Deleterious Mutations and the Evolution of Sex

Peter D. Keightley[1] and Adam Eyre-Walker[2]

It has been suggested that sexual reproduction is maintained because it reduces the load imposed by recurrent deleterious mutations. If rates of deleterious mutation per diploid genome per generation (U) exceed 1, and mutations interact synergistically, then sexuals can overcome their inherent twofold disadvantage. We have tested this hypothesis by estimating genomic point mutation rates for protein-coding genes in a range of animal taxa. We find a positive linear relationship between U and generation time. In species with short generation times, U is predicted to be far below 1, suggesting that sex is not maintained by its capacity to purge the genome of deleterious mutations.

Almost all species indulge in at least a little sex, the exchange of genetic material between individuals. In higher plants and animals, this generally involves anisogamy, the production of gametes of different sizes, and in its most derived form, the production of two sexes. Anisogamy incurs a cost, since resources directed towards male gametes, or male offspring, could be directed towards female gametes, or female offspring. For example, a parthenogenetic female who only produces female offspring will have a twofold advantage over her sexual conspecifics. It is this cost of sexual reproduction which has so troubled evolutionary biologists; why do so many species, what we will call "obligate sexuals," produce two different types of gametes, or different sexes, every generation?

Sexual populations can have two principal advantages over asexuals: they can adapt more rapidly to changing environments and they are less prone to the accumulation of deleterious mutations (*1*). Sexuals accumulate fewer deleterious mutations because they avoid Muller's ratchet (*2*), but the ratchet is unlikely to maintain obligate sexuality, since it operates very slowly in large populations and can be avoided with very little sex. However, much attention has been focused on the potential benefits of sex in reducing the mutation load if mutations interact synergistically (*3*), for three reasons. First, the effect is independent of population size, a property which has led to the theory being called the "the mutational deterministic" (MD) hypothesis (*1*). Second, the benefits of sex can be large; if U exceeds 1, then sexual populations can overcome their twofold disadvantage over pure asexuals. This is a minimum condition, since U needs to be rather greater than 1 if stochastic factors are taken into account (*4, 5*), or if obligate sexuals are to have an advantage over facultative sexuals. And third, the hypothesis is testable; if many obligate sexual populations have deleterious mutation rates below 1, then obligate sex is unlikely to be maintained by selection against deleterious mutations. Here, we estimate deleterious mutation rates for a variety of organisms in an attempt to test the MD hypothesis.

Laboratory mutation accumulation experiments provide estimates for rates of mutations with moderate phenotypic effects, but such experiments cannot be used to test the MD hypothesis, since mutations of small effect are missed (*6, 7*). A radically different approach is to compare the genomes of related species (*8*). In a neutral segment of DNA, the rate of nucleotide substitution is expected to equal the mutation rate. The substitution rate in randomly sequenced stretches of DNA is expected to be lower than the neutral rate, due to selective constraint on functional elements. The difference between the substitution rates in truly neutral and randomly sequenced segments therefore could estimate the rate of selective elimination of mutations from the randomly sequenced segments (*8*). Here, we apply a simplified approach that restricts the analysis to protein-coding gene sequences (*9*). Under the assumption that synonymous mutations are neutral, the nucleotide mutation rate can be estimated from the synonymous substitution rate (K_s). The amino acid mutation rate can then be estimated from the product of K_s and the number of sites in the gene that, if changed, lead to an amino acid substitution. The number of selectively eliminated amino acid mutations is the difference between the estimated amino acid mutation rate and the observed number of amino acid substitutions. The estimate can be scaled up to the whole genome, if the number and average length of protein-coding genes are known, and expressed per generation, if estimates are available for the evolutionary divergence time and generation interval.

We obtained estimates for the genomic amino acid mutation rate (M) and genomic deleterious mutation rate (U) for six pairs of vertebrate species and three pairs of insects for which divergence dates can be inferred from the fossil record, biogeographical data, or a locally calibrated molecular clock (*10*). We corrected for the effect of selection on synonymous codon use in *Drosophila* by regressing K_s against codon bias (*11*). For *D. melanogaster/D. simulans*, the regression coefficient is nonsignificant, and this procedure has little effect on our estimate of U; for *D. picticornis/D. silvestris*, there are insufficient data, but for *D. melanogaster/D. pseudoobscura*, the regression is significant ($P < 0.05$), and the corrected estimate is approximately 1.5-fold greater as a consequence. Other evidence (*12*) suggests that selection on synonymous codon bias is sufficiently weak in *Drosophila*, at least, to have little effect on the synonymous substitution rate.

Surprisingly, the estimates of mutation rates for the protein-coding genome per year are approximately constant across a broad range of taxa (Table 1). With the exception of mouse/rat, for which there is debate about the divergence time (*13, 14*), there is just over threefold variation in estimates of M per year (approximately sixfold in U per year). This does not just reflect the molecular clock, because the total length of the protein-coding sequence in insects is considerably smaller than that in vertebrates. Instead, it seems to reflect an inverse proportionality between the nucleotide mutation rate per cell division and the total length of protein-coding sequences in the genome (*15*). However, the estimates of M and U per generation vary by ~90-fold and ~50-fold, respectively (Table 1). Rates for M and U within mammals also vary substantially by ~11-fold and ~6-fold, respectively. The amino acid mutation rates agree with independent estimates of M obtained from the rate of appearance of new electrophoretic mutations [2.2 and 0.09 mutations per genome per generation in humans and *Drosophila*, respectively (*6*)].

The constancy of the mutation rate per year leads to a highly significant positive correlation ($P < 0.001$) between both M and U per generation and generation time (Table 1 and Fig. 1); the relationship appears to be linear on both an untransformed and a log-log scale. This parallels a positive relationship between new mutational variance for quantitative traits per generation (expected to be linear in M) and generation time (*7*). The relationships between U and M and generation time (Fig. 1) are insensitive to generation time estimates, because U and M are expressed per generation. In mammals with long generation times, and particularly primates, U is sufficiently high to maintain obligate sexuality, if they were able to become

[1]Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. [2]Centre for the Study of Evolution and School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK.

E-mail: p.keightley@ed.ac.uk (P.D.K.) and a.c.eyre-walker@sussex.ac.uk (A.E.-W.)

Table 1. Estimates of K_s, the synonymous divergence; M' and U', genomic amino acid and deleterious mutation rates, respectively, per year; and M and U, genomic amino acid and deleterious mutation rates, respectively, per generation. We assume 80,000 genes, averaging 1500 bp in length, for vertebrates (*34*) and 13,600 genes, averaging 1770 bp in length, for insects (*20*).

Species	Number of genes	Divergence time (My) (reference)	Generation time (years)*	K_s (SE)	M'	U'	M	U
Human/chimpanzee	63	6 (*30*)	25	0.0179 (0.0021)	0.25	0.12	6.3	3.0
Macaque/New World monkey	28	40 (*31*)	11	0.147 (0.014)	0.31	0.18	3.4	1.9
Sheep/cow	50	25 (*31*)	6	0.0677 (0.0045)	0.23	0.15	1.4	0.90
Dog/cat	35	38 (*31*)	4	0.240 (0.017)	0.54	0.40	2.2	1.6
Chicken/old world quail	28	34†	2	0.120 (0.013)	0.30	0.25	0.60	0.49
Mouse/rat	423	13 (*13*)	0.5	0.179 (0.003)	1.2	0.99	0.59	0.50
D. picticornis/D. silvestris‡	4	3.7 (*32*)	0.2	0.101 (0.029)	0.47	0.35	0.095	0.071
D. melanogaster/D. simulans	42	2.5 (*33*)	0.1	0.100 (0.008)	0.70	0.58	0.070	0.058
D. melanogaster/D. pseudoobscura	29	35 (*33*)	0.1	0.937 (0.072)	0.75	0.70	0.075	0.070

*See suplementary data at *Science* Online (*35*). †Assumes that the divergence of *Anseriformes-Galliformes* occurred in the late Cretaceous [~75 My ago (*31*)], and uses this to calibrate a local molecular clock, assuming a linear relationship between DNA melting point (*36*) and date of divergence. ‡*Drosophila picticornis/D. silvestris* are Hawaiian *Drosophila*. For *D. melanogaster/D. pseudoobscura*, the values of M and U are corrected for variation among genes in codon bias [see (*11*)].

asexual, but genomic imprinting probably prevents this. However, many obligate sexuals have generation times on the order of 1 month or less [e.g., many insects (*16*), copepods (*17*), and nematodes (*18*)], implying deleterious amino acid mutation rates of the order of 0.05, some 20 times lower than the MD hypothesis requires. The deleterious mutation rate in protein-coding sequences is therefore predicted to be so low in many obligate sexuals that the MD hypothesis can be seriously questioned.

The estimates for U (Table 1 and Fig. 1) are for point mutations in protein-coding genes, but mutations in noncoding DNA and insertion/deletion mutations also contribute to the deleterious mutation rate. However, these additional sources do not appear to be sufficient to increase U to levels which would be consistent with the MD hypothesis. Currently, the only estimate for the level of genome-wide selective constraint in noncoding DNA comes from a comparison of the nematodes *Caenorhabditis elegans* and *C. briggsae* (*19*). Mutation events in noncoding DNA seem to contribute a similar total number of deleterious mutations as mutations that occur in coding DNA. If this finding is general, then the estimates of U in Table 1 need to be multiplied by 2 to account for mutations in noncoding DNA. U is less than 1 in *Drosophila*, even if we assume that all point mutations throughout the genome are deleterious {U = 0.8, obtained by multiplying the average estimate for the synonymous substitution rate per generation [corrected (*11*)], $K_s' = 2.2 \times 10^{-9}$, by 3.6×10^8, the number of nucleotides in the diploid *Drosophila* genome (*20*)}.

Small insertion and deletion mutations are relatively infrequent in all taxa that have been studied and are estimated to contribute an additional ~10% to the estimates of U (*21, 22*). However, transposable element (TE) insertions are a major source of mutation in some organisms. For example, it has been estimated that spontaneous TE movement in *Drosophila* produces up to 0.2 new insertions per diploid per generation (*23, 24*). Most of these appear to be deleterious, because few TE insertions are fixed within *Drosophila* species (*24*). However, it is thought that most TE insertions are deleterious because of deletions caused by ectopic recombination, and not because they disrupt gene sequences (*24*). If meiosis is suppressed in new asexual lineages, as often appears to be the case (*25*), then ectopic exchange will be eliminated, and an asexual lineage will gain a benefit relative to sexuals. Thus, the deleterious effects of TE insertions brought about by ectopic exchange may favor asexuals over sexuals. Furthermore, the rate of transposition is expected to evolve to a lower rate in asexual lineages, because of the elimination of horizontal transfer.

The variation across taxa in the amino acid mutation rate per year is a little less than the variation in the deleterious mutation rate per year, reflecting a significant (Spearman's rank correlation coefficient $r_s = -0.854$, $P < 0.02$) negative correlation between the fraction of amino acid mutations eliminated by selection (U/M) and generation time (Fig. 2). The low value of U/M for human/chimpanzee genes is partly due to unusual properties of the genes that have been sequenced (*6, 26*); however, the rank correlation remains significant when the human/chimpanzee point is excluded ($r_s = -0.790$, $P < 0.04$). The negative relationship between U/M and generation time may reflect higher rates of fixation of slightly deleterious mutations in organisms with small population sizes (*27*), as population size is negatively correlated to generation time (*28*).

The MD hypothesis has stood out as one of the few hypotheses for the evolution of obligate sexuality that could be tested simply (*1*). However, based on our deleterious mutation rate estimates, the hypothesis is not supported as a general mechanism. While this leaves many other hypotheses to test, they all share a common feature: it is adaptive evolution that principally drives the evolution of sex, perhaps in combination with other mechanisms (*29*).

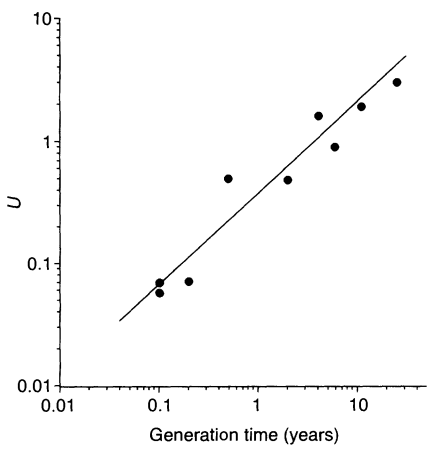

Fig. 1. Relationship between the genomic deleterious mutation rate per generation (U) and generation time with best-fitting linear regression line. The linear regression is also significant for the mammals plus bird subset ($P < 0.02$).

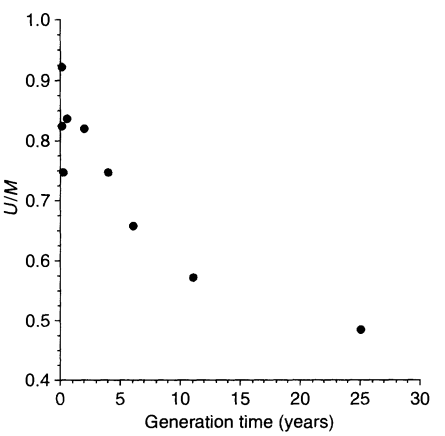

Fig. 2. Relationship between estimated fraction of amino acid mutations eliminated by selection (U/M) and generation time.

References and Notes

1. A. S. Kondrashov, *J. Hered.* **84**, 372 (1993).
2. H. J. Muller, *Mutat. Res.* **1**, 2 (1964).
3. M. Kimura and T. Maruyama, *Genetics* **54**, 1337 (1966).
4. B. Charlesworth, *Genet. Res.* **55**, 199 (1990).
5. R. S. Howard, *Theor. Popul. Biol.* **45**, 313 (1994).
6. P. D. Keightley and A. Eyre-Walker, *Genetics* **153**, 515 (1999).
7. M. Lynch *et al.*, *Evolution* **53**, 645 (1999).
8. A. S. Kondrashov and J. F. Crow, *Hum. Mutat.* **2**, 229 (1993).
9. A. Eyre-Walker and P. D. Keightley, *Nature* **397**, 344 (1999).
10. Calculation of *M* and *U*. Homologous DNA sequences were collected from Genbank for several pairs of species using Entrez (www.ncbi.nlm.nih.gov/Entrez/). Species were chosen for which estimates of the evolutionary divergence date were available from the literature that were not based on a global molecular clock. For mouse/rat, we used the data set described in W. Makalowski and M. S. Boguski [*Proc. Natl. Acad. Sci. U.S.A.* **95**, 9407 (1998)]. Sequences were aligned as described in (*9*). For each pair of genes, we calculated the rates of synonymous and nonsynonymous substitution. To calculate the rate of synonymous substitution, we only used codons that did not differ by a nonsynonymous substitution. We estimated the rate of transition (K_{ts4}) and transversion (K_{tv}) substitution at fourfold degenerate codons using the method of K. Tamura [*Mol. Biol. Evol.* **9**, 678 (1992)], and the rate of transition (K_{ts2}) substitution at twofold degenerate codons using the method of M. G. Bulmer, K. H. Wolfe, and P. M. Sharp [*Proc. Natl. Acad. Sci. U.S.A.* **88**, 5974 (1991)]. These methods take into account base composition bias. For each gene, we then calculated an overall rate of synonymous transition substitution (K_{ts}) by taking a weighted (by the number of sites) average of the two- and fourfold transition rates, and an overall rate of synonymous substitution (K_s) by summing K_{ts} and K_{tv}. The nonsynonymous substitution rate per codon was calculated as $K_n = -ln(1 - P_n)$, where P_n is the proportion of amino acids that differ between the two sequences. For each pair of homologous sequences, we calculated the average proportion of transitions (N_{ts}) and transversions (N_{tv}) that would change an amino acid. The genomic amino acid (*M*) and deleterious (*U*) mutation rates were calculated as $M = Z(\overline{K_{ts}N_{ts}} + \overline{K_{tv}N_{tv}})$, and $U = M - Z\bar{K}_n/3$, where averages were unweighted averages across genes and *Z* is a constant that converts the per site estimate to a per genome per generation estimate: e.g., for human/chimpanzee, $Z = 2$(genomes) $\times$ 80,000 (genes) $\times$ 1500 (base pairs for a gene) $\times$ 25 (years for a generation)/[12 $\times$ 10^6 (years of divergence)]. In the *D. melanogaster/D. pseudoobscura* comparison, we restricted our analysis to genes that were greater than 500 base pairs (bp) in length, since the divergence was relatively high (Table 1), and synonymous sites were close to saturation, leading to difficulty in accurately inferring K_{ts}; unfortunately, we were still unable to estimate K_{ts} for two genes, *G6PD2* and *tpi*; these genes were not analyzed.
11. The synonymous substitution rate is correlated to the level of synonymous codon bias in some *Drosophila* species. This may be due to selection on synonymous codon use or a downward bias in the methods used to correct for multiple hits in sequences with high codon bias. To correct our mutation rate estimates, we calculated the average (across the two species) χ_L value for each gene, where χ_L is the departure from expected codon usage assuming no codon bias calculated as a weighted χ^2 statistic [D. C. Shields, P. M. Sharp, D. G. Higgins, F. Wright, *Mol. Biol. Evol.* **5**, 704 (1988)], summed over amino acids, assuming an empirical genomic G and C base content of 40%. We then regressed the estimates of *M* and *U* obtained from each gene against the average χ_L value taking the *y* intercepts as our corrected *M* and *U* values; these are the predicted values of *M* or *U* for no synonymous codon bias.
12. H. Akashi, *Genetics* **139**, 1067 (1995).
13. J.-J. Jaeger, H. Tong, C. Denys, *C. R. Acad. Sci. Paris Ser. II* **302**, 917 (1986).
14. D. Yoder and Z. Yang, *Mol. Biol. Evol.* **17**, 1081 (2000).
15. J. W. Drake, *Ann. N.Y. Acad. Sci.* **870**, 100 (1999).
16. Examples: *Diptera* [G. C. Varley, G. R. Gradwell, M. P. Hassell, *Insect Population Ecology: An Analytical Approach* (Blackwell Scientific, Oxford, 1973)]; parasitoid wasps [T. M. Blackburn, *J. Anim. Ecol.* **60**, 151 (1991)]; planthoppers [R. F. Denno, K. L. Olmstead, E. S. McCloud, *Ecol. Entomol.* **14**, 31 (1989)].
17. D. J. Allan, *Am. Nat.* **110**, 165 (1976).
18. W. G. Chitwood and M. B. Chitwood, *Introduction to Nematology* (University Park Press, Baltimore, MD, 1974).
19. S. A. Shabalina and A. S. Kondrashov, *Genet. Res.* **74**, 23 (1999).
20. M. D. Adams *et al.*, *Science* **287**, 2185 (2000).
21. D. Petrov, E. R. Lozovskaya, D. L. Hartl, *Nature* **384**, 346 (1996).
22. R. Ophir and D. Graur, *Gene* **205**, 191 (1997).
23. Some studies have inferred that the genome-wide rate of transposition in *Drosophila* is as high as 0.4 [S. V. Nuzhdin and T. F. C. Mackay, *Mol. Biol. Evol.* **12**, 180 (1995)], but as X. Maside, S. Assimacopoulos, and B. Charlesworth [*Genet. Res.* **75**, 275 (2000)] argue, the true transposition rate is likely to be below 0.2 if element families that had become activated in individual lines and transposed at unusually high rates are disregarded.
24. B. Charlesworth and C. H. Langley, *Annu. Rev. Genet.* **23**, 251 (1989).
25. G. Bell, *The Masterpiece of Nature* (Univ. of California Press, San Francisco, CA, 1982).
26. M. Steiper *et al.*, personal communication.
27. T. Ohta, *J. Mol. Evol.* **1**, 150 (1972).
28. L. Chao and D. E. Carr, *Evolution* **47**, 688 (1993).
29. S. A. West, C. M. Lively, A. F. Read, *J. Evol. Biol.* **12**, 1003 (1999).
30. M. Goodman *et al.*, *Mol. Phylogenet. Evol.* **9**, 585 (1998).
31. R. L. Carroll, *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
32. R. C. Fleisher, C. E. McIntosh, C. L. Tarr, *Mol. Ecol.* **7**, 533 (1998).
33. J. R. Powell and R. DeSalle, *Evol. Biol.* **28**, 87 (1995).
34. F. Antequera and A. Bird, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11995 (1993).
35. Supplementary material is available at www.sciencemag.org/feature/data/1053250.shl
36. C. G. Sibley and J. E. Alquist, *Phylogeny and Classification of Birds* (Yale Univ. Press, New Haven, CT, 1990).

37. We thank A. Peters, J-L. Hartenberger, W. Hill, J. Crow, J. Peck, S. West, and S. Otto for helpful discussions and comments, and N. Smith for the data set of mouse/rat genes. Supported by the U.K. Royal Society.

19 June 2000; accepted 24 August 2000

Localized Rac Activation Dynamics Visualized in Living Cells

Vadim S. Kraynov,[1]* Chester Chamberlain,[1]* Gary M. Bokoch,[1,2] Martin A. Schwartz,[3] Sarah Slabaugh,[1] Klaus M. Hahn[1]†

Signaling proteins are thought to be tightly regulated spatially and temporally in order to generate specific and localized effects. For Rac and other small guanosine triphosphatases, binding to guanosine triphosphate leads to interaction with downstream targets and regulates subcellular localization. A method called FLAIR (fluorescence activation indicator for Rho proteins) was developed to quantify the spatio-temporal dynamics of the Rac1 nucleotide state in living cells. FLAIR revealed precise spatial control of growth factor–induced Rac activation, in membrane ruffles and in a gradient of activation at the leading edge of motile cells. FLAIR exemplifies a generally applicable approach for examining spatio-temporal control of protein activity.

Rac is a member of the Ras superfamily of small guanosine triphosphatase (GTPase) proteins (*1*) and plays a critical role in diverse processes, such as control of cell morphology, actin dynamics, transcriptional activation, and apoptosis signaling (*2*). The broad range of events controlled by this GTPase requires regulation of its interactions with multiple downstream targets. The effects of Rac may in part be controlled by regulating the subcellular localization of its activation. GTP exchange factors (GEFs), which regulate nucleotide exchange on Rho GTPases, contain a variety of localization domains and may modulate downstream signaling from Rac (*3*). Rac induces localized actin rearrangements to generate polarized morphological changes (*4*), but it has been difficult to explore how Rac activation produces localized actin behavior in an intact cell. We developed a method based on fluorescence resonance energy transfer (FRET) that quantifies the timing and location of Rac activation in living cells. Here, it was used to study activation of the Rac1 isoform in cell motility and extracellular signal-induced cytoskeletal changes.

Sensing the Rac nucleotide state required introducing a fluorescently labeled biosensor into a cell together with a fusion protein comprising Rac and green fluorescent protein (GFP) (Fig. 1A) (*5*). This protein biosensor was labeled with the acceptor dye Alexa 546, which can undergo FRET with GFP. Because the

Departments of [1]Cell Biology, [2]Immunology, and [3]Vascular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: khahn@scripps.edu