Supplementary Material

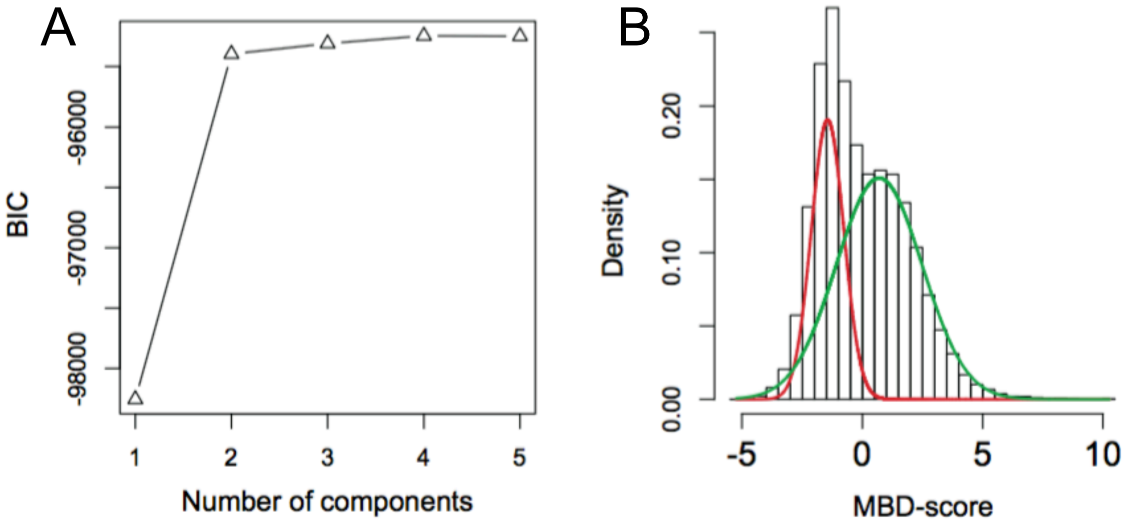


Figure S1: Fitting of Gaussian mixture components to distribution of MBD-scores. (A) Plot of Bayesian Information Criteria for models of the distribution of MBD-scores using different numbers of Gaussian components. (B) Traces of the two-component model overlaid on the distribution.

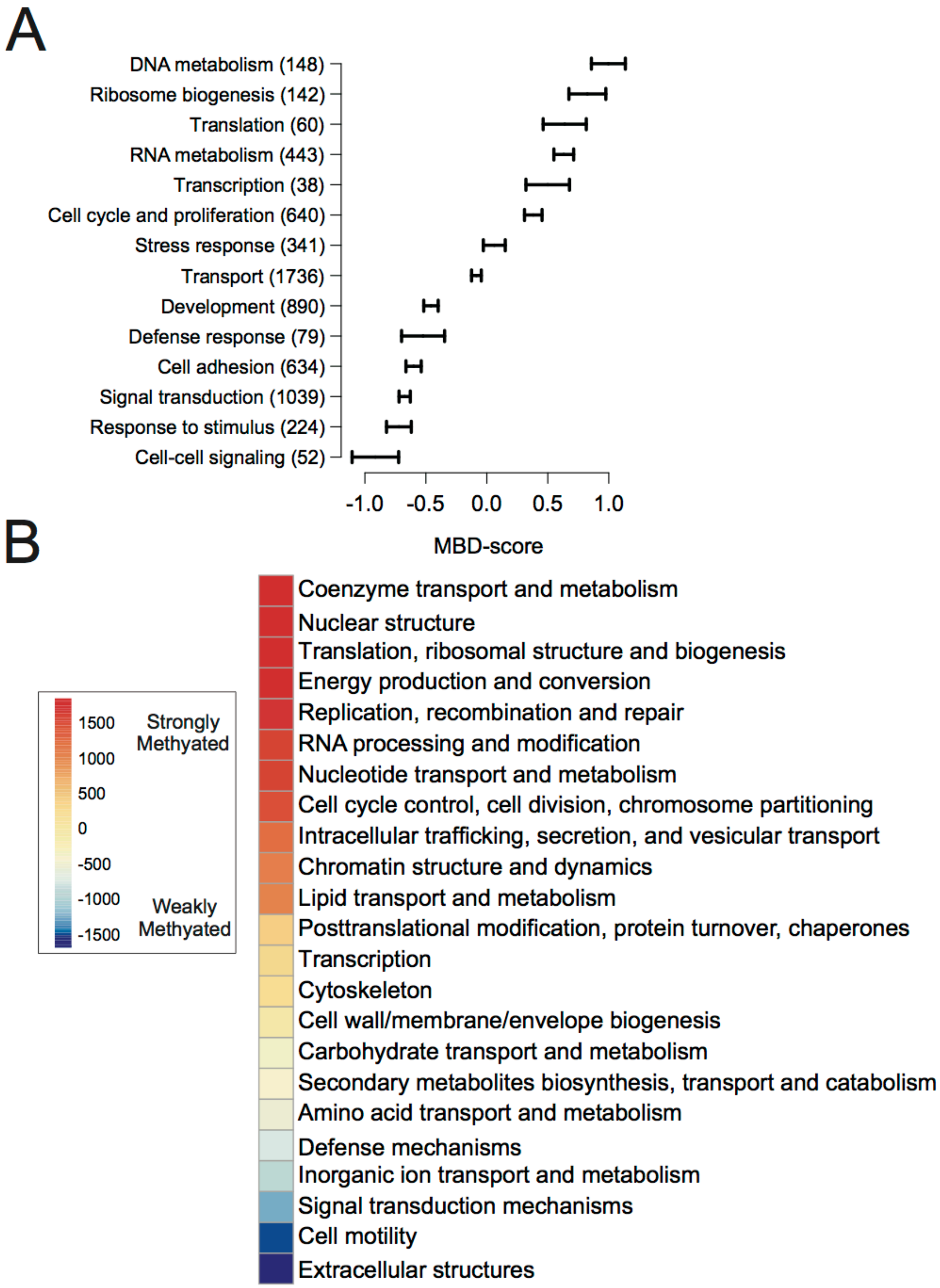


Figure S2: Relationship between gene functional categories and MBD-score. (A) Mean MBD-score for a selected set of Gene Ontology (GO) terms for biological processes. Error bars indicate standard error. (B) Enrichment of KOG terms based on Mann-Whitney U tests implemented in the R package KOGMWU as in Dixon et al. (2015).

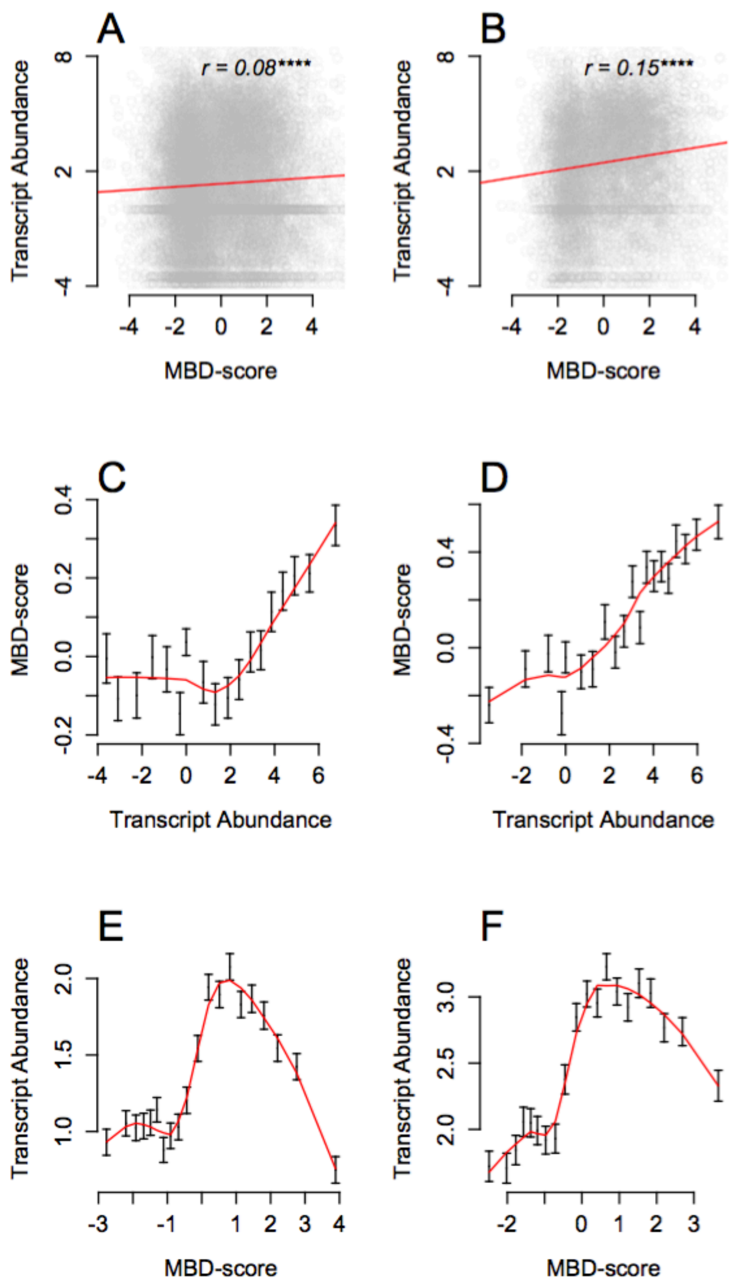


Figure S3: Relationships between transcript abundance and MBD-score. Figures are paired to illustrate interrelationship with gene length. Left panels show relationships for all coding sequences, right panels for coding sequences longer than 800 bp. (A-B) Correlation between MBD-score and normalized transcript abundance. Correlation is given as Spearman's Rho (r). Asterisks denote significance based on Spearman's rank tests. Red line traces least squared linear regression. (C-D) Highly expressed genes tend to be strongly methylated. Mean MBD-score was plotted for 12 quantiles of genes ranked by transcript abundance. Error bars indicate standard error. (E-F) MBD-score generally predicts higher expression, but the most strongly methylated genes show lower expression. This effect is especially true for shorter genes, an effect also described in *Arabidopsis* (Zilberman et al. 2007). Significance notation: ns > 0.05; * < 0.05; ** < 0.01; ***<0.001; **** < 0.0001.

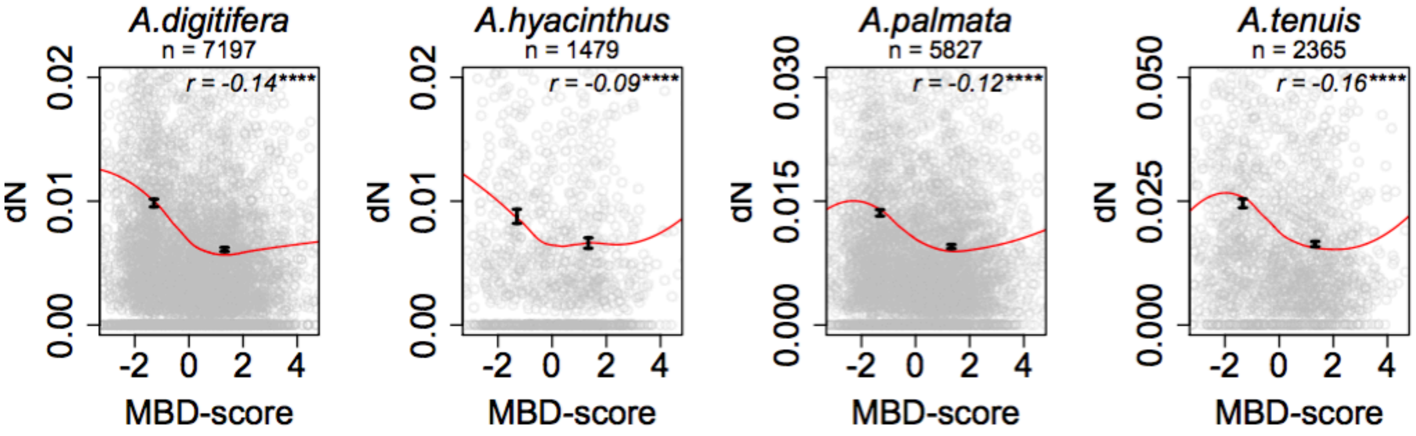


Figure S4: Relationship between MBD-score and pairwise dN between *A. millepora* and other *Acropora* species. Genes were divided into 20 quantiles based on MBD-score. Error bars show mean and standard error for the weakly methylated and strongly methylated genes. The number of orthologs (n) is given at the top of each panel. Asterisks indicate significance based on Spearman's rho (r). Red lines are smoothed traces of the relationship across all points.

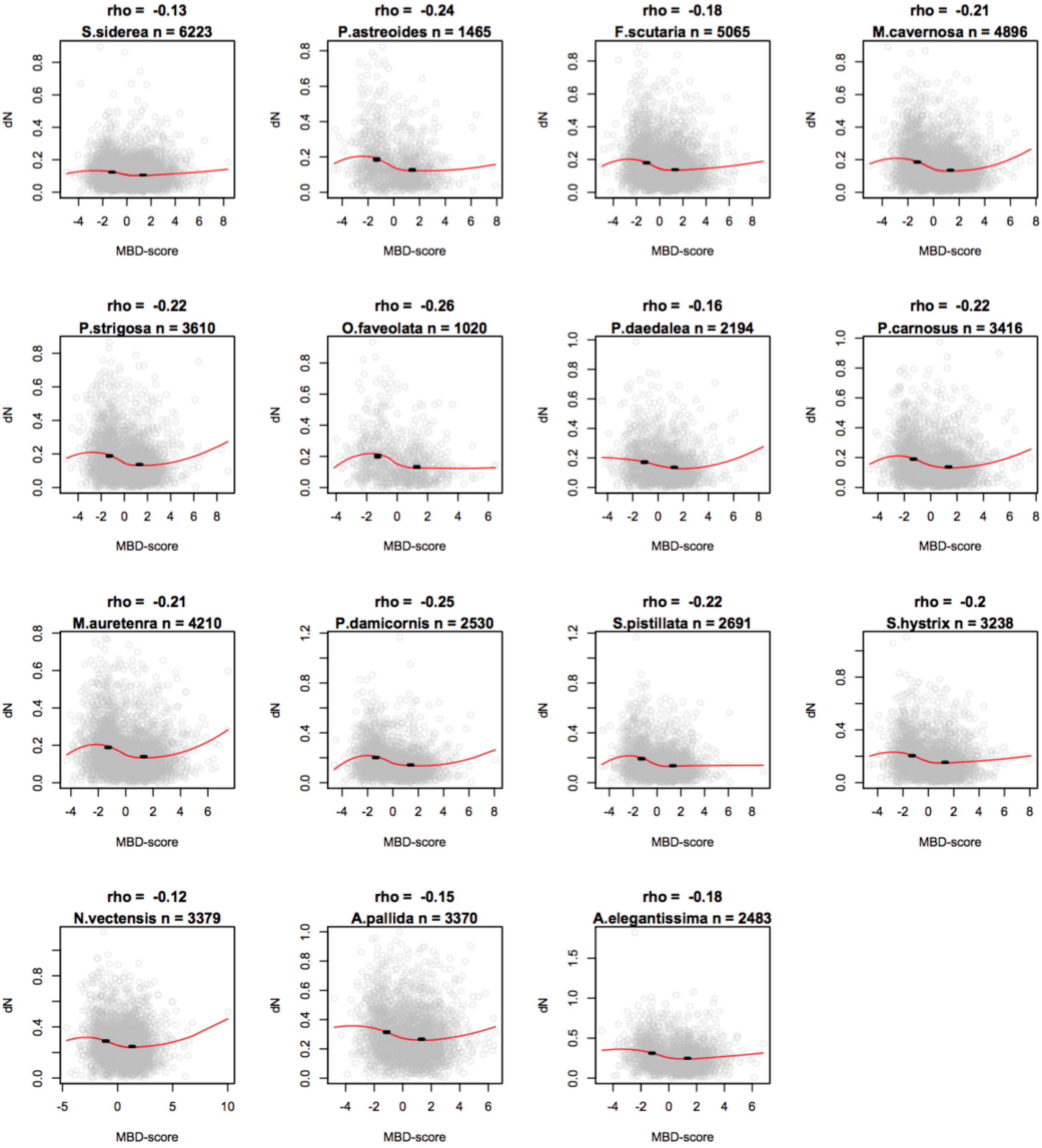


Figure S5: Relationship between nonsynonymous substitution rate (dN) and MBD-score across all species outside of *Acropora*. The two error bars in each panel display mean dN and standard error for the strongly methylated (MBD-score >= 0) and weakly methylated (MBD-score < 0) genes. Correlations are given as Spearman's Rho. All p values for Spearman's rank correlation test were < 0.0001. Red lines are smoothed traces with a span of 0.8.

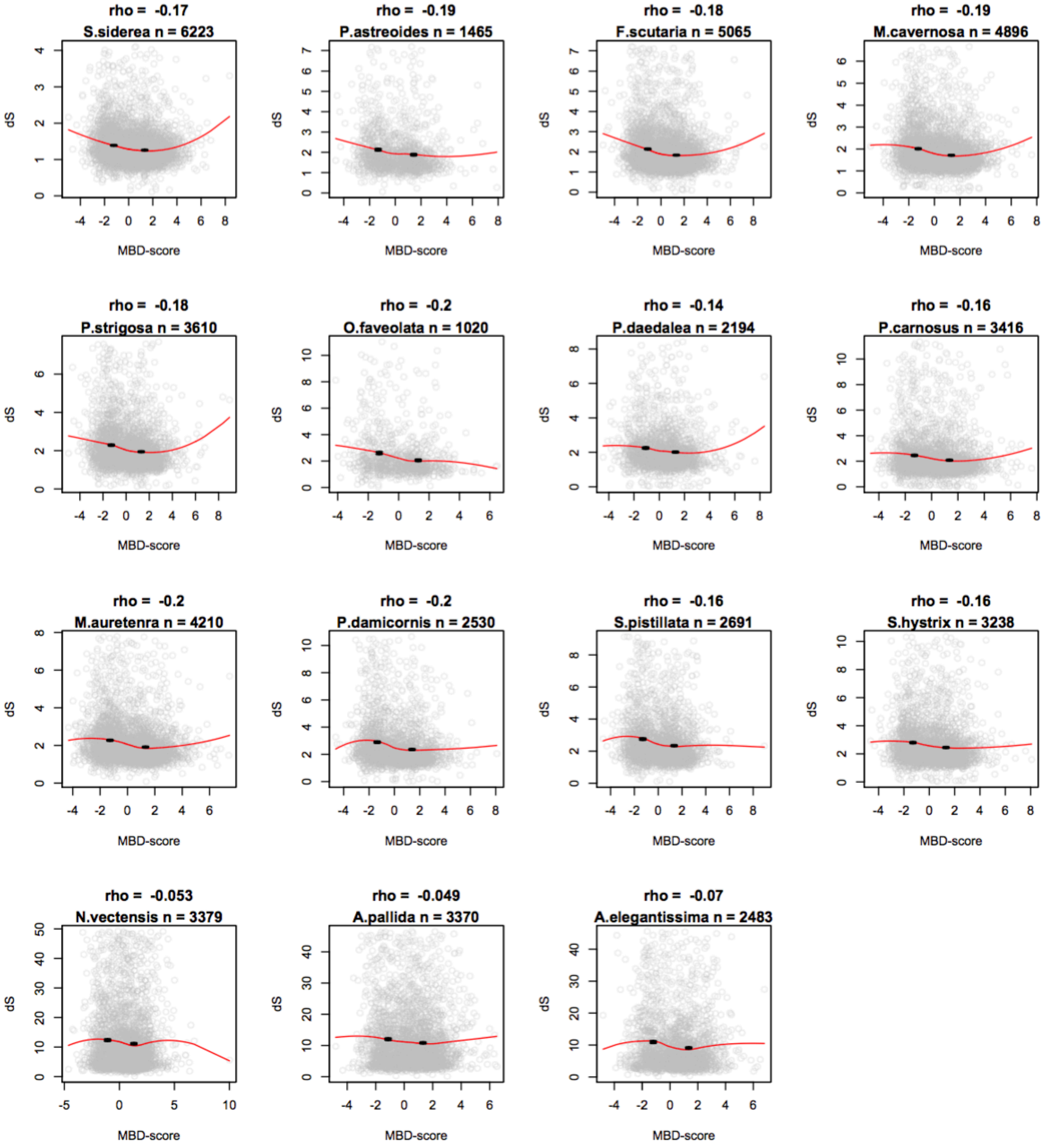


Figure S6: Relationship between synonymous substitution rate (dS) and MBD-score across all species outside of *Acropora*. The two error bars in each figure display mean dS and standard error for the strongly methylated (MBD-score >= 0) and weakly methylated (MBD-score < 0) genes. Correlation is given as Spearman's Rho. All p values for Spearman's rank correlation test were < 0.0001. Red lines are smoothed traces with a span of 0.8.

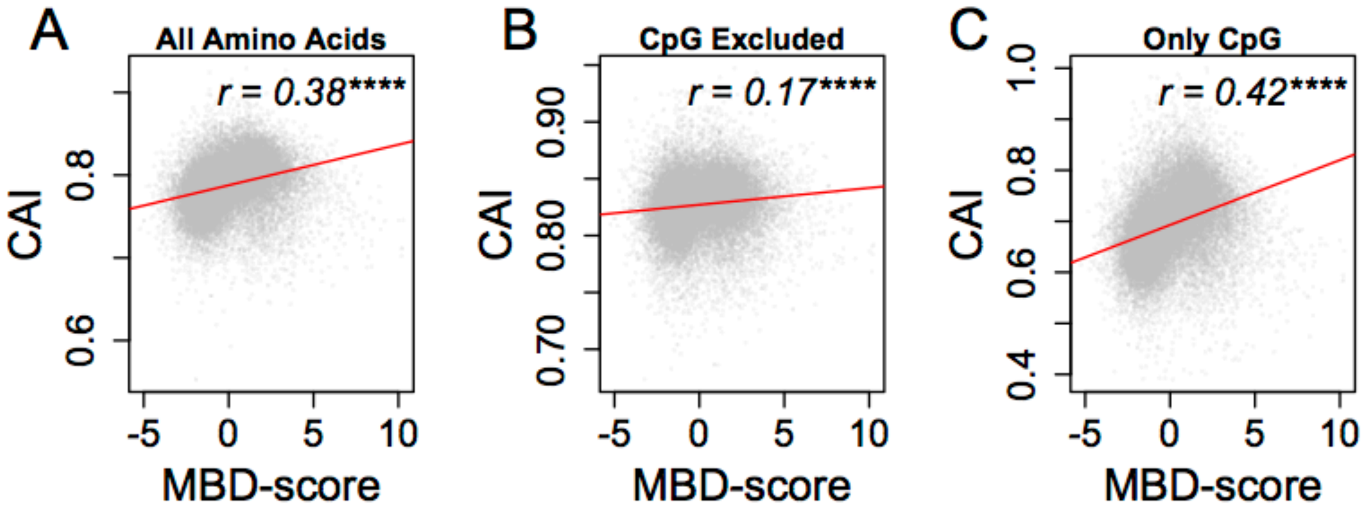


Figure S7: Correlation between codon adaptation index (CAI) and MBD-score with and without amino acids coded for by codons with CpG dinucleotides (Serine, Proline, Threonine, Alanine and Arginine). (A) Correlation between CAI and MBD-score with all amino acids included. (B) Calculating CAI with Serine, Proline, Threonine, Alanine and Arginine severely reduces correlation. (C) Calculating CAI based solely on Serine, Proline, Threonine, Alanine and Arginine increases strengthens correlation. Asterisks indicate significance based on Spearman's rank tests. Red lines trace least squared regression.

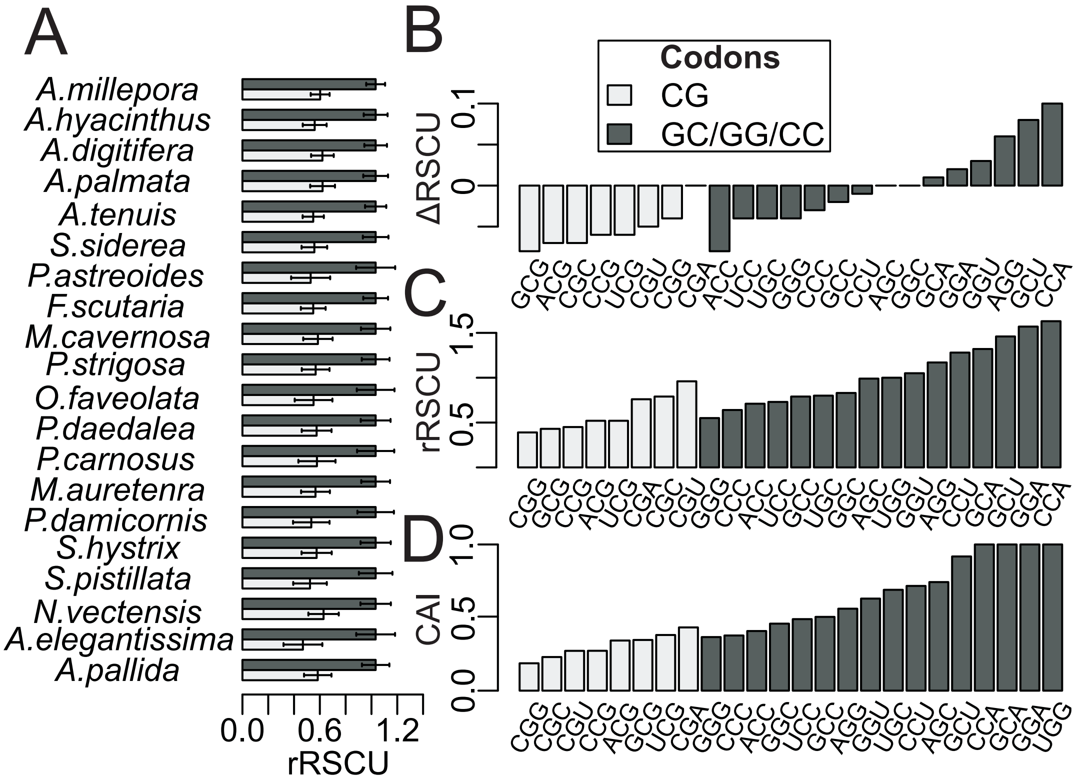


Figure S8: Codons bearing CpG dinucleotides are underrepresented in highly expressed genes. A) Comparison of mean Relative Synonymous Codon Usage for CG bearing codons compared to all other codons bearing GC, GG, or CC dinucleotides in ribosomal genes. Error bars show standard error. A value of 1 for this metric indicates no bias. B) Comparison of ΔRSCU for codons bearing CG, GC, GG or CC dinucleotides in *A. millepora*. ΔRSCU is the difference in RSCU between the top 5% most highly expressed genes and bottom 5%. Negative values indicate underrepresentation in highly expressed genes. C) Comparison of RSCU for CG, GC, GG, or CC codons in ribosomal genes from *A. millepora*. Values less than one indicate underrepresentation in ribosomal genes. D) Comparison of relative adaptiveness (*W*) for CG, GC, GG, or CC codons in *A. millepora*. Here a value of 1 indicates that the codon is optimal for its amino acid. No CpG codons were optimal, and were all less than half as frequent as the optimal codon.

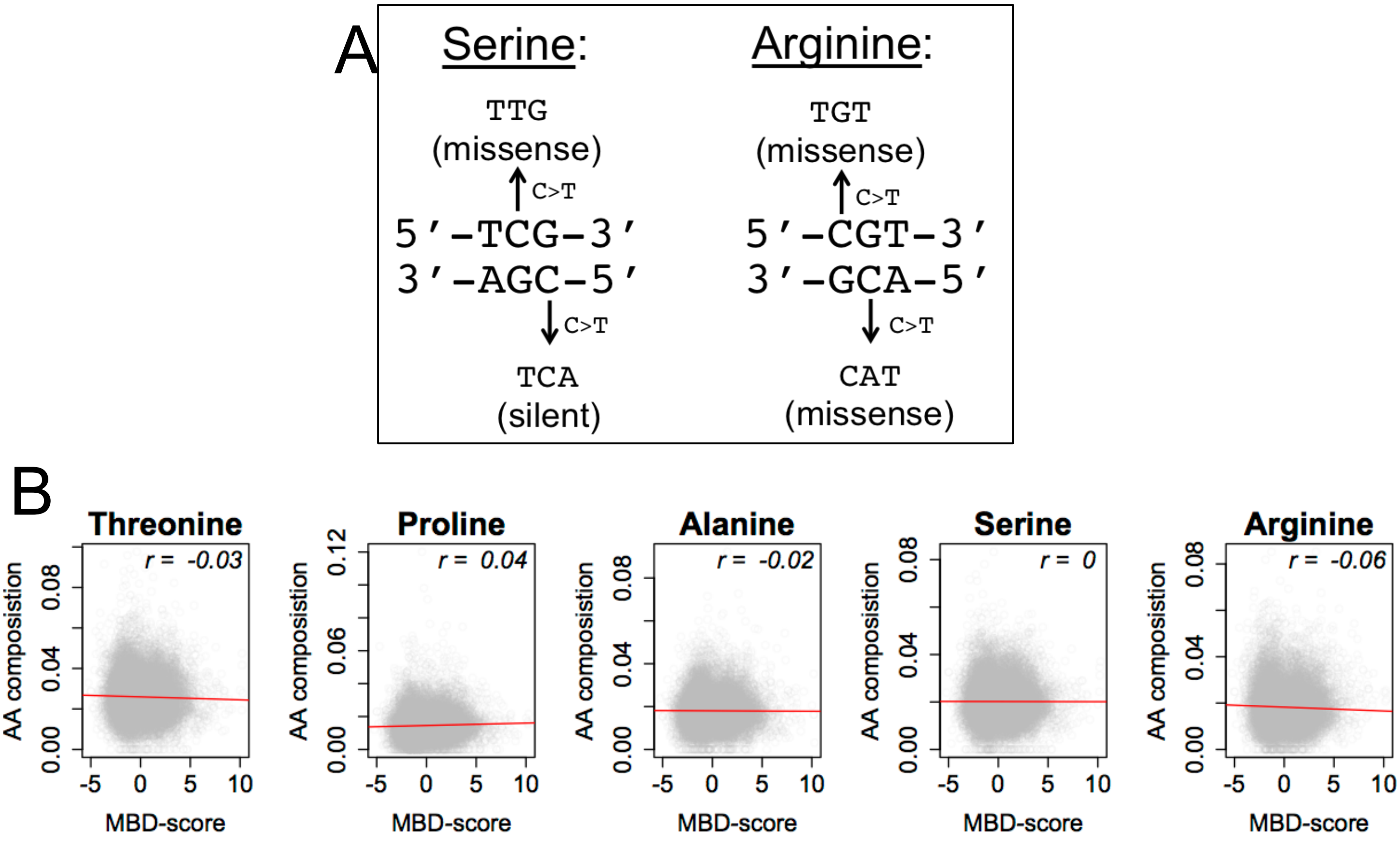


Figure S9: Loss of CpG bearing codons occurs through silent C>T substitution on the antisense stand. Methylated cytosines tend to be substituted for thymine (Shen et al. 1994). (A) On the sense strand, 5mC>T substitutions result in amino acid changes, whereas 5mC>T substitutions on the antisense strand are silent. (B) MBD-score shows little correlation with amino acid content, indicating that purifying selection counteracts most nonsynonymous 5mC>T substitutions. Although the correlation is weak, arginine content shows a stronger negative correlation than of the other amino acid. This is consistent with the fact for CGN codons, 5mC>T substitutions on either strand will replace the arginine.

Table S1: Optimal codons identified based on correspondence analysis of codon usage implemented in CodonW. X^2 indicates the chi-square statistic describing the enrichment of the synonymous codons (see supplemental methods below). All NCA codons were identified as optimal and tended toward higher X^2. No NCG codons were optimal.

codon	**amino acid**	X^2
AGA	Arg	562.906
UUU	Phe	389.064
ACA	Thr	309.689
GCA	Ala	286.519
AGG	Arg	277.184
UCA	Ser	188.606
AAU	Asn	143.76
AGU	Ser	132.991
CCA	Pro	132.423
UAU	Tyr	111.065
AUA	Ile	95.299
GAU	Asp	91.223
CAU	His	85.921
UGU	Cys	59.489
UUA	Leu	49.41
GGA	Gly	48.125
AUU	Ile	37.656
GUG	Val	31.472
CUA	Leu	27.806
CUG	Leu	19.063
GGG	Gly	18.105
GUA	Val	17.354
AAG	Lys	8.718
GAA	Glu	4.369

Table S2: Relative adaptiveness of codons in *Acropora millepora* (see supplemental methods below). NCA codons are highlighted in green and tend to have values equal to close to the maximum of 1.00. NCG codons are highlighted in red and always have the lowest relative adaptiveness value for their respective amino acids.

codon	amino acid	wi	rscu
GCU	Ala	1.00	1.39
GCA	Ala	0.96	1.33
GCC	Ala	0.58	0.80
GCG	Ala	0.34	0.47
AGA	Arg	1.00	1.98
AGG	Arg	0.57	1.13
CGA	Arg	0.48	0.96
CGU	Arg	0.41	0.82
CGC	Arg	0.34	0.67
CGG	Arg	0.22	0.44
AAU	Asn	1.00	1.09
AAC	Asn	0.83	0.91
GAU	Asp	1.00	1.21
GAC	Asp	0.65	0.79
UGU	Cys	1.00	1.14
UGC	Cys	0.75	0.86
CAA	Gln	1.00	1.06
CAG	Gln	0.89	0.94
GAA	Glu	1.00	1.21
GAG	Glu	0.65	0.79
GGA	Gly	1.00	1.56
GGU	Gly	0.71	1.11
GGC	Gly	0.52	0.81
GGG	Gly	0.34	0.53
CAU	His	1.00	1.14
CAC	His	0.75	0.86
AUU	Ile	1.00	1.4
AUC	Ile	0.67	0.94
AUA	Ile	0.48	0.67
UUG	Leu	1.00	1.45
CUU	Leu	0.97	1.40
CUG	Leu	0.79	1.15
UUA	Leu	0.57	0.82
CUC	Leu	0.48	0.69
CUA	Leu	0.34	0.50
AAA	Lys	1.00	1.09
AAG	Lys	0.83	0.91
AUG	Met	1.00	1.00
UUU	Phe	1.00	1.22
UUC	Phe	0.64	0.78
CCA	Pro	1.00	1.63
CCU	Pro	0.78	1.27
CCC	Pro	0.39	0.64
CCG	Pro	0.28	0.46
UCA	Ser	1.00	1.36
AGU	Ser	0.88	1.20
UCU	Ser	0.88	1.20
AGC	Ser	0.70	0.95
UCC	Ser	0.56	0.76
UCG	Ser	0.39	0.53
ACA	Thr	1.00	1.53
ACU	Thr	0.78	1.19
ACC	Thr	0.49	0.75
ACG	Thr	0.35	0.53
UGG	Trp	1.00	1.00
UAU	Tyr	1.00	1.02
UAC	Tyr	0.96	0.98
GUU	Val	1.00	1.37
GUG	Val	0.89	1.22
GUC	Val	0.57	0.78
GUA	Val	0.47	0.64

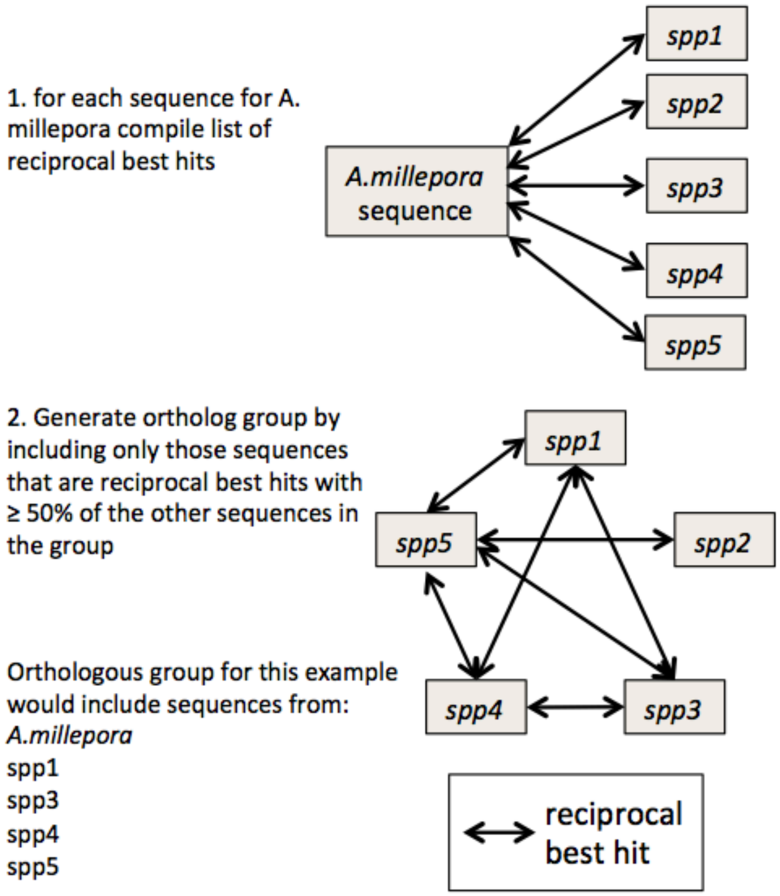
Table S3: Spearman's rank correlations between gene characteristics: Codon adaptation index (CAI), Effective number of codons (Nc), Frequency of optimal codons (Fop), log2 fold difference between methylation binding domain captured and flow-through fractions (MBD-score), transcript abundance (mRNA), length of the coding region (length), normalized CpG content (CpGo/e), and GC content (GC).

Variable	CAI	Nc	Fop	MBD-score	mRNA	length	CpGo/e	GC
CAI	1.00	-0.52	0.33	0.38	0.16	0.05	-0.71	-0.61
Nc		1.00	-0.24	-0.31	-0.02	0.17	0.44	0.34
Fop			1.00	0.17	0.18	0.09	-0.33	-0.01
MBD-score				1.00	0.08	0.05	-0.51	-0.22
mRNA					1.00	0.36	-0.12	0.00
length						1.00	-0.08	0.11
CpGo/e							1.00	0.33
GC								1.00

Table S4: Anthozoan species with publicly available reference transcriptomes used in this study.

Order	**Family**	**Genus**	**Species**	**Citation**	**URL**
Actiniaria	Actiniidae	*Anthopleura*	*elegantissima*	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Actiniaria	Aiptasiidae	*Aiptasia*	*pallida*	Sunagawa et al. 2009	http://pringlelab.stanford.edu/projects.html
Actiniaria	Edwardsiidae	*Nematostella*	*vectensis*	Nordberg et al. 2014	http://genome.jgi-psf.org/Nemve1/Nemve1.download.ftp.html
Scleractinia	Acroporidae	*Acropora*	*palmata*	Polato *et al.* 2011	http://www.personal.psu.edu/ibb3/Research.htm#Data
Scleractinia	Acroporidae	*Acropora*	*hyacinthus*	Willette et al. 2014	http://www.bio.utexas.edu/research/matz_lab/matzlab/Data.html
Scleractinia	Acroporidae	*Acropora*	*tenuis*	none	http://www.bio.utexas.edu/research/matz_lab/matzlab/Data.html
Scleractinia	Acroporidae	*Acropora*	*millepora*	Moya et al. 2012	http://www.bio.utexas.edu/research/matz_lab/matzlab/Data.html
Scleractinia	Acroporidae	*Acropora*	*digitifera*	Shinzato et al. 2011	http://marinegenomics.oist.jp/genomes/downloads?project_id=3
Scleractinia	Astocoeniidae	*Madracis*	*auretenra*	none	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Montastraeidae	*Montastraea*	*cavernosa*	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Faviidae	*Platygyra*	*carnosus*	Sun et al. 2013	http://www.comp.hkbu.edu.hk/~db/PcarnBase/
Scleractinia	Faviidae	*Platygyra*	*daedalea*	none	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Fungiidae	*Fungia*	*scutaria*	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Merulinidae	*Orbicella*	*faveolata*	Schwarz et al. 2008	http://www.compagen.org/
Scleractinia	Mussidae	*Pseudodiploria*	*strigosa*	none	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Pocilloporidae	*Pocillopora*	*damicornis*	Lubinski & Granger 2013	http://cnidarians.bu.edu/PocilloporaBase/cgi-bin/pdamdata.cgi
Scleractinia	Pocilloporidae	*Seriatopora*	*hystrix*	Kitchen et al. 2015	http://people.oregonstate.edu/~meyere/data.html
Scleractinia	Pocilloporidae	*Stylophora*	*pistillata*	Maor-Landaw et al. 2014	http://data.centrescientifique.mc/Data/
Scleractinia	Poritidae	*Porites*	*astreoides*	Kenkel et al. 2013	http://www.bio.utexas.edu/research/matz_lab/matzlab/Data.html
Scleractinia	Siderastreidae	*Siderastrea*	*siderea*	Davies et al. 2015	https://sarahwdavies.wordpress.com/data/

*See main text for full citations



1. for each sequence for A. millepora compile list of reciprocal best hits

2. Generate ortholog group by including only those sequences that are reciprocal best hits with ≥ 50% of the other sequences in the group

Orthologous group for this example would include sequences from:
A.millepora
spp1
spp3
spp4
spp5

Figure S10: Schematic representation of ortholog assignment method. Sequences from *A. millepora* were used as anchors. For each sequence, reciprocal best hits from each other species were assembled as candidate orthologs. This group of candidates was then subset by iteratively removing sequences that were reciprocal best its with < 50% of other sequences within the group.

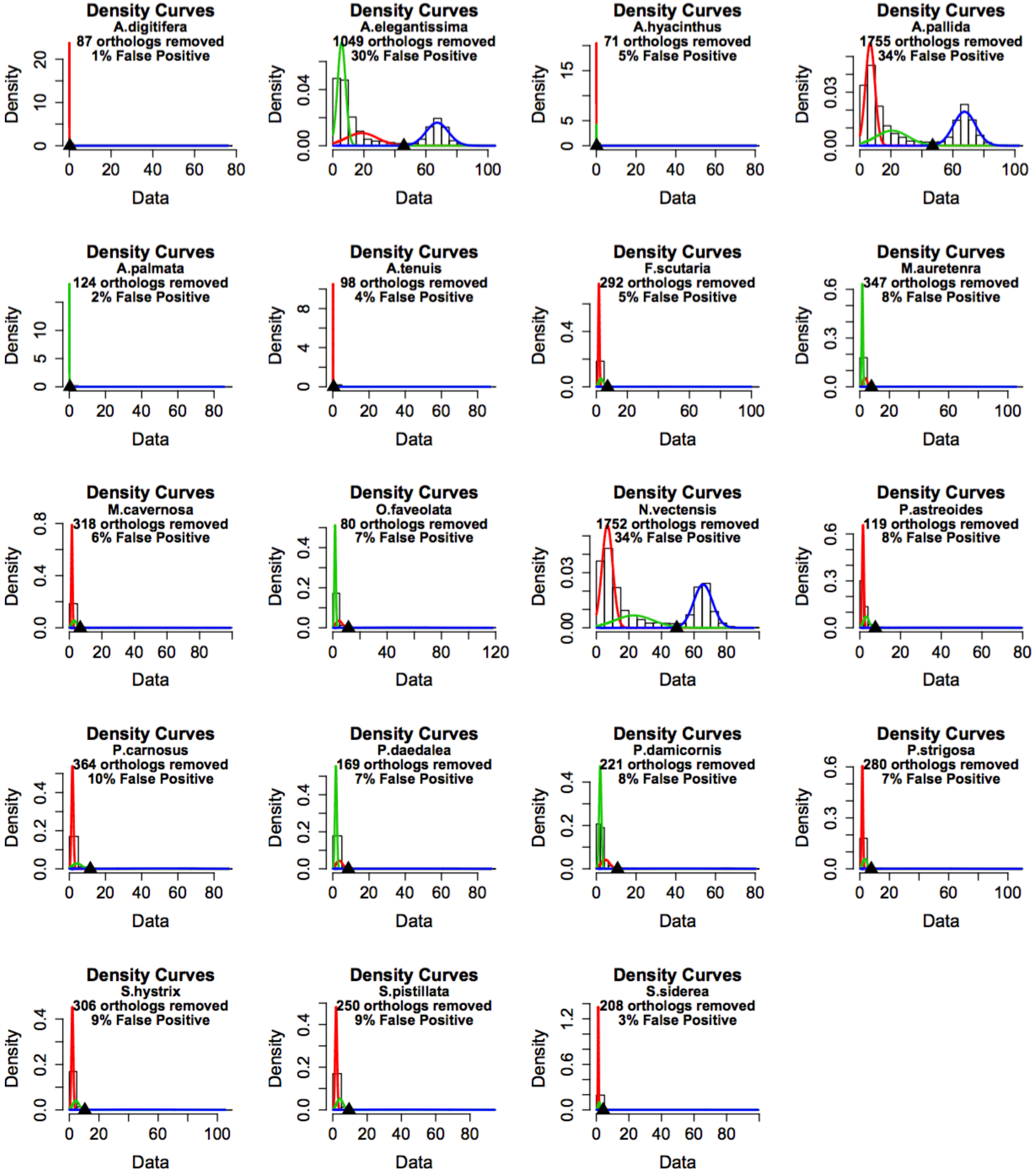


Figure S11: Identification and removal of false-positive ortholog calls. A three component Gaussian mixture model was fitted to the pairwise dS estimates with *A. millepora* for each species. The third component (blue above) was assumed to represent false positives. These orthologs (to the right of the black triangle) were removed from further analysis. The number and percentage of false positives removed is given in the title for each figure. The three anemone species, (*A. elegantissima*, *A. pallida*, and *N. vectensis*) displayed much greater rates of false positives.