Title
Molecular clock or erratic evolution? A tale of two genes.

Journal
Proceedings of the National Academy of Sciences of the United States of America, 93(21)

ISSN
0027-8424

Authors
Ayala, FJ
Barrio, E
Kwiatowski, J

Publication Date
1996-10-15

DOI
10.1073/pnas.93.21.11729

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 11729–11734, October 1996
Evolution

Molecular clock or erratic evolution? A tale of two genes

(neutral theory of evolution/glycerol-3-phosphate dehydrogenase/superoxide dismutase/*Drosophila*/*Ceratitis*)

FRANCISCO J. AYALA*†, ELADIO BARRIO*‡, AND JAN KWIATOWSKI*

*Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; and ‡Department of Genetics, University of Valencia, 46100-Burjassot, Valencia, Spain

Contributed by Francisco J. Ayala, July 19, 1996

ABSTRACT **We have investigated the evolution of glycerol-3-phosphate dehydrogenase (*Gpdh*). The rate of amino acid replacements is 1×10^{-10}/site/year when *Drosophila* species are compared. The rate is 2.7 times greater when *Drosophila* and *Chymomyza* species are compared; and about 5 times greater when any of those species are compared with the medfly *Ceratitis capitata*. This rate of 5×10^{-10}/site/year is also the rate observed in comparisons between mammals, or between different animal phyla, or between the three multicellular kingdoms. We have also studied the evolution of Cu,Zn superoxide dismutase (*Sod*). The rate of amino acid replacements is about 17×10^{-10}/site/year when comparisons are made between dipterans or between mammals, but only 5×10^{-10} when animal phyla are compared, and only 3×10^{-10} when the multicellular kingdoms are compared. The apparent decrease by about a factor of 5 in the rate of SOD evolution as the divergence between species increases can be consistent with the molecular clock hypothesis by assuming the covarion hypothesis (namely, that the number of amino acids that can change is constant, but the set of such amino acids changes from time to time and from lineage to lineage). However, we know of no model consistent with the molecular clock hypothesis that would account for the increase in the rate of GPDH evolution as the divergence between species increases.**

Glycerol-3-Phosphate Dehydrogenase (GPDH): Erratic Evolution

The NAD-dependent cytoplasmic GPDH (EC 1.1.1.8) plays a crucial role in metabolism through its keystone position in the glycerophosphate cycle, which in *Drosophila* and other dipterans provides energy for flight in the thoracic muscles (1, 2). In *Drosophila melanogaster*, the *Gpdh* gene consists of eight exons and produces three isozymes by differential splicing of the last three exons (3–5). We have analyzed 768 nucleotides (nt) from exons 3–6, coding for 256 amino acids in 27 dipteran species (only 759 nt in 12 species of the *Drosophila obscura* group, and 729 nt in the medfly *Ceratitis capitata*) (6, 7). The species studied are from two families: *Tephritidae*, represented by *Ceratitis capitata*; and *Drosophilidae* represented by three genera: *Scaptodrosophila* (one species), *Chymomyza* (two species), and *Drosophila* (23 species, classified within five subgenera: *Dorsilopha*, *Drosophila*, *Hirtodrosophila*, *Sophophora*, and *Zaprionus*).

The phylogeny of the dipteran genera and subgenera is shown in Fig. 1. The topology represented in Fig. 1 is statistically superior to any alternative topology (6). In any case, the branching sequence among the *Drosophila* subgenera or between *Chymomyza* and *Scaptodrosophila* are the only issues that might be in question, and they are of no material consequence for the present purposes.

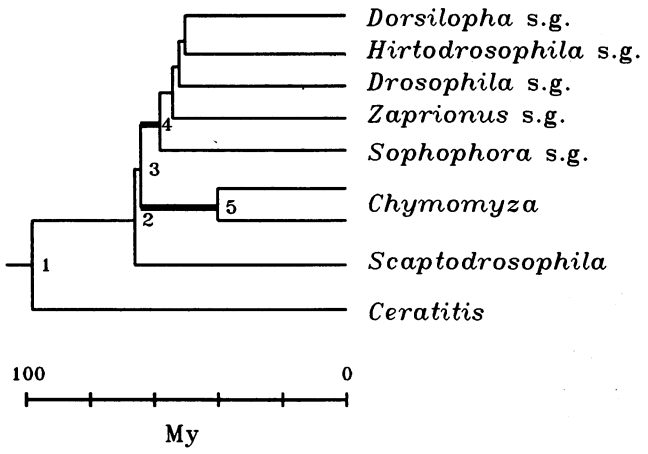

FIG. 1. Phylogeny of the genera and subgenera (s.g.), according to ref. 6. The thicker branches indicate hypothetical instances of fast GPDH evolution. *Ceratitis capitata* and *Scaptodrosophila lebanonensis* are the only representatives of their genera. *Chymomyza* includes two species, *Chymomyza amoena* and *Chymomyza procnemis*. The genus *Drosophila* is represented by five subgenera: *Dorsilopha* (*Drosophila busckii*), *Hirtodrosophila* (*Drosophila pictiventris*), *Drosophila* (*Drosophila hydei* and *Drosophila virilis*), *Zaprionus* (*Zaprionus tuberculatus*), and *Sophophora* (represented by three species groups: the *melanogaster* group, including *Drosophila melanogaster*, *Drosophila simulans*, and *Drosophila teissieri*; the *willistoni* group, including *Drosophila willistoni*, *Drosophila paulistorum*, and *Drosophila nebulosa*; and the *obscura* group including the Palearctic *Drosophila ambigua*, *Drosophila bifasciata*, *Drosophila guanche*, *Drosophila madeirensis*, *Drosophila obscura*, and *Drosophila subobscura*, plus the Nearctic *Drosophila pseudoobscura*, *Drosophila persimilis*, *Drosophila miranda*, *Drosophila affinis*, *Drosophila azteca*, and *Drosophila tolteca*, of which the last three are included in the *affinis* subgroup). *Zaprionus* has traditionally been considered a separate genus, but phylogenetically belongs within the genus *Drosophila* and we are, therefore, including it as a subgenus of *Drosophila*. *Scaptodrosophila* was originally described as a subgenus of *Drosophila* but phylogenetically falls outside the genus and has been formally raised to a genus (8). The time scale is based on data from refs. 6, 9, and 10; the date for the divergence between the two *Chymomyza* species [45 million years (My)] is based on a more limited data set than other dates, but it seems unlikely that it would be smaller than 40 My or greater than 55 My.

Fig. 2 *Left* plots the number of amino acid replacements between species against the time since their divergence. The amino acid replacements are calculated by counting the number of differences between aligned sequences and then applying an algorithm that increments the observed number by taking into account superimposed and back replacements. We have used the algorithm known as PAM, for accepted point mutations (13), but other algorithms give similar[2] results (6). For the divergence times, we have used crude consensus of published estimates (9, 10, 14) (see Table 1), but the conclu-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GPDH, glycerol-3-phosphate dehydrogenase; My, million years; PAM, accepted point mutations; SOD, superoxide dismutase.

†To whom reprint requests should be addressed.

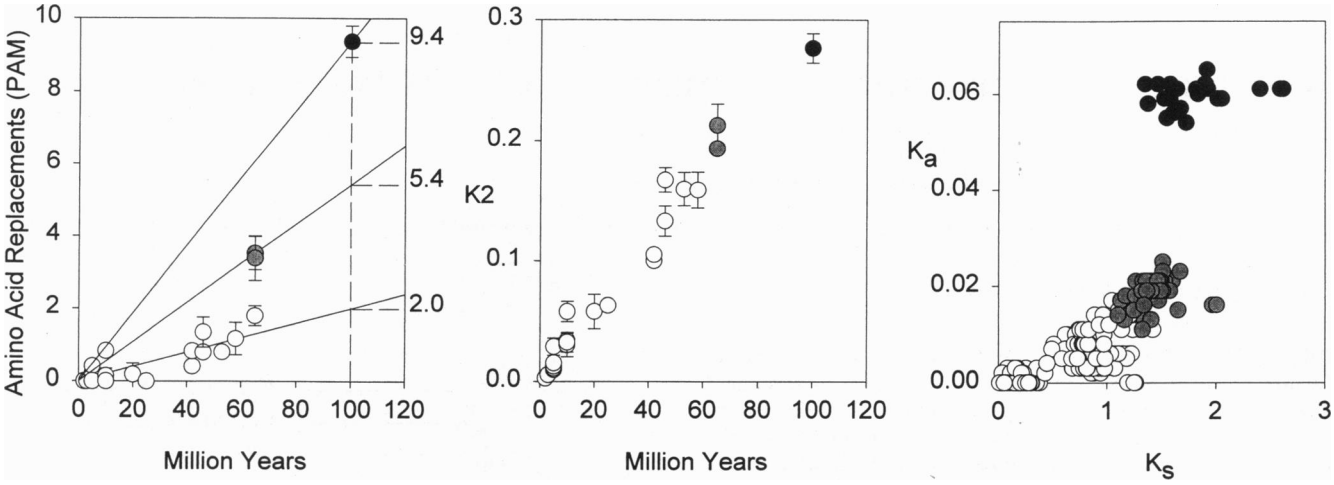

FIG. 2. Rate of *Gpdh* evolution in dipterans. (*Left*) Number of amino acid replacements between species. Open circles represent comparisons between species of *Drosophila* and/or *Scaptodrosophila* (each open circle often represents numerous identical values); gray circles, between species of *Chymomyza* and *Drosophila* or *Scaptodrosophila*; solid circles, between *Ceratitis* and all other species. Data are from (6, 7). It is apparent that a single straight line cannot provide a reasonable fit for all points. The three straight lines, drawn by eye, represent rates of amino acid replacement between species per 100 residues per 100 My. (*Center*) Rate of nucleotide substitutions (K2) between species. The units are substitutions per site, estimated by Kimura's (11) two-parameter method. (*Right*) Synonymous (K_s, abscissa) versus nonsynonymous (K_a, ordinate) rate of substitutions in *Gpdh*, estimated according to Li (12).

sions to be drawn do not require that these estimates be very accurate.

It is apparent that the rate of amino acid replacements is not constant over time. The average number of amino acid replacements per 100 residues (Table 1) is 0.8–1.2 between species from different *Drosophila* groups or subgenera, which diverged 45–55 My ago, but 3.0 between species from different genera, which diverged only somewhat earlier (≈60 My ago). When comparisons are made between the medfly *Ceratitis capitata* and the Drosophilids, the number of amino acid replacements becomes 9.4 ± 0.1, for a time elapsed only twice as long as between the *Drosophila* genera or subgenera. The rates of amino acid replacements between species for the three comparisons are shown in Fig. 2 *Left*; the lineage rates are half as large. That is, the rate of amino acid replacements observed in *Drosophila* lineages is $0.9–1.1 \times 10^{-10}$/site/year, a very slow rate (comparable to the rates of 0.25–1.7 observed for histone proteins; ref. 15). But when *Ceratitis* and *Drosophila* species are compared, the lineage rate becomes 5 times greater, 4.7×10^{-10}/site/year (slightly slower than the rates observed in some intracellular enzymes, such as 5.3×10^{-10} for triosephosphate isomerase, or 6.7×10^{-10} for cytochrome *c*).

The different rates of evolution displayed in Fig. 2 become even more disparate when we take into account that these rates apply to largely overlapping lineages. Consider, for example (Fig. 1), the evolution from node 3 to a *Chymomyza* species and a *Drosophila* species (say, *D. melanogaster*). The average rate of amino acid evolution is 2.7×10^{-10}/site/year (Table 1, line 4) over the 120 My of evolution separating these two species (60 My from node 3 to the *Chymomyza* species and 60 My from node 3 to the *Drosophila* species). But during 100 of the 120 My (55 My from node 4 to the *Drosophila* species and 45 My from node 5 to the *Chymomyza* species), the rate of evolution is $0.9–1.1 \times 10^{-10}$ (Table 1, lines 2 and 3). Thus, the acceleration in rate of evolution could only have occurred over the 20 My of evolution indicated by thick lines in Fig. 1 (between nodes 3 and 4 and between nodes 3 and 5). The rate of GPDH evolution during those 20 My must have been 11×10^{-10}/site/year, or more than 10 times faster than for the remainder 100 My, to give an average of 2.7×10^{-10} over the 120 My. (As noted in Table 1, the total number of amino acid replacements between the *Drosophila* subgenera, or between the two *Chymomyza* species, is mostly 2, whereas between *Chymomyza* and *Drosophila* is 8 or 9.)

However, the rate of divergence between *Scaptodrosophila* and *Drosophila* species is not greater than between the *Drosophila* subgenera (6). This implies that GPDH evolution between nodes 3 and 4 (Fig. 1) has occurred at the prevailing *Drosophila* rate $\approx 1 \times 10^{-10}$/site/year. Therefore, to account for the much larger average rate of divergence between

Table 1. Rate of *Gpdh* evolution for increasingly divergent species

Comparison	My	Amino acid replacements		Nucleotide substitutions	
		$\bar{x}$	Per 100 My	$\bar{x}$	Per 100 My
1. Within *Drosophila* groups	5–25	0.0–0.8	0–2.1	1.0–8.5	10–21
2. Between *Drosophila* groups	45 ± 10	0.8 ± 0.0	0.9	16.3 ± 0.1	18.1
3. Between *Drosophila* subgenera	55 ± 10	1.2 ± 0.0	1.1	15.8 ± 0.2	14.2
4. Between dipteran genera	60 ± 10	3.0 ± 0.1	2.7	20.3 ± 0.2	16.9
5. Between mammals	70 ± 15	7.4 ± 1.6	5.3	—	—
6. Between dipteran families	100 ± 20	9.4 ± 0.1	4.7	27.7 ± 0.2	13.8
7. Between animal phyla	650 ± 100	54.7 ± 0.3	4.2	—	—
8. Between kingdoms	1100 ± 200	87.0 ± 0.8	4.0	—	—

The species compared are listed in the legends for Figs. 1 (Dipterans), 3, and 4. The plus/minus values are crude estimates of error for My, but are standard deviations for replacements and substitutions. $\bar{x}$ values are per 100 residues for differences between species; the "Per 100 My" are lineage values. Replacements are corrected according to ref. 13; nucleotide substitutions are estimated according to ref. 11. The matrices of amino acid and nucleotide differences between dipteran species (rows 1–4 and 6) are given in refs. 6 and 7; the total numbers of amino acid differences between *Drosophila* species are mostly 0 (within groups) or 2 (between groups or subgenera, and between the two *Chymomyza* species), whereas they are mostly 8 or 9 between *Chymomyza* and *Drosophila*, and 20–22 between *Ceratitis* and any other species.

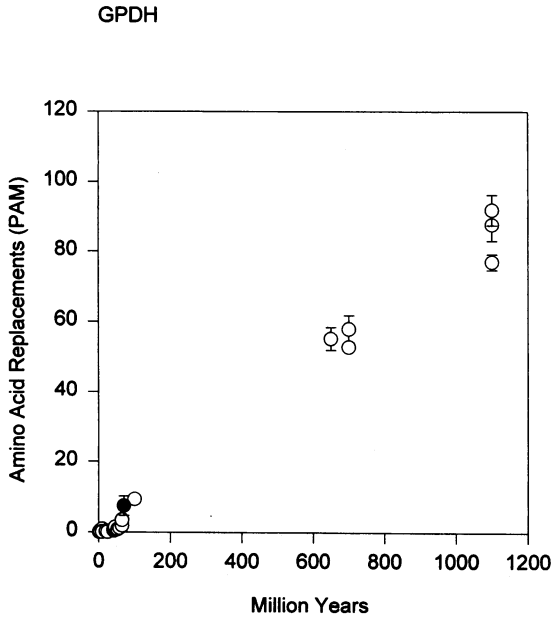

FIG. 3. Long-term rate of GPDH evolution. The points on the lower left represent all those included in Fig. 2 *Left* plus comparisons between three mammals: human, mouse, and rabbit (solid circle). Points in the center are for comparisons between species from different animal phyla: fruitflies, mammals, and the nematode *Caenorhabditis elegans*. Points on top right are for comparisons between species from different kingdoms: animals, a plant (*Cuphea lanceolata*), and two fungi (*Saccharomcyces cerevisiae* and *Schizosaccharomyces pombe*). The rates of amino acid replacement corresponding to various comparisons are given in Table 1.

Chymomyza and *Drosophila*, the acceleration would be limited to the 15 My separating nodes 3 and 5 in the *Chymomyza* lineage, during which period the rate of GPDH evolution must have been 15×10^{-10}/site/year, or 15 times greater than the rate prevailing in the evolution of all the other Drosophilid lineages.

Similarly, the rate of GPDH divergence between the medfly *Ceratitis* and *Drosophila* is 4.7 times greater (Table 1, line 6, and Fig. 2 *Left*) than between *Drosophila* species. But the accelerated rate of evolution could only have occurred for a fraction of the 200 My elapsed (between nodes 1 and 2 and between node 1 and *Ceratitis*), during which the rate of GPDH evolution must have been much more than 5 times faster than between *Drosophila* species. The evolution of GPDH in dipterans is not clocklike at all. (As noted in Table 1, the total number of amino acid replacements between the *Drosophila* subgenera or between the two *Chymomyza* species is mostly 2, whereas between *Ceratitis* and any of the other species is 20–22.)

Fig. 2 *Center* displays the nucleotide distances between the *Gpdh* nucleotide sequences [corrected with the two-parameter algorithm of Kimura (11)] for the same set of species as in Fig. 2 *Left*. It appears that at the nucleotide level *Gpdh* is evolving with the regularity expected from a molecular clock: a straight line drawn from the origin to the average *Ceratitis–Drosophilid* value would pass through or near most intermediate points. (This observation, by the way, indicates that the discrepancies observed at the amino acid level are not caused by errors in the branching sequence of the taxa or in the assumed times of divergence.)

Fig. 2 *Right* shows a plot of K_a, the rate of nonsynonymous substitutions (i.e., those that result in amino acid replacements) against K_s, the rate of synonymous substitutions. The two rates are not closely correlated. It follows that the apparent regular rate of nucleotide evolution, manifest in the middle panel (Fig. 2), is made of two components: K_s, which evolves fairly regularly, and K_a, which evolves spastically, as also manifested when the amino acid differences are directly observed (Fig. 2 *Left*).

Fig. 3 displays the number of amino acid replacements against time for the dipteran comparisons, as well as for comparisons between three mammals, three animal phyla, and three kingdoms. The relationship is fairly linear, corresponding to a rate of amino acid replacement evolution of $4.0–5.3 \times 10^{-10}$/site/year (Table 1, lines 5, 7, 8), fairly similar to the rate observed between *Ceratitis* and the *Drosophilids*. At this scale, the rate of evolution of GPDH appears fairly clock-like.

GPDH

```
D. melanogaster IPNFCKQLLG KIKPNAIAIS LIKGFDKAEG GGIDLISHII TRHLKIPCAV LMGANLANEV AEGNFCETTI GCTDK-KY-- ---G---KVL  90
Ceratitis       ......T...  ......V...  ..........  ..........  ........S.  ..........  .L........  ..R.I.....  ...T....P.
Human           .GKI.D..K.  HL.A.PTG..  ....V.EG-P  N.LK...EV.  GER.G..MS.  .....I.S..  .DEK......  ..K.PAQ-..  .......QL.
Mouse           .GKI.D..K.  HL.A.T.G..  ....V.EG-P  N.LK...EV.  GER.G..MS.  .....I.S..  ..EK......  ..K.PAQ-..  .......QL.
Rabbit          .GKI.DEIK.  HL.A...G..  ....VNEG-P  K.LK...EV.  GE..G..MS.  .....I.S..  .DEK......  ..K.QAQ-..  .......QL.
C. elegans      VKGI.EK.V.  ..PADTQ...  ....VSTEKR  ..LK...EE.  KEI...EVS.  .......P..  .ND....A..  ..KR.AED..  .......PL.
Cuphea          MEGI..R.E.  ..QEG.Q.L.  ....MEVKME  .PC-M..SL.  SDL.G.N.C.  .....I...I  .VEK.S.A.V  .FRENRDI..  ...A...EKW
S. pombe        .ERVWH.MV.  L.R.G.VG..  C...VAVSKE  .S--.Y.EV.  SEK.G.Y.G.  .S...V....  .REQ......  .FNPPNEV..  ...DIPREQI
S. cerevisiae   L.RI.S..K.  HVDSHVR...  CL...EVG-A  K.VQ.L.SY.  .EE.G.Q.GA  .S...I.T..  .QEHWS...V  AYHIPKDFRG  EGKDVDH...

D. melanogaster RDLFQANHFR VVVVDDADAV EVCGALKNIV ACGAGFVDGL KLGDNTKAAV IRLGLMEMIR FVDVFYPGS- KLSTFFESCG VADLITTCYG 180
Ceratitis       ......PY..  ....E.S...  ..........  .T........  G.........  .........V  ..........  ..........  ..........
Human           KE.M.TPN..  IT..QEV.T.  .I......V.  .V....C...  GF........  .........A  .AKL.CS.PV  SSA..L....  ..........
Mouse           K..M.TPN..  IT..QEV.T.  .I........  .V....C...  GF........  .........A  .AKL.CS.IV  SSA..L....  ..........
Rabbit          KQ.M.TPN..  I..TQEVNT.  .I.....DL.  .V....C..I  GF........  .........A  .AKL.CS.PV  SPA..L....  ..........
C. elegans      KK..HTDN..  IN..E..HT.  .L......V.  ..A...T...  GY........  .......TTK  ..EHY.....  N.Q.......  I.........
Cuphea          VQ..STPY.M  .SA.E.VEG.  .L..T.....  .IA.......  EM.N.....I  M.I..R..KA  .SKLLF.-.V  .DT.......  .......L.
S. pombe        AAVSDRPY.S  ..S...VAG.  ALG.....V.  .MAV..A...  EW.G.....I  M.R..L..QK  .ATT.FDSDP  R-TMVEQ...  I...V.S.L.
S. cerevisiae   KA..HRPY.H  .S.IE.VAGI  SI......V.  .L.C...E..  GW.N.AS..I  Q.V..G.I..  .GQM.F.E.R  EETYYQ..A.  ........A.

D. melanogaster GRNRRVSEAF VTSG--KTIE ELEKEMLNGQ KLQGPPTAEE VNYMLKNKGL EDKFPLFTAI HKICTNQLKP NDLI 254
Ceratitis       ......A...  .KT....S..  ..........  ..........  ..........  ..........  .........V  Q...
Human           ....K.A...  ART....S..  Q....L....  .....E..R.  LYSI.QH...  V......M.V  Y.V.YEGQPV  GEF.
Mouse           ....K.A...  ART....S..  Q.........  .....Q..R.  LHSI.QH...  V.........V  Y.V.YEGQPV  GEF.
Rabbit          ....K.A...  ART....S..  Q.........  .....E..R.  LHSI.QH...  V.W....M.V  Y.V.YQGQPV  GEF.
C. elegans      ....K.C...  .KT....SMA  .V...L....  SA...L....  .YL.MHKT..  DA.......V  ....AGEM..  AE.V
Cuphea          ....K.A...  AKN.GKRSFD  D..A...R..  ....VS..K.  .YEV.GHR.W  LEL....STV  .E.S.GR.P.  SAIV
S. pombe        ...N.CA...  .KT....SL.  T....L.G..  L...AA.SKD  .HEF.LT.DM  VKD......V  YN.SYEDMD.  K...
S. cerevisiae   ...VK.AR-L  MATSGKDAW.  C-...L....  SA..LI.CK.  .HEW.ETC.S  VED....E.V  YQ.VY.NYPM  KN.P
```

FIG. 4. GPDH amino acid sequence alignment between *D. melanogaster*, the medfly *Ceratitis capitata*, three mammals, a nematode, a plant, and two fungi. Dots indicate identical amino acids to those of *D. melanogaster*. Dashes indicate gaps introduced to accommodate extra amino acids found in some, but not other, species. The GenBank *Gpdh* accession numbers for these species, respectively from top to bottom, are J04567, L36960, L34041, M25558, P08507 (from the Swiss-Prot Protein Sequence Data Base), Z22180, X79677, X56162, and Z24454.

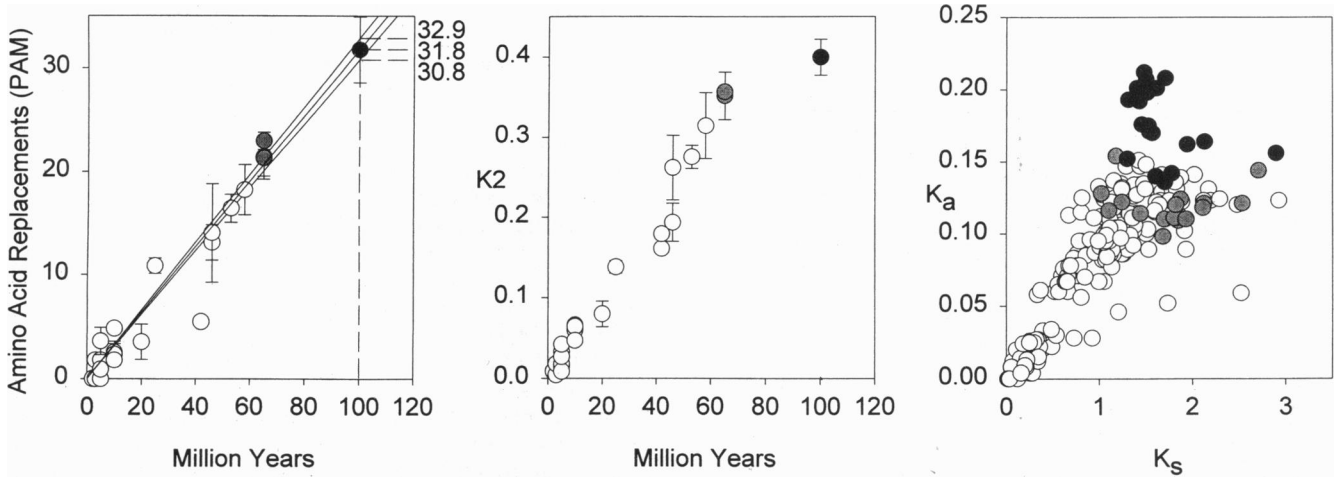

FIG. 5. Rate of *Sod* evolution in dipterans. (*Left*) Number of amino acid replacements between species. (*Center*) Rate of nucleotide substitutions. (*Right*) Correlation between synonymous (K_s) and nonsynonymous (K_a) substitutions. The species compared are the same as for *Gpdh*, and are listed in the legend to Fig. 1. Comparisons are based on 342 nt of the *Sod* coding sequence (consisting of 459 nt). Data are from (7, 9). Symbols are as in Fig. 2. The three rate lines on the *Left* are drawn for comparison with Fig. 2. The rate of 30.8 is for comparisons between *Drosophila* species, 32.9 is for comparisons between *Drosophila* and *Chymomyza*, and 31.8 is for comparisons between *Ceratitis* and *Drosophila* or *Chymomyza*.

Nevertheless, this "global" rate would yield very erroneous conclusions if it were used for timing events or for deciding branching topologies in the evolution of dipterans. The amino acid sequences used in this figure are shown in Fig. 4 for non-dipterans; for dipterans, see Kwiatowski *et al.* (6) and Barrio and Ayala (7). It can be seen in Table 1 that the apparent rate of amino acid replacements is fairly constant for comparisons between species diverged >70 My ago (i.e., 4.0–5.3 $\times$ 10^{-10}/site/year), but this is 2–5 times as large as between species diverged <60 My ago.

Evolution of Superoxide Dismutase (SOD): A Constrained Clock

The superoxide dismutases defend the organism against the toxicity of oxygen. The Cu,Zn SOD of *Drosophila* is a dimer molecule made up of two identical polypeptides, each consisting of 151 amino acids. Ayala (16; see also ref. 17) pointed out the seemingly erratic evolution of SOD. Comparisons between PAM-corrected amino acid sequences of mammals (lineages separated ≈70 My ago) indicated a rate of replacement 5 times greater than between mammals and *D. melanogaster* (divergence ≈600 My), and nearly 10 times greater than between fungi and animals (divergence ≈1100 My).

Fig. 5 gives the *Sod* rate of amino acid and nucleotide evolution for the same set of species compared in Fig. 2. The rate of amino replacement (*Left*) is fairly linear over time, about 31 $\times$ 10^{-10} amino acid replacements/site/year, although there are some conspicuously divergent points. Fairly linear are also the rate of nucleotide substitution (*Center*) and the relationship between the rate of nonsynonymous (K_a) and synonymous (K_s) substitutions (*Right*). The linearity of SOD evolution in these dipteran lineages contrasts with the sharp increase observed earlier in GPDH evolution when *Chymomyza* or *Ceratitis* are compared with the *Drosophila* species.

Fig. 6 displays the SOD amino acid replacements for the same set of species compared for GPDH in Fig. 3 (except that the plant is now *Ipomoea batatas* rather than *Cuphea lanceolata*); the aligned amino acid sequences are shown in Fig. 7 (also, see Table 2). As noted earlier by Ayala (ref. 16; but see also ref. 17), the rate of evolution of SOD seems to slow down as the lineages compared become increasingly remote. This state of affairs is precisely the reverse of the pattern observed in GPDH, where the rate of amino acid replacement is slower when the species compared are closely related.

Fitch and Ayala (10) have shown that the evolution of SOD can be made to fit the expectations of a molecular clock by introducing the concept of *covarions* (*co*ncomitantly *vari*able cod*ons*), which asserts that there is a limited number of amino acid sites that can be replaced at any time in any given lineage. The number of sites that can vary remains constant, but the composition of the set of variable sites changes through time and between lineages. The application of this assumption to a particular protein requires that one determines (*i*) the size of the covarion set—i.e., the number of sites at which replacements can occur at any given time in a given lineage; (*ii*) the

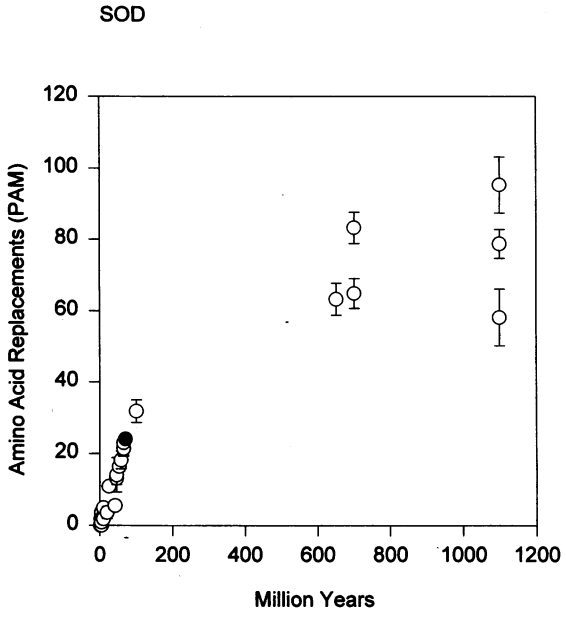

FIG. 6. Amino acid replacements in SOD. The species compared are the same as for GPDH (see Fig. 3; except that here the plant is *Ipomoea batatas*). Note that the rate of SOD evolution between dipterans or mammals (solid circle), which diverged <100 My ago, is much greater than between species diverged >600 My ago, just the opposite of what is observed in GPDH (Fig. 3). The rates estimated for different comparisons are given in Table 2. Data are from refs. 10 and 16.

SOD

```
D. melanogaster INGDA--KGT VFFEQESSGT PVKVSGEVCG LA-KGLHGFH VHEFGDNTNG CMSSGPHFNP YGKEHGAPVD ENRHLGDLGN IEATGDCPTK  90
Ceratitis       ....V..... .H...QDAKS ..L.T...N. .......... .......... .T.A..... ..NS....S. L......... ...S..GA..
Human           LK..GPVQ.I IN...KE.NG ....W.SIK. .T.E..... ........A. .T.A..... LSRK..G.K. .E..V..... VT.DK.GVAD
Mouse           LK..GPVQ.. IH...KA..E ..VL..QIT. .T.E.Q.... ..QY....Q. .T.A..... HS.K..G.A. .E..V..... VT.GK.GVAN
Rabbit          LK..GPVEA. IH...KGT.- ..V.K.RIT. .T.E...E.. ..Q....RQ. .T.A..... LS.K..G.K. .E..V..... VT.GSNGVAD
C.elegans       LR.ET.VT.. IWIT.K.END QAVIE..IK. .T.P..... ..QY..S... .I.A..... F..T..G.KS .I..V..... V..GA.GVA.
Ipomoea         SSSEG.VS.. I..S..GD.- .TT.T.N.S. .K.P..... ..AL..T... ...T..... A.......G. D...A..... .TVGE.GTAS
S.pombe         LR..SKVS.V .T...VDQNS Q.S.IVDLV. NDANAKR... I.Q..... .T.A..... E..T..DRTA AV..V..... L.SDAQGNI.
S. cerevisiae   LK...GVS.V .K...A.ESE .TT..Y.IA. NSPNAER... I.....A... .V.A..... FK.T....T. .V..V..M.. VKTDENGVA.

D. melanogaster VNITDSKITL FGADSIIGRT VVVHADA 117
Ceratitis       .E.S.KL... ..EN..V... I.....P
Human           .S.E..V.S. S.DHC..... L...EK.
Mouse           .S.E.RV.S. S.EH..... M...EKQ
Rabbit          .L.E..V.S. S.DM.V.... L...EKE
C.elegans       IKL..TLV.. Y.PNTVV..S M....GQ
Ipomoea         FT...KQ.P. T..N.V...A ....G.P
S.pombe         TTFS..V.S. ...N..... I.I..GE
S. cerevisiae   GSFK..L.K. I.PT.VV..S ..I..GQ
```

FIG. 7. SOD amino acid sequence alignment between *D. melanogaster*, *Ceratitis*, three mammals, a nematode, a plant, and two fungi. Conventions as in Fig. 4. The GenBank accession numbers for these species, respectively from top to bottom, are X13780, M76975, K00065, M35725, Z22644, L20135, L36229, X66722, and J03279.

total number of sites that are invariable—i.e., the number of sites, if any, at which amino acid replacements can never occur in any lineage; (*iii*) the number of different amino acids that can occur at the variable sites, which may range from only two different amino acids tolerated at some sites (for example, at a site that must have a negative charge, only aspartate and glutamate can occur) to all 20 amino acids; (*iv*) the persistence of the covarion set—i.e., the rate at which one site in the set becomes replaced by another site; and (*v*) the rate of amino acid replacements.

Fitch and Ayala (10) have analyzed a set of 67 SOD sequences from very diverse organisms from all three multicellular kingdoms to estimate the parameter values corresponding to the five assumptions just mentioned. The parameter values that maximize the fit between the observed and expected number of amino acid replacements are as follows: (*i*) the number of covarions is 28; (*ii*) the number of codons that are permanently invariable across animals, plants, and fungi is 44; (*iii*) the average number of amino acids that can occur at a variable site is 2.5; (*iv*) the persistence of the covarion set is 0.01 (that is, there is a 0.01 probability that one site of the covarion set will change whenever one amino replacement has occurred); and (*v*) the rate of amino acid replacement for the whole polypeptide is 4×10^{-10}/site/year. A reasonably good fit is obtained (see table 1 in ref. 10) between the number of amino acid differences actually observed and the number expected if the constraints defined by these parameter values obtain. The conclusion drawn by Fitch and Ayala (10) is that SOD may be evolving at a constant rate as postulated by the molecular clock, even though it may appear at first quite erratic—i.e., when the constraints under which the clock operates have not been taken into account.

Discussion

The hypothesis of the molecular clock of evolution was put forward by Zuckerkandl and Pauling (18), who conjectured that most amino acid and nucleotide substitutions may be of little functional consequence. Molecular sequence differences between species would then reflect the time elapsed since the divergence of their ancestral lineages. The hypothesis of the molecular clock was mathematically formulated by Kimura (19, 20), and became the keystone of his neutral theory of molecular evolution. Under the assumptions of the neutral theory, the rate of molecular evolution is stochastically constant and Poisson-distributed, with variance identical to the mean, which is simply specified by the rate of neutral mutation. A general empirical observation, however, is that the variance-to-mean ratio is larger than one, whether it be measured in terms of amino acid or nucleotide substitutions (21). To account for this deviation, the neutral theory has been modified by assuming that most molecular evolution involves slightly deleterious substitutions rather than strictly neutral ones (20, 22), or by assuming that there is a generation-time effect (e.g., ref. 23), and in other ways. From a theoretical, as well as applied, perspective these modifications have the discomforting consequence that they involve additional empirical parameters, often not easy to estimate. It is of great epistemological significance that the neutral theory (*i*) is highly predictive and (*ii*) is, therefore, eminently testable. These two properties, really two sides of the same coin, become diluted in the modified versions of the neutral theory of molecular evolution. Nevertheless, it is commonly assumed that molecular evolution is sufficiently regular over time and across lineages, so that a molecular clock hypothesis can be assumed to be applicable to test phylogenetic hypotheses, or to estimate the time of remote evolutionary events.

The combined consideration of GPDH and SOD evolution in the same set of species is disquieting. The covarion hypothesis becomes helpful to account for the apparent slower rate of evolution that obtains when the species compared become increasingly remote. But the covarion model that Fitch and Ayala (10) successfully applied to SOD cannot be extended to

Table 2. Rate of *Sod* evolution for increasingly divergent species

Comparison	My	Amino acid replacements		Nucleotide substitutions	
		$\bar{x}$	Per 100 My	$\bar{x}$	Per 100 My
1. Within *Drosophila* groups	5–25	0.0–11.4	0–22.7	0.9–14.3	8.8–28.7
2. Between *Drosophila* groups	45 ± 10	15.0 ± 0.2	16.6	23.9 ± 0.5	26.6
3. Between *Drosophila* subgenera	55 ± 10	17.8 ± 0.3	16.2	31.0 ± 0.4	28.1
4. Between dipteran genera	60 ± 10	21.4 ± 0.2	17.8	34.2 ± 0.4	28.5
5. Between mammals	70 ± 15	24.0 ± 0.4	17.2	—	—
6. Between dipteran families	100 ± 20	31.8 ± 0.6	15.9	40.0 ± 0.4	20.0
7. Between animal phyla	650 ± 100	68.3 ± 0.9	5.3	—	—
8. Between kingdoms	1100 ± 200	72.5 ± 1.2	3.3	—	—

Species compared and other conventions are the same as in Table 1.

GPDH, where the apparent rate of evolution increases as the organisms compared become more remote. The hypothesis of the generation–time effect cannot account either for the divergent patterns of evolution of both GPDH and SOD, since the same set of species are compared in both cases, and thus identical generation times have been involved at all times in the evolution of these species. Similarly, the postulate of slightly deleterious mutations or other subsidiary hypotheses may be adjusted to account for the evolution of one or the other protein, GPDH and SOD, but not for both, without stretching ad hoc their elasticity to make the molecular clock hypothesis universally applicable to any possible empirical state of affairs and, therefore, without any predictive power and untestable—i.e., empirically empty (24, 25).

It may be possible to ascertain the biological processes that account for the unclock-like patterns of evolution of a particular gene or protein; they will usually be constraints with natural selection implications. In the case of GPDH evolution in *Drosophila*, it has been ascertained that amino acid replacements are tolerated at very few sites, and that only two or three amino acids are selectively accepted at those sites (6, 26). The empirical observation is "homoplasy;" that is, identical amino acid replacements repeatedly occur in independently evolving lineages. But the issue is not whether biologically ascertainable processes are at work, which of course they are. The issue rather is whether the processes are of such regularity that a molecular clock may be assumed to be, at least approximately, at work. The stark contrast between the pattern of evolution of GPDH and SOD may be an aberration rather than representative of prevailing erratic modes of protein evolution. This may very well be the case, since so often protein evolution seems to behave in a clock-like manner. But the congruence between observations and the clock predictions may often be simply due to the fact that the data collected do not have sufficient resolution to exhibit detectable discrepancies (and thus to provide meaningful tests of the clock hypothesis).

The congruence between empirical observations and the clock predictions may also occur because of the convergence associated with the "law of large numbers." One might observe that people take on average half as long to travel from Los Angeles to Vancouver than from Los Angeles to Toronto (which is twice as distant), an observation derived from a very large sample of traveling individuals, some of whom might be traveling by car, others by plane, some might go directly, and others stop along the way. It would be unwarranted to infer from that observation that two particular persons who had traveled for equal time lengths had also traveled the same distances, or more generally, to infer distances between two places by noting how long any one person had taken to travel from one to the other. The rate of GPDH evolution in the Drosophilids differs between lineages by a factor of 15, as noted earlier. This is the same proportional difference that exists between a traveler leisurely driving his car at 40 miles per hour on a country road, and another flying over the Atlantic at 600 miles per hour. To postulate that these two rates are random variations of the same constant rate, would stretch the elasticity of the most accommodating molecular clock hypothesis to the breaking point.

Surely, protein evolution is typically more regular than the extremes detected in GPDH. But it may also be that protein evolution may be subject to more functional constraints than has often been assumed in the past, and that these constraints may substantially vary between times and lineages (27). However, evolution at the nucleotide level, particularly with respect to synonymous codons or other sites unlikely to be subject to important functional constraints, is likely to be more regular and, thus, more dependable as a molecular clock. This is, indeed, the case for *Gpdh* and *Sod*, where the rate of synonymous nucleotide substitutions (K_s) is fairly linear over the times examined, in contrast to the rates of nonsynonymous (K_a) substitution or amino acid replacement.

We are very grateful to Walter M. Fitch and Richard R. Hudson for valuable comments. This research is supported by National Institutes of Health Grant GM42397.

1. O'Brien, S. J. & MacIntyre, R. J. (1978) in *The Genetics and Biology of Drosophila*, eds. Ashburner, M. & Wright, T. R. F. (Academic, New York), pp. 395–551.
2. Bewley, G. C. (1983) in *Isozymes: Current Topics in Biological and Medical Research*, eds. Rattazzi, M. C., Scandalios, J. G. & Whitt, G. S. (Liss, New York), pp. 33–62.
3. Bewley, G. C., Cook, J. L., Kusakabe, S., Mukai, T., Rigby, D. L. & Chambers, G. K. (1989) *Nucleic Acids Res.* **17,** 8553–8567.
4. von Kalm, L., Weaver, J., DeMarco, J., MacIntyre, R. J. & Sullivan, D. T. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5020–5024.
5. Cook, J. L., Bewley, G. C. & Shaffer, J. B. (1988) *J. Biol. Chem.* **263,** 10858–10864.
6. Kwiatowski, J., Krawczyk, M., Jaworski, M., Skarecky, D. & Ayala, F. J. (1996) *J. Mol. Evol.*, in press.
7. Barrio, E. & Ayala, F. J. (1996) *Mol. Phylogenet. Evol.*, in press.
8. Grimaldi, D. (1990) *Bull. Am. Mus. Nat. Hist.* **197,** 1–139.
9. Kwiatowski, J., Skarecky, D., Bailey, K. & Ayala, F. J. (1994) *J. Mol. Evol.* **38,** 443–454.
10. Fitch, W. M. & Ayala, F. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6802–6807.
11. Kimura, M. (1980) *J. Mol. Evol.* **16,** 111–120.
12. Li, W.-H. (1993) *J. Mol. Evol.* **36,** 96–99.
13. Dayhoff, M. D. (1978) *Atlas of Protein Sequences and Structure* (Natl. Biomed. Res. Found., Washington, DC).
14. Kwiatowski, J., Skarecky, D. & Ayala, F. J. (1992) *Mol. Phylogenet. Evol.* **1,** 72–82.
15. Wilson, A. C., Carlson, S. S. & White, T. J. (1977) *Annu. Rev. Biochem.* **46,** 573–639.
16. Ayala, F. J. (1986) *J. Hered.* **77,** 226–235.
17. Lee, Y. M., Friedman, D. J. & Ayala, F. J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 824–828.
18. Zuckerkandl, E. & Pauling, L. (1965) in *Evolving Genes and Proteins*, eds. Bryson, V. & Vogel, H. J. (Academic, New York), pp. 97–166.
19. Kimura, M. (1968) *Nature (London)* **217,** 624–626.
20. Kimura, M. (1983) *The Neutral Allele Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge).
21. Gillespie, J. H. (1991) *The Causes of Molecular Evolution* (Oxford Univ. Press, New York).
22. Ohta, T. (1976) *Theor. Popul. Biol.* **10,** 254–275.
23. Li, W.-H., Ellsworth, D. L., Krushkal, J. K., Chang, B. H.-J. & Hewett-Emmett, D. (1996) *Mol. Phylogenet. Evol.* **5,** 182–187.
24. Popper, K. R. (1959) *The Logic of Scientific Discovery* (Hutchinson, London).
25. Ayala, F. J. (1994) *Hist. Philos. Life Sci.* **16,** 205–240.
26. Wells, R. S. (1996) *Proc. R. Soc. London Ser. B* **263,** 393–400.
27. Fitch, W. M. (1976) in *Molecular Evolution*, ed. Ayala, F. J. (Sinauer, Sunderland, MA), pp. 160–178.