J Mol Evol (1989) 28:279–285

Journal of Molecular Evolution

Deletions in Processed Pseudogenes Accumulate Faster in Rodents than in Humans

Dan Graur,[1] Yuval Shuali,[1] and Wen-Hsiung Li[2]

[1] Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel
[2] Center for Demographic and Population Genetics, University of Texas Health Science Center at Houston, PO Box 20334, Houston, Texas 77225, USA

Summary. The relative rates of point nucleotide substitution and accumulation of gap events (deletions and insertions) were calculated for 22 human and 30 rodent processed pseudogenes. Deletion events not only outnumbered insertions (the ratio being 7:1 and 3:1 for human and rodent pseudogenes, respectively), but also the total length of deletions was greater than that of insertions. Compared with their functional homologs, human processed pseudogenes were found to be shorter by about 1.2%, and rodent pseudogenes by about 2.3%. DNA loss from processed pseudogenes through deletion is estimated to be at least seven times faster in rodents than in humans. In comparison with the rate of point substitutions, the abridgment of pseudogenes during evolutionary times is a slow process that probably does not retard the rate of growth of the genome due to the proliferation of processed pseudogenes.

Key words: Processed pseudogenes — Rate of substitution — Deletions — Insertions — Genome size

Introduction

Pseudogenes are DNA sequences that bear structural similarity to functional genes but are prevented from being expressed properly due to mutational defects (Proudfoot 1980; Proudfoot and Maniatis 1980; Vanin 1985). Because, with few exceptions, all pseudogenes lack function, one can assume that all mutations occurring in pseudogenes are free from selective constraints, and thus, may be randomly fixed in populations. Therefore, the pattern and rate of substitution in pseudogenes should reflect the pattern and rate of spontaneous mutations (Gojobori et al. 1982).

Pseudogenes arise in evolution by one of three pathways. The first pathway involves the duplication of a functional gene and the subsequent accumulation of inactivating mutations in one of the resulting duplicates (Lacy and Maniatis 1980; Proudfoot and Maniatis 1980; Brisson and Verma 1982; Proudfoot et al. 1982; Hutchinson et al. 1983; Vanin 1983). The second pathway is the duplication of a preexisting pseudogene. Some examples of such pseudogenes exist in the literature (e.g., Cleary et al. 1981). The third mechanism of creating pseudogenes involves the reverse transcription of a processed messenger RNA, and the subsequent incorporation of the resulting complimentary DNA (cDNA) back into the genome. These retropseudogenes or processed pseudogenes, so called because they lack introns, possess oligo-A sequences at their 3′ ends, are usually unlinked to their functional homologs, and sometimes bear remnants of post-transcriptional modifications (Chen et al. 1982; Hollis et al. 1982; Reilly et al. 1982; Vanin 1985), are ubiquitous throughout the mammalian genome.

Processed pseudogenes have a unique property that makes them ideal for evolutionary studies. They are, with very few exceptions (e.g., McCarrey and Thomas 1987), dead-on-arrival sequences, i.e., they are devoid of function from the the moment they are incorporated back into the genome. Thus, their sequences can be used to infer the rates of spontaneous deletions and insertions and also to compare

Offprint requests to: D. Graur

these rates with the rate of point substitution (mutation). In comparison, when dealing with nonprocessed pseudogenes, one must take into account the so-called nonfunctionalization time, i.e., the time it takes from the moment of duplication to the time one copy loses its function (Li et al. 1981). Without taking this factor into account, one is prone to underestimate the rate of evolution in pseudogenes. However, in order to estimate the nonfunctionalization time, one must have independent knowledge of the time of divergence between either the gene and its conspecific pseudogene, or between two orthologous functional genes, i.e., the time of divergence between two species (Li et al. 1981). This type of information is usually not available.

The aim of the present paper is to study the rate of molecular evolution of processed pseudogenes. We are interested in both point substitution and gap events (deletions and insertions). In this study we incorporate the findings of Wu and Li (1985) and Li et al. (1987) to the effect that the rates of nucleotide substitution are higher in rodents than in humans.

Data and Methods

Twenty-two human, 14 murine, and 16 rat processed pseudogene sequences were gathered from the literature and aligned to their conspecific functional homologs by the methods of Needleman and Christian (1970) and Wilbur and Lipman (1985). Lack of all the introns that are present in the functional homologs was the only criterion used in determining whether a pseudogene is processed or not, and consequently, whether it should be included in the present study or not. For this reason, pseudogenes derived from intronless functional genes, e.g., the human interferon α pseudogenes, were not included. We also disregarded truncated pseudogenes, regardless of the size of their 5′ or 3′ truncation, and despite the fact that most truncated pseudogenes are processed. The reason is that in truncated genes it is difficult to know how much of the total length of the truncation was deleted as a result of the initial processing of the mRNA and how much is due to the accumulation of subsequent deletions adjacent to the 5′ and 3′ termini.

The number of nucleotide differences between a pseudogene and its conspecific functional homolog was translated into the number of substitutions per site between the two sequences by the method of Jukes and Cantor (1969), and denoted as d. Because our sample of genes contains only conservative ones, the great majority of the nucleotide substitutions should have occurred in the pseudogenes. Therefore, d may be regarded as an approximate indicator of the age of the pseudogene. An alternative to using d would be to compute the number of substitutions at fourfold degenerate sites (Li et al. 1985a) that vary less between genes. This procedure, however, would reduce the number of sites drastically and increase the error of the estimate.

With respect to gap events, it was assumed that all insertions and deletions have occurred in the pseudogene and not in the functional gene, where they could have caused frameshifts in the coding sequence with consequent deleterious effects. In 36 out of 52 cases, we could compare the conspecific functional homolog with a homologous gene from a different species (an ortholog), and thus ascertain with greater confidence in which lineage the gap occurred, and whether it was a deletion or an insertion. (In all the cases marked with an asterisk in Table 1, we have independent corroboration that the gaps indeed occurred in the pseudogene line.) The number and size of gaps were recorded. For each pseudogene, we also recorded the total lengths of deletions and insertions as percentages of the length of the respective functional gene. For reasons that will become apparent later, the data from mouse and rat were pooled together. The genes and the pseudogenes that constituted the data base, as well as the number of point nucleotide substitutions (d) and the percentage of deletions and insertions are listed in Table 1, together with the bibliographical sources.

Results

Our sample of pseudogenes contains 75 deletion events in human pseudogenes and 54 events in rodent pseudogenes. Insertions occur much less frequently, for our sample contains only 11 insertions in humans and 16 insertions in rodents. The mean length of deletions is significantly lower in humans (2.36 ± 0.32 nucleotides per deletion) than in rodents (4.54 ± 0.87). Fifty-seven percent of all deletions in humans involve single-base deletions. In rodents, 50% of all deletions are single-base ones. In both humans and rodents, deletion events significantly outnumbered insertions (Wilcoxon signed-rank matched-pair test, $P < 0.01$ and $P < 0.05$, respectively). On the average, there were 3.41 ± 0.71 deletion events and 0.50 ± 0.15 insertion events per human pseudogene, and 1.80 ± 0.44 deletions and 0.53 ± 0.20 insertions per rodent pseudogene, the ratio of deletions to insertions being approximately 7:1 and 3:1 for human and rodent pseudogenes, respectively. In addition, the total length of deletions was greater than that of insertions. In fact, about 1.2% of its total length is deleted, whereas only 0.1% is inserted in an average human processed pseudogene. The corresponding numbers in the rodent are 2.3% and 1.1%. The differences between the total lengths of deletions and insertions are highly significant for the human lineage (paired $t = 4.299$, $P < 0.0001$). The difference was not statistically significant in rodents due to the presence of a 125-base insertion in the rat α-tubulin pseudogene. Our pseudogene data support de Jong and Rydén's (1981) hypothesis that the preponderance of deletions among gap events is inherent in the mutational process, rather than being a by-product of selection.

Next, we looked for a relationship between the rate of deletions and insertions and the rate of point nucleotide substitution in processed pseudogenes. We find that the number of deletion events is significantly correlated with d in both the human and rodent lineages ($r = 0.677$ and $r = 0.597$, respectively). Similarly, the percentage of deleted DNA is positively correlated with d ($r = 0.477$ and 0.642, for humans and rodents, respectively). Moreover,

Table 1. The number of point substitutions per nucleotide site (d), and the relative lengths of gaps (insertions and deletions) between human, rat, and mouse processed pseudogenes and their functional conspecific homologs. Continued on next page

Pseudogene	Length of coding region	d	Lengths of insertions (total %)	Lengths of deletions (total %)	Source[a]
Human					
Metallothionein II*	180	0.028	None	None	(1)
Apoferritin H-133	429	0.036	None	None	(2)
β-tubulin-21B*	1332	0.042	None (0.000)	1 (0.075)	(3)
Glycerate-3-P dehydrogenase-X*	1002	0.042	None (0.000)	3 (0.299)	(4)
Apoferritin H-123	429	0.045	None	None	(2)
Phosphoglycerate kinase-X	1248	0.054	None (0.000)	1, 1, 5 (0.561)	(5)
Argininosuccinate synthetase-1	1233	0.071	3, 6 (0.728)	1, 1, 3, 9 (1.127)	(6)
Dihydrofolate reductase-1*	558	0.076	1 (0.179)	1, 4 (0.896)	(7)
Argininosuccinate synthetase-3	1233	0.077	None (0.000)	1, 1, 5 (0.649)	(6)
Triosephosphate isomerase 19A*	747	0.079	1 (0.134)	1, 1, 1, 2, 3, 4 (1.163)	(8)
Chromosomal protein HMG17-28	267	0.081	1 (0.375)	12 (4.494)	(9)
Triosephosphate isomerase 5A*	747	0.087	1 (0.134)	1, 3, 3, 4, 5 (2.142)	(8)
Chromosomal protein HMG17-60	267	0.092	None (0.000)	1 (0.375)	(9)
Lactate dehydrogenase A-H463*	993	0.097	None (0.000)	3, 3 (0.604)	(10)
β-actin-2*	1128	0.101	3, 3 (0.532)	1, 1, 2, 3, 19 (2.305)	(11)
β-tubulin-14B*	1332	0.104	None (0.000)	1, 1 (0.376)	(3)
β-actin-1*	1128	0.109	2 (0.178)	1, 1, 1, 3 (0.535)	(11)
β-tubulin-7B*	1332	0.134	None (0.000)	1, 1, 1, 1, 1, 1, 2, 2, 2 (3.228)	(3)
Metallothionein I*	180	0.141	None (0.000)	1 (0.556)	(12)
Cu/Zn superoxide dismutase-69.1	459	0.177	1 (0.218)	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3 (3.484)	(13)
Cu/Zn superoxide dismutase-71.4	459	0.191	1 (0.218)	1, 1, 1, 2, 3 (1.743)	(13)
β-tubulin-11B*	1332	0.227	None (0.000)	1, 1, 2, 2, 3, 4, 9 (1.652)	(3)
Mouse					
Ribosomal protein L32-4A	402	0.002	None	None	(14)
Ribosomal protein L30-2*	342	0.009	None	None	(15)
Ribosomal protein S16-P	123	0.029	1, 3 (3.252)	None (0.000)	(16)
Ribosomal protein L30-4*	342	0.032	None	None	(15)
Cellular tumor antigen P53*	1173	0.035	1, 1, 6 (0.682)	1, 2, 5 (0.682)	(17)
Cytochrome c-MC3*	312	0.038	None (0.000)	16 (5.128)	(18)
Lactate dehydrogenase A-M11*	993	0.059	None (0.000)	1, 1, 3, 3 (0.806)	(19)
Cytochrome c-MC4*	312	0.075	None (0.000)	1, 3 (1.282)	(17)
Ribosomal protein L30-3*	342	0.083	1 (0.292)	None (0.000)	(15)
Lactate dehydrogenase A-M14*	993	0.093	None (0.000)	1, 1, 1, 1, 1, 1, 2, 11 (1.854)	(19)

Table 1. Continued from previous page

Pseudogene	Length of coding region	*d*	Lengths of insertions (total %)	Lengths of deletions (total %)	Source[a]
Lactate dehydrogenase A-M10*	993	0.095	None (0.000)	1, 1, 4 (0.604)	(19)
Cytochrome c-MC2*	312	0.110	None (0.000)	1, 3, 7 (3.352)	(17)
α-globin-3*	423	0.175	1, 2, 6, 9 (4.255)	3, 3, 3, 28 (8.745)	(20)
α-globin-30.5*	993	0.216	None (0.000)	1, 1, 1, 1, 3, 5, 21 (8.511)	(21)
Rat					
Ribosomal protein L35A-A	327	0.000	None	None	(22)
Ribosomal protein L35A-B	327	0.000	None	None	(22)
Metallothionein I-B*	180	0.000	1, 3 (0.566)	None (0.000)	(12)
Apoferritin L-31*	345	0.017	None	None	(23)
Cytochrome c-RC13*	312	0.017	None (0.000)	11 (3.526)	(24)
Apoferritin L-45*	345	0.019	None	None	(23)
Cytochrome c-RC5*	312	0.019	None	None	(24)
Cytochrome c-RC9*	312	0.019	None	None	(24)
α-tubulin	1350	0.025	1, 1, 125 (9.407)	1, 1, 1, 1, 5 (0.667)	(25)
Cytochrome c-RC6*	312	0.036	None	None	(24)
Cytochrome c-RC8*	312	0.036	None	None	(24)
Metallothionein I-C*	180	0.036	None (0.000)	21 (11.677)	(12)
Glutathione-S-transferase	630	0.068	None (0.000)	1, 1, 1, 1, 1, 1, 3 (1.429)	(26)
Cytochrome c-RC10*	312	0.072	None (0.000)	2, 9, 9 (6.410)	(24)
Metallothionein I-A*	180	0.100	None (0.000)	9 (5.000)	(12)
Ribosomal protein L35A-G	327	0.147	45 (13.761)	24 (8.257)	(22)

None = no recorded gap events; * = independent corroboration that gaps occur in the pseudogene (see text)

[a] Sources for the nucleotide sequences of the pseudogenes: (1) Karin and Richards 1982; (2) Constanzo et al. 1986; (3) Wilde et al. 1982, Gwo-Shu Lee et al. 1983; (4) Benham et al. 1984, Hanauer and Mandel 1984; (5) Michelson et al. 1985; (6) Anagnou et al. 1984, Masters et al. 1983; (7) Freytag et al. 1984; (8) Brown et al. 1985; (9) Srikantha et al. 1987; (10) Tsujibo et al. 1985; (11) Moos and Gallwitz 1983; (12) Varshney and Gedamu 1984; (13) Danciger et al. 1986, Delbar et al. 1987; (14) Dudov and Perry 1984; (15) Wiedemann and Perry 1984; (16) Wagner and Perry 1985; (17) Zakut-Houri et al. 1983, Benchimol et al. 1984; (18) Limbach and Wu 1985; (19) Fukasawa et al. 1986, Fukasawa et al. 1987; (20) Vanin 1985; (21) Vanin et al. 1980; (22) Kuzumaki et al. 1987; (23) Leibold et al. 1984; (24) Scrapulla 1983, Scrapulla and Wu 1983; (25) Lemischka and Sharp 1982; (26) Okada et al. 1987

we can show that a linear relationship exists between the total percentage of deletions and the number of substitutions. In both lineages, the slopes of the linear regression lines of percent deletion on *d* were found to be significantly different from zero ($P < 0.0001$ and $P < 0.025$ in rodents and humans, respectively). Because we found no statistically significant difference between the rates of deletions in mouse and rat, the data from these two species were pooled together. The relationship between the rates of nucleotide substitution and percentage DNA loss is shown in Fig. 1. The slope of regression of the cumulative loss of DNA due to deletions on *d* is 11.41 for humans and 40.10 for rodents. Thus, the slope for rodents is about 3.5 times steeper than that for the human pseudogenes. The absolute values of the slopes in Fig. 1 are underestimates because the *d* value includes changes not only in the pseudogene but also in the functional gene, whereas almost all deletions occurred in the pseudogene. The ratio between the two rates, on the other hand, is not affected. Because rodent genes accumulate point mutations at least two times faster than human ones (Wu and Li 1985; Britten 1986; Li et al. 1987), we may deduce that the rate of DNA loss from processed pseudogenes due to deletion events is at least seven times higher in rodents than in humans. One must note, however, that our sample of genes in rodents is different from that in humans. Because of the fact that functional genes are known to have

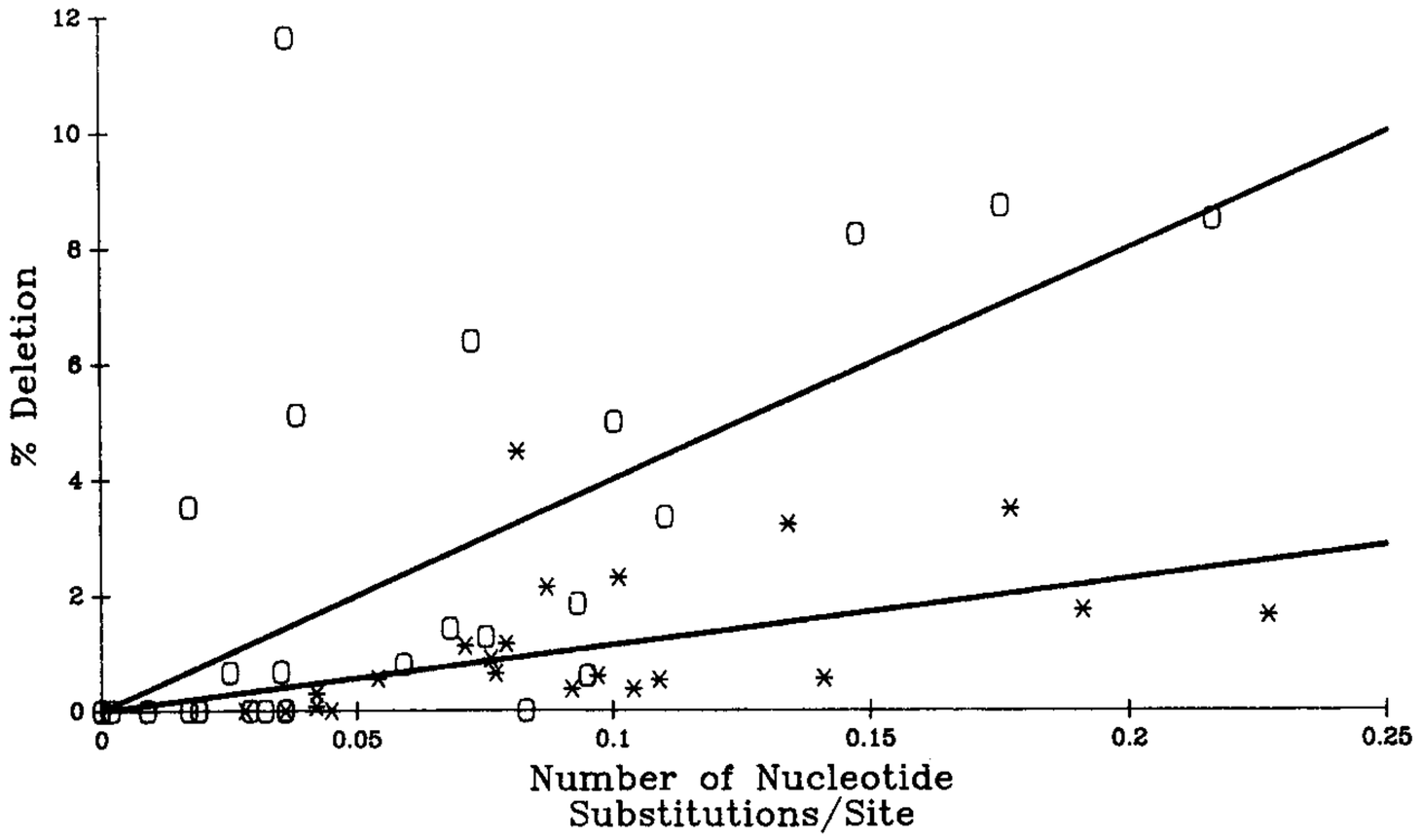

Fig. 1. The relationship between the total length of deletions and the number of nucleotide substitutions per site between a processed pseudogene and its conspecific functional homolog. Upper line = rodents; lower line = humans.

different rates of substitution, one must use the same set of genes in both groups in order to obtain an accurate estimate of the ratio. Thus, the difference in the slopes for humans and rodents may, in reality, be either larger or smaller than that in Fig. 1. To be on the conservative side, let us assume that the difference is smaller. However, because the majority of point substitutions should have occurred in the pseudogenes, the difference in slopes is unlikely to be explained completely by the fact that different genes were used.

Despite the fact that insertions clearly occur more frequently in rodents than in humans, and that the differences between rodents and humans were statistically significant whether we used the number of insertion events or the percentage of length inserted into the processed pseudogene as variables, insertions are much more infrequent than deletions for a detailed picture to emerge. Much more data than presently available is needed to estimate the rate and pattern of insertion in processed pseudogenes.

Discussion

There are two kinds of processes that affect the evolution of processed pseudogenes. The first involves the very rapid accumulation of point mutations that, being unconstrained by function, are fixed in the population according to the rules of stochastic substitution of neutral alleles. This accumulation of molecular changes eventually obliterates the similarity between the pseudogene and its functional homolog that evolves much more slowly. Indeed, identifying old pseudogenes is quite a difficult task (e.g., Hardison et al. 1986). We call this process *compositional assimilation,* whereby a pseudogene loses similarity to the functional gene and melts into the background of its surrounding DNA. From the pattern of substitution in pseudogenes, it was estimated that the nucleotide composition of pseudogenes will eventually become A+T-rich (Gojobori et al. 1982). Functional genes, on the other hand, will tend to accumulate C and G residues at degenerate sites (Ticher and Graur, in press).

The second process is characterized by the pseudogene becoming progressively shorter compared to the functional gene through the accumulation of deletions. Interestingly, rodents are at least seven times more efficient at getting rid of this particular type of junk DNA than are humans. This process, which we call *length abridgment,* is by far a much slower process than the one of compositional assimilation. Previously we noted that the relative rate of deletions to substitutions is approximately 40% length deletion per nucleotide substitution per site in rodents and about 10% in humans. Assuming that during long-term evolution the relative rate of deletions to substitutions is 25%, and by using Li et al.'s (1985b) estimate for the rate of substitution in pseudogenes (5.0×10^{-9} substitutions/site/year), we can estimate that a processed pseudogene loses half of its DNA in 400 million years. This means, for instance, that the human genome still contains major chunks of most of the pseudogenes that were found in the ancestor of all mammals, including monotremes, marsupials, and eutherians. Obviously, these ancient pseudogenes had lost any recognizable similarity to the functional genes much earlier. Possible mechanisms involved in the deletion process are summarized in Levinson and Gutman (1987) and in Walsh (1987). Initial results show that most deletions in pseudogenes are in fact reductions in the number of repeated elements within short simple tandem arrays (Landan, Li, and Graur, unpublished).

P. Leder (cited in Lewin 1981) referred to the generation of processed pseudogenes as the Vesuvian model, whereby a functional locus continuously pumps out defective copies of itself and dis-

perses them all over the genome. As a consequence, the genome size increases during evolutionary times due to the accumulation of junk DNA. Our findings on the relative inefficiency of the deletion process in mammalian genomes, in conjunction with the fast accumulation of nucleotide substitutions, indicates that the mammalian genome is literally littered with processed pseudogenes in various stages of compositional assimilation. Because in many multigene families the number of processed retropseudogenes far exceeds that of the functional homologs (for the most extreme example to date, see Srikantha et al. 1987), we believe that processed pseudogenes are created at a much faster rate than they are obliterated by the process of pseudogene abridgment. Thus, the growth of the genome is not significantly retarded by the occurrence of deletions. The old processed pseudogenes are too divergent from their functional paralogs to be recognizable as such by either molecular probes or by computer searches for similarity in DNA data banks. On the other hand, the difference in the rates of deletion between rodents and humans may be one of the factors accounting for the fact that the nuclear genome of rodents is approximately 2.94×10^9 bp in size, whereas the human genome is a little larger, about 3.43×10^9 bp (Cavalier-Smith 1985).

Acknowledgments. We thank Aharon Ticher and Robert Schwartz for their advice on handling computerized nucleotide data banks. This work was supported in part by research grants from the Foundation for Basic Research, Tel Aviv University, the Hertz Foundation, and by NIH grant GM30998.

References

Anagnou NP, O'Brien SJ, Shimada T, Nash WG, Chen MJ, Nienhuis AW (1984) Chromosomal organization of the human dihydrofolate reductase genes: dispersion, selective amplification, and a novel form of polymorphism. Proc Natl Acad Sci USA 81:5170–5174

Benchimol S, Jenkins JR, Crawford LV, Leppard K, Lamb P, Williamson NM, Pim DC, Harlow E (1984) Molecular analysis of the gene for the p53 cellular tumor antigen. Cancer Cells 2:383–391

Benham FJ, Hodgkinson S, Davies KE (1984) A glyceraldehyde-3-phosphate dehydrogenase pseudogene on the short arm of the human X chromosome defines a multigene family. EMBO J 3:2635–2640

Brisson N, Verma DPS (1982) Soybean leghemoglobin gene family: normal, pseudo and truncated genes. Proc Natl Acad Sci USA 79:4055–4059

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Brown JR, Dear IO, Krug JR, Masquat LE (1985) Characterization of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family. Mol Cell Biol 5:1694–1706

Cavalier-Smith T (1985) Eukaryote gene numbers, non-coding DNA and genome size. In: Cavalier-Smith T (ed) The evolution of genome size. Wiley, London, pp 69–103

Chen MJ, Shimada T, Moulton AD, Harrison M, Nienhuis AW (1982) Intronless human dihydrofolate reductase genes are derived from processed RNA molecules. Proc Natl Acad Sci USA 79:7435–7439

Cleary ML, Schon EA, Lingrel JB (1981) Two related pseudogenes are the result of gene duplication in the goat β-globin locus. Cell 26:181–190

Constanzo F, Colombo M, Staempfli S, Santoro C, Marone M, Frank R, Delius H, Cortese R (1986) Structure of gene and pseudogenes of human apoferritin H. Nucleic Acids Res 14: 721–736

Danciger E, Dafni N, Bernstein Y, Laver-Rodich Z, Neer A, Groner Y (1986) Human copper-zinc superoxide dismutase gene family molecular structure and characterization of four copper-zinc superoxidase related pseudogenes. Proc Natl Acad Sci USA 83:3619–3623

de Jong WW, Rydén L (1981) Causes of more frequent deletions than insertions in mutations and protein evolution. Nature 290:157–159

Delbar JM, Nicole A, Davriol L, Meunier-Rotival M, Galibert F, Sinet PM, Jerome H (1987) Cloning and sequencing of a rat copper-zinc superoxide dismutase complementary DNA. Eur J Biochem 166:181–188

Dudov KP, Perry RP (1984) The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed pseudogene. Cell 37:457–468

Freytag SO, Bock HGO, Beaudet AL, O'Brien WE (1984) Molecular structure of human argininosuccinate synthetase pseudogenes. Evolutionary and mechanistic implications. J Biol Chem 259:3160–3166

Fukasawa KM, Li WH, Yagi K, Luo CC, Li SSL (1986) Molecular evolution of mammalian lactate dehydrogenase-A genes and pseudogenes: association of a mouse processed pseudogene with a B1 repetitive sequence. Mol Biol Evol 3:330–342

Fukasawa KM, Tanimura M, Sakai I, Sharief FS, Chung FZ, Li SSL (1987) Molecular nature of spontaneous mutations in mouse lactate dehydrogenase-A processed pseudogenes. Genetics 115:177–184

Gojobori T, Li WH, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18:360–369

Gwo-Shu Lee M, Lewis SA, Wilde CD, Cowan NJ (1983) Evolutionary history of a multigene family: an expressed human β-tubulin gene and three processed pseudogenes. Cell 33:477–487

Hanauer A, Mandel JL (1984) The glyceraldehyde-3-phosphate dehydrogenase gene family: structure of a human cDNA and of an X chromosome linked pseudogene; amazing complexity of the gene family in the mouse. EMBO J 3:2627–2633

Hardison RC, Sawada I, Cheng JF, Shen CKJ, Schmid CW (1986) A previously undetected pseudogene in the human alpha globin gene cluster. Nucleic Acids Res 14:1903–1911

Hollis GF, Hieter PA, McBride OW, Swan D, Leder P (1982) Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. Nature 296:321–325

Hutchinson CA, Brown BA, Davis MG, Hardies SC, Hill A, Padgett RW, Phillips SJ, Timmons BEL, Weaver SG, Edgell MH (1983) β homologous structures in the β globin locus of the mouse. In: Goldwasser E (ed) Regulation of hemoglobin biosynthesis. Elsevier, New York, pp 51–68

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Karin M, Richards RI (1982) Human metallothionein genes—primary structure of the metallothionein-II gene and a related processed pseudogene. Nature 299:797–802

Kuzumaki T, Tanaka T, Ishikawa K, Ogata K (1987) Rat ri-

bosomal protein L35a multigene family: molecular structure and characterization of three L35a-related pseudogenes. Biochim Biophys Acta 909:99–106

Lacy E, Maniatis T (1980) The nucleotide sequence of a rabbit β-globin pseudogene. Cell 21:545–553

Leibold EA, Aziz N, Brown AJP, Munro HN (1984) Conservation in rat liver of light and heavy subunit sequences of mammalian ferritin: presence of unique octopeptide in the light subunit. J Biol Chem 259:4327–4334

Lemischka I, Sharp PA (1982) The sequences of an expressed rat α-tubulin gene and a pseudogene with an inserted repetitive element. Nature 300:330–335

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Lewin R (1981) Evolutionary history written in globin genes. Science 214:426–429

Li WH, Gojobori T, Nei M (1981) Pseudogenes as a paradigm of neutral evolution. Nature 292:237–239

Li WH, Luo CC, Wu CI (1985a) A new method for estimating synonymous and nonsynonymous rates of substitution considering the relative likelihood of nucleotide and codon changes. J Mol Evol 2:150–174

Li WH, Luo CC, Wu CI (1985b) Evolution of DNA sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 1–94

Li WH, Tanimura M, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J Mol Evol 25:330–342

Limbach J, Wu R (1985) Characterization of a mouse somatic cytochrome c gene and three cytochrome c pseudogenes. Nucleic Acids Res 13:617–630

Masters JN, Yang JK, Cellini A, Attardi G (1983) A human dihydrofolate reductase pseudogene and its relationship to multiple forms of specific messenger RNA. J Mol Biol 167: 23–36

McCarrey JR, Thomas K (1987) Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature 326:501–505

Michelson AM, Bruns GAP, Morton CC, Orkin SH (1985) The human phosphoglycerate kinase family. HLA-associated sequences and an X-linked locus containing a processed pseudogene and its functional counterpart. J Biol Chem 260:6982–6992

Moos M, Gallwitz D (1983) Structure of two human β-actin related processed genes one of which is located next to a simple repetitive sequence. EMBO J 2:757–761

Needleman SB, Christian DW (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Okada A, Sakai M, Muramatsu M (1987) The structure of the rat glutathione S-transferase P gene and related pseudogenes. J Biol Chem 262:3858–3863

Proudfoot N (1980) Pseudogenes. Nature 286:840–841

Proudfoot N, Maniatis T (1980) The structure of a human α-globin pseudogene and its relationship to a α-globin gene duplication. Cell 21:537–544

Proudfoot N, Gil A, Maniatis T (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 31:553–563

Reilly JG, Ogden R, Rossi JJ (1982) Isolation of a mouse pseudo tRNA gene encoding CCA—a possible example of reverse flow of genetic information. Nature 300:287–289

Scrapulla RC (1983) Processed pseudogenes of rat cytochrome c are preferentially derived from one of three alternate mRNAs. Mol Cell Biol 4:2279–2288

Scrapulla RC, Wu R (1983) Nonallelic members of the cytochrome c multigene family of the rat may arise through different messenger RNAs. Cell 32:473–482

Srikantha T, Landsman D, Bustin M (1987) Retropseudogenes for human chromosomal protein HMG-17. J Mol Biol 197: 405–413

Ticher A, Graur D (1989) Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein coding genes. J Mol Evol (in press)

Tsujibo H, Tiano HF, Li SSL (1985) Nucleotide sequence of the cDNA and an intronless pseudogene for human lactate dehydrogenase-A isozyme. Eur J Biochem 147:9–15

Vanin EF (1983) Globin pseudogenes. In: Goldwasser E (ed) Regulation of hemoglobin biosynthesis. Elsevier, New York, pp 69–88

Vanin EF (1985) Processed pseudogenes: characteristics and evolution. Annu Rev Genet 19:253–272

Vanin EF, Goldberg GI, Tucker PW, Smithies O (1980) A mouse α-globin related pseudogene lacking intervening sequences. Nature 286:222–226

Varshney U, Gedamu L (1984) Human metallothionein MT-I and MT-II processed genes. Gene 31:135–145

Wagner M, Perry RP (1985) Characterization of the multigene family encoding the mouse S16 ribosomal protein: strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. Mol Cell Biol 5:3560–3576

Walsh JB (1987) Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics 115:553–567

Wiedemann LM, Perry RP (1984) Characterization of the expressed gene and several processed pseudogenes for the mouse ribosomal protein L30 gene family. Mol Cell Biol 4:2518–2528

Wilbur WJ, Lipman DJ (1985) Rapid similarity searches of nucleic acid and protein data banks. Proc Natl Acad Sci USA 80:726–730

Wilde CD, Crowther CE, Cripe TP, Gwo-Shu Lee M, Cowan NJ (1982) Evidence that a human β-tubulin pseudogene is derived from its corresponding mRNA. Nature 297:83–84

Wu CI, Li WH (1985) Evidence for higher rates in rodents than in man. Proc Natl Acad Sci USA 82:1741–1745

Zakut-Houri R, Oren M, Bienz B, Lavie V, Hazum S, Givol D (1983) A single gene and a pseudogene for the cellular tumor antigen p53. Nature 306:594–597

Received March 21, 1988/Revised and accepted August 18, 1988