

Supplementary Table 1. List of 3097 human and 1106 mouse putative SA gene pairs (examples)^a

Genome	Genomic locus (chromosome_start_end)	Subtype (i.e., overlapping pattern)	Evolutionarily conserved?	Same- overlapping- pattern conserved?	Co- expressed?	Inversely expressed?
Human	Chr1_100310729_100355822	3'-to-3'			YES	YES
Human	Chr1_109070564_109113188	5'-to-5'			YES	YES
Human	Chr1_111441837_111459073	5'-to-5'			YES	
Human	Chr1_112929084_112955403	embedded				
Human	Chr1_142173665_142565230	embedded			YES	YES
Human	Chr1_142996653_143005120	3'-to-3'			YES	
Human	Chr1_143065032_143078007	3'-to-3'			YES	YES
Human	Chr1_147767626_147793914	3'-to-3'	YES	YES	YES	
Human	Chr1_150444225_150456552	3'-to-3'	YES	YES	YES	YES
Human	Chr1_152442305_152521871	5'-to-5'			YES	
Human	Chr1_152528661_152640081	3'-to-3'				
Human	Chr1_152976342_152995660	5'-to-5'				
Human	Chr1_157071842_157141815	5'-to-5'			YES	YES
Human	Chr1_157900963_157915531	3'-to-3'			YES	YES
Human	Chr1_162363106_162410863	5'-to-5'				
Mouse	Chr1_106511158_106687250	embedded			YES	
Mouse	Chr1_125703338_125743510	3'-to-3'	YES	YES		
Mouse	Chr1_127568817_127613413	3'-to-3'				
Mouse	Chr1_127788756_127938649	embedded			YES	YES
Mouse	Chr1_131880183_131940253	3'-to-3'	YES	YES	YES	YES
Mouse	Chr1_161999680_162493023	embedded			YES	YES
Mouse	Chr1_163945094_164010548	3'-to-3'	YES	YES	YES	YES
Mouse	Chr1_164325043_164475058	3'-to-3'	YES	YES	YES	YES
Mouse	Chr1_170274606_170292777	5'-to-5'				
Mouse	Chr1_171285202_171293746	3'-to-3'	YES	YES		YES
Mouse	Chr1_171574027_171575982	5'-to-5'				
Mouse	Chr1_178334982_178339881	3'-to-3'	YES	YES	YES	YES
Mouse	Chr1_180217468_180268637	3'-to-3'			YES	
Mouse	Chr1_191218119_191289206	3'-to-3'			YES	YES
Mouse	Chr1_193067170_193095188	3'-to-3'	YES	YES		

^aThese 30 putative SA gene pairs are presented here as examples. The full list is available from the authors upon request.

Supplementary Table 2. Analysis of the sense-antisense (SA) pairs of which both partners have roughly full-length, curated reference sequences (RefSeq).

a. Comparison of the prevalence of different subtypes of the putative SA pairs

Genome	Total pairs	3'-to-3' (percentage)	5'-to-5' (percentage)	Embedded (percentage)	Ratio of 3'-to-3' / 5'-to-5' (χ^2 -test)
Human	693	419 (60.5%)	160 (23.1%)	114 (16.4%)	2.6 ($P < 10^{-4}$)
Mouse	385	274 (71.2%)	83 (21.6%)	28 (7.2%)	3.3 ($P < 10^{-4}$)

b. Comparison of the evolutionary conservation rates of different subtypes of the putative SA pairs^a

Genome	Putative SA subtype	Rate of conservation	Rate of same-overlapping-pattern conservation	Ratio of 3'-to-3' / 5'-to-5' in conservation rate (χ^2 -test)
Human	3'-to-3'	44.4% (186/419)	43.9% (184/419)	3.4; 5.4 ($P < 10^{-4}$; $P < 10^{-4}$)
	5'-to-5'	13.1% (21/160)	8.1% (13/160)	
	embedded	8.8% (10/114)	2.6% (3/114)	
	whole	31.3% (217/693)	28.9% (200/693)	
Mouse	3'-to-3'	63.1% (173/274)	60.2% (165/274)	3.1; 5.0 ($P < 10^{-4}$; $P < 10^{-4}$)
	5'-to-5'	20.5% (17/83)	12.0% (10/83)	
	embedded	21.4% (6/28)	7.1% (2/28)	
	whole	50.9% (196/385)	46.0% (177/385)	

^aNote that, all p -values far less than 10^{-4} are also shown as $P < 10^{-4}$. Rate of conservation refers to the percentage of human or mouse putative SA pairs conserved as putative SA in the mouse or human genome. Rate of same-overlapping-pattern conservation refers to the percentage of human or mouse putative SA pairs conserved as putative SA with the same overlapping pattern (i.e., 3'-to-3', 5'-to-5' or embedded) in the mouse or human genome.

Supplementary Table 3. Comparison of subtype proportions in different putative SA pair sets in which at least the antisense in a given SA pair belongs to non-coding RNA^a

Putative SA pair set	Genome	Total pairs (SA pairs in which the sense also belongs to non-coding RNA) ^b	3'-to-3' (percentage)	5'-to-5' (percentage)	Embedded (percentage)	Ratio of 3'-to-3' / 5'-to-5' ^c
The whole	Human	1953 (209)	507 (26.0)	363 (18.6)	1083 (55.4)	1.4
	Mouse	469 (46)	152 (32.4)	101 (21.5)	216 (46.1)	1.5
Inversely expressed	Human	399 (20)	127 (31.8)	72 (18.1)	200 (50.1)	1.8
	Mouse	105 (12)	43 (41.0)	20 (19.0)	42 (40.0)	2.2
Co-expressed	Human	487 (21)	158 (32.4)	69 (14.2)	260 (53.4)	2.9
	Mouse	127 (18)	59 (46.5)	25 (19.7)	43 (33.8)	2.4
Conserved	Human	99 (0)	62 (62.6)	17 (17.2)	20 (20.2)	3.6
	Mouse	82 (0)	65 (79.3)	7 (8.5)	10 (12.2)	9.3
Conserved & inversely expressed	Human	27 (0)	17 (63.0)	4 (14.8)	6 (22.2)	4.3
	Mouse	22 (0)	18 (81.8)	2 (9.1)	2 (9.1)	9.0
Conserved & co-expressed	Human	32 (0)	27 (84.4)	2 (6.3)	3 (9.3)	13.5
	Mouse	33 (0)	29 (87.9)	3 (9.1)	1 (3.0)	9.7

^aThe co-expressed putative SA pairs were defined as those in which the two partners are coordinately expressed in the same tissues more often than expected by chance; the inversely expressed putative SA pairs were defined as those with both an inverse expression pattern between two partners and a significantly greater change of the relative expression ratio of sense to antisense between two states of the same tissues than expected by chance; the conserved putative SA pairs are the pairs conserved as putative SA form in both human and mouse genomes. See Materials and Methods for more details. Because many non-coding antisense genes have only partial EST sequences, the proportion of “embedded” putative SA pairs is seriously overestimated here. In fact, a large proportion of the “embedded” putative SA pairs are also mainly overlapping at the 3' UTRs of the sense genes (data not shown).

^bAccording to our definition (see Materials and Methods), for the SA pairs in which one member has CDS while the other lacks CDS, we define the first one as the S and the latter one as the A. Thus, in a given SA pair, once the sense belongs to non-coding RNA, the antisense partner must be non-coding RNA as well.

^cThe slight differences between the “Ratios of 3'-to-3' / 5'-to-5'” in this Table and those in Table 3 probably due to the fact that many non-coding antisense genes have only partial EST sequences and thus are included in “embedded” putative SA pairs, although a large proportion of them are indeed overlapping at the 3' UTRs of the sense genes.

Supplementary Table 4. Analysis of orthologous genes between mammals and birds (and/or amphibians) that contain highly conserved regions (HCRs) in their noncoding sequences^a

Gene list	Gene symbol	Mammalian genes ^b					Orthologous bird and/or amphibian genes (chicken; African clawed frog) ^b	
		Gene ID ^c	Species	Gene category	Overlapping pattern of putative SA pair (Genomic locus)	Overlapping area of the target gene includes HCR? (overlapping area; target gene with overlapping partner)	Gene ID ^c	Gene category
3'-HCR-containing genes								
1	ATP2A2	X53754	pig	unknown; human ortholog NBD			M66385	NBD
2	DMD	M18533	human	NOB			X13369	NBD
3	BTG1	X61123	human	NBD			X64146	NBD
4	Ncam1	X15051	mouse	SA	3'-to-3' (chr9_49625979_49925346)	No (CDS; NM_010875 with BC049643)	M15861; M25696	NBD; NBD
5	Nmyc1	X53674	Wood chuck	unknown; mouse ortholog NBD			D90071	NBD
6	CALD1	M64110	human	NOB			J04968	NBD
7 [same gene as 87]	FOS	K00650	human	NBD			M37000	NBD
8	ATP1B1	X03747	human	SA	3'-to-3' (Chr1_16580760_4_166068844)	Yes^d (3' UTR; X03747 with NM_013330)	J02787	SA (J02787 overlapping with BX936183 at 3' UTR including HCR; 3'-to-3')
9	H3f3b	X13605	mouse	NBD			M11393	NBD
10	Calm2	M17069	rat	NBD; mouse ortholog NBD			L00101; K01945	NBD; NBD
11	HMG2	X13546	human	SA	embedded (Chr1_26483055_26486605)	Yes (3' UTR; NM_005517 with BU674882)	J03229	NBD
12	H3F3A	X05857	human	NBD			M11667	NBD
13	Marcks	M24638	cow	NBD; human ortholog NBD			M31650	NBD
14	COL1A2	K01078	human	NBD			M25984	NBD
15	NEF3	Y00067	human	SA	5'-to-5' (chr8_24826429_24832512)	No (5' UTR+CDS; NM_005382 with BC002421)	X17102	NBD
16	Tfrc	M58040	rat	NBD; mouse ortholog NBD			X13753	NBD
17 [same gene as 76]	JUN	J04111	human	NBD			X15547	NBD
18	GATA-3	X58072	human	NBD			; M76565	; NBD
19	CDH2	X57548	human	NBD			X07277	NBD
20	CALB1	X06661	human	NBD			X06633	NBD
21	Nsep1	M57299	rat	NBD; mouse ortholog NBD			; M59453	; NBD
22	Lamp1	M32015	mouse	NBD			M59365	NBD

23	ODC1	X16277	human	SA	embedded (chr2_10531098_10539051)	No (CDS; NM_002539 with BG185876)	X64710; X56316	NBD; NBD
24	Nfia	D90173	mouse	NOB			X51486	NOB
25	COL2A1	X06268	human	NBD			L00063; M63596	NBD; NBD
26	VIM	M25246	human	SA	5'-to-5' (chr10_17296250_17319891)	No (5' UTR+CDS; NM_003380 with BC078172)	V00447; X16843	NBD; NBD
27	EEF1A1	J04617	human	SA	embedded (chr6_74283957_74287476_M)	Yes (3' UTR+CDS; NM_001402 with AB019568)	; M25697	; NBD
28	HSPA5	M19645	human	SA	embedded (chr9_125076680_125083181_M)	Yes (3' UTR; NM_005347 with BM969352)	M27260	NBD
29	Actg	X52815	rat	NBD; mouse ortholog NBD			M26111 (goose)	NBD (chicken ortholog)
30	Pam	M25732	rat	NBD; mouse ortholog NBD			; X62771	; NBD
31	Cntn1	X14943	mouse	SA	5'-to-5' (chr15_92367855_92760915)	No (5' UTR; NM_007727 with BY707228)	Y00813	NBD
32	TCF12	M83233	human	NBD			M87337	NBD
33	HMGN1	M21339	human	SA	5'-to-5' (chr21_39635840_39642917)	No (5' UTR; AK056033 with CB853997)	X52708	NBD
34	PDGFA	X06374	human	NBD			; M23238	; NBD
35 [same gene as 83]	RARB	Y00291	human	SA	3'-to-3' (chr3_25444013_25680792)	No (3' UTR; AI564510 with NM_001068)	X57340	NBD
36	HAPLN1	Y00165	pig	unknown; human ortholog NBD			M13212	NBD
37	ACTC	J00073	human	NBD			M10607	NBD
38	MYB	M15024	human	SA	embedded (chr6_135544145_135582003)	Probably Yes^e (3' UTR; AJ606321 with BX090126)	X51569	NBD
39	ACTA1	M20543	human	NBD			J00805	NBD
40	MAPK1	M84489	human	NBD			; X59813	; NBD
41 [same gene as 78]	Tpm1	M18135	rat	NBD; mouse ortholog NBD			M69145	NBD
42	MYH10	X62659	(sequence deleted)				M26510	NBD
43 [same gene as 77]	ACTB	X00351	human	SA	3'-to-3' (chr7_5340018_5343473)	Yes (3' UTR; NM_001101 with CA439838)	X00182	NBD
44	HMGB2	M83665	human	NBD			M83235	NBD
45	Npm1	J03969	rat	NBD; mouse ortholog NBD			X17200; X56039	NBD; NBD
46	FGFR1	M37722	human	NBD			M24637; M55163	NBD; NBD
47	POU2F1	X13403	human	NBD			M29972; X57165	NBD; NBD
48	NCL	M60858	human	NBD			X17199; X63091	NBD; NBD
49	CSNK2A1	M55265	human	NBD			M59456	NBD
50	GNAI1	X03642	cow	NBD; human ortholog NBD			; X56089	; NBD

51	ACTN1	X55187	human	NBD			M74143	NBD
52	Twist1	M63650	mouse	NBD			; M27730	; NBD
53	Ngfb	M35075	mouse	NBD			X04067; X55716	NBD; NBD
54 [same gene as 66]	TGFB1	X02812	human	NBD			X12373 (pig, not birds)	-
55	MSX1	M76732	human	NBD			X61922	NBD
56	TRA1	X15187	human	SA	3'-to-3' (chr12_10282662 6_102861940)	No (3' UTR; N92597 with BC044618)	X04961	NBD
57	Hoxb4	M36654	mouse	NBD			; M26884	; NBD
58	TNC	X56160	human	NBD			M23121	NBD
59	P4HA1	M24486	human	SA	embedded (chr10_74435550 _74526626)	No (CDS; NM_000917 with BG212501)	M26217	NBD
60	FN1	X00739	human	SA	3'-to-3' (chr2_216050137 _216126297_M)	Yes (3' UTR; X00739 with AY450391)	; M77820	; NBD
61	ITGB1	X07979	human	NBD			M14049; M20140	NBD; NBD
62	NPY	K01911	human	NBD			M87295	NBD
63	Cyr61	M32490	mouse	NBD			J04496	NBD
64 [same gene as 81]	CKB	X15334	human	NBD			M33714	NBD
5'-HCR-containing genes								
65	IGF1	M14155	human	NBD			M32791; M29857	NBD; M29857
66 [same gene as 54]	TGFB1	X02812	human	NBD			X12373 (pig, not birds)	-
67	ACTA2	D00618	human	SA	3'-to-3' (chr10_9062960 2_90724890)	No (3' UTR+CDS; NM_001613 with AK093340)	M13756	NBD
68	Col1a2	K01832	mouse	NBD			M25963	NBD
69	Nfib	J04122	golden hamster	unknown; mouse ortholog NBD			X51485	NBD
70	MYL1	J05026	human	NBD			K02610	NBD
71	HOXD4	X17360	human	NBD			X52671	NBD
72	Myb	X16390	mouse	NBD			X12495	NBD
73	Gabrg2	M62374	mouse	NBD			X54944	NBD
74	Hoxb7	X06762	mouse	NBD			; X06592	; NBD
75	TGFB3	M60556	human	NBD			X58127	NBD
76 [same gene as 17]	JUN	J04111	human	NBD			M57467	NBD
77 [same gene as 43]	ACTB	M10277	human	SA	3'-to-3' (chr7_5340018_ 5343473)	No (3' UTR; NM_001101 with CA439838)	X00182	NBD
78 [same gene as 41]	Tpm1	M15474	rat	NBD; mouse ortholog NBD			M69145	NBD
79	Hoxc6	X16511	mouse	NBD			X12500	NBD
80	Dcn	X53929	mouse	NBD			X63797	NBD

81 [same gene as 64]	CKB	X15334	human	NBD			M33714	NBD
82	FGFR2	X52832	human	NBD			M35196	NBD
83 [same gene as 35]	RARB	Y00291	human	SA	3'-to-3' (chr3_25444013_25680792)	No (3' UTR; AI564510 with NM_001068)	X57340	NBD
84	FTH1	X03488	human	SA	3'-to-3' (chr11_6147393_1_61491681)	No (3' UTR; NM_002032 with NM_004183)	M16343; M15655	NBD; NBD
85	COL3A1	M26939	human	SA	embedded (chr2_18966460_4_189703013)	No (3' UTR; NM_000090 with CN367234)	K01481	NBD
86	Hoxa7	M17192	mouse	NBD			M59714 (common quail)	NBD (chicken ortholog)
Genes with HCRs in introns								
87 [same gene as 7]	FOS first intron	K00650	Human	NBD			M18043	NBD

^aA total of 87 highly conserved regions (HCRs) in the 3' UTR, 5' UTR and intron of the relevant genes were listed in 'Table 1' of the reference (34), in which a highly conserved region (HCR) was defined as sharing at least 70% identity over 100 nt between corresponding orthologs among vertebrates that diverged more than 300 million years (34,35). As noted in our table, 7 genes have two different HCR regions, so a total of 80 unique mammalian genes have HCRs. These HCR regions are proposed to form long perfect duplexes with (*cis*-encoded) antisense transcripts (23). Our analysis of these HCRs supports this proposal (at least partially).

^bFor genes in the human and mouse genomes, their gene categories were determined by our own gene cluster sets as well as UCSC 'BLAT Search Genome' website (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>; Assembly: human May 2004; mouse May 2004). Gene categories of the genes in the other genomes were determined by the UCSC website alone (Assembly: rat Jun 2003; cow Sep. 2004; Chicken Feb. 2004; X. tropicalis Oct. 2004) based on the annotated mRNA sequences. Pig and woodchuck genomes are not available in UCSC website yet. Birds besides chicken (*Gallus gallus*) and amphibian besides African clawed frog (*X. laevis*) were indicated. As *X. laevis* genome is not available yet, *X. tropicalis* genome was used in BLAT searching. SA: sense-antisense gene pair; NOB: non-exon-overlapping bidirectional gene pair; NBD: non-bidirectional gene (14-16). Due to the facts that mouse putative SA pairs were seriously underestimated caused by the limitation of qualified sequences (Sun *et al.* unpublished), and that the putative SA pairs in the remaining non-human genomes are more seriously underestimated because only mRNAs shown in the Blat searching-result plots were considered, the rate of HCRs belonging to or including putative SA overlapping regions as well as the rate of putative SA pairs must be seriously underestimated. Of the 51 human genes with HCRs, 17 form putative SA pairs with their partners (seven 3'-to-3' putative SA pairs, three 5'-to-5' putative SA pairs and seven 'embedded' putative SA pairs). Five out of the seven 'embedded' putative SA pairs are overlapping at the 3' UTRs of the analyzed genes. Indeed, 12 (70.6%) of the 17 human putative SA genes are targeted at their 3' UTRs, of which, 58.3% (7 out of 12) contain HCRs in their putative SA overlapping regions. Notably, ATP1B1 gene form 3'-to-3' putative SA pair with its partner in both human and chicken genomes, whose putative SA overlapping region include the relevant HCR in both genomes. In contrast, only 3 (17.6%) are targeted to their 5' UTRs, but none of them contain 5'-HCRs. Same as some highly conserved miRNAs whose expression levels are very low (38), many antisense genes may be expressed at a very low level and thus could not be detected in the human genome yet. Indeed, even many of the detected human antisense genes have only one or two qualified sequences in our putative SA set, while their sense partners are much more abundant (data not shown). Thus, the rate of putative SA pairs and of HCRs belonging to or including putative SA overlapping regions might be underestimated in the human genome as well. Thus, this study suggests that the 3' UTR-targeted antisense-regulation modes might be much more conserved than 5' UTR-targeted ones, during short (75 million years diverged between humans and mice) or long (over 300 million years diverged between birds and mammals) evolutionary periods.

^cThe GenBank IDs were listed in "Table 1" of the reference (34). As some of them are genomic DNA sequences, we chose their cDNA sequences as representative sequences in putative SA pairs.

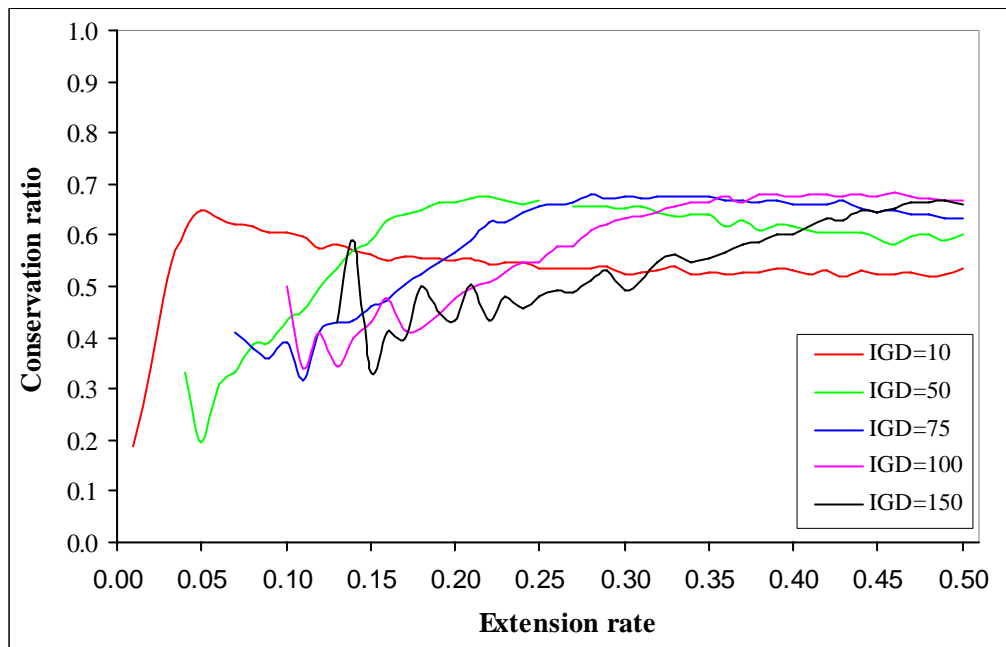
Sun et al.

^dThe HCR conserved between human mRNA X03747 and chicken mRNA CO421259 (but not J02787) covers all the SA-overlapped 3' UTR of X03747 (about 200 nt).

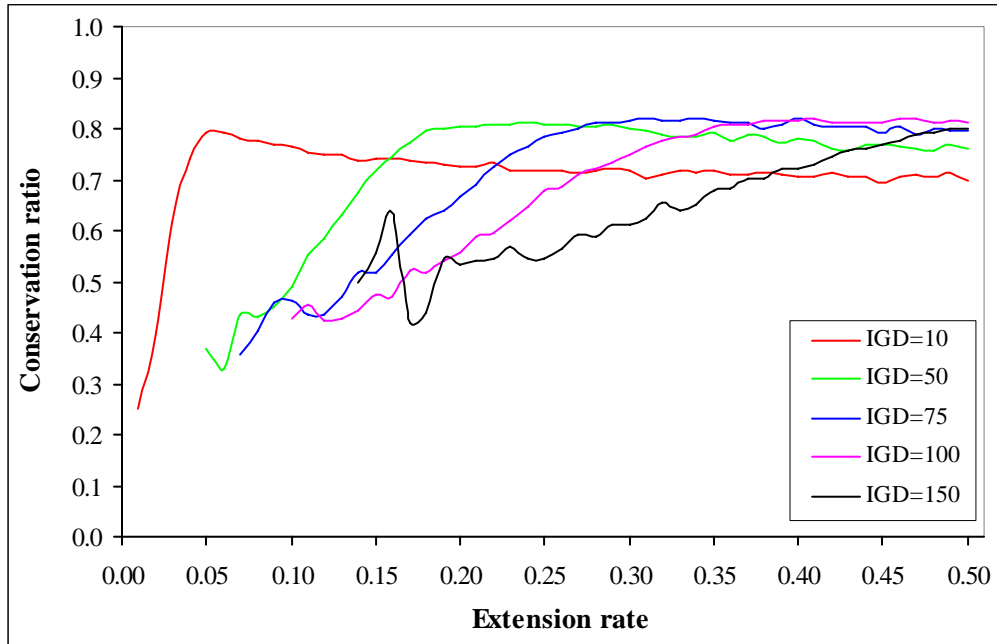
^eBecause BX090126 is an incomplete EST sequence, it is expected that its full-length mRNA sequence would overlap with the 3' UTR of the sense partner gene, including the HCR region.

Supplementary Figure 1. The proportion of SA pairs found in both lineages in simulation as a function of the extension rate for different values of intergenic distance. For Supplementary Figure 1a the re-arrangement rate per unit time is set to 1/500. For Supplementary Figure 1b it is set to 1/1000 and for Supplementary Figure 1c it is 1/5000. We also consider what happens if, for a given re-arrangement rate ($p=1/1000$), the extension rate is different for the two lineages. We consider the extreme possibility that the rate in one lineage might be double that in another (Supplementary Figure 1d). Red lines are for intergenic distance (IGD) of 10 units, green for 50 units, blue for 75 units, magenta for 100 units and black for 150 units. Only instances in which more than 10 overlapping pairs out of 10,000 simulants are incorporated in the analysis.

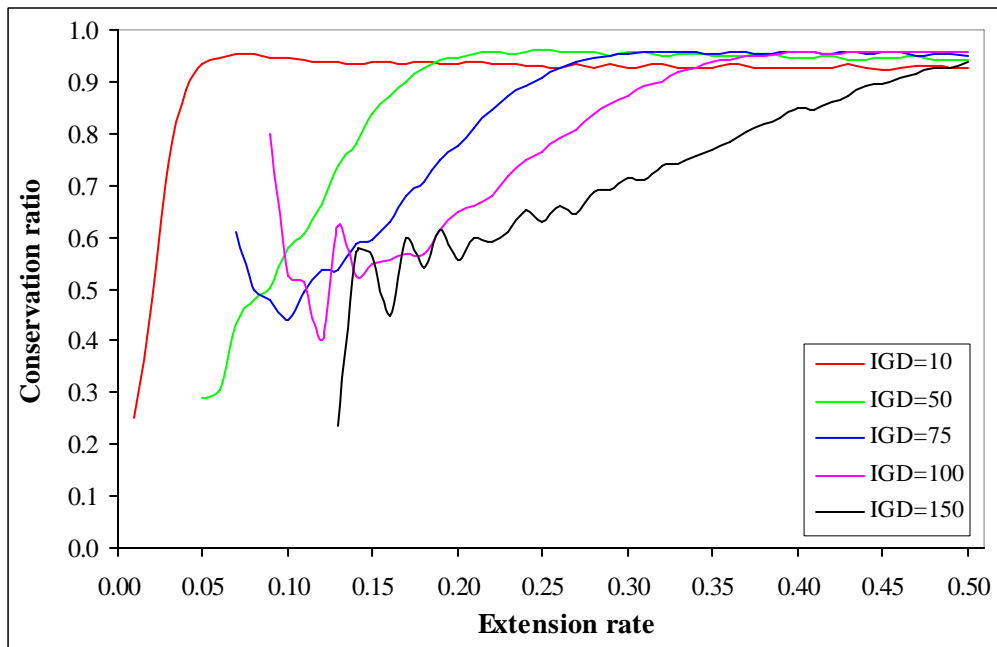
Supplementary Figure 1a.



Supplementary Figure 1b.



Supplementary Figure 1c.



Supplementary Figure 1d.

