunt BG, Brisson JA, Yi SV, Goodisman MAD (2010) Functional conservation of DNA methylation in the pea aphid and the honeybee. *Genome Biol Evol* 2:719–728.
4. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352.
5. Carmel L, Rogozin IB, Wolf YI, Koonin EV (2007) Evolutionarily conserved genes preferentially accumulate introns. *Genome Res* 17:1045–1050.
6. Wurm Y, et al. (2011) The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci USA* 108:5679–5684.
7. Lyko F, et al. (2010) The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biol* 8:e1000506.
8. Jørgensen FG, Schierup MH, Clark AG (2007) Heterogeneity in regional GC content and differential usage of codons and amino acids in GC-poor and GC-rich regions of the genome of *Apis mellifera*. *Mol Biol Evol* 24:611–619.
9. Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R (2009) Epigenetic regulation of the honey bee transcriptome: Unravelling the nature of methylated genes. *BMC Genomics* 10:472.
10. Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H (2008) Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *Am Nat* 171:35–43.
11. Grozinger CM, Fan YL, Hoover SER, Winston ML (2007) Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol Ecol* 16:4837–4848.
12. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919.

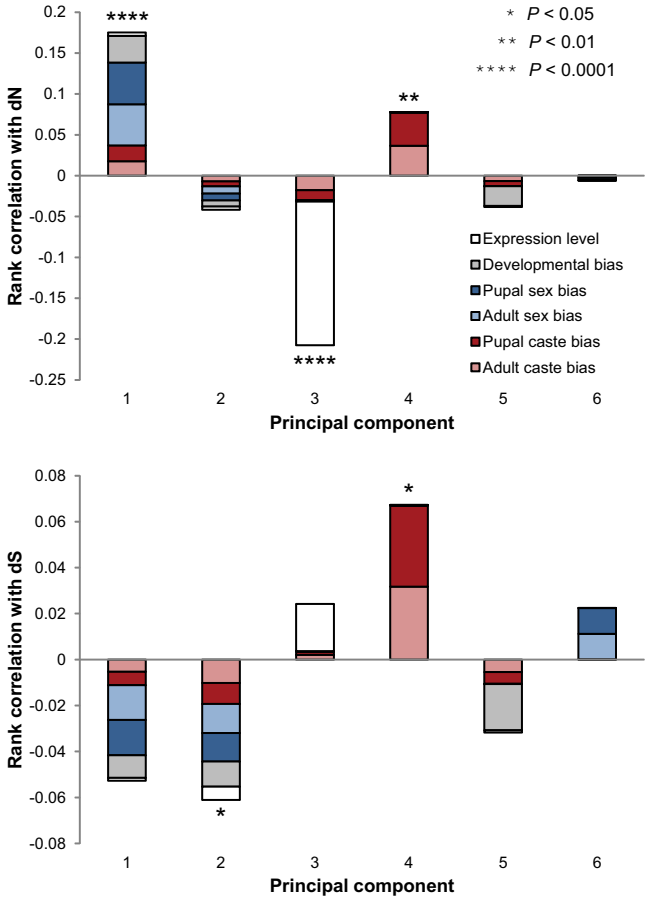

Fig. S1. Principal component analysis of gene expression measures and evolutionary rates in *S. invicta*. Bars represent the relative contribution of each gene expression measure to each principal component when normalized by the Spearman's rank correlation coefficient of each principal component with *S. invicta* dN and dS. The composition of principal components is illustrated in Fig. 2*A*.

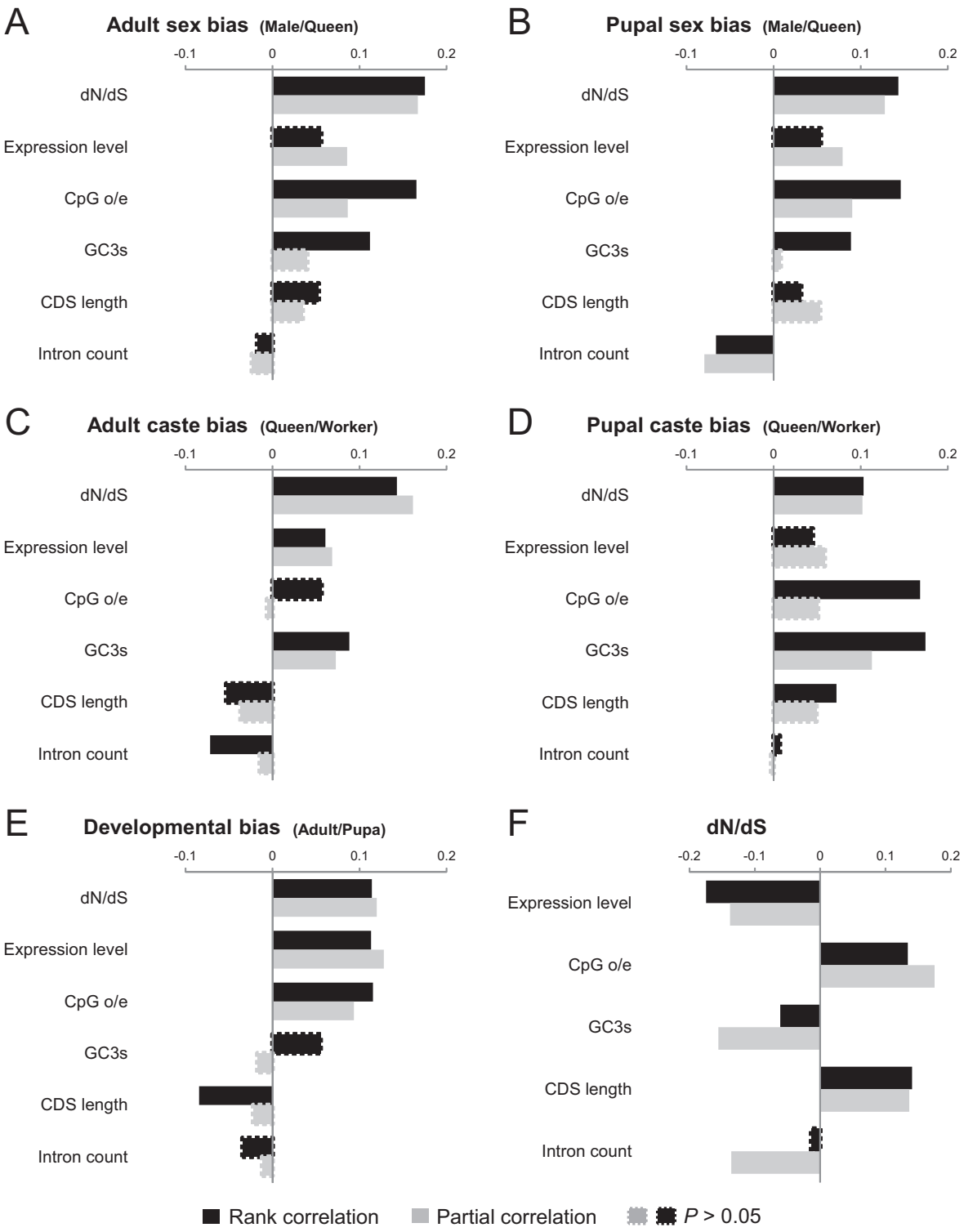

Fig. S2. Correlations between morph-biased gene expression, protein evolutionary rate dN/dS, gene expression level, normalized CpG dinucleotide content (CpG o/e, a negative correlate of DNA methylation), third codon position synonymous GC content (GC3s), coding sequence (CDS) length, and intron count in *S. invicta*. Correlations between dN/dS and measures of (*A*) adult and (*B*) pupal sex-biased gene expression, (*C*) adult and (*D*) pupal caste-biased gene expression, and (*E*) developmentally-biased gene expression are each significant, as are partial correlations. (*F*) Correlations between dN/dS and several *S. invicta* gene characteristics. Black bars represent Spearman's rank correlations. Gray bars represent partial correlations controlling for all other variables. Solid outlines on bars indicate $P < 0.05$ and wavy outlines indicate $P > 0.05$.

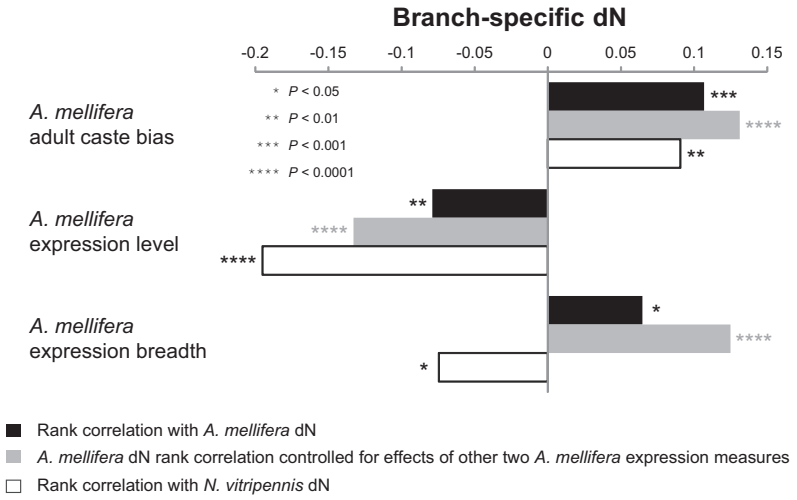

Fig. S3. Relationship between protein evolutionary rate, caste bias, gene expression level, and gene expression breadth in the honey bee *A. mellifera*. Branch-specific dN was determined for a three-species phylogeny between *A. mellifera*, *S. invicta*, and *N. vitripennis*, as in Fig. 4*A*. A total of 1,152 ortholog groups were analyzed with brain gene expression measures among castes (11), worker whole-body gene expression levels according to RNA-seq analysis (12), and the number of tissues with observed expression (ranging from 1 to 6) (9). Bars represent Spearman's rank correlations and partial correlations.

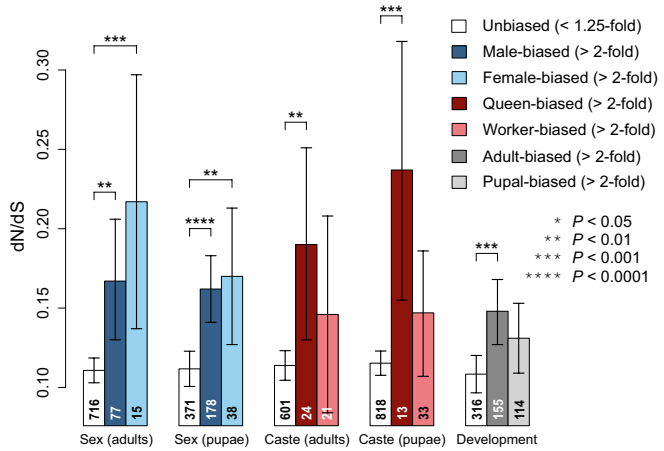

Fig. S4. Differences in relaxation of selective constraint according to phenotype in *S. invicta*. Genes with greater than twofold expression difference, categorized according to the phenotypic group (significance determined by Bonferroni-corrected Wilcoxon signed-rank tests). Means with 95% confidence intervals are plotted; the text in bars denotes the number of genes in each bin.

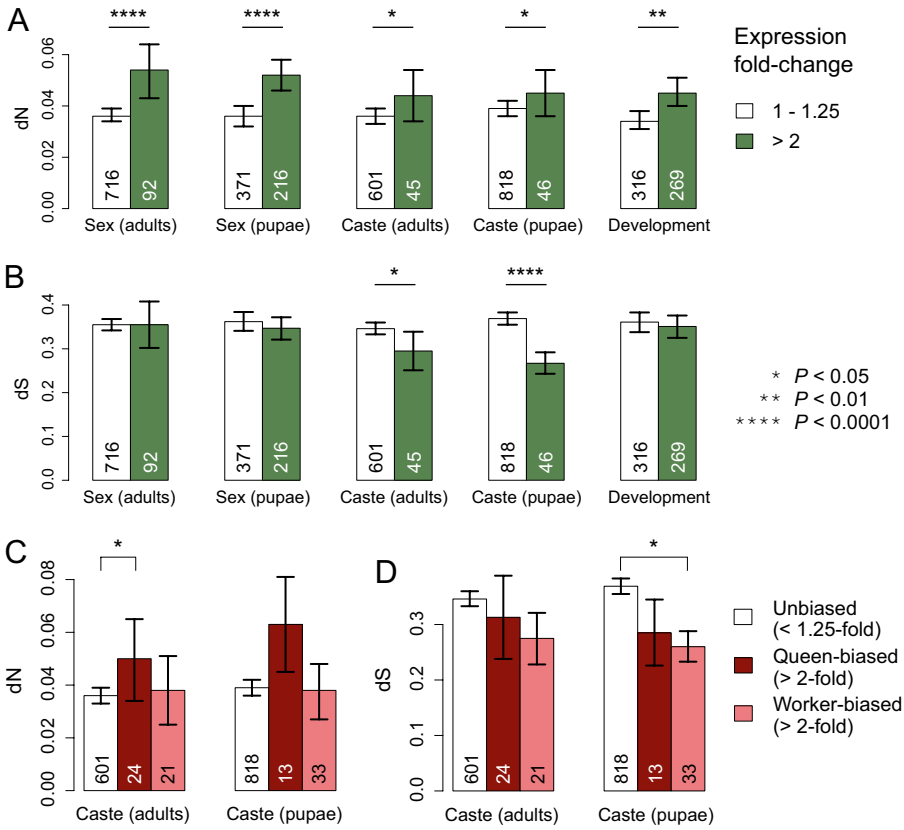

Fig. S5. dN and dS in relation to differential gene expression in *S. invicta*. (*A*) dN denotes rates of nonsynonymous substitution. (*B*) dS denotes rates of synonymous substitution. (*C*) Dissection of the unique properties of caste-biased genes with respect to dN and dS. Significance denotes results of Wilcoxon rank sum tests, subject to Bonferroni correction in *C* and *D*. Means with 95% confidence intervals are plotted.

Table S1. Genes under positive selection as determined by substitution rates averaged over all codons in a given gene

S. invicta gene ID	N·dN	S·dS	dN/dS	>Twofold expression	AmiGO* BLAST protein similarity	Gene ontology biological process
SI2.2.0_02232	28.8	0	∞	Male pupa up-regulation vs. queen pupa	No hits (P < 1e-5)	
SI2.2.0_12283	17.7	0	∞	None	*Mus musculus* 1110059E24Rik (P = 2.9e-21)	No biological process annotation
SI2.2.0_09898[†]	2	0	∞	None	*Drosophila melanogaster* mago nashi (P = 1.7e-71)	Cell-cell signaling (GO:0007267), nuclear mRNA splicing via spliceosome (GO:0000398), oogenesis (GO:0048477), 11 other terms

*Results generated using AmiGO (1) gene ontology database release 2010-11-20.
[†]Note that the low number of substitutions in this gene severely limits the inference of positive selection.

1. Carbon S, et al. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics* 25(2):288–289.

Table S2. Genes under positive selection as determined by branch-site analysis

S. invicta gene ID	Branch-site MUSCLE *P* value	Branch-site PRANK *P* value	>Twofold expression	AmiGO* BLAST protein similarity	Gene ontology biological process
SI2.2.0_02629	1.88e-11	1.19e-11	Male adult up-regulation vs. queen adult, male pupa up-regulation vs. queen pupa	No hits ($P < 1e-5$)	
SI2.2.0_05545	9.09e-7	8.07e-7	Male pupa up-regulation vs. queen pupa	*Drosophila melanogaster* Bifunctional phosphopantetheine adenylyltransferase-dephospho-CoA kinase ($P = 1.5e-95$)	CoA biosynthetic process (GO:0015937), imaginal disk-derived wing morphogenesis (GO:0007476), ovarian follicle cell migration (GO:0007297), two other terms
SI2.2.0_11797	4.39e-6	1.94e-5	None	*Gallus gallus* Uncharacterized protein ($P = 3.6e-249$)	Nucleobase, nucleoside, nucleotide, and nucleic acid metabolic process (GO:0006139)
SI2.2.0_12264	4.90e-10	1.37e-4	None	*Pan troglodytes* Dyslexia susceptibility 1 candidate gene 1 protein homolog ($P = 8.7e-56$)	Neuron migration (GO:0001764), regulation of estrogen receptor signaling pathway (GO:0033146), regulation of proteasomal protein catabolic process (GO:0061136)

*Results generated using AmiGO (1) gene ontology database release 2010-12-04.

1. Carbon S, et al. (2009) AmiGO: Online access to ontology and annotation data. *Bioinformatics* 25(2):288–289.

Table S3. Polymorphism and divergence summary for genes with unbiased and biased expression in *S. invicta*

Category	No. genes	Direction of selection*	D_n†	D_s†	P_n†	P_s†	D_n/D_s	P_n/P_s	*P* value‡
Less than 1.25-fold bias in all comparisons	23	−0.128	564	2,816	11	26	0.200	0.423	0.054
Less than 1.5-fold bias in all comparisons	99	−0.125	2,139	11,323	48	120	0.189	0.400	4.3e-5
Less than twofold bias in all comparisons	223	−0.126	5,814	24,779	124	268	0.235	0.463	3.4e-9
Greater than twofold bias in one or more of five comparisons	158	−0.050	5,752	16,839	92	209	0.342	0.440	0.048
Greater than twofold bias between adult sexes	34	0.004	1,783	3,701	20	43	0.482	0.465	0.898
Greater than twofold bias between pupal sexes	80	−0.024	3,919	8,669	52	102	0.452	0.510	0.487
Greater than twofold bias between adult castes	18	−0.005	668	1,938	11	31	0.345	0.355	0.935
Greater than twofold bias between pupal castes	12	−0.039	357	807	7	13	0.442	0.538	0.685
Greater than twofold bias between developmental stages	100	−0.050	3,050	10,156	54	137	0.300	0.394	0.101

*Direction of selection $[D_n/(D_n + D_s) - P_n/(P_n + P_s)]$ is calculated according to Stoletzki and Eyre-Walker (1), where a negative value indicates purifying selection and a positive value indicates positive selection.
†D_s represents the total number of synonymous fixed differences, P_s represents the total number of synonymous polymorphisms, D_n represents the total number of nonsynonymous fixed differences, and P_n represents the total number of nonsynonymous polymorphisms (*Methods*).
‡*P* value denotes the results of the McDonald–Kreitman (2) test according to a *G* test of independence (with the Williams correction for continuity).

1. Stoletzki N, Eyre-Walker A (2011) Estimation of the neutrality index. *Mol Biol Evol* 28:63–70.
2. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.

Table S4. Associations between *S. invicta* codon use, CpG depletion, expression level, and evolutionary rates

	Principal component†		
	1	2	3
Percent variance explained in CpG o/e, expression level, Fop, and GC3s	60.923	24.715	14.265
Rank correlation with *S. invicta* dN/dS	−0.02	−0.19****	0.18****
Rank correlation with *S. invicta* dN	−0.09**	−0.17****	0.18****
Rank correlation with *S. invicta* dS	−0.17****	0.03	0.01
Percent contributions to principal component			
CpG o/e	22.4	2.7	74.9
Expression level	1.3	97.1	1.6
Fop	38.2	0.1	11.8
GC3s	38.2	0.1	11.7

†Principal component 4 explains 0.10% of the variance in CpG o/e, expression level, Fop, and GC3s, and it is not presented. CpG o/e, normalized CpG dinucleotide content; GC3s, third codon position synonymous GC content.
**$P < 0.01$; ****$P < 0.0001$.