

Supplementary Materials

1. Summary of surveyed loci.
2. Removing the effect of K_s using partial regression.
3. Investigating effects of possible recombination rate variation.
4. Effects of local gene density.
5. Spatial autocorrelation in the data.
6. Estimating the fraction of adaptive protein divergence.
7. Codon preference table for *D. melanogaster*.

Supplement 1. Summary of surveyed loci

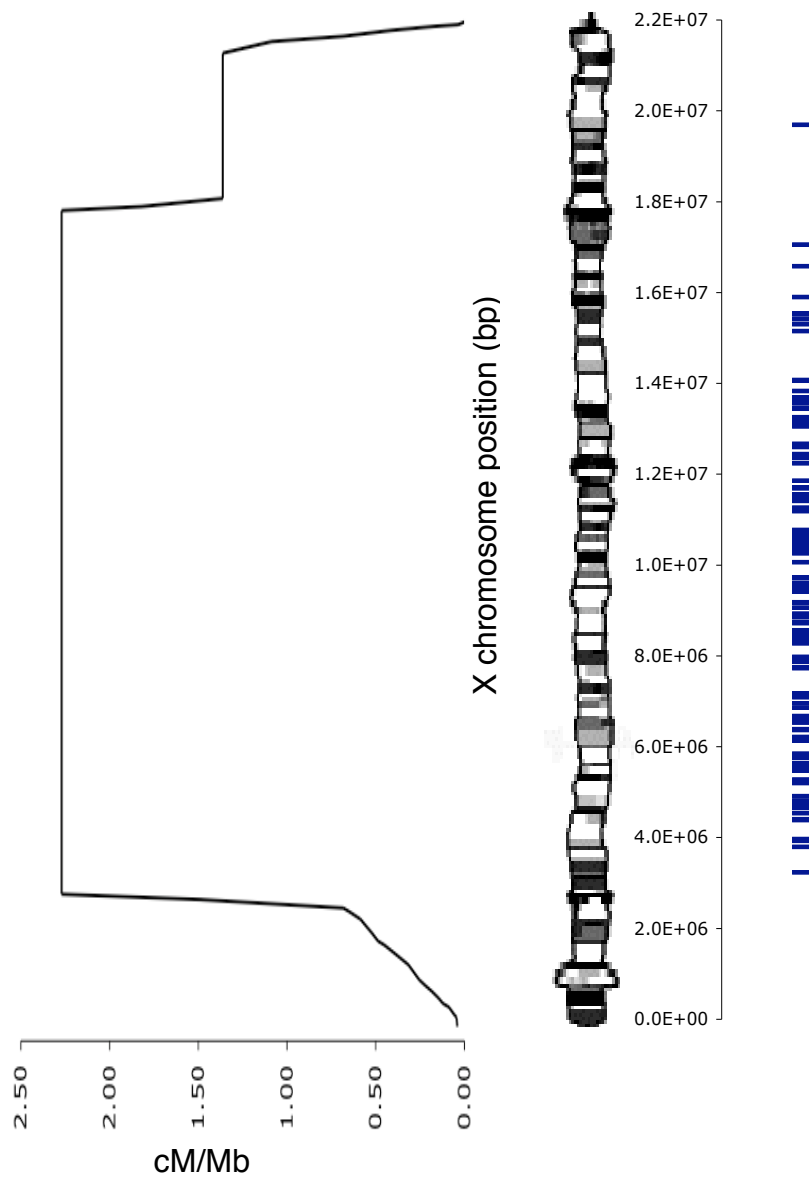


Figure S1.1. Positions of the 137 coding regions surveyed in this study. Estimated recombination rates based on (Charlesworth 1996) and (Andolfatto and Przeworski 2001) are plotted to the left.

TableS1. Information on loci surveyed in this study.

locus	cytology	CNum	For primer	Rev primer	Dsim	start position	end position
3D2_CG10798	3D2	CG10798	CGACAGATAACACCCAGTCAAATC	CATATCGTTGCAGCGCGTGTAGC	MD99	3231623	3232321
3F9_CG5014	3F9	CG5014	GATATATTCTCGCCTTACAGGTCCG	CGCAGCGCTTTGATCTCTCCGG	MD99	3796686	3797295
4A4_CG32790*	4A4	CG32790	ATCAGGAGACATCCATCAAGC	CAGCGGCATAGCCATAGTAAG	MD99	3923335	3923966
4A5_CG6379	4A5	CG6379	TCCGCAACGTTTTTGATAGTC	TATTCGATGAAAGCAACCAAG	MD99	3944235	3944942
4A6_CG6121	4A6-B1	CG6121	CGAAGCTTGGACAACTCGTAGC	AGCTTGTGAATGGCAACAATC	MD73	3960765	3961469
4C7_CG11427	4C7	CG11427	CAGATCGAGTAGCAGGTCCAG	CTCGCTATGATCCCAACTACG	MD260	4382804	4383574
4C7_CG3568*	4C7	CG3568	GCGAAGAGATGTGTGTGGTTC	GCCAAATGAAACCATGAGAAG	MD260	4388525	4389205
4C7_CG3546	4C7	CG3546	AACAGAAGAACAGCGACAGC	CTTCGGAAGAGACAACAGTGG	MD73	4404390	4405109
4C12_CG2984	4C12	CG2984	CCGCCAATAACAGGAGTAGCACC	GGCCACCAGAACTGATCCTCC	MD99	4530632	4531302
4D1_CG2861	4D1-2	CG2861	GGCGATGAAGTAACCTGACCCAC	CGCCGGGCTCTACAAGTCC	Dsec***	4647541	4648321
4D3_CG12681	4D3	CG12681	CCCATCGAATATGCTGGTACACATG	CTTTCGCGGTTTTAGTCGCTGCC	wt501 mosaic**	4721277	4721981
4D5_CG4041	4D5-6	CG4041	CGTAAACTTGTGATCTCTCGC	AAGCTCAAGCTCTCGAATGTG	MD99	4762000	4762738
4E2_CG32767	4E2	CG32767	CTGTTGTAGCTGCTGCTGTTG	ACATCATCTGCTAAACGGACG	MD73	4805261	4806020
4D7_CG15468	4D7	CG15468	CGCCTTCTTTATTCGATCTC	AATCACGCTAGTCAGCCAGAG	MD260	4805361	4806073
4E2_CG6824*	4E2	CG6824	CCAAGAGCCCAAGGACTTATC	TCTCATGAAACGGAGGCAATG	MD63	4910099	4910829
4F5_CG10706	4F5-6	CG10706	TTAACGACACCAAACTCGG	TTGCAGGTGATACTGTTGCTG	MD99	5197104	5197862
4F10_CG15784	4F10	CG15784	AAGGCCAAGAAGGACAAGAAG	AGGAAGAAGTGGAGCTGGAAAC	MD73	5276283	5276908
5A8_CG32763*	5A8	CG32763	ATCTCGTTCTATGTTGCGGTG	GCTGATTCATGAGAAGATCGC	MD99	5468482	5469253
5A8_CG11473	5A8-9	CG11473	TCGATGATGAACGGATAGAGG	AATGTGCTATGCTCAAGCTGGTG	MD73	5475661	5476392
5A10_CG4119	5A10-11	CG4119	CGCGAGGAAGACATAAAGATG	CTGTAGGCGGATCTCCTTTTC	wt501 mosaic**	5512759	5513488
5A12_CG3171	5A12	CG3171	GCCTAAGTTCATAACCTGCAGCTG	CAGCGTGAACAGCGGCGTGC	MD99	5519290	5519945
5B1_CG33080*	5B1	CG33080	GTCTGGCGGCACTGATAGTAG	TATTTCACTCACTCGATCGGC	MD63	5554995	5555717
5B6_CG4078	5B6	CG4078	CAAGACCCCTAAGCTTGCTGTG	GAGCAGTGTAAAGTGTGCTCCAG	MD105	5593314	5594018
5C2_CG3108	5C2	CG3108	CTGCTTATCACTGGCTCTG	CGGCTATGCAACTCGTAGAAC	wt501**	5625667	5626362
5C2_CG12239_A*	5C2	CG12239	AGCCACCAATGATGAGACAAC	GTCTTCTGTCAGCTTCATGTC	MD260	5637516	5638203
5C2_CG12239_B	5C2	CG12239	GCGAATCGGACAGGCGTGAC	CTTCTCTTCGACAGCCATCGC	MD99	5638314	5639045
5C7_CG3011	5C7	CG3011	AATCGGTTATCTTGGGACTGC	ATTACGATAAGTTGGCGGAGG	MD260	5758573	5759281
5D1_CG4766	5D1	CG4766	GAGCTTGGTGGAACCTCGAAGT	GCGAAGGTGCCATCTGTCACAG	MD99	5845157	5845371
5E5_CG3458	5E	CG3458	GCGGAGAAGGATACCTCAAAAC	CTACATCAGCTATCCGCGAAC	wt501**	6130754	6131551
5E5_CG15897	5E5	CG15897	ATCAGTACGACTGCGTTGAGG	AATTTCTTTGGGTATGTGGG	MD99	6133938	6134646
5E7_CG12219	5E7	CG12219	TCGTTCTCTTTCGAGCCTTTTC	CAACTGGAGAGCTTCCACAAC	MD99	6145397	6146131
5F4_CG3869*	5F4	CG3869	TAAAGGTGAGCAACGAGAAGG	CCCCAGCTGAACCTTAAACTCC	MD63	6213872	6214570
6B1_CG3918*	6B1	CG3918	CAAGAAGCATGGCAAGAAGAG	TCATGGAGAAGTTGTTTGGG	MD99	6370521	6370958
6B1_CG3342	6B1	CG3342	AATGTGTAGCTGCATTTTGGC	CTGCCAGTGGAGCACTATCTC	MD73	6371731	6372403
6B1_CG3929	6B1-2	CG3929	CCACCAATCTACGCCAGATAC	TTTTGCTGGGCAATTATCATC	MD63	6381242	6381967
6C6_CG4523*	6C6	CG4523	TCATGTGAACAAGTCGACAG	TGGAAAACACATTTCCGACTC	MD63	6530894	6531618
6C6_CG3184	6C6	CG3184	TGCATCTGGTGTGCAAGAGAC	ATGTTTACAAACGCTCCACAGC	MD63	6533253	6533918
6C11_CG14439	6C11-12	CG14439	ACCATTGGTTATATCCTGGGCG	TGCCGGATTCTTGATAAATTG	sim6**	6599123	6599776
6C12_CG14438	6C12	CG14438	AAGGTAAAGGCCAAGCTGGT	CGCGTGGTGATGTGTTTCTG	MD03	6606538	6606091
6D3_CG32743	6D3-4	CG32743	GGCTACAGTGAGCTGGTCAAG	TCCTTCTCCTTCAGGCTTCTC	MD99	6674108	6674816
6E4_CG4593*	6E4	CG4593	AACAAGGTGAACAACCTGGA	TGAACCAAAAAGATGGCAATT	MD57	6859476	6859838
6F5_CG4626*	6F5	CG4626	TAAGGAAGACACTAGCGCAGG	GCTCAATAGACACTCCATGC	MD63	6953924	6954646
7A2_CG9650*	7A2	CG9650	ATTATCCACAGAGGACCCAC	AGGCTTCAGCTTCAGGTTTTTC	MD255	7083716	7084449
7A8_CG1677	7A8-B1	CG1677	TTGTCTCAAAGTCCATGTCC	CGAAACAAACAGACCAAGAG	MD99	7128224	7128928
7B1_CG2079	7B1	CG2079	CCACGCAAAAGCCGACAAACCC	CACCTCGTCTCGGTTTGCTGACC	MD99	7172034	7172736
7C1_CG15327*	7C1	CG15327	GCGGATATTAAAGTCTTGCCC	ATACACACCCAGGAGCTGTTT	MD260	7722728	7723436
7C2_CG1422	7C2	CG1422	GTTTGAGCTCTCACAGGATGC	GAACTCTCGTTTCATCGTGCTC	MD63	7740214	7740963
7D2_CG15333	7D2	CG15333	ATCGCCCACTGTACACATTTTC	AGCCAAAAGGACAAGAGCTTC	MD63	7865264	7865935
7D3_CG1560*	7D3	CG1560	GTGCAGACCTCTAAGTGCAGC	CTGGACACAATCCTTCAGCTC	MD99	7912906	7913613
7D12_CG2116	7D12	CG2116	GAAAGGGTGGAAGAGGATCTG	ATGCAATCTCTACGATGACG	MD260	7975619	7976368
7E7_CG18009*	7E7	CG18009	ATGCAAAAACGATATGGTCAGC	AAATCGAGGCAGTGGTGTATG	wt501 mosaic**	8272355	8273054
7E11_CG1474	7E11	CG1474	AAATGTAGCAACACCCCAATTG	TTTGTTTTGGATTTCGCATCC	MD63	8307389	8308081
7F3_CG11219	7F3	CG11219	CTTGTAACAAGCTGGAGCTGG	AATAATCAAAACGCGACAACG	wt105 mosaic**	8372263	8372923
7F4_CG11265	7F4	CG11265	GTTTCAGAAGGAACAGGAGGG	GTGGGCGAGGACTACTGATAG	MD63	8408677	8409411
7F7_CG12737*	7F7	CG12737	TTTATCCGTTTTCATTGAGGAG	ATCTGGCTGATCGTGGTACTG	MD260	8441222	8441926
8A1_CG12772*	8A1	CG12772	TTTAGCATATCGCCCAAACTG	CAGTCTCCTGCTCCACTTGAC	MD99	8523561	8524301
8A2_CG1994	8A2-3	CG1994	AGTGCAGGACACATAAATGG	TCTGCAAGACCTACGATCAGG	MD63	8563779	8564510
8B4_CG12075	8B4	CG12075	ATCCATTTCGAACAAGTCGTG	GTTCCCTGGTCTTGGGAGTAG	MD99	8715613	8716308
8B6_CG1885	8B6	CG1885	ATCCGATGATGATTTGTGGAG	GATTTCAATCCCGTCTTTGTG	MD106TS**	8744465	8745175
8B8_CG10966*	8B8-C1	CG10966	CGAGTGAGGATGAGGATGAAG	TATGATGACTGGAGGAGGCAC	MD63	8846307	8847052
8C4_CG7246	8C4	CG7246	AATACAGAATGACACTGCGCC	CATTATGCTGAAAAGGCTGC	MD62	8924769	8925464
8D1_CG15368	8D1	CG15368	GCGAAGCGCAAGCGTACCGAATATC	CCCAAAATTAACAATGCCAAGGCCAC	MD105	9061168	9061584
8D7_CG12149	8D7	CG12149	CTGCTGCAGGAATCTCTCATC	CCTTATGGCTGGATCTCTTC	MD63	9167586	9168326
8D7_CG32705	8D7-8	CG32705	GAAGAAGTCCCAGAGATTGGC	CAATTCAAGTTTCAATGGCCTG	MD63	9174664	9175293
8E10_CG32702	8E10	CG32702	GAAGTGTAGGACAAAGCTGACG	TTCCAGTTGAACCTACGGGATG	MD260	9412574	9413296
8E10_CG3001*	8E10	CG3001	AACGAGGATGTTGTCCAGTTG	CGAGATCTTCTCGACCATCAG	MD63	9432212	9432923
8F7_CG15319_B	8F7-9	CG15319	ACACACACTGGTTATCCCTGTC	ATCAATACTCGATGGCGAATG	MD99	9514930	9515532

TableS1. ...cont.

locus	cytology	CGnum	For primer	Rev primer	Dsim	start position	end position
8F7_CG15319_A	8F7-9	CG15319	GTGTGTCGCTATTTAGACCGC	CAGTTAGCCAAACAGTGATCC	MD99	9516613	9517348
9A1_CG1343*	9A1	CG1343	CTCGCGAGTAATTTACGATC	TCITTTCCATACACCTTGCCAC	MD99	9598798	9599514
9A2_CG2045	9A2	CG2045	ACCCAGGATCAGTAATGGGAC	TAGACATATCGATTGCCGCTC	MD63	9711285	9711998
9A3_CG15247	9A3	CG15247	GTGGAGAGTGTCTTTCTCTGC	GCCAATTTGCTGAGTTTGAAG	MD62	9720161	9720892
9A3_CG1468	9A3	CG1468	GCTTCCGACAGATTTTGGTTCCG	GTAGCGCAGTTCCGCTATTGCC	Dsec***	9723724	9724404
9B4_CG2962	9B4	CG2962	GCGGCTCCCCAGATCTTCAAGG	GTGACCGATGGGCAGGACTG	MD99	10060563	10061023
9B14_CG32683	9B14	CG32683	CACCGAGCCGGAGGAGATCAC	CACATTGATGACTTTGGGACTGC	MD99	10258127	10258739
9D2_CG33175*	9D2	CG33175	ATGAAACGGACACAGAAGTGG	TTGGAAGTGGTCAAGTCGTTT	MD63	10343855	10344559
9D3_CG1619	9D3	CG1619	TATCGGCACTTCGGAGCTGTGTC	GCCGCGCGCTTGGGCAAAATTC	MD99	10439128	10439891
9D3_CG2124	9D3-4	CG2124	ACTCCATGTTCTGTGCAATC	CTGTGCTTGCTTCTCGGTAAAC	MD63	10442133	10442834
9D4_CG9817	9D4	CG9817	TTCCGTTTCCCATTAGTAG	AGCAGTCGTTACCATCACAC	MD99	10449385	10450091
9E1_CG2111	9E1	CG2111	GAGATCTGGCTGAAGAAGCTG	ACACGATGGATGAGACGAAAC	MD73	10504042	10504792
9E1_CG32677*	9E1	CG32677	ACAATATGGCTGCTCAGTG	GAGCGATTTCTCACTGTGCTC	MD63	10550069	10550766
9E1_CG32676	9E1-2	CG32676	GACACTTGCTAGGTGAGCTGG	ATCCTCAACGTCATCTGTCATC	MD62	10561407	10562177
9E1_CG1799	9E1	CG1799	TATATACGGCCAGAACCGCTGG	GCCACCAATGACCTGTAGTTCCGG	MD99	10591638	10592263
9F2_CG1691	9F2	CG1691	CATCCCTTAAAAAGCCCCCTCG	CGATGACGACGCCGACGACAC	MD99	10642605	10643142
9F5_CG32669	9F5	CG32669	ATTGTTGTCCCATTTCAAGCAG	CCATTCTGTGCTGCTTATGTC	MD73	10688550	10689278
9F5_CG2202	9F5	CG2202	GTTGTTTCCCAGTTAGAGCC	CGTATTGCATCGAAAGGCTTC	MD260	10694220	10694948
9F8_CG15208	9F8	CG15208	AGTCTCTGCCTCAAGATGCCGTAC	CAACAGATGTCGGGTGGGGTCTC	wt501 mosaic**	10745875	10746557
9F11_CG2145*	9F11	CG2145	GTGGAGAGCAGCAACCAAGAG	TACCGAGGAAGAGCAACGAAGG	MD63	10773006	10773706
10B3_CG11759	10B3	CG11759	ATGTCGAAGCAATATCGCATC	ACGCGATTGCTGTGAAATATC	MD63	11180614	11181345
10B3_CG1737*	10B3	CG1737	ACTTCATTTCCACAGCCATTG	AAGACCTCCTCCTCTGTGACG	MD260	11197707	11198415
10B10_CG12622	10B10	CG12622	ATGCGAGTGGGCAAGTTGTGC	CCGTGAGCAGCATCTTCAAGC	MD99	11257634	11258301
10C6_CG1554	10C6	CG1554	TTCAATACGTAGCAACCAAGG	ACGGATGATGTGATCAAGGAG	MD63	11410951	11411691
10C9_CG1561	10C9	CG1561	CAGCAGCTCAAAGTACGATCC	ACACCGAAGATGAACGAAATG	MD260	11428067	11428792
10D1_CG1703*	10D1	CG1703	CAGGCCAAGAGGTAAGAAAG	ATTGTTGTGCAAGTCCGAGTG	MD99	11450211	11450897
10D1_CG1705	10D1	CG1705	ATCTACCAATCCCAGTCTGC	TGCTGTAAATGGTCAAGCCC	MD63	11455020	11455785
10D4_CG1817	10D4	CG1817	GTGGGAAGGTCGGGCACTTCT	CCGAATGTCACTATCCGGTGATC	MD99	11511017	11511697
10D5_CG1830	10D5	CG1830	CTAATGGCCAAGGACGAGGAGG	ACTATGATAGGATGCCCATGATCC	MD99	11534837	11535365
10D6_CG2446	10D6	CG2446	TTTCGTGCTCCAAATCAACAC	TACCAACATTCATTAGCAGGC	MD99	11550603	11551334
10F1_CG10347	10F1	CG10347	CCTGATTATAGCTGGTCCCAACGG	CAGATTGTTTGGCGTGGTTATGCC	MD99	11686929	11687601
10F1_CG1886	10F1	CG1886	ACATTGAGGATGCCAAGTGAC	TATATTCTCGGCCGTCAACAC	MD73	11689342	11690052
10F2_CG1847*	10F2	CG1847	TCCGATCTGGAGTTTCATCAT	CGATGGACTGAGCTCCTTA	MD57	11703862	11704290
10F2_CG11802*	10F2	CG11802	TTCCACGAGGTGATTCTGTTC	AGGAATCGGATCGATAAATGG	MD63	11705745	11706476
11A1_CG1796*	11A1	CG1796	TCTGGTTAAGTACAATCGGG	GATGGATGGAGGACTATGCTG	MD255	11855467	11856187
11A9_CG32651	11A9	CG32651	GCGTTTGCAAAATGGGTGCTCCC	GACATCGCCACCCGTTGACCAC	Dsec***	12240749	12241447
11A12_CG11146	11A12	CG11146	TGATGATGATGATGGTGATGG	GACTCGCAAGATGCCCAAGTCG	MD63	12354666	12355344
11B2_CG32654_A*	11B2	CG32654	ACAGCCATTGAATCAGCAATC	TAGTACAGGAATGGGTGCAG	MD63	12432012	12432705
11B2_CG32654_B	11B2-3	CG32654	CGCTAAAGCAGCCACTGTTCAC	TTGTTGCTCTTTTGTGCTGTC	MD73	12440348	12441016
11B16_CG12717	11B16	CG12717	TCATCACTGAGTGTGCTGTGC	TTTCATCTGGCAATAAAGCAGGG	MD03	12595789	12596541
11C4_CG3754	11C4	CG3754	TTTTATACTGCCACCAATGCC	CAAGAGCATCCGAGTGAAAG	MD99	12663902	12664644
11E3_CG1771	11E3-8	CG1771	CACAGTCTTTCTGGAAGGTGTC	ACACACTCAAGTTGACGGGAC	MD63	13051319	13052044
11E8_CG15742	11E8	CG15742	GCATTAAGCCGCCCTATCGCTC	GAACCTCTCAGTCCAGGTAGATGC	wt501 mosaic**	13091573	13092084
11E9_CG1618	11E9	CG1618	CTCCAGGAACCTATGACTC	ATGCCGATAACGAGAATGTTG	MD63	13093786	13094478
11F1_CG10617	11F1	CG10617	TTCCGATACCATCGACAAGGAG	ACTTTCTCTCTCCACGAAGGG	MD99	13197625	13198301
11F1_CG1662	11F1	CG1662	CTAAGTCCAGAACCCGAAAG	TAGAAGATGCGGTAATGGGTG	MD73	13221948	13222545
11F4_CG1716	11F4-5	CG1716	GAACGAATCGAGCAGATGAAG	CTGAATGAGCTCGTTTCCAG	MD99	13252916	13253686
12A8_CG11178	12A8	CG11178	TTGGATATTCTTTTGCAGGC	CTCGAGGGAGTCAGTGAAATC	wt501 mosaic**	13440381	13441115
12A8_CG12175	12A8	CG12175	TTGGAATTGTATTGGCTACGG	TATGATCCGCGCTATGGTAAC	MD99	13448377	13448889
12B4_CG9941	12B4	CG9941	CTGTAGCAGGTAGCACAAGCC	TGCTTACAAGTATCCACCGAG	MD63	13554785	13555528
12B4_CG10988	12B4	CG10988	TGACCTTCATTAGTTCTGTCGC	TGAAGAAGGATGTGGTAACGG	MD63	13576319	13577020
12B8_CG10990*	12B8	CG10990	CCAAATGTCTTAGCTCTCTCC	AGGATTAAAGCGTAAGGCCAAG	MD260	13584907	13585622
12C1_CG11111	12C1-4	CG11111	AACAAGCGGCAGTTACGTTT	CGATCCGAACATGAAGAAGTC	MD99	13605449	13606183
12C5_CG10996*	12C5	CG10996	GTGAGAGCAAAAGCAACATC	TCTTTGGTATTGCGCAACATC	MD99	13630805	13631510
12C6_CG11092	12C6	CG11092	AGAACGAGCGGGAGAACGCC	GTACTCCAACCTTAGTGTACCCCTC	MD99	13657347	13658047
12C7_CG11071	12C7	CG11071	GTAATGATGAATGCCGTCGGAG	TCAGCCAGATGCTGAAGAGTC	MD99	13689201	13689917
12D2_CG18319	12D2	CG18319	GCAGCTCAACGAGAAGGTCCAG	GTATTCAATTTACCCACCCGCCCG	MD99	13832097	13832549
12E3_CG32604	12E3-5	CG32604	GACATGACCTGAAGCTGATCG	ATCCACAACCTGGTACAGCGAG	MD99	14056879	14057595
12E5_CG33249	12E5	CG33249	GGAGCTCATCAACTTGGTGTC	GAGAGCAAACTCGTGAAACTG	MD260	14063838	14064533
12E5_CG1839	12E5	CG1839	GCCGACGCTACTGGTCGATTTTAC	TGGTCTTCTTCACTGACGGCGC	MD99	14070563	14071175
13B6_CG6211	13B6	CG6211	TTTATGCAAGTTGGCTCGTG	ATAGCCCGTAAAGTGACTCG	MD57	15150055	15150285
13C4_CG9201	13C4	CG9201	CTGGTCAGGAGCCTGTGCGG	GGCTGCCCTTTAAGTTGTTGAGGC	MD99	15304988	15305818
13C4_CG9203	13C4-5	CG9203	GCGTTTTAGCGAAGATGCTTCC	CATGATCTCCAGGCCACATCTG	wt501 mosaic**	15306959	15307651
13D4_CG8097	13D4	CG8097	GGAAAGCCGGATGAGTCTGG	GTGGGCGACCAAAAGAGATCC	MD106	15417548	15418327
13E6_CG15599	13E6-7	CG15599	CGAACGATACGCCACCGAAAGC	CAAGTTCCCAAGTATGCCACCCG	MD99	15520015	15520781
13E8_CG9245	13E8	CG9245	CCAAAGGTCACGCTGCGGTAC	TCGGGCTCTTTAGGGCGTGCC	wt501**	15537098	15537562
14A6_CG32580	14A6	CG32580	AAGCTCACAGGTAACTGTTTACG	ACGCACCGTTTCATTGTTGGGCTC	MD99	15893918	15894454
15A3_CG32568	15A3	CG32568	TGAGACTGAACATGAGCAACTGCT	AGCATTGAAAAGATTAGCCCGATTGC	MD99	16575512	16576109
15F4_CG5172	15F4	CG5172	TGTTGCTGCTGATCGCTTAACC	CACCTCTCCGTTCTCTGCC	Dsec***	17055458	17055739
18F4_CG11943	18F4	CG11943	CAAGAACCTCTGCCAGTGATC	CGCGGAGTTGTTGCGCGG	MD99	19691293	19692060

MD - Madagascar population of *Drosophila simulans*

* surveyed by Andolfatto (2005)

** wt501 mosaic - Release 1.0 *Drosophila simulans* assembly, <http://genome.wustl.edu/home.cgi>.*** *Drosophila sechellia*, <http://rana.lbl.gov/drosophila/>

Supplement 2. Removing the effect of K_s using partial regression

There is considerable heterogeneity in K_s among the 137 surveyed loci and a significant positive correlation between K_s and K_a (Figure S2.1). If these patterns result from mutation rate variation, such variation would tend to obscure a negative correlation between π_s and K_a expected under a recurrent hitchhiking model. Shown in Figure S2.1 are the observed relationships between π_s , K_a and K_s .

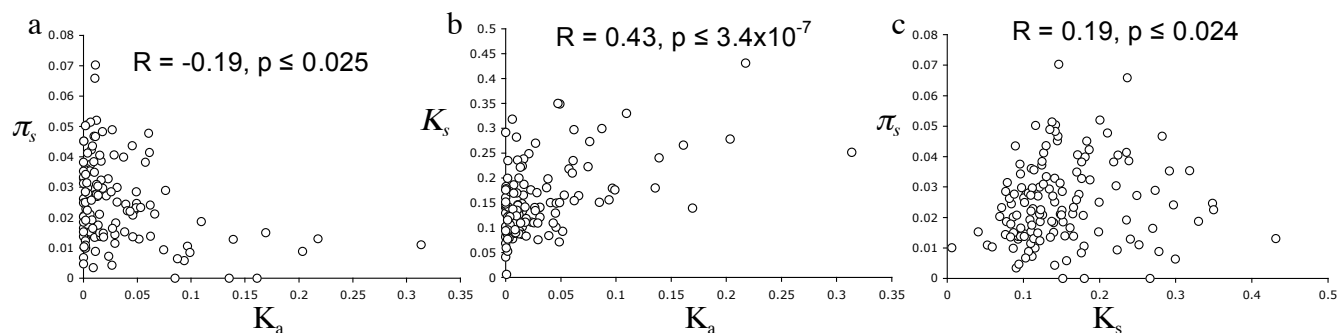


Figure S2.1. Correlations between π_s , K_a and K_s . Correlation coefficients and P -values are given for a non-parametric rank correlation test.

To remove the effect of K_s on the relationship between π_s and K_a , I performed a partial regression using the software JMP v5.1 (www.jmpin.com). K_s was found to have a significant effect on the correlation between π_s and K_a ($P=0.001$). As a result, the strengths of correlations between π_s vs K_a and π_s vs K_s improved significantly in partial correlations (Figure S2.2).

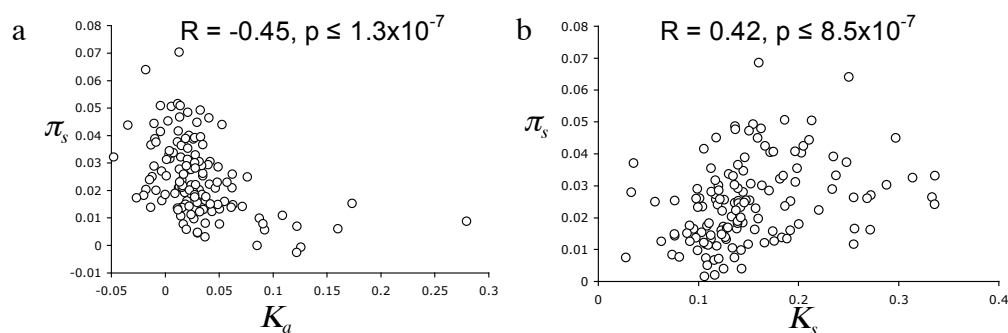


Figure S2.2. Partial correlations. **a)** Residuals of π_s vs K_a controlling for the effect of K_s **b)** Residuals of π_s vs K_s controlling for the effect of K_a . Correlation coefficients and P -values are given for a non-parametric rank correlation test.

Supplement 3. Investigating effects of possible recombination rate variation.

In this study, I surveyed 137 loci from cytological positions 3C3 to 18F4 on the X chromosome of *D. melanogaster*. There is little evidence for variation in recombination rates in this interval (Charlesworth 1996), and was chosen to minimize possible effects of recombination rate variation. Several other methods have used sliding window approaches (Begun et al. 2007) or high-order polynomials (Comeron and Kreitman 2000; Kliman and Hey 1993) to document heterogeneity in recombination rates in the *D. melanogaster* genome. While none of these studies shows that there is indeed significant heterogeneity in recombination rates, I nonetheless tested whether π_s , K_a and K_s correlated with recombination rates inferred using these methods. In Table S3.1, I examine the relationship between π_s and K_a vs. three proposed measures of recombination rate. Results are not particularly consistent across estimators. No significant effect of recombination was found after correcting for multiple tests. I also looked for an effect of recombination rate variation on the correlation between π_s and K_a by partial regression as implemented in JMP5.1 for all three estimates of recombination rate (see Supplementary Materials 2). In these analyses, K_a and K_s and recombination rate were included as model effects, but no significant effects of recombination were found after correcting for multiple tests.

Table S3.1. The effect of recombination rates on levels of synonymous diversity, nonsynonymous divergence in surveyed fragments.

Statistic	Recombination rate estimate	R	Z	P-value
π_s	KH93	0.025	0.29	0.77
	CK00	0.042	0.49	0.63
	B08	0.173	2.02	0.044
K_a	KH93	0.105	1.22	0.22
	CK00	0.057	0.66	0.51
	B08	-0.098	-1.15	0.25

B08 – (Begun et al. 2007); CK00 - (Comeron and Kreitman 2000); KH93 - (Kliman and Hey 1993). P-values are given for a non-parametric rank correlation test.

Supplement 4. Effects of local gene density

Genes are targets for both adaptive and deleterious mutations. It follows that the intensity of genetic hitchhiking and background selection may be expected to correlate with local gene density. To test for effects of gene density, I used the *D. melanogaster* genome annotation (Release 4.7), and quantified gene density as the number of genes within a 20, 50 or 100 kilobase (kb) window surrounding each surveyed locus. I then looked for correlations between gene density and π_s or K_a . Though trends point in the right direction, no significant correlations were found (see Table S4.1). I also looked for an effect of gene density variation on the correlation between π_s and K_a by partial regression as implemented in JMP5.1 (see Supplementary Materials 2). In these analyses, K_a , K_s and local gene density were included as model effects. No significant effects of gene density were found using all three window sizes ($P \geq 0.05$).

Table S4.1. The effect of gene density on levels of synonymous diversity, nonsynonymous divergence in surveyed fragments.

Statistic	Window	R	Z	P-value
π_s	20 kb	-0.061	-0.71	0.48
	50 kb	-0.118	-1.37	0.17
	100 kb	-0.041	-0.48	0.63
K_a	20 kb	0.008	0.09	0.93
	50 kb	0.062	0.72	0.47
	100 kb	0.065	0.76	0.45

Correlation coefficients and P -values are given for a non-parametric rank correlation test.

Supplement 5. Spatial autocorrelation in the data.

In this study, most of the rank correlation tests and the fit of the recurrent hitchhiking model to the data assume that the surveyed loci have independent evolutionary histories. The dense coverage of the X chromosome (~3 surveyed coding regions per centimorgan) raises the issue that this may not be the case. I investigated the possibility of spatial autocorrelation in the data using the *acf* function as implemented in the R statistical package (R Development Core Team 2006). No evidence for significant spatial autocorrelation in the data was found.

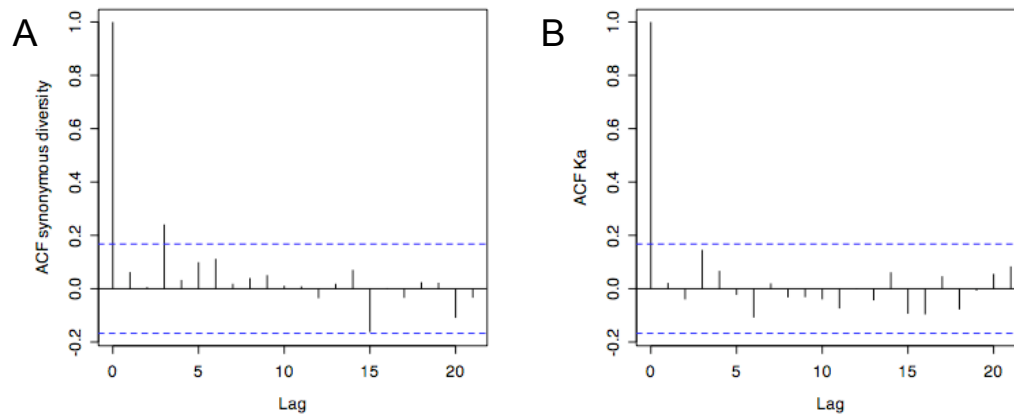


Figure S5.1. Spatial autocorrelation in the data for **A**) synonymous site diversity (π_s) and **B**) nonsynonymous divergence (K_a) per site. The dotted lines indicate approximate 95% confidence intervals under the null (i.e. ACF=0).

Supplement 6. Estimates of the fraction of adaptive divergence.

While my results suggest that levels of synonymous site variability in regions of high recombination are being shaped by recurrent hitchhiking associated with adaptive amino acid substitutions, the strength of selection and the frequency of adaptive substitutions are conflated parameters in the model I fit to the data. To estimate the strength of selection, it is necessary to independently estimate the fraction of amino acid divergence driven to fixation by positive selection. A framework to do this is provided by the McDonald-Kreitman test (Bierne and Eyre-Walker 2004; Fay et al. 2002; McDonald and Kreitman 1991; Smith and Eyre-Walker 2002). The presence of segregating deleterious amino acid mutations will bias estimates of adaptive divergence downward (Andolfatto 2005; Eyre-Walker 2002; Fay et al. 2001; Templeton 1996). In the analysis below, I partly circumvent this problem by excluding low frequency polymorphisms from both the synonymous and nonsynonymous classes. A frequency cutoff of 2/12 maximizes the inferred fraction of adaptive divergence.

Table S6.1. Estimates the fraction of adaptive amino acid divergence using three proposed methods.

Approach	Polymorphisms included	α	95% CI
FWW02	all	0.56	0.43 - 0.66
	f>1/12	0.65	0.56 - 0.73
	f>2/12	0.68	0.59 - 0.76
	f>3/12	0.66	0.57 - 0.74
SEW02	all	0.51	0.39 - 0.61
	f>1/12	0.59	0.45 - 0.71
	f>2/12	0.57	0.46 - 0.68
	f>3/12	0.54	0.36 - 0.68
BE04	all	0.33	0.23 - 0.42
	f>1/12	0.46	0.34 - 0.54
	f>2/12	0.50	0.38 - 0.59
	f>3/12	0.46	0.31 - 0.56

NOTE - FWW02 – (Fay et al. 2002); SEW02 – (Smith and Eyre-Walker 2002); BE04 – (Bierne and Eyre-Walker 2004); f - unpolarised allele frequency; α – fraction of adaptive divergence.

Supplement 7. Codon preference table for *D. melanogaster*.

Previous classification of preferred and unpreferred coding in *D. melanogaster* (Akashi 1995) involved the analysis of relatively few genes. Following Bachtrog's (2007) analysis of *D. pseudoobscura*, I reclassified codon preferences in *D. melanogaster* based on whole genome annotation (see Methods).

Figure S7.1 Inferred codon preferences in *D. melanogaster* based on whole genome analysis.

	T	C	A	G
T	TTT Phe TTC Phe P	TCT Ser TCC Ser P	TAT Tyr TAC Tyr P	TGT Cys TGC Cys P
	TTA Leu TTG Leu	TCA Ser TCG Ser P	TAA TAG	TGA TGG Trp
	CTT Leu CTC Leu P	CCT Pro CCC Pro P	CAT His CAC His P	CGT Arg nP CGC Arg P
	CTA Leu CTG Leu P	CCA Pro CCG Pro P	CAA Gln CAG Gln P	CGA Arg CGG Arg P
C	ATT Ile ATC Ile P	ACT Thr ACC Thr P	AAT Asn AAC Asn P	AGT Ser AGC Ser P
	ATA Ile ATG Met	ACA Thr ACG Thr P	AAA Lys AAG Lys P	AGA Arg AGG Arg
	GTT Val GTC Val P	GCT Ala GCC Ala P	GAT Asp GAC Asp P	GGT Gly GGC Gly P
	GTA Val GTG Val P	GCA Ala GCG Ala P	GAA Glu GAG Glu P	GGA Gly GGG Gly nP

NOTE - "Preferred" codons (P) were identified as those whose frequencies (within a synonymous family) show positive correlations with the degree of bias as quantified by the scaled X^2 measure (Akashi 1995). All other codons show significant negative correlations and are classified as "unpreferred", except Arg:CGT and Gly:GGG, which are labeled as nP (not Preferred).

References.

- Andolfatto, P., 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*. 437: 1149-1153.
- Andolfatto, P., and M. Przeworski. 2001. Regions of Lower Crossing Over Harbor More Rare Variants in African Populations of *Drosophila melanogaster*. *Genetics* **158**: 657-665.
- Bachtrog, D., 2007, Reduced selection for codon usage bias in *Drosophila miranda*. *J Mol Evol*. 64:586-90.
- Begun, D. J., Holloway, A.K., and 11 others. 2007. Population genomics: Whole genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*.
- Bierne, N. and A. Eyre-Walker. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol Biol Evol* **21**: 1350 - 1360.
- Charlesworth, B. 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet. Res.* **68**: 131-149.
- Cameron, J. and M. Kreitman. 2000. The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. *Genetics* **156**: 1175 - 1190.
- Eyre-Walker, A. 2002. Changing Effective Population Size and the McDonald-Kreitman Test. *Genetics* **162**: 2017-2024.
- Fay, J.C., G.J. Wyckoff, and C.I. Wu. 2001. Positive and negative selection on the human genome. *Genetics* **158**: 1227-1234.
- Fay, J.C., G.J. Wyckoff, and C.I. Wu. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**: 1024-1026.
- Kliman, R. and J. Hey. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol* **10**: 1239 - 1258.
- McDonald, J.H. and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652-654.
- Smith, N.G. and A. Eyre-Walker. 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022-1024.
- R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Templeton, A. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* **144**: 1263 - 1270.