5. Charlesworth, B. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63,** 213–227 (1994).
6. Fay, J., Wycoff, G. J. & Wu, C.-I. Positive and negative selection on the human genome. *Genetics* **158,** 1227–1234 (2001).
7. McDonald, J. H. & Kreitman, M. Adaptive evolution at the *Adh* locus in *Drosophila*. *Nature* **351,** 652–654 (1991).
8. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134,** 1289–1303 (1993).
9. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23,** 23–35 (1974).
10. Begun, D. J. & Aquadro, C. F. levels of naturally occuring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356,** 519–520 (1992).
11. Begun, D. The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol. Biol. Evol.* **18,** 1343–1352 (2001).
12. Kliman, R. Recent selection on synonymous codon usage in *Drosophila*. *J. Mol. Evol.* **49,** 343–351 (1999).
13. Adams, M. D. *et al*. The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195 (2000).
14. Powell, J. R. & DeSalle, R. Drosophila molecular phylogenies and their uses. *Evol. Biol.* **28,** 87–138 (1995).
15. Haldane, J. B. S. The cost of natural selection. *J. Genet.* **55,** 511–524 (1957).
16. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217,** 624–626 (1968).
17. Thompson, J. D., Higgins, D. G. & Gibson, T. J. ClustalW—improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22,** 4673–4680 (1994).
18. Xia, X. *Data Analysis in Molecular Biology and Evolution* (Kluwer Academic, London, 2000).
19. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15,** 174–175 (1999).
20. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13,** 555–556 (1997).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We thank B. Charlesworth, C.-I. Wu, S. Otto, M. Whitlock, T. Johnson, P. Awadalla, J. Gillespie, G. McVean and P. Keightley for helpful discussions, and E. Moriyama for help with data collection. N.G.C.S. was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and A.E.-W. is funded by the Royal Society and the BBSRC.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.E.-W. (e-mail: a.c.eyre-walker@sussex.ac.uk).

Testing the neutral theory of molecular evolution with genomic data from *Drosophila*

Justin C. Fay*†, Gerald J. Wyckoff*† & Chung-I Wu*‡

** Committee on Genetics, University of Chicago, Chicago, Illinois 60637, USA*
‡ Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA

Although positive selection has been detected in many genes, its overall contribution to protein evolution is debatable[1]. If the bulk of molecular evolution is neutral, then the ratio of amino-acid (*A*) to synonymous (*S*) polymorphism should, on average, equal that of divergence[2]. A comparison of the *A*/*S* ratio of polymorphism in *Drosophila melanogaster* with that of divergence from *Drosophila simulans* shows that the *A*/*S* ratio of divergence is twice as high—a difference that is often attributed to positive selection. But an increase in selective constraint owing to an increase in effective population size could also explain this observation, and, if so, all genes should be affected similarly. Here we show that the difference between polymorphism and divergence is limited to only a fraction of the genes, which are also evolving more rapidly, and this implies that positive selection is responsible. A higher *A*/*S* ratio of divergence than of polymorphism is also observed in other species, which suggests a rate of adaptive evolution that is far higher than permitted by the neutral theory of molecular evolution.

The neutral theory holds that the bulk of DNA divergence between species is driven by mutation and drift, rather than by positive darwinian selection[3]. But because the effect of positive selection is often masked by negative selection[4], detecting positive selection is a challenging task. A rate of amino-acid substitution greater than that of synonymous substitution can be explained only by positive selection[5], but such a criterion is very stringent as negative selection lowers the rate of amino-acid substitution. A high rate of amino-acid substitution is limited mostly to genes that are involved in resistance to disease or in sexual reproduction, where there is continual room for improvement[6,7].

The McDonald–Kreitman test can detect positive selection even in the presence of negative selection through a ratio of amino-acid divergence to synonymous divergence greater than that of polymorphism[2]. The *A*/*S* ratio of divergence is inflated above polymorphism by advantageous amino-acid mutations, which quickly sweep through a population but have a cumulative effect on divergence. The McDonald–Kreitman test has been applied to many genes individually, but only a few have yielded a significant excess of amino-acid divergence (*Drosophila* genes are reviewed in refs 8, 9). This may in part be caused by a lack of power in detecting positive selection in individual genes unless a large number of adaptive substitutions have occurred.

For those genes that have yielded a significant McDonald–Kreitman test result, the *A*/*S* ratio of divergence is more than twice as great as polymorphism[10–12]. The effects of positive selection may also be obscured by slightly deleterious amino-acid mutations that inflate the *A*/*S* ratio of polymorphism but not divergence. The effects of slightly deleterious mutations can be removed by comparing common polymorphism with divergence, because deleterious amino-acid mutations are kept at low frequency in the population[4]. This can only be done when the data from a large number of genes are combined; individual genes rarely contain more than a few common amino-acid polymorphisms.

An important but rarely appreciated assumption of the McDonald–Kreitman test is that the selective constraint on a gene remains constant over time. The selective constraint on a gene is determined by the proportion of amino-acid mutations that are deleterious[3], $2Ns < -1$, so both a change in the selection coefficient (s) and a change in effective population size (N) can result in a change in selective constraint. Although it is well known that selective constraint is not static across phylogenetic lineages[13,14], this assumption is rarely justified in applications of the McDonald–Kreitman test. Whereas the strength of selection on each gene might fluctuate over time depending on the genetic or environmental background, a genome-wide change in constraint, such as that caused by a change in effective population size, should produce a consistent increase or decrease in the *A*/*S* ratio across all genes. Alternatively, under positive selection each gene might be affected to a different degree and some genes might not be affected at all.

To compare genomic patterns of amino-acid and synonymous

Table 1 Polymorphisms in *D. melanogaster* and divergence from *D. simulans*

Gene*	Class	Amino-acid polymorphism, *A*	Synonymous polymorphism, *S*	*A*/*S*
X-linked	Rare (≤12.5%)	4	67	0.06
	Common (>12.5%)	6	46	0.13
	Divergence	42	189	0.22
Autosomal	Rare	79	126	0.63
	Common	44	118	0.37
	Divergence	421	521	0.81

* There are 5 X-linked and 31 autosomal genes with a sample size of eight or greater (see text for the data from all 45 genes).

† Present addresses: Department of Genome Sciences, Lawrence Berkeley National Laboratory, Berkeley, California 94720 (J.C.F.); Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA (G.J.W).

Table 2 African and non-African common polymorphism and divergence

Class	Population	Amino-acid polymorphism, *A*	Synonymous polymorphism, *S*	*A/S*
Polymorphism	Non-African	48	124	0.39
	African	40	159	0.25
Divergence		413	663	0.62

site evolution, we tabulated polymorphism in *D. melanogaster* and divergence from *D. simulans* from 45 gene surveys (Methods). If all amino-acid and synonymous variation is neutral, then the *A*/*S* ratio of polymorphism and divergence should be constant. The *A*/*S* ratio of divergence (598/950 = 0.63) is significantly greater than that of common polymorphism (65/224 = 0.29; $P < 10^{-6}$). We compared divergence with the common rather than the total polymorphism because deleterious mutations at low frequency inflate the *A*/*S* ratio of polymorphism. For the 36 genes with sample sizes of eight or greater, there is a significant excess of rare over common amino-acid variation in autosomal genes ($P = 0.022$; Table 1), as is observed in humans[4]. The absence of a difference in X-linked genes suggests that the deleterious mutations are partially recessive and are more readily eliminated from the X chromosome.

Both positive selection and an increase in selective constraint on amino-acid changes can produce a higher *A*/*S* ratio of divergence than of polymorphism. But only under certain restrictive conditions is a genome-wide change in constraint possible. One such condition is an increase in effective population size that is neither too distant nor too recent in the evolutionary past. If this possibility can be ruled out, positive selection may be the only viable explanation for the high rate of amino-acid divergence.

If an increase in selective constraint resulted from a population size increase associated with the spread of *D. melanogaster* outside Africa[15], it might be more appropriate to compare the *A*/*S* ratio of the African population with that of divergence. Table 2, which includes the 32 genes for which both African and non-African populations were surveyed, shows that there is a significantly larger *A*/*S* ratio of divergence than of polymorphism in either population. If a recent increase in effective population size increased constraint on amino-acid polymorphism in both African and non-African populations, then patterns of synonymous polymorphism might be skewed towards rare variants. Neither African or non-African populations show this pattern[16]. Finally, if there has been a decrease in effective population size along the *D. melanogaster* lineage[17,18], the *A*/*S* ratio of polymorphism should be greater than that of divergence between the two species.

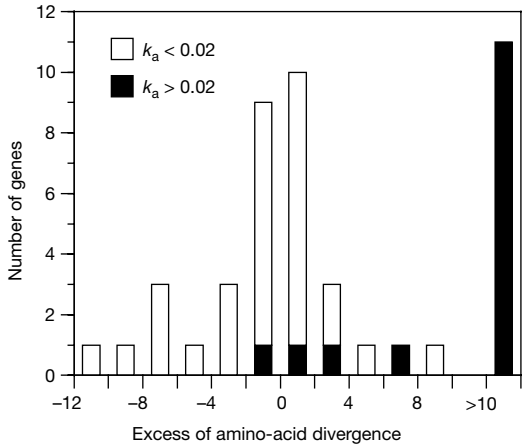

Figure 1 The distribution of the excess of amino-acid divergence contributed by each gene. For reference, fast and slowly evolving genes are denoted by a rate of amino-acid substitution (k_a) greater than (filled bars) or less than (open bars) 2%.

Table 3 Polymorphism and divergence in neutral and fast genes

Genes*	Class	Amino-acid polymorphism, *A*	Synonymous polymorphism, *S*	*A/S*
Neutral	Rare	31	90	0.34
	Common	16	69	0.23
	Divergence	65	247	0.26
Fast	Rare	48	36	1.33
	Common	28	49	0.57
	Divergence	356	274	1.30

*X-linked genes are excluded.

If an increase in effective population size has produced a genome-wide increase in selective constraint, the *A*/*S* ratio of all genes should be affected. In Fig. 1, the distribution of each gene's contribution to the excess of amino-acid divergence suggests that there are two classes of gene: neutral and rapidly evolving. The neutral class comprises 34 genes that deviate by less than 10 amino-acid substitutions from that expected on the basis of the *A*/*S* ratio of all common polymorphism. The remaining 11 genes all have a higher *A*/*S* ratio of divergence than of polymorphism, and account for the whole difference in the *A*/*S* ratio of polymorphism and divergence. These genes are *Acp26Aa*, *Acp29Ab*, *anon1A3*, *anon1E9*, *anon1G5*, *ci*, *est-6*, *Ref2P*, *Rel*, *tra* and *Zw*. As expected under positive selection, which increases the rate of protein evolution, these 11 genes have a high rate of amino-acid substitution (Fig. 1).

Can the pattern in Fig. 1 be explained by selection or demography? Table 3 shows that, in the rapidly evolving genes, the *A*/*S* ratios of divergence and of rare polymorphism are much higher than the *A*/*S* ratio of the common polymorphism. This is expected if the genes are under positive selection. Although a large increase in population size in the recent past could account for the difference between the *A*/*S* ratio of divergence and that of common polymorphism, this explanation is incompatible with the very small difference found in the 26 neutral genes. Because both the neutral and rapidly evolving genes have a higher *A*/*S* ratio of rare polymorphism than of common polymorphism, both should have been affected by an increase in effective population size.

If positive selection is common, other species should also have an *A*/*S* ratio of divergence greater than that of polymorphism. In addition, any demographic scheme is not likely to be shared by several species. In a study of eight genes in *D. simulans*, *Drosophila mauritiana* and *Drosophila sechellia*, the *A*/*S* ratio of polymorphism (*A*/*S* = 32/183) is 34% that of divergence (28/55)[19]. In a study of 42 genes with polymorphism in both *D. melanogaster* and *D. simulans*, the *A*/*S* ratio of polymorphism is 65% that of divergence (N. G. C. Smith and A. Eyre-Walker, personal communication). In another study of 23 genes, the *A*/*S* ratio of polymorphism (45/305) is 30% that of divergence along the *D. simulans* lineage (65/133)[20]. In humans, the *A*/*S* ratio of common polymorphism (70/122) found in 181 genes is 65% that of divergence (3,660/4,151) found in a different set of 182 human and Old World monkey genes[4].

Although these genomic patterns of variation are not explained easily by the neutral theory, slightly deleterious mutations must clearly be accounted for in attempting to measure positive selection. In humans, 38% of amino-acid polymorphism was estimated to be slightly deleterious[4], and in *D. melanogaster* the estimate is 26%, $(0.63 - 0.37) \times 126/123$, from the combined neutral and rapidly evolving genes (Table 3). These slightly deleterious mutations, which are emphasized by the nearly neutral theory[21], could become effectively neutral and fixed during a population bottleneck of sufficient severity, providing a burst of amino-acid substitutions and an increase in the *A*/*S* ratio of divergence. We control for the impact of these slightly deleterious mutations by comparing the rapidly evolving class of gene to the neutral class (Fig. 1, Table 3). Additional genomic data from other species will be needed to estimate the general impact of these slightly deleterious mutations on protein evolution. □

Methods

Data

A literature search yielded 45 genes for which polymorphism had been surveyed in *D. melanogaster* and for which an outgroup sequence was available. Of these, 36 had a sample size of eight or greater, 32 had been surveyed in at least two African and two non-African individuals and 10 were of X-linked genes. The 45 genes and their references are listed in Supplementary Information.

Analysis

Polymorphism data was tabulated by hand or from GenBank accession numbers using SITES[21] or DNASP[22]. For each polymorphic site, the minor allele was classified as rare ($\leq 12.5\%$) or common ($> 12.5\%$). The cutoff of 12.5% was chosen to exclude deleterious mutations from the common frequency class and to include those genes with samples of eight or more in the analysis of rare compared to common polymorphism. Cutoffs of 10 and 15% produce similar results. We treated three alleles segregating at a single nucleotide as two segregating sites and excluded complex variations. Divergence data was obtained by comparing a randomly chosen sequence of *D. melanogaster* with that of *D. simulans* or, if unavailable, either *D. mauritiana* or *D. sechellia*. The number of amino-acid and synonymous substitutions between species was estimated using Kimura's two-parameter model to correct for multiple hits.

The contribution of each gene to the excess number of amino-acid substitutions was calculated as the excess number of amino-acid substitutions minus the excess number of amino-acid polymorphisms found in each gene. The excess for polymorphism and divergence is $A - S \times (65/224)$, where A and S are the number of amino-acid and synonymous substitutions, respectively, and 65/224 is the total number of amino-acid polymorphisms divided by synonymous polymorphisms. (Ideally, the excess of amino-acid divergence in each gene should be calculated using only polymorphism and divergence in that gene but there is rarely sufficient polymorphism in a single gene for comparison with divergence.) We also calculated the contribution to the excess separately for three groups of genes sorted by their rate of amino-acid divergence. The two methods produced a similar distribution so the simpler method using a single group of genes was used.

Received 27 June; accepted 4 December 2001.

1. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
2. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351,** 652–654 (1991).
3. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
4. Fay, J. C., Wyckoff, G. J. & Wu, C.-I. Positive and negative selection on the human genome. *Genetics* **158,** 1227–1234 (2001).
5. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267,** 275–276 (1977).
6. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15,** 496–503 (2000).
7. Wyckoff, G. J., Wang, W. & Wu, C.-I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403,** 304–309 (2000).
8. Weinreich, D. M. & Rand, D. M. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* **156,** 385–399 (2000).
9. Moriyama, E. N. & Powell, J. R. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13,** 261–277 (1996).
10. Eanes, W. F., Kirchner, M. & Yoon, J. Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. *Proc. Natl Acad. Sci. USA* **90,** 7475–7479 (1993).
11. Begun, D. J. & Whitley, P. Adaptive evolution of *relish*, a *Drosophila* NF-κB/IκB protein. *Genetics* **154,** 1231–1238 (2000).
12. Tsaur, S. C., Ting, C. T. & Wu, C. I. Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*: II. Divergence versus polymorphism. *Mol. Biol. Evol.* **15,** 1040–1046 (1998).
13. Langley, C. H. & Fitch, W. M. An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3,** 161–177 (1974).
14. Ohta, T. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40,** 56–63 (1995).
15. Lachaise, D. M., Cariou, M.-L., David, J. R., Lemeunier, F. & Tsacas, L. The origin and dispersal of the *Drosophila melanogaster* subgroup: a speculative paleogeographic essay. *Evol. Biol.* **22,** 159–225 (1988).
16. Andolfatto, P. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18,** 279–290 (2001).
17. Akashi, H. Codon bias evolution in *Drosophila*: Population genetics of mutation-selection drift. *Gene* **205,** 269–278 (1997).
18. McVean, G. A., Vieira, J. Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics* **157,** 245–257 (2001).
19. Kliman, R. M. *et al.* The population genetics of the origin and divergence of the *Drosophila* simulans complex species. *Genetics* **156,** 1913–1931 (2000).
20. Begun, D. J. The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol. Biol. Evol.* **18,** 1343–1352 (2001).
21. Ohta, T. Slightly deleterious mutant substitutions during evolution. *Nature* **246,** 96–98 (1973).
22. Hey, J. & Wakeley, J. A coalescent estimator of the population recombination rate. *Genetics* **145,** 833–846 (1997).
23. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15,** 174–175 (1999).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

This work was supported by grants from the NIH and NSF to C.-I.W. and a Genetics Training Grant and a Department of Education PhD fellowship to J.C.F.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C.F. (e-mail: jcfay@lbl.gov).

Brain potential and functional MRI evidence for how to handle two languages with one brain

Antoni Rodriguez-Fornells*, Michael Rotte†, Hans-Jochen Heinze†, Tömme Nösselt† & Thomas F. Münte*

** Department of Neuropsychology, Otto von Guericke University, Universitätsplatz 2, Gebäude 24, 39106 Magdeburg, Germany*
† Klinik für Neurologie 2, Otto von Guericke University, Leipzigerstrasse 44, 39120 Magdeburg, Germany

Bilingual individuals need effective mechanisms to prevent interference from one language while processing material in the other[1]. Here we show, using event-related brain potentials and functional magnetic resonance imaging (fMRI), that words from the non-target language are rejected at an early stage before semantic analysis in bilinguals. Bilingual Spanish/Catalan and monolingual Spanish subjects were instructed to press a button when presented with words in one language, while ignoring words in the other language and pseudowords. The brain potentials of bilingual subjects in response to words of the non-target language were not sensitive to word frequency, indicating that the meaning of non-target words was not accessed in bilinguals. The fMRI activation patterns of bilinguals included a number of areas previously implicated in phonological and pseudoword processing[2–5], suggesting that bilinguals use an indirect phonological access route to the lexicon of the target language to avoid interference[6].

High-proficiency bilingual subjects manage to understand and speak one of their languages without apparent interference from the other. This is a remarkable ability in the face of the fact that neuroimaging studies have revealed, at least for high-proficiency bilinguals, that neuro-anatomical representations of both languages are

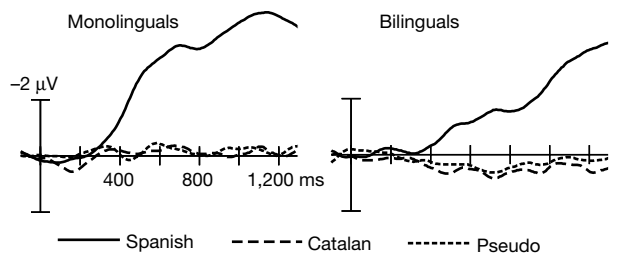

Figure 1 Lateralized readiness potentials (LRPs) from the main experiment indicating the preparation of motor responses. The onset latency of the LRP to Spanish words, estimated by the time at which the amplitude was significantly different from zero for at least 4 consecutive time points (sequential *t*-tests)[14], was 408 ms in the monolingual and 520 ms in the bilingual group. No LRP activity is observed for Catalan words, indicating an effective blocking of 'word' (go) responses in the bilingual group.