J Mol Evol (1998) 47:369–371

JOURNAL OF MOLECULAR EVOLUTION

Letter to the Editor

Early Metazoan Divergence Was About 830 Million Years Ago

Xun Gu

Institute of Molecular Evolutionary Genetics and Department of Biology, the Pennsylvania State University, University Park, PA 16802, USA

Received: 21 November 1997 / Accepted: 18 February 1998

Abstract. From a total of 22 nuclear genes, we estimate that the divergence time between *Drosophila* and vertebrates was about 830 million years ago (mya), which is significantly (1% level) earlier than the Cambrian explosion indicated by the early triploblastic fossils (<600 mya).

Key words: Molecular clock — Divergence time estimation — Cambrian explosion — Triploblastic metazon divergence

Molecular clock analyses have challenged the conventional wisdom about the emergence of several metazoan phyla in late Vendian and early Cambrian periods [<600 million years ago (mya)] (e.g., Runnager 1982; Doolittle et al. 1996; Wray et al. 1996; Nikoh et al. 1997). However, the earliest triploblastic metazoan divergences have been dated from 600 up to 1200 mya. These results have prompted considerable debate about the time of emergence of the earliest triploblasts, the gradual or explosive emergence of diverse body forms, and the conflict between molecular evidence and fossil records (Runnager 1982; Doolittle et al. 1996; Wray et al. 1996; Bell 1997; Morris 1997; Vermeij 1996).

The estimation of species divergence times from molecular data is a statistical problem. It is known that estimates based on one or only a few genes may be subject to large sampling variances and greatly influenced by gene-specific deviations from rate constancy. This problem can be addressed by two approaches: one is to find universal-clock genes for time estimation (e.g., Nikoh et al. 1997), and the other is to use a large number of independent genes to improve the accuracy of time estimation (Doolittle et al. 1996; Hedges et al. 1996; Gu 1997). In this Letter, we have used a larger number of independent constant-rate genes to date the Arthropoda–Vertebrate split, which constitutes one of the earliest divergences in triploblastic animal evolution.

From GenBank, as well as the data set compiled by Doolittle et al. (1996), we obtained the amino acid sequences of the nuclear gene sequences from *Drosophila,* vertebrates, and an outgroup species (e.g., fungus, bacterium). For each protein, we conducted relative rate tests by Gu and Li's (1992) method, and genes in which the assumption of rate consistency was not rejected were included in the final analysis. The final data set contained a total of 22 nuclear genes. (Table 1). Kimura's (1983) protein distance was used for estimating the sequence divergence.

Time estimation requires the calibration time *t;* in the current study it is based on the vertebrate fossil record (vertebrate clock) and/or the animal–fungus divergence time (animal–fungus clock). Then, the arthropod–vertebrate divergence time (T) can be simply estimated by the formula

$$T = (D/d)t \quad (1)$$

where d is the average distance between the species (e.g., human–bird) used for calibration and D is the average

Correspondence to: X. Gu; *e-mail:* xungu@imeg.bio.psu.edu

Table 1. Estimates of arthropod–vertebrate divergence time

Gene	Length (aa)[a]	Divergence time (mya)	
		Internal clock[b]	External clock[c]
Inosine monophosphate dehydrogenase	337	—	693
HMG–CoA reducase	403	985	1375
Glucose 6-phosphate dehydrogenase	483	617	814
Cu–Zn superoxide dismutase	153	274	935
Glyceraldehyde-3-phosphate dehydrogenase	292	554	748
Dihydroorotate oxidase	291	—	1012
Methionine adenosyl transferase	393	—	935
Phosphoglycerate kinase	407	1171	1001
DNA-directed RNA polymerase II	284	—	506
Trypsin	217	333	—
Orotidine phosphate decarboxylase	237	638	1232
Phospho*eno*lpyruvate carboxykinase	504	789	—
Enolase	366	600	638
Peptidyl prolyl isomerase	161	888	781
Triose phosphate isomerase	229	1248	726
DNA topoisomerase I	463	1600	781
Glutamate–ammonia ligase	323	995	—
Hsp70	567	837	594
α-Carbonic anhydrase	260	485	—
Cytochrome *c*	85	820	671
β-Globin	96	995	—
Elongation factor G/2	382	—	693
Weighted average		851 ± 81	830 ± 55
Simple average		813 ± 82	831 ± 55

[a] aa, amino acids.
[b] Based on within-vertebrate divergence times.
[c] Based on animal–fungus divergence times.

distance between the vertebrates and *Drosophila*. For each gene, we used divergence times of 100 mya (human/mouse) and/or 310 mya (mammal/bird) for calibrating the vertebrate clock (Hedge et al. 1995). For 17 genes (Table 1), the estimate of *Drosophila*–vertebrate divergence time based on the vertebrate clock was 853 ± 81 mya, weighted by the sequence length, or 813 ± 82 mya, based on a simple arithmetic average.

On the other hand, our 17 gene comparisons have shown that, on average, the *Drosophila*–vertebrate divergence time is ~75% of that between animal and fungus. Thus, when a divergence time (*t*) of 1100 mya was used for calibration based on the animal–fungus clock (Doolittle et al. 1996), the weighted-average *Drosophila*–vertebrate divergence time was 830 ± 55 mya, and the simple average divergence time was 831 ± 55 mya (Table 1). Note that the estimates from HMG–CoA reductase and orotidine phosphate decarboxylase calibrated by the animal–fungus clock are larger than 1100 mya (Table 1), indicating that the evolutionary rate may not be constant between animals and fungi. However, the rate-constant assumption cannot be rejected by using bacterial sequences as outgroups, because of the low statistical power of the test when the outgroup is distantly related. This problem can be addressed in the future if many genes from primitive eukaryotes are available.

It is known that Kimura's (1983) protein distance is a good approximate for the Dayhoff model but does not consider the rate variation among sites sufficiently. Therefore, we examined this problem using the distance measure developed by Gu (1997) and found similar results (not shown).

Thus the molecular time estimates for Arthropoda–vertebrate divergence based on internal (vertebrate) and external (animal–fungus) clocks are quite similar (~830 mya), and are about 280 million years older than those suggested by the earliest triphoblastic fossils (~550 mya); the difference is statistically significant at the 1% level. Therefore, the diversification of the metazoan phyla began much earlier in the Precambrian than previously thought. Incidentally, our estimates are quite close to those of Conway Morris's (1997) "compromise view," in which the marked increase in the oxygen levels in the ocean and atmosphere is considered to play a role in the Precambrian evolution of animal phyla (Vermeij 1996; Knoll 1996).

Wray et al. (1996) suggested that the early triploblastic metazoan divergences occurred about 1200 mya and that the emergence of diverse body forms was gradual rather than "explosive" as is currently believed. Although our study supports Wray and co-workers' (1996) notion, it seems that our estimate for the vertebrate–

invertebrate divergence time is more reliable than theirs, because mitochondrial genes they used are not evolving in a clocklike fashion in vertebrates (Nikoh et al. 1997).

A number of factors can cause different time estimates from different molecular data (for detailed discussions, see Nikoh et al. 1997; Gu 1997). First, it is unclear how to determine the weighting for minimizing the sampling variance and estimation bias. Although a weighting inversely proportional to the sampling variance is theoretically desirable, it can cause some biases because the "estimated" sampling variance is positively correlated with the estimate. Instead, we simply used the sequence length for weighting, as it is inversely proportional to the sampling variance. Nevertheless, it seems that different weightings do not change our conclusion significantly. Second, the orthologous relationship is not always unambiguous. We have minimized the bias caused by this problem by conducting phylogenetic analyses for all available sequences and comparing our sequence data with those from previous studies (e.g., Doolittle et al. 1996). Another related issue concerns gene duplications in chordate evolution. Many genes in vertebrate genomes may have more than one copy (e.g., enolase, HSP70). In our study we simply chose one rate-constant copy, because our preliminary result has shown that using another gene copy would give a similar result (not shown). Third, it is desirable to estimate the divergence time from not only many genes but also many species, so that both gene-specific and lineage-specific effects can be taken into account. As sequence data (from different organisms) are not sufficient for this purpose, several alternative approaches were suggested as compromises between using multigenes and using multiorganisms (e.g., Wray et al. 1996; Doolittle et al. 1996; Nikoh et al. 1997). In our study, we used many genes but relatively few calibration points because, for time estimation, gene-specific effects may be more serious than lineage-specific effects (Hedge et al. 1996; Gu 1997). Fourth, there are several options for converting evolutionary distances to divergence times. In Doolittle and co-workers' (1996) study, each point (comparison) in the linear regression is the average difference over genes; e.g., the distance of mammals was from 43 genes but that of tetrapod–fish was from only 4 genes. Since the linear regression by Doolittle et al. (1996) included eukaryotes and prokaryotes, which may overall underestimate divergence times overall [see Gu (1997) for a detailed discussion], their estimate (~670 mya) can be regarded as the low bound of the divergence time between vertebrates and *Drosophila.*

Since these problems are still controversial and require further study, interpretation of molecule dating should be cautious (Gu 1997). Keeping this in mind, we conclude that the earliest triploblastic animal phyla emerged in the early or middle rather than the late neoproterozic (about 830 mya). This result is consistent with Nikoh and co-workers' (1997) conclusion that the divergence time between cephalochordates (amphioxus) and vertebrates was about 700 mya. Whether the animal phyla in the Precambrian diverged over a span of tens of millions of years (explosive radiation) or hundreds of millions of years (gradual evolution) can be examined from molecular data only if the early deuterostome and protostome animal divergences can be dated from a large number of constant-rate genes. At present, the apparent paucity of orthologous gene sequences from most early-diverging animal phyla prevents us from conducting reliable molecular-clock analyses to examine the relative merits of these two hypotheses.

Acknowledgments. The author thanks S. Kumar, M. Nei, and S. B. Hedges for critical comments on an early version of the manuscript. This work was supported by NSF and NIH grants to Masatoshi Nei.

References

Bell MA (1997) Origin of metazoan phyla: Cambrian explosion or proterozoic slow burn? Trends Ecol Evol 12:1–2

Conway Morris S (1997) Molecular clocks: defusing the Cambrian explosion? Curr Biol 7:R71–R74

Doolittle RF, Feng DF, Tsang S, Chao G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477

Gu X (1997) The age of common ancestor of eukaryotes and prokaryotes: statistical inferences. Mol Biol Evol 14:861–866

Gu X, Li WH (1992) Higher rates of amino acid substitution in rodents than in human. Mol Phylogenet Evol 1:211–214

Hedges SB, Parker PH, Sibley CG, Kumar S (1996) Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226–229

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Knoll AH (1996) Breathing room for early animals. Nature 382:111–112

Nikoh N, Iwabe N, Kuma K, Ohno M, Sugiyama T, Watanabe Y, Yasui K, Zhang S, Hori K, Shimura Y, Miyata T (1997) An estimate of divergence time of Parazoa and Eumetazoa and that of Cephochordata and Vertebrata by aldolase and triose phosphate isomerase clocks. J Mol Evol 45:97–106

Runnager B (1982) A molecular-clock date for the origin of animal phyla. Lethaia 15:199–205

Vermeij G (1996) Animal origins. Science 274:525–526

Wray GA, Levinton JS, Shapiro LH (1996) Molecular evidence for deep Precambrian divergences among metazoan phyla. Science 274:568–573