Supplementary Material on Materials and Methods

Selection of genes for visual inspection, 12 species alignments

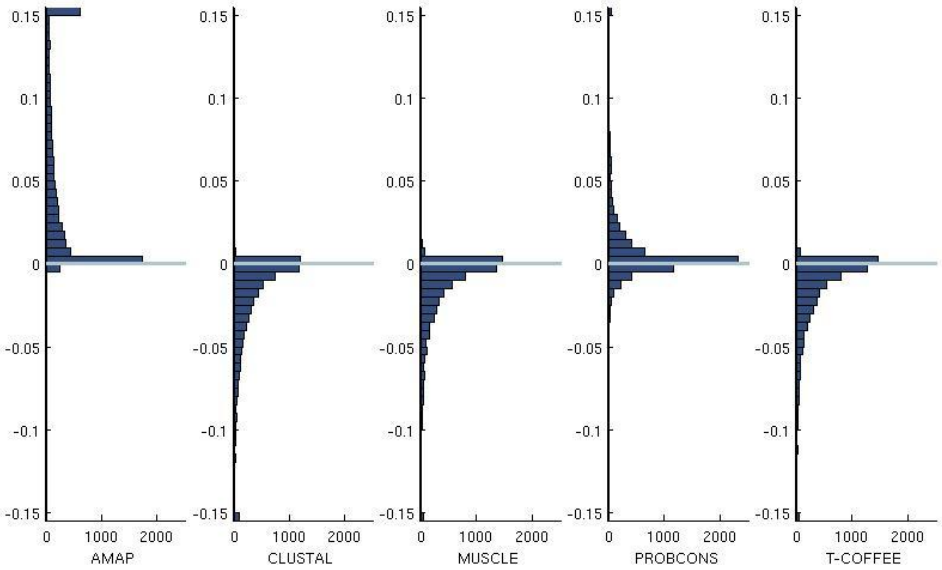
For the purposes of picking which genes to be visually inspected, we sliced the data set in the following three different ways. We first picked 25 genes that had a positively selected site with the cutoff posterior probability set at 95% such that five of them had such a site in all five alignments, five genes had such a site in exactly four alignments, and so on (Table 2S-A). This set was meant to give a sense of whether the consensus among aligners was correlated with the rate of true positives. Second, for each aligner we selected five to seven genes inferred to have a positively selected site based on that alignment alone independently of the results based on the other four alignments (Table 2S-B), for a total of 30 genes. This set was intended to estimate the rate of false positives for each separate aligner and clarify whether any of the aligners performed significantly better than others. Lastly, to further compare the five aligners, we selected a similar number of genes per aligner randomly but required that the results based on the other four alignments did not lead to inference of positive selection at any site in this gene (Table 2S-C) (28 genes). Over the three sets, for each of the picked genes we visually inspected all sites that were inferred to be positively selected at a 95% cutoff posterior probability in any of the five alignments.

PAML evolutionary analysis (in addition to as described in Materials and Methods)

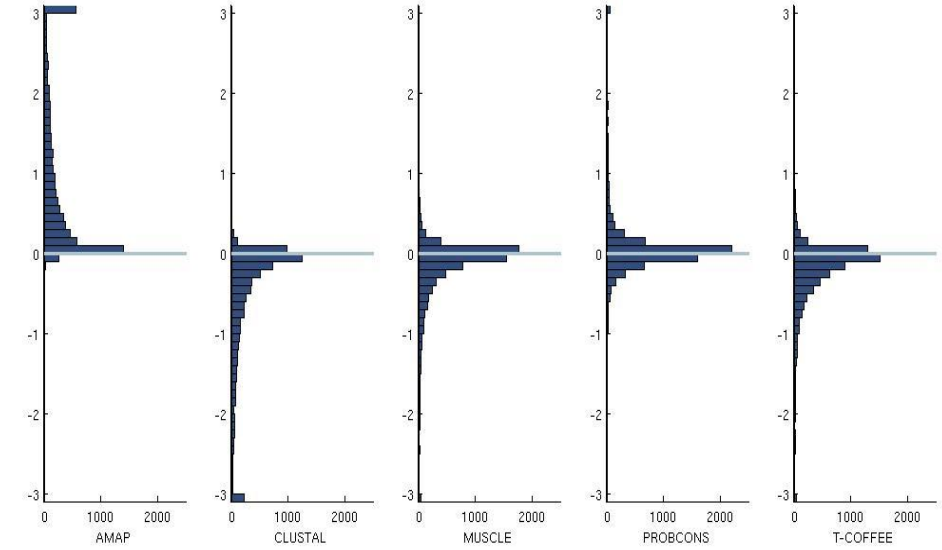
Initialization of the trees for PAML was done via the following procedure: We randomly picked a set of 100 genes and ran model M0 (using the same 100 genes for all aligners) on their concatenated alignments with initial omega values of 0.001, 0.4 and 2. For each aligner, we then used the best estimated tree and branch lengths as initial values for all single gene CODEML runs. Initially we ran each model with three different initial omegas (0.001, 0.4 and 2). Based on the results on the T-Coffee and ClustalW alignments, we concluded that the differences due to the different omega values were minor, and thus for computational reasons we did not exhaustively run all three cases for the remaining aligners.

Supplementary Material Figures

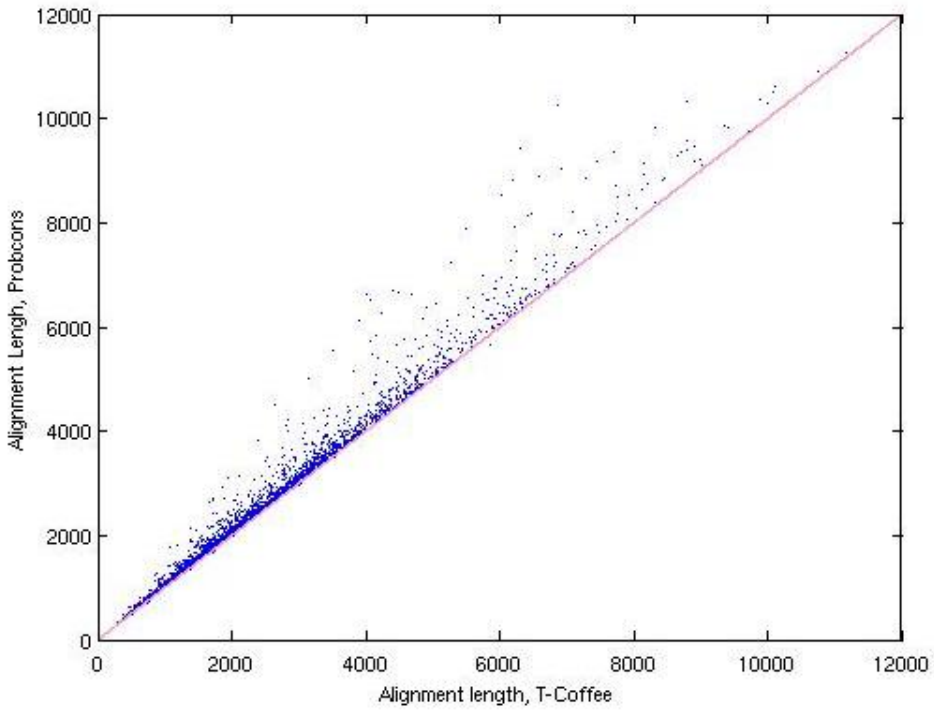
Figure 1S. (a) Frequency distribution of the difference per gene between alignment length based on a given aligner and the average alignment length among all 5 aligners; normalized by the average length; (b) Same for percentage identity, not normalized; (c) Length of T-Coffee vs. ProbCons alignments (each dot represents a gene), (d) Percentage identity of AMAP vs. ClustalW (each dot represents a gene).



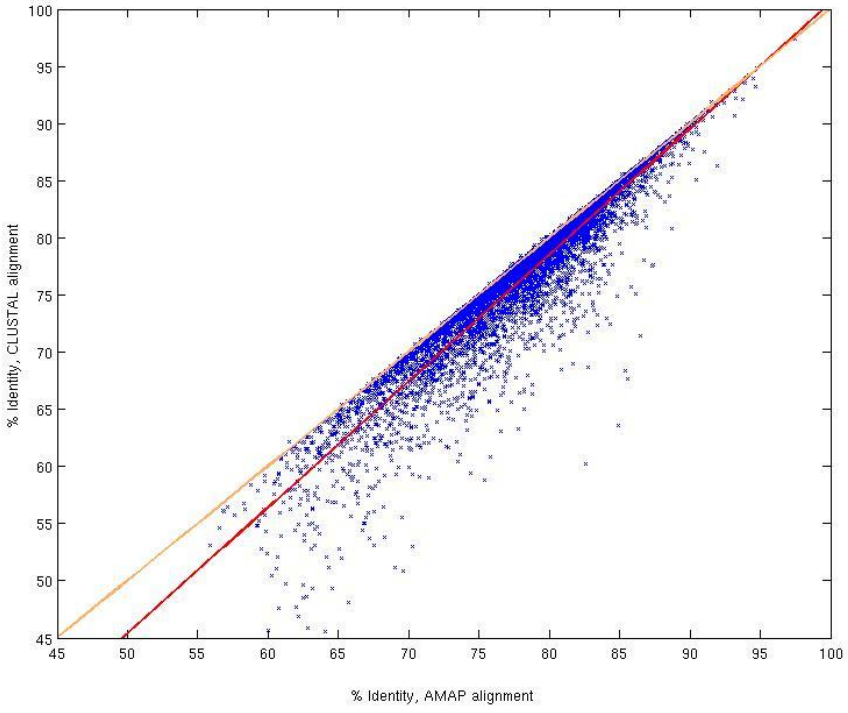
(a)



(b)

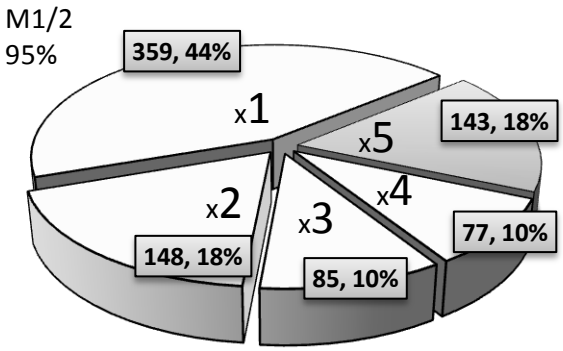

Alignment Lengh, Probcons
Alignment length, T-Coffee
0
2000
4000
6000
8000
10000
12000

(c)

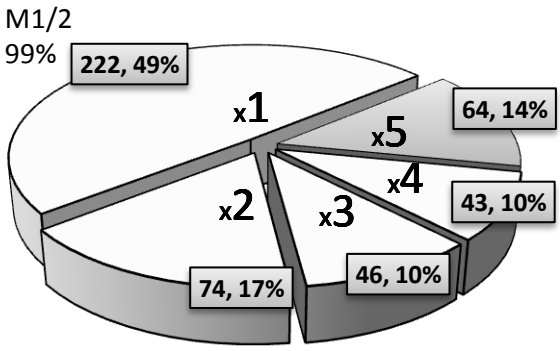

% Identity, CLUSTAL alignment
% Identity, AMAP alignment
45
50
55
60
65
70
75
80
85
90
95
100

(d)

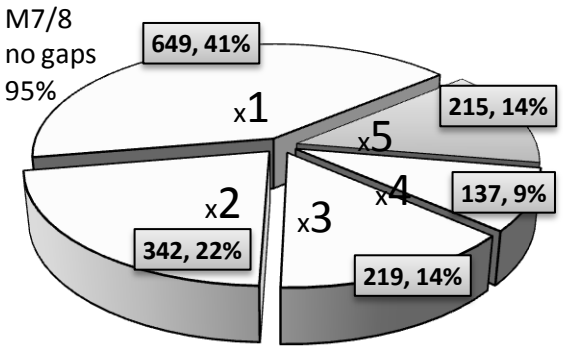
Figure 2S. The number of genes predicted to have a positively a positively selected site by exactly 1, 2, 3, 4 or 5 aligners (out of the genes with at least 1 such prediction). (a,b) models M1a and M2a, at 95% and 99% cutoffs, (c,d) models M7 and M8 after removing all locations belonging to indel in any of the species, at 95% and 99% cutoffs, (e,f) models M7 and M8 for *melanogaster* group species alignments, at 95% and 99% cutoffs.



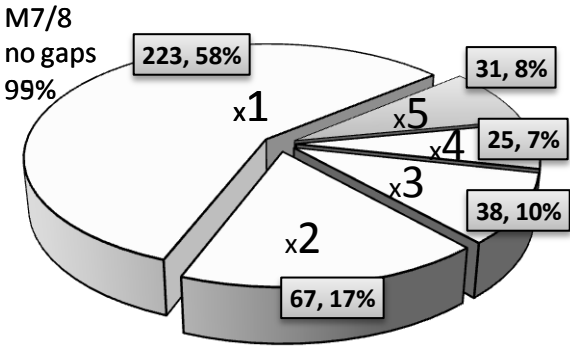
(a)



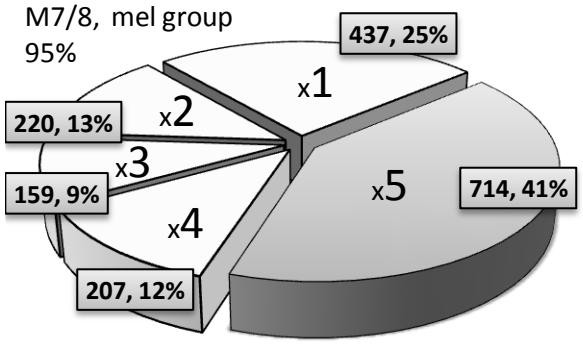
(b)



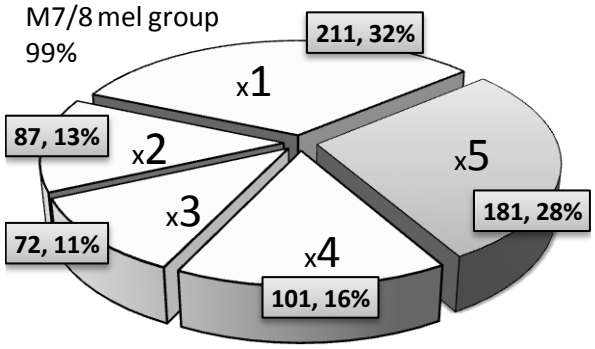
(c)



(d)



(e)



(f)

Supplementary Material Tables

Table 1S. Relative overlaps among the genes with inferred selected sites of any two aligners [12 species alignments, with models M7 and M8] at cutoff of (a) 95% and (b) 99%. Both the number of genes and percentage (relative to the aligner in that column) are shown.

(a)

M 7/8 95%	AMAP (817)		MUSCLE (1043)		ProbCons (1013)		T-Coffee (1290)		ClustalW (902)	
AMAP	817	100%	702	67%	620	61%	624	48%	531	59%
MUSCLE	702	86%	1043	100%	697	69%	804	62%	646	72%
ProbCons	620	76%	697	67%	1013	100%	761	59%	630	70%
T-Coffee	624	76%	804	77%	804	79%	1290	100%	688	76%
ClustalW	531	65%	646	62%	630	62%	688	53%	902	100%

(b)

M 7/8 99%	AMAP (817)		MUSCLE (1043)		ProbCons (1013)		T-Coffee (1290)		ClustalW (902)	
AMAP	817	100%	702	67%	620	61%	624	48%	531	59%
MUSCLE	702	86%	1043	100%	697	69%	804	62%	646	72%
ProbCons	620	76%	697	67%	1013	100%	761	59%	630	70%
T-Coffee	624	76%	804	77%	804	79%	1290	100%	688	76%
ClustalW	531	65%	646	62%	630	62%	688	53%	902	100%

Table 2S A-F. Summary of visual inspection observations.

The alignments visualized in each of 2S A-F were selected at random among the genes with inferred positive selection such that

2S-A) include 5 genes for each of the 5 levels of inferred selection inference coverage in the five 12 species-, unmasked- alignments (PRANK alignments are not included, see Materials and Methods for details of the selection procedure). (25 genes)

2S-B) the gene had inferred positively selected sites in more than one of its five (excluding PRANK) 12 species-, unmasked- alignments. (30 genes)

2S-C) the gene had inferred positively selected sites in exactly one of its five (excluding PRANK) 12 species-, unmasked- alignments (28 genes)

2S-D) the gene had inferred positively selected sites in its PRANK 12 species-, unmasked- alignments. In this analysis alignments other than ones generated by PRANK were not visualized. (20 genes)

2S-E) were among the *melanogaster* group, masked alignments genes that were inferred to be under positive selection at FDR of 0.01 in Drosophila 12 Genomes Consortium (2007). (25 genes)

Details of the selection process are described in the Materials and Methods section of the article.

The table contents are as follows (only the applicable items are shown in each of the tables):

(a) Flybase *D. melanogaster* ID.

(b) Number of aligners with whose alignment the gene contained sites inferred to be under positive selection at cutoff of 95%, among the five aligners other than PRANK.

(c) Aligners contributing to the number in (b); "*" means that this program's alignment does not have any positively selected sites at this cutoff.

(d) Are the sites inferred to be under positive selection after the different alignment treatments the same? Y – yes; SR – not the same site (at least not in *D. melanogaster*) but they are in the same neighborhood; N – no, and it's not SR.

(e) What is the quality of the alignment at the sites inferred to be under selection? M - most likely it is a misaligned codon inappropriate for PAML analysis; C – most likely it is a correctly aligned codon; "~" indicates lower confidence in the C/M conclusion. In 2S-E, additionally S indicates that there are some well aligned as well as some not well aligned codons, and this would likely affect the significance (and the gene might not pass the applied FDR cutoff once the misalignments are accounted for)

(f) Number of sites inferred to be under positive selection at cutoff 0.5 with BEB analysis after the different alignment treatments.

(g) Number of sites inferred to be under positive selection at cutoff 0.95 with BEB analysis.

(h) Number of sites in the positive ($\omega>1$) PAML class.

(i) Characteristics of the selected sites: S – single site in a well conserved neighborhood; R – a site belonging to a fast evolving region (compared to the rest of the gene); "i" – presence of insertions and deletions affecting the alignment of the site.

(j) Interesting features related to the selected sites: S – at the start of the coding sequence; E – at the end of the coding sequence; GM – a gross misalignment where a reasonably long conserved sequence is misaligned; R – presence of "repeats" affecting the selected codon's alignment. ; P – issues related to the PAML analysis (for example inferred selection at a site with only Serine amino acids); DT – the selected codons belong to different transcripts or other similar annotation issue making PAML analysis inappropriate.

(k) Are the inferred selected sites masked in the masked version of GLEAN-R (Drosophila 12 Genomes Consortium 2007)? Y - yes, at the sites inferred to be selected, N – not at the sites inferred to be selected, P - partial masking. In addition, for 2S-E only (where changes at other locations in the genes might affect the FDR and thus selection

inference substantially): Nc – masking closeby to these sites; Ne – masking elsewhere in the gene; N – no masking anywhere in the gene; I – there are cases of bad masking.

(l) Is positive selection inferred for this gene in Prank? [Used in the analyses that do not account for the PRANK alignment results during gene selection]: Y - yes, N - no.

(m) Number of sites inferred to be under positive selection at cutoff 0.5 with NEB analysis

(n) Number of sites inferred to be under positive selection at cutoff 0.95 with NEB analysis;

(o) Are the sites inferred to be under positive selection in the PRANK alignment located at sites and regions that are classified by IUpred as disordered (IUpred score >= 0.5)? Y – yes, otherwise - no.

Underscript, such as in $M_{1,}$ was used to distinguish between the features of different selected sites within the same gene (in some, but not all, of the cases when there were multiple sites and their features or annotation differed).

Table 2S-A.

Flybase *D. melanogaster* ID	#	A	C	M	P	T	same sites?	selected codons quality	# BEB sites	# BEB sites P>=95%	# (w>1)	type of site(s)	features	masked?	Pran k?
(a)	(b)	(c)					(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
FBgn0000316	5	A	C	M	P	T	Y	C	12-16	1	31-34	S		N	Y
FBgn0022960	5	A	C	M	P	T	SR	M	4-35	2-25	8-37	R	S, GM	N	
FBgn0030935	5	A	C	M	P	T	SR	M	57-88	33-52	58-204	Ri	E	PN	
FBgn0040696	5	A	C	M	P	T	SR	M	9-13	5-9	13-19	Ri	E	PN	
FBgn0036058	5	A	C	M	P	T	Y	C, M	4-5	1-3	8-14	R, Ri	S, GM	N, PY	Y
FBgn0036089	4	A	C	*	P	T	Y	M~	5-9	1	42-53	Ri	E	Y	
FBgn0030914	4	A	C	M	*	T	Y	C~	8-24	1	30-51	R		N	Y
FBgn0051212	4	A	*	M	P	T	N	M	27-46	1-3	99-111	Ri	E	P	
FBgn0026210	4	*	C	M	P	T	SR	M	13-149	17-22	38-171	Ri	E, R, DT	P	
FBgn0032408	4	A	*	M	P	T	SR	M~	4-10	1-2	12-16	Ri	R	Y	
FBgn0039025	3	*	*	M	P	T	SR	M	8-19	1-3	18-28	Ri	R, DT	N	
FBgn0034792	3	A	*	M	P	*	Y	M~	6-10	1	25-30	Ri	R	na	
FBgn0032629	3	*	C	M	*	T	SR	M~	19-73	2-12	92-111	Ri	R	PY	
Fbgn0002932	3	*	C	M	*	T	SR	M	12-24	1-2	42-44	Ri	S, R, DT	N	
FBgn0031478	3	*	*	M	P	T	Y	M~	7-12	1	27-32	Ri		N	
FBgn0036707	2	*	*	M	*	T	SR	M	5-20	1	71	Ri	E	PN	
FBgn0034434	2	*	*	M	P	*	Y*	M	5-8	1	17-20	Ri	R	na	Y
FBgn0037094	2	*	C	*	*	T	SR	M	4-26	1	22-43	Ri	S, R	N	
FBgn0050190	2	A	*	*	*	T	SR, N	M	11-18	1-3	23-47	Ri	S, DT	PN	
FBgn0030395	2	*	*	M	*	T	N	M	6-18	5	42-45	Ri	R, E, DT	P, N	
FBgn0010114	1	*	*	M	*	*	NA	M	36-38	2	74	Ri	R	na	
FBgn0004380	1	*	*	*	*	T	NA	M	2-4	1	10	R	S	PN	
FBgn0025833	1	*	*	M	*	*	NA	M	8-29	1	59	Ri	R	na	Y
FBgn0037580	1	*	*	*	P	*	NA	M	8-18	1	21	Ri		na	
FBgn0035941	1	*	*	M	*	*	NA	M	12-14	2	26	Ri		na	
FBgn0035048	1	*	C	*	*	*	NA	M	17	1	34	Ri	R	na	Y
FBgn0002466	1	A	*	*	*	*	NA	M	17	1	53	Ri	R	na	Y

Table 2AS-B.

	Flybase *D. melanogaster* ID	#	A	C	M	P	T	same sites?	selected codons quality	# BEB sites P>=95%	type of site(s)	features	masked ?	Prank?
	(a)	(b)	(c)					(d)	(e)	(g)	(i)	(j)	(k)	(l)
A	**FBgn0052405**	5	A	C	M	P	T	Y	C	1	Si		N	Y
	FBgn0032058	4	A	C	M	*	T	SR	M~	5	Ri	R	PY	
	FBgn0025140	2	A	*	*	P	*	Y	M	2	Ri	R	N	
	FBgn0036202	4	A	*	M	P	T	SR	M	4-10	Ri	S	PN	Y
	FBgn0010651	3	A	*	*	P	T	SR,N	M	1-3	Ri	R~	Y	
	FBgn0035008	5	A	C	M	P	T	Y	M_1,C_2	1-2	Ri_1, S_2	S_1	N	Y
	FBgn0037521	5	A	C	M	P	T	SR	M	1-2	Ri	R	PN	
C	**FBgn0035729**	4	*	C	M	P	T	SR	M	15-33	Ri	S_1, DT_1, E_2, DT_2	N_1, PN_2	
	FBgn0027570	5	A	C	M	P	T	Y,N	C	1-2	S		N	Y
	FBgn0015721	5	A	C	M	P	T	SR	M	2-3	Ri	S, DT	N	
	FBgn0034259	3	*	C	M	*	T	Y	M	1	Ri	DT, S	Y	
	FBgn0039225	4	*	C	M	P	T	SR	M	4-30	Ri	DT, S	PN	
M	**FBgn0013717**	3	*	*	M	P	T	SR	M	1-13	Ri	S, DT~, R	PN	
	FBgn0039531	3	*	C	M	*	T	SR	M	2-5	Ri	S, DT	Y	
	FBgn0040823	5	A	C	M	P	T	SR,N	M	31-77	Ri	E_2, DT_2	PN	Y
	FBgn0033068	4	*	C	M	P	T	SR	M	5	Ri	S, DT	N	
	FBgn0004595	2	*	*	M	*	T	SR	M	3-4	Ri	GM_1, R_1, S_1, R_2, S_2	N_1, PN_2	Y
P	**FBgn0051665**	3	A	*	*	P	T	SR,N	M~	1-40	Ri		N	
	FBgn0004649	3	A	*	*	P	T	N	S	1-21	Ri		N	Y
	FBgn0050376	5	A	C	M	P	T	Y	C~	1	Ri		PN	Y
	FBgn0033607	5	A	C	M	P	T	Y	M	1	Si	DT, E	N	Y
	FBgn0010278	3	A	C	*	P	*	SR	S	1-2	Ri	DT	PN	
	FBgn0033313	5	A	C	M	P	T	SR	M	3-10	Ri	S, DT	PN	
	FBgn0035518	5	A	C	M	P	T	Y	M	1	Ri	S~, DT	N	
T	**FBgn0020647**	3	*	C	M	*	T	SR	M	1-2	Ri	R, ~DT, S	PN	
	FBgn0036844	5	A	C	M	P	T	SR	M	1	Ri	R	N	
	FBgn0030204	5	A	C	M	P	T	Y_1, SR_2	C_1, M_2	1-2	S_1, Ri_2	S_2, DT_2	N_1, PN_2	Y
	FBgn0035873	5	A	C	M	P	T	SR,N	M	2-9	Ri	S, DT	PN	Y
	FBgn0050164	5	A	C	M	P	T	SR	M	16-33	R, Ri	GM, DT, S, E	PN	
	FBgn0037254	3	*	C	*	P	T	SR	M	1-2	Ri	S, DT	Y	

Table 2S-C.

Flybase *D. melanogaster* ID	#	A	selected codons quality	# BEB sites P>=95%	type of site(s)	features	masked?	Prank?
(a)	(b)	(c)	(e)	(g)	(i)	(j)	(k)	(l)
FBgn0003392	1	A	M	3	Ri	DT	N	
FBgn0038603	1	A	M	1	Ri	R	N	
FBgn0000562	1	A	M	1	Ri	R	N	
FBgn0039197	1	A	M~	1	Ri	S, DT~	PN	
FBgn0034099	1	A	M	1	Ri	E	PN	
FBgn0002466	1	A	M	1	Ri	R	Y	Y
FBgn0039157	1	C	M	1	Ri		N	
FBgn0034994	1	C	M	1	Ri	R	PN	
FBgn0004102	1	C	M	4	Ri	R, E, DT	na	
FBgn0034674	1	C	M	1	Ri	GM, DT~	P	Y
FBgn0015567	1	C	M	2	Ri	GM, S, DT	PN	
FBgn0035048	1	C	M	1	Ri	R	PN	Y
FBgn0029928	1	M	M	1	Ri	R(short)	PN	
FBgn0040376	1	M	M	1	Ri		N	
FBgn0011829	1	M	M	2	Ri	S_1, DT_1	PN_1, N_2	
FBgn0032850	1	M	M	1	Ri		PN	
FBgn0013467	1	M	M	1	Ri	R(short)	N	
FBgn0025833	1	M	M	1	Ri	DT~	N	Y
FBgn0036512	1	P	S	1	Ri		N	
FBgn0030774	1	P	M~	1	Ri	S, DT~	PN	
FBgn0035523	1	P	M~	1	Ri		PN	
FBgn0030993	1	P	M~	1	Ri	S, DT~	N	
FBgn0036206	1	P	M	2	Ri	E, DT	N	
FBgn0032131	1	T	M	1	S, i	DT	N	
FBgn0037807	1	T	M	7	Ri	S, DT	Y	
FBgn0025383	1	T	M	1	Ri	S, DT	PY	
FBgn0038131	1	T	M	1	Ri	S, DT~	N	
FBgn0031844	1	T	M	1	Ri	E, DT	Y	

Table 2S-D.

[(e), (f), (g), (i), (k), (o) refer to only as observed in the PRANK alignment]

Flybase *D. melanogaster* ID	#	A	C	M	P	T	same sites?	selected codons quality	# BEB sites	# BEB sites P>=95%	type of site(s)	features	masked?	Disorder?
(a)	(b)	(c)					(d)	(e)	(f)	(g)	(i)	(j)	(k)	(o)
FBgn0032923	5	A	C	M	P	T	Y	~C	4	1	Ri	F	N	N
FBgn0000575	2	A	*	*	P	*	Y	~C	2	1	Ri	~R	N	Y
FBgn0016794	1	*	*	*	P	*	N	M	23	1	Ri			Y
FBgn0016694	5	A	C	M	P	T	N	M	32	1	Ri	R, F	N	Y
FBgn0034271	5	A	C	M	P	T	Y	~M	13	1	Ri	~R	N	Y
FBgn0036738	1	*	*	*	P	*	Y	M	8	1	i	E, ~DT	N	Y
FBgn0035192	1	*	C	*	*	*	Y	M	9	1	Ri	R	N	Y
FBgn0029836	3	A	*	M	P	*	SR	M	2	1	Ri	S, ~DT	N	Y
FBgn0028961	2	A	C	*	*	*	N	M	16	1	Ri	S, ~DT	N	N
FBgn0051660	5	A	C	M	P	T	SR	M	36	1	Ri	S, ~DT, ~R	P	Y
FBgn0010415	3	A	*	*	P	T	SR	~$C_{1,5}$, ~$M_{2\text{-}4}$	17	5	Ri	S_1, ~DT_1; R_2, F_2	$N_{1\text{-}4}$, PY_5	Y
FBgn0035873	5	A	C	M	P	T	SR	?S	7	1	Ri	S	PN	Y
FBgn0034398	5	A	C	M	P	T	Y	~M	8	1	Ri	S	P	N
FBgn0000451	0	*	*	*	*	*	*	~C	23	1	Ri	~R	N	Y
FBgn0010611	5	A	C	M	P	T	Y	C	1	1	S		N	N
FBgn0040936	1	A	*	*	*	*	SR	C_1, M_2	6	2	Ri	S	N	N
FBgn0051666	3	A	*	*	P	T	N	~M	15	1	Ri	P(Ser)	N	Y
FBgn0030323	2	*	*	M	P	*	SR	?S	3	1	Ri	E, ~DT	N	Y
FBgn0027585	0	*	*	*	*	*	n/a	~M	33	1	Ri		N	Y
FBgn0043458	5	A	C	M	P	T	Y	C	5	1	S		N	N

Table 2S-E.

Flybase D *melanogaster* ID	selected codons quality	# BEB sites	# BEB sites P>=95%	# NEB sites	# NEB sites, P>=95%	# (w>1)	type of site(s)	features	masked?
(a)	(e)	(f)	(g)	(m)	(n)	(h)	(i)	(j)	(k)
FBgn0000039	M	5	2	6	5	6	R	S	Y
FBgn0036686	M	6	0	17	3	~20	Ri	R, P	N
FBgn0011676	S	27	1	23	5	38	S, Ri	R	Ne
FBgn0036195	M	2	1	2	1	3	S	E, DT~	N
FBgn0050166	M	51	36	38	33	54	Ri_1, R	E	Nc
FBgn0051145	M	71	14	26	7	140	Ri_1	S, DT, GM	N
FBgn0029710	S	18	2	24	2	28	S, R		Y
FBgn0034295	S	12	1	44	14	35	S, R	R	Y
FBgn0039246	M	11	1	13	2	40	S, Ri	E, DT	Y, Nc
FBgn0033418	M	4	1	4	4	4	R_1	E, DT~	Y
FBgn0043854	C	30	0	15	1	~23	S, Ri		Y, Nc
FBgn0037940	M	12	7	0	0	15	R		N
FBgn0053249	C	12	1	10	3	17	S, Ri		Y
FBgn0036522	S	37	1	16	2	123	S, R, Ri		Ne
FBgn0050325	S	26	0	133	28	-	S, R, Ri		Y
FBgn0032627	M	94	50	70	30	99	R	S, E, DT~, GM	I, Y, Ne
FBgn0025815	M	3	1	2	1	6	Ri	R~	Y, I
FBgn0033607	M	1	1	1	1	1	S	DT	N
Fbgn0030998	S	8	1	10	3	10	S, Ri	S	N
FBgn0027563	C	22	1	21	3	78	S, R, Ri	G~	Y
FBgn0031155	C	122	0	148	47	199	S, R, Ri	R	Y
FBgn0031196	M	8	0	9	4	~11	Ri_1		Ne
FBgn0036254	C	24	1	0	0	35	R	P, F intron	Y
FBgn0039862	S	7	0	4	1	-	S, Ri	R	Ne
FBgn0033942	C	4	2	3	3	6	S		N

Table 3S. Summary of data from the visual inspection of inferred positively selected sites which are located in regions that appear to be evolving much faster than the rest of the gene.

(a) Flybase *D. melanogaster* ID.

(b) Number of aligners with whose alignment the gene contained sites inferred to be under positive selection at cutoff of 95%, among the five aligners other than PRANK.

(c) Aligners contributing to the number in (b); "*" means that this program's alignment does not show any positively selected sites at this cutoff.

(d) Is/are the site(s) inferred under positive selection located immediately next to an exon start or end in *D. melanogaster*? Y – yes, N – no.

(e) For sites that are not close to an exon start or end in *D. melanogaster*, is there a record that the particular sequence is part of a *D. melanogaster* full length cDNA or 5` EST sequence? Y - yes, N – no.

(f) Does the positive selection inference appear to be due to annotation differences, as opposed to an actual change in the coding sequence? Y – likely annotation differences, N – could be a change in the coding sequence.

(g) Is positive selection inferred for this gene in Prank? [Used in the analysis that do not account for the PRANK alignment results during gene selection]: Y – yes, otherwise - no.

Flybase *D. melanogaster* ID	**#**	**A**	**C**	**M**	**P**	**T**	**At Exon Border**	**cDNA or 5`EST, *D. melanogaster***	**Annotation issue**	**Prank?**
(a)	(b)	(c)					(d)	(e)	(f)	(g)
FBgn0037521	5	A	C	M	P	T	N	Y	N	
FBgn0015721	5	A	C	M	P	T	Y	na		
FBgn0040823	5	A	C	M	P	T	N_1, Y_2		Y_2	Y
FBgn0033607	5	A	C	M	P	T	Y	na		Y
FBgn0036844	5	A	C	M	P	T	N	Y	N	
FBgn0032058	4	A	C	M	*	T	N	N	N	
FBgn0035729	4	*	C	M	P	T	Y	na	Y	
FBgn0013717	3	*	*	M	P	T	Y	na	Y	
FBgn0051665	3	A	*	*	P	T	N, Ytc	Y	N	
FBgn0020647	3	*	C	M	*	T	N	Y	N	
FBgn0003392	1	A	*	*	*	*	Y	na	Y	
FBgn0039157	1	*	C	*	*	*	N	Y	N	
FBgn0029928	1	*	*	M	*	*	N	Y	N	
FBgn0036512	1	*	*	*	P	*	N	Y	N	
FBgn0037807	1	*	*	*	*	T	Y	na	Y	
FBgn0032131	1	*	*	*	*	T	Y	na	Y	
FBgn0035518	5	A	C	M	P	T	Y	na	Y	

Table 4S. Number of genes for which exactly the given number of alignments, out of the AMAP, ClustalW, ProbCons, MUSCLE and T-Coffee alignments, were significant for positive selection as inferred based on (a and b) comparing LRT with the χ^2 distribution, and (c and d) q-values and FDR (see Materials and Methods). The significance cutoffs are indicated in the second row.

number of alignments with positive selection	χ^2		FDR	
	0.01 (a)	0.05 (b)	0.01 (c)	0.1 (d)
1	719	794	665	806
2	355	439	291	444
3	286	401	221	372
4	302	418	241	378
5	584	909	361	755
Total	2246	2961	1779	2775

Table 5S. Number and percentage of the genes inferred to be under positive selection based on the "masked and trimmed" alignments (see Materials and Methods) for which exactly the given number of unmasked alignments were also significant for positive selection, out of (a, b) the AMAP, ClustalW, MUSCLE and ProbCons alignments, and (c, d) the PRANK, AMAP, ClustalW, MUSCLE and ProbCons alignments. Positive selection was inferred based on the χ^2 test and a p-value cutoff of 0.05, approximately equivalent to the 10% FDR cutoff in Drosophila 12 Genomes Consortium (2007).

number of alignments with positive selection	out of 4 aligners		out of 4 aligners plus PRANK	
	number of genes (a)	% (b)	number of genes (c)	% (d)
0	129	14.5	93	10.4
1	69	7.75	86	9.66
2	85	9.55	70	7.87
3	95	10.67	70	7.87
4	512	57.52	169	19
5	n/s	n/a	402	45.2

Table 6S. Percent of (a, d) sites which are disordered (IUpred score of >=0.5) among all sites inferred to be under selection (models M7 & M8, cutoff 0.95); (b, e) genes for which at least one of the sites inferred to be under selection is disordered; (c, f) percent of sites which are disordered (IUpred score of >=0.5) among all coding sequence sites considered.

For (a, b, c) disorder was determined only based on the *D. melanogaster* sequence. For (d, e, f) disorder was determined based on the maximum IUpred score in each of the 12 species.

	D. *melanogaster* disorder			12 species disorder		
	among selected sites, % (a)	in selected genes, % (b)	disordered sites, % (c)	among selected sites (d)	in selected genes, % (e)	disordered sites, % (f)
AMAP	36	49	31	75	76	48
ClustalW	38	58	31	83	81	45
MUSCLE	41	60	31	84	83	45
ProbCons	42	59	31	79	82	46
Prank	41	43	31	61	62	50
T-Coffee	38	62	31	85	86	45