

Estimating selection on non-synonymous mutations

Laurence Loewe, Brian Charlesworth, Carolina Bartolomé and Véronique Noël

Genetics (Accepted Nov 2005)

Supplementary Online Material

Genes used in the study:

D. miranda: *ade3*¹, *Adh*², *amd*¹, *bcd*¹, *Bruce*¹, *Ddc*¹, *Eno*¹, *Gapdh2*¹, *Gld*¹, *hyd*¹,
*Lam*¹, *rosy*¹, *scute*², *sesB*², *sisA*², *sry-alpha*², *Uro*¹

1. Bartolomé, C., X. Maside, S. Yi, A. L. Grant and B. Charlesworth, 2005 Patterns of selection on synonymous and non-synonymous variants in *Drosophila miranda*. *Genetics* 169: 1495-1507.
2. Yi, S., D. Bachtrog and B. Charlesworth, 2003 A survey of chromosomal and nucleotide sequence variation in *Drosophila miranda*. *Genetics* 164: 1369-1381.

D. pseudoobscura: *Adh*^{1,2}, *Amy1*³, *bcd*⁴, *EcR*⁵, *Est-5A*⁶, *Est-5B*⁷, *Est-5C*⁶, *eve*⁵, *exu-1*^{5,12}, *Hsp82*⁸, *per*⁹, *rh1*⁴, *run*¹⁰, *rosy*¹¹ (Current names for the genes *Hsp82* and *rh1* are *Hsp83* and *ninaE*, respectively.)

1. Schaeffer, S. W., and E. L. Miller, 1992 Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Genetics* 132: 163-178.
2. Schaeffer, S. W., 2002 Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*. *Genet. Res.* 80: 163-175.

3. Popadic, A., and W. W. Anderson, 1994 The history of a genetic system. Proc. Natl. Acad. Sci. USA 91: 6819-6823.
4. Machado, C. A., R. M. Kliman, J. A. Markert and J. Hey, 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. 19:472-88.
5. Schaeffer, S.W., M. P. Goetting-Minesky, M. Kovacevic, J. R. Peoples, J. L. Graybill, J. M. Miller, K. Kim, J. G. Nelson, W. W. Anderson, 2003 Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. Proc. Natl. Acad. Sci. USA 100: 8319-8324.
6. King, L. M., The role of gene conversion in determining sequence variation and divergence in the *Est-5* gene family in *Drosophila pseudoobscura*. Genetics 148: 305-315.
7. Veuille, M, and L. M. King, 1995 Molecular basis of polymorphism at the esterase-5B locus in *Drosophila pseudoobscura*. Genetics 141: 255-262.
8. Wang R. L., J. R. Wakeley and J. Hey, 1997 Genetics 147: 1091-1161.
9. Wang R. L., and J. Hey, 1996 The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the *period* locus. Genetics 144: 1113-1126.
10. Kovacevic, M., S.W. Schaeffer, 2000 Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. Genetics 156: 155-172.
11. Begun, D. J., and P. Whitley, 2002 Molecular population genetics of *Xdh* and the evolution of base composition in *Drosophila*. Genetics 162: 1725-1735.
12. Luk, S. K., Kilpatrick, M., Kerr, K. & Macdonald, P. M., 1994 Components acting in localization of *bicoid* mRNA are conserved among *Drosophila* species. Genetics 137: 521-30.

Supplementary Table 1: Summary of the data for *D. pseudoobscura*.

Columns denote:

A: Gene name, total number of sequences analyzed, number of codons, total number of synonymous base pairs, number of polymorphic synonymous base pairs, total number of non-synonymous base pairs, number of polymorphic non-synonymous base pairs, total number of silent sites, number of polymorphic silent sites, chromosome where the gene is located.

B: Average pairwise diversities (P_i) at synonymous sites (S), non-synonymous sites (A) and silent sites (s) are given together with the ratios of uncorrected $P_i(A)/P_i(S)$.

C: WATTERSON'S estimator θ_w (Theta). Rest, see **B**.

StDev gives the standard deviation and StErrMean the standard error of the arithmetic mean.

A	No of sqces	No codons analyzed	No of S sites analyzed	N(S)	No of A sites analyzed	N(A)	No of s sites analysed	N(s)	Chr
Adh/Adh-r	139	529	385	132	1202	16	1948	439	4
Amy1	7	494	350	24	1132	9	825	52	3
bcd	21	379	278	33	859	12	475	45	2
EcR	103	87	63	10	198	0	132	21	3
Est-5A	8	546	390	24	1248	19	1283	49	2
Est-5B	17	545	385	69	1250	34	909	88	2
Est-5C	8	545	382	33	1253	13	1066	53	2
eve	101	101	76	14	227	1	147	24	3
exu-1	106	77	52	3	179	0	176	14	3
Hsp82	12	240	158	6	565	1	1305	34	X
per	16	314	228	26	735	7	548	53	XL
rh1	18	312	220	16	716	2	725	85	2
run	40	41	29	5	94	0	316	48	XL
rosy	9	520	390	45	1170	16	390	45	2

B	Pi(S)	Pi(A)	Pi(A)/Pi(S)	Pi (s)
Adh/Adh-r	0.0269	0.0013	0.0480	0.0181
Amy1	0.0269	0.0033	0.1218	0.0261
bcd	0.0246	0.0024	0.0974	0.0186
EcR	0.0167	0.0000	0.0000	0.0173
Est-5A	0.0217	0.0043	0.1983	0.0144
Est-5B	0.0401	0.0068	0.1687	0.0211
Est-5C	0.0318	0.0037	0.1157	0.0180
eve	0.0249	0.0001	0.0036	0.0236
exu-1	0.0025	0.0000	0.0000	0.0074
Hsp82	0.0079	0.0005	0.0685	0.0059
per	0.0277	0.0021	0.0754	0.0227
rh1	0.0149	0.0003	0.0208	0.0189
run	0.0199	0.0000	0.0000	0.0168
rosy	0.0338	0.0041	0.1200	0.0338
Mean	0.0229	0.0021	0.0742	0.0188
Variance	0.000101	0.000004	0.004310	0.000050
StDev	0.0100	0.0021	0.0656	0.0070
StErrMean	0.0027	0.0006	0.0175	0.0019

C	Theta(S)	Theta(A)	Theta(A)/Theta(S)	Theta(s)
Adh/Adh-r	0.0622	0.0024	0.0386	0.0409
Amy1	0.0280	0.0033	0.1162	0.0257
bcd	0.0330	0.0039	0.1176	0.0263
EcR	0.0303	0.0000	0.0000	0.0305
Est-5A	0.0237	0.0059	0.2474	0.0147
Est-5B	0.0530	0.0081	0.1519	0.0286
Est-5C	0.0333	0.0040	0.1200	0.0192
eve	0.0357	0.0009	0.0238	0.0315
exu-1	0.0111	0.0000	0.0000	0.0152
Hsp82	0.0126	0.0006	0.0468	0.0086
per	0.0309	0.0027	0.0868	0.0291
rh1	0.0211	0.0008	0.0384	0.0341
run	0.0399	0.0000	0.0000	0.0357
rosy	0.0425	0.0050	0.1185	0.0425
Mean	0.0327	0.0027	0.0790	0.0273
Variance	0.000197	0.000006	0.005071	0.000099
StDev	0.0140	0.0025	0.0712	0.0099
StErrMean	0.0037	0.0007	0.0190	0.0027