

Supplementary material for:

Positive selection acting on splicing motifs reflects compensatory evolution

Shengdong Ke^{*}, Xiang, H.-F. Zhang^{*†}, and Lawrence A. Chasin[‡]

Department of Biological Sciences
Columbia University
New York, NY 10027 USA

Three tables and three figures:

Table S1

Table S2

Table S3

Fig. S1

Fig. S2

Fig. S3

Genome Research 2008

Table S1. Rates of predicted splicing regulatory motif changes between human and macaque

	Disruptions				Creations			
	ESE	non-ESE Control	ESS	non-ESS Control	ESE	non-ESE Control	ESS	non-ESS Control
Constitutive exons:								
1. K_s disrupt or create	0.019	0.022	0.029	0.029	0.017	0.015	0.007	0.009
2. K_a disrupt or create	0.004	0.005	0.005	0.006	0.004	0.003	0.002	0.002
3. K_i disrupt or create	0.027	0.026	0.030	0.033	0.016	0.017	0.013	0.013
4. Odds Ratio of K_s : K_i	0.702	* 0.827	0.979	* 0.883	1.091	* 0.903	0.538	* 0.721
5. 95% confidence interval	0.691-0.714	0.814-0.841	0.959-1.000	0.865-0.912	1.0757-1-1066	0.891-0.916	0.530-0.546	0.711-0.731
6. Odds ratio of K_a : K_i	0.192	* 0.207	0.166	* 0.205	0.219	* 0.225	0.221	* 0.240
7. 95% confidence interval	0.188-0.195	0.203-0.211	0.161-0.171	0.200-0.210	0.215-0.223	0.221-0.229	0.216-0.226	0.236-0.245
Cassette exons:								
8. K_s disrupt or create	0.014	0.017	0.021	0.022	0.012	0.013	0.006	0.007
9. K_a disrupt or create	0.004	0.004	0.006	0.006	0.004	0.003	0.002	0.002
10. K_i disrupt or create	0.027	0.026	0.030	0.033	0.016	0.017	0.013	0.013
11. Odds Ratio of K_s : K_i	0.517	* 0.633	0.700	0.651	0.769	0.776	0.464	* 0.573
12. 95% confidence interval	0.493-0.541	0.604-0.665	0.655-0.748	0.613-0.692	0.734-0.806	0.744-0.808	0.442-0.487	0.547-0.600
13. Odds ratio of K_a : K_i	0.134	* 0.174	0.160	* 0.177	0.241	* 0.205	0.120	* 0.174
14. 95% confidence interval	0.131-0.137	0.169-0.180	0.154-0.166	0.172-0.183	0.236-0.246	0.201-0.209	0.117-0.123	0.170-0.178

* An asterisk between two values indicates that the difference between the experimental and control values is statistically significant.

Table S2. Variance of the distribution of differences in number of motifs between macaque and human orthologous exons

Change	Variance		p-value
	Real set	Simulated Set	
ESE	5.96	6.32	$< 10^{-12}$
ESS	1.72	1.97	$< 10^{-19}$
Splicing promotion	8.24	9.19	$< 10^{-24}$

*Random positioning of macaque mutations relative to human

Table S3. Merged hexamers of ESEs and ESSs

This set was used for compensatory mutation screening. See Methods in main text for details.

hexESES							
aaaaag	aaaacc	aaaaga	aaaagc	aaacag	aaacca	aaacct	aaacga
aaactc	aaagaa	aaagac	aaagag	aaagat	aaagca	aaagct	aaagga
aaatcc	aacaac	aacaag	aacacc	aacaga	aacatg	aaccaa	aaccag
aacctg	aacgaa	aactac	aactgg	aacttc	aagaaa	aagaac	aagaag
aagaat	aagaca	aagacc	aagact	aagaga	aagagg	aagatc	aagatg
aagcaa	aagcag	aagcca	aagcta	aagctc	aagctg	aaggaa	aaggac
aaggag	aaggat	aatacg	aatcaa	aatcag	aatcca	aatgaa	aatgac
aatgga	acaaag	acaacg	acaact	acaaga	acagaa	acagag	acagga
acatct	acatga	accaac	accaag	accaga	accagc	accagt	accatc
accag	accctg	acctga	acctgc	acctgg	acgaaa	acgaag	acgact
acgaga	acgcaa	actaca	actgaa	actgga	acttca	acttcg	agaaaa
agaaac	agaaag	agaaca	agaacc	agaact	agaaga	agaagc	agaagg
agaagt	agaatc	agaatg	agacaa	agacat	agacca	agacct	agacga
agagaa	agagat	agagga	agatcc	agatga	agatgc	agatgt	agcaaa
agcaag	agcact	agcaga	agcatc	agctga	agctgg	agctgt	aggaaa
aggaac	aggaag	aggaca	aggacc	aggact	aggaga	aggagc	aggagg
aggagt	aggatc	aggatt	agtgaa	agtgac	agtgga	atcaaa	atcaag
atcaat	atcaga	atcagc	atccaa	atcctg	atctga	atctgc	atcttc
atgaaa	atgaag	atgaga	atgatg	atgcaa	atggaa	atggat	atggcg
atggtc	attcag	attgga	caaaac	caaaag	caaaca	caaaga	caactt
caagaa	caagac	caagat	caagga	caagta	caatca	cacaga	cacctg
cactga	cagaaa	cagaac	cagaag	cagaat	cagaga	cagagg	cagatc
cagcag	cagctg	caggaa	caggag	cagtga	catcag	catctg	cattga
ccaaca	ccacaa	ccagaa	ccagca	ccagcc	ccctga	ccgaag	cctcga
cctctg	cctgaa	cctgac	cctgag	cctgat	cctgcc	cctgct	cctgga
cctgtg	cgaaaa	cgaaca	cgaaga	cgaatg	cgacga	cgactg	cgagga
cggaga	cgtatg	cgtcgc	ctacat	ctcaag	ctccat	ctcgaa	ctctgc
ctgaaa	ctgaac	ctgaag	ctgacc	ctgacg	ctgaga	ctgagg	ctgatc
ctgcag	ctgcct	ctgctc	ctgctg	ctggaa	ctggac	ctggag	ctggcg
ctgtgc	ctgtgg	cttcaa	cttcac	cttcag	cttgaa	cttgga	gaaaaa
gaaaac	gaaaag	gaaaat	gaaaca	gaaacc	gaaacg	gaaact	gaaaga
gaaagc	gaaatc	gaacaa	gaacag	gaacat	gaacca	gaacct	gaacga
gaactg	gaactt	gaagaa	gaagac	gaagag	gaagat	gaagca	gaagcg
gaagct	gaagga	gaagta	gaagtt	gaatca	gacaaa	gacaat	gacatc
gaccag	gaccat	gacctg	gacgaa	gacgac	gacttc	gagaaa	gagaac
gagaag	gagaca	gagacc	gagaga	gagagg	gagatc	gagatg	gagcac
gagctg	gaggaa	gaggac	gaggag	gaggat	gatatc	gatatg	gatcaa
gatcat	gatcct	gatgaa	gatgag	gatgat	gatgca	gatgga	gattca
gcaaaa	gcaaga	gcactg	gcagaa	gcagct	gcatct	gccctg	gcgaat
gctgaa	gctgac	gctgag	gctgga	gctgtt	ggaaaa	ggaaac	ggaaca

ggaaga	ggacac	ggacca	ggaccc	ggacct	ggactg	ggagaa	ggagac
ggagat	ggagca	ggagct	ggagga	ggatca	ggatcc	ggattc	ggctga
ggtgga	gtcaag	gtcctg	gtcgac	gtgaac	gtgaag	gtgacc	gtggac
gtggag	gtgtcg	gttgga	tacaag	tacaga	tacctg	tatgaa	tatgga
tcaacg	tcaaga	tcagaa	tcagca	tcagga	tcatca	tcctga	tcctgc
tcgtcg	tctgac	tctgct	tcttca	tgaaac	tgaaag	tgaaca	tgaaga
tgaagc	tgaagg	tgaagt	tgacca	tgacct	tgacga	tgactg	tgagaa
tgagga	tgatga	tgcaac	tgcaga	tgctct	tgctga	tgctgt	tggaaa
tggaac	tggag	tggaat	tggacc	tggaga	tggagg	tggatc	tgtgga
ttcaga	ttcagc	ttcgaa	ttgaag	ttgcga	ttggaa	ttggat	tttgga
hexESSs							
aaaaaa	aaaaat	aaaata	aaaatg	aaatat	aaatga	aaatgc	aaatgg
aaatgt	aatggt	aatgta	aatgtt	acacac	acacat	actagg	aggggg
aggggt	agggta	agggtg	agggtt	aggtaa	aggtag	aggtag	aggtta
aggttt	agtagg	agtcgt	agtgtg	agtgtt	agttag	agttcg	agttct
agttgt	agttta	agtttg	ataaaa	ataaat	ataatg	atacat	atagtt
atatat	atatgt	atatta	atattt	atgcat	atgctt	atgggg	atggtt
atgtaa	atgtat	atgtca	atgtct	atgtgt	atgtta	atgttt	attaat
atttaa	atttat	atttta	attttt	cacaca	cataaa	catata	catgca
cctggg	cgttcc	ctaggg	ctattt	cttagg	cttttt	gaaatg	gatggg
ggatgg	gggagg	gggatg	ggggag	ggggat	ggggga	gggggg	gggggt
ggggta	ggggtg	ggggtt	gggtaa	gggtag	gggtgg	gggtta	gggttt
ggtaag	ggtagg	ggtagt	gggtata	gggtggg	gggttag	gggttta	gggtttg
ggtttt	gtaagg	gtaagt	gtaggg	gtaggt	gtagtt	gtatag	gtatat
gtatgt	gtattt	gtcgtt	gtgggg	gtgggt	gtgtag	gttagg	gttagt
gttatt	gttcct	gttcgt	gttctg	gttggt	gttttag	gtttat	gtttgg
gtttgt	gttttt	taaaaa	taaaat	taaatg	taacta	taagtg	taatgt
tagggg	tagggt	taggta	taggtt	tagtgg	tagtgt	tagtta	tagttt
tataaa	tataat	tatact	tatagg	tatagt	tatata	tatatg	tatatt
tatgct	tatgta	tatgtc	tatgtt	tattaa	tattat	tattgt	tattta
tatttg	tatttt	tcgttc	tctagg	tcttta	tctttt	tgcata	tggggg
tgggtg	tgggtg	tggttt	tgtaat	tgtatg	tgtatt	tgtctt	tgttag
tgttat	tgttcc	tgttta	tgtttt	ttaaaa	ttaata	ttaatt	ttaggg
ttaggt	ttagta	ttagtg	ttagtt	ttatat	ttatgt	ttatta	ttattt
ttcata	ttcatt	ttcctg	ttcctt	ttcgtt	ttctag	ttctat	ttcttt
ttgggg	ttgttc	ttgttt	tttaaa	tttaat	tttact	tttagg	tttagt
tttata	tttatg	tttatt	tttcat	tttctt	tttggg	tttggt	ttttaa
tttttag	ttttat	ttttct	ttttgg	ttttgt	ttttta	tttttc	tttttg
tttttt							

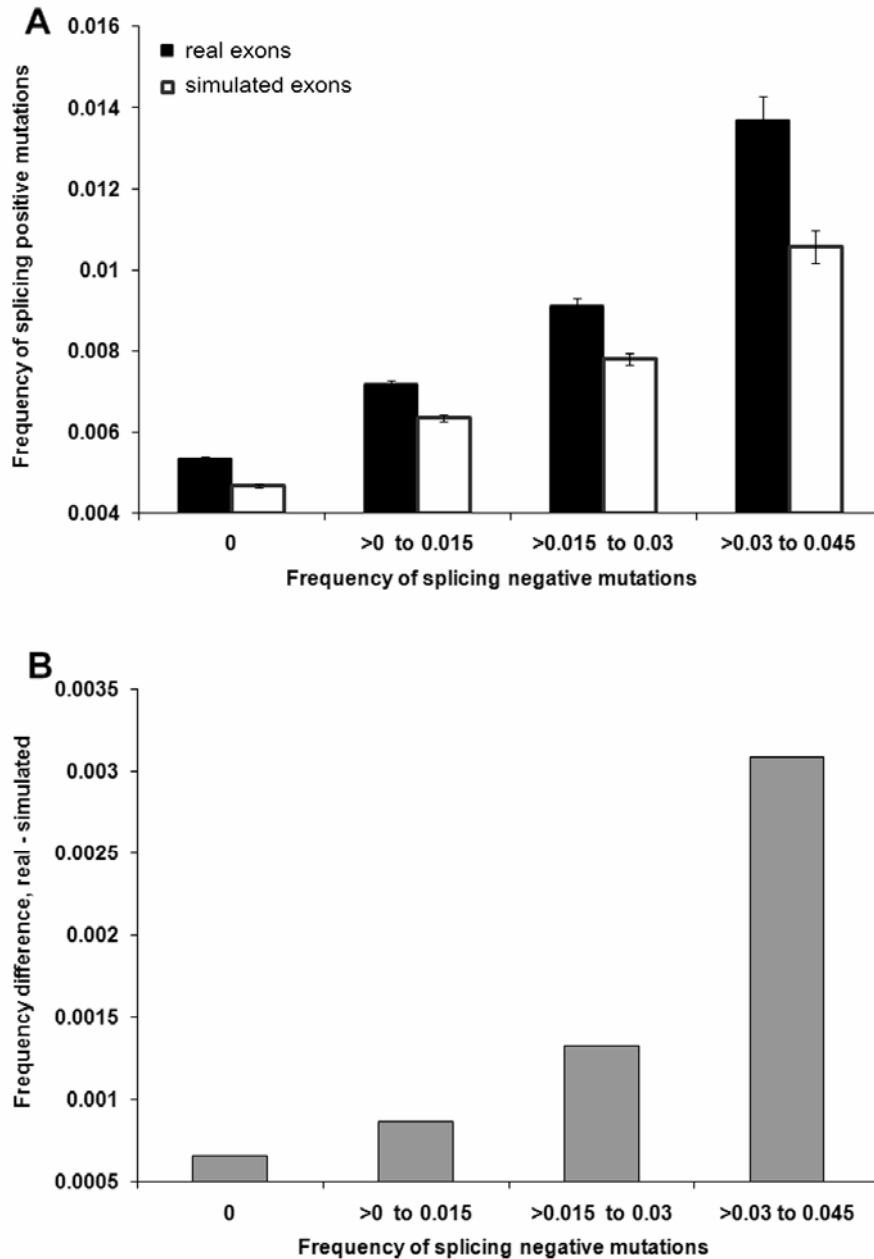


Fig. S1. Evolutionary compensation of exonic splicing negative motif changes by exonic splicing positive motif changes (comparing macaque to human). A. Exons were separated into 4 groups according to their frequency of exonic splicing negative events (events per nt) as indicated. Exons with frequencies of 0 to 0.015, 0.015 to 0.030, and .030 to 0.045 generally contain 1 or 2, 3 or 4, and 5 or 6 splicing negative mutations, respectively. Exon pairs with frequencies higher than 0.045 have been omitted as many of them are poorly aligned. Black bars: real exons; white bars, simulated exons. B. The net difference between the real and simulated sets is shown for each splicing negative mutation category.

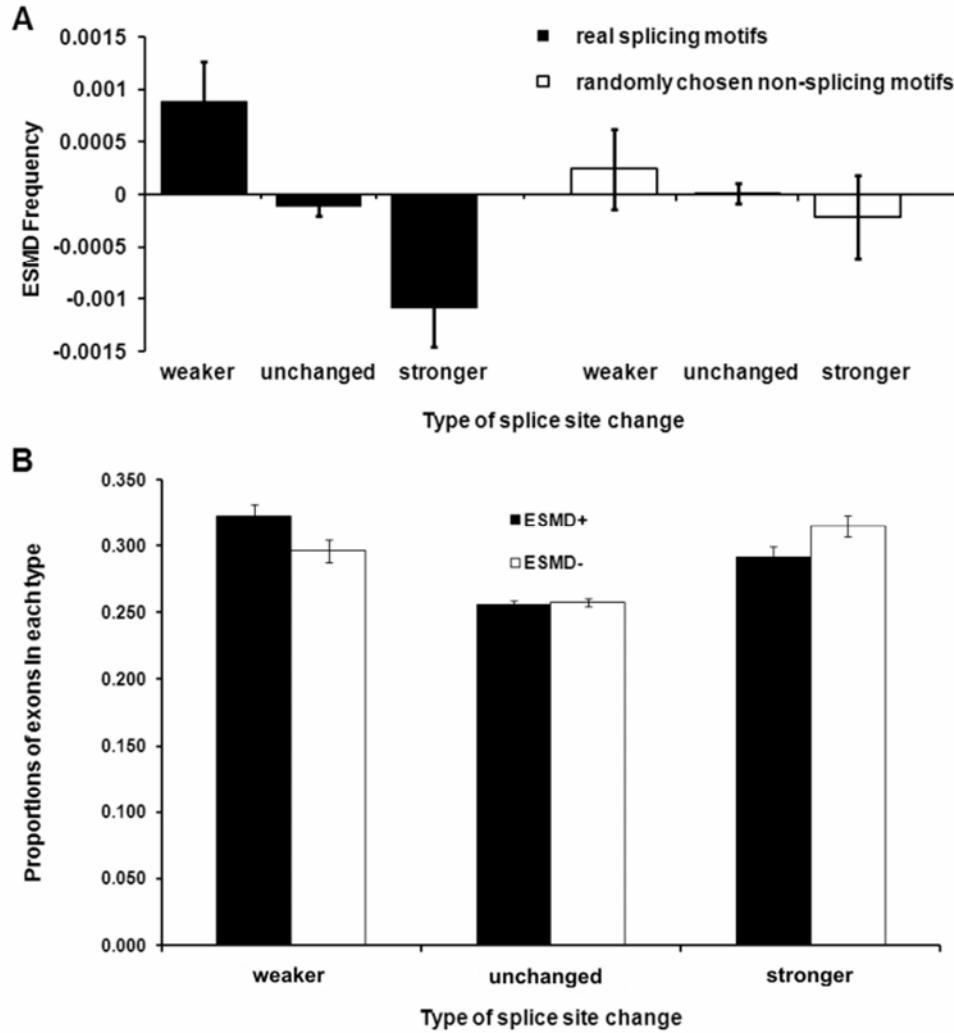


Fig. S2. Compensation between splice site differences and exonic splicing motif differences (comparing macaque to human). A. Splicing positive splice site differences correlate with splicing negative motif differences, and vice versa. The Exonic Splicing Motif Difference (ESMD) frequency is defined as the frequency of splicing positive differences (ESE creations and ESS disruptions) minus the frequency of splicing negative differences (ESS creations and ESE disruptions) in macaque relative to human. A positive ESMD is predicted to promote splicing while a negative ESMD would discourage splicing. Exons in the “weaker” set have one splice site at which the CV score (consensus values) has decreased by at least 5 on a CV scale of 0 to 100 and in which the other splice site has not increased by more than 5. Exons in the “stronger” set are defined in the opposite way. Exons in the “unchanged” set show no change at all in CV score for both 3’ and 5’ splice sites. The results of a control using non-ESE/ESS motifs, as described in the legend to Fig. 3D, are shown on the right. B. The proportion of exons that have undergone the changes in splice site sequence and splicing motifs reflects compensation. Standard errors are indicated in both panels.

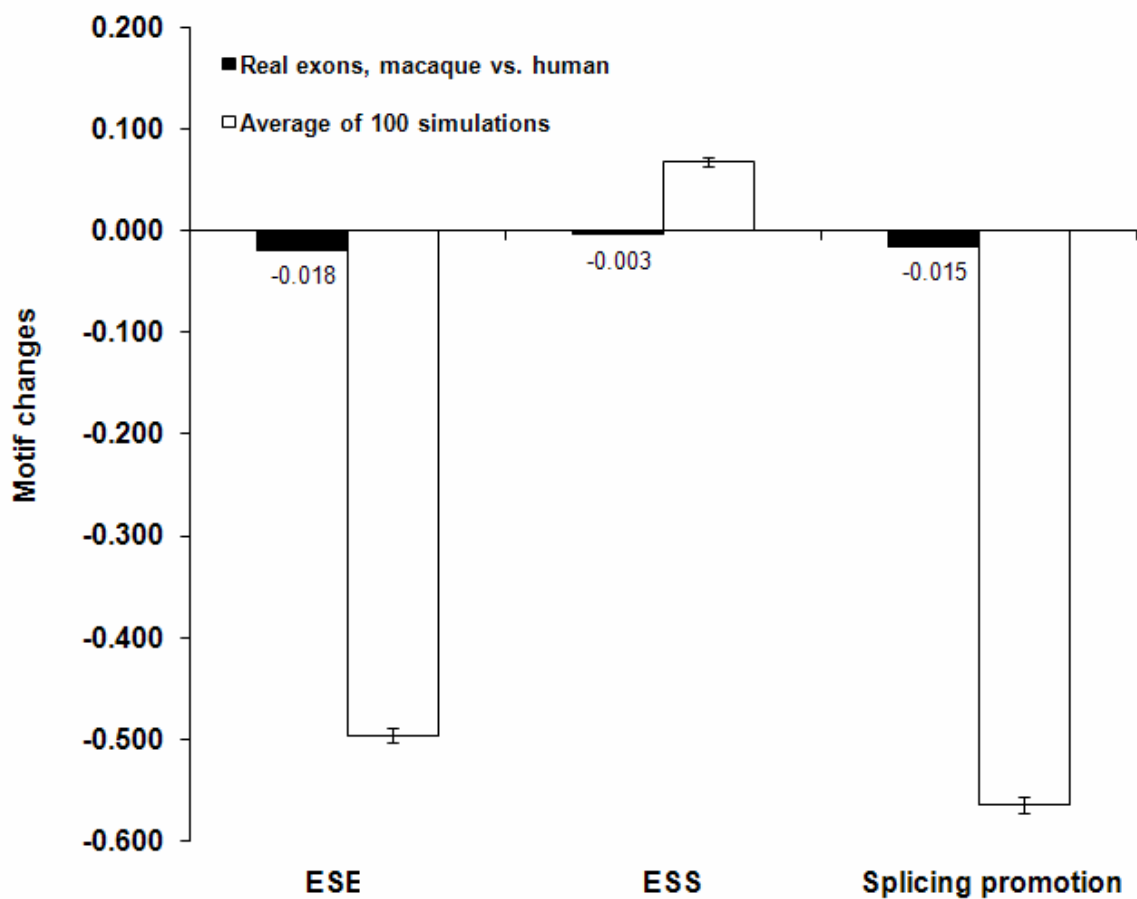


Fig. S3. Global maintenance of exonic splicing motifs (comparing macaque to human).

Motif change is defined as the number of ESE or ESS hexamers in a human exon minus the number in its macaque orthologue. Splicing promotion for an exon is defined as the ESE number minus the ESS number for that exon. In the simulated sets, the same number and type of differences seen in macaque exons were placed randomly among synonymous sites of human exons, as described in the text.