J Mol Evol (1989) 29:208–211

Journal of
Molecular Evolution

Rates of Synonymous Substitution in Plant Nuclear Genes

Kenneth H. Wolfe,[1,2] Paul M. Sharp,[1] and Wen-Hsiung Li[2]

[1] Department of Genetics, Trinity College, Dublin 2, Ireland
[2] Center for Demographic and Population Genetics, University of Texas, P.O. Box 20334, Houston, Texas 77225, USA

Summary. The rate of synonymous nucleotide substitution in nuclear genes of higher plants has been estimated. The rate varies among genes by a factor of up to two, in a manner that is not immediately explicable in terms of base composition or codon usage bias. The average rate, in both monocots and dicots, is about four times higher than that in chloroplast genes. This leads to an estimated absolute silent substitution rate of 6×10^{-9} substitutions per site per year that falls within the range of average rates ($2–8 \times 10^{-9}$) seen in different mammalian nuclear genomes.

Key words: Plant molecular evolution — Molecular clock — Chloroplast DNA

The rate of nucleotide substitution at silent codon positions in genes is of general interest in elucidating mechanisms of DNA evolution and of particular relevance to the molecular clock hypothesis. Thus, it is desirable to know this rate in as many diverse genomes as possible. Unfortunately, the nucleotide sequence data from which these rates can be calculated are still limited to a rather small subset of organisms: approximate estimates of substitution rates are known only for various mammalian orders, *Drosophila,* enterobacteria, influenza virus, retroviruses, and some organelle genomes. Recently, we have provided evidence for large differences in the rates of evolution among higher plant nuclear, mitochondrial, and chloroplast genomes (Wolfe et al. 1987) but were unable to obtain a reliable estimate of the rate in plant nuclear genes because very few single-copy genes had been sequenced in more than one species. Since that study was completed, many plant nuclear sequences have been published and we are now able to quantify their substitution rates. We first calculate the ratio between the substitution rates in plant nuclear and chloroplast genes, before applying fossil data to estimate absolute rates.

Most nuclear genes are saturated with synonymous substitutions when compared between monocots and dicots; so instead, we compare genes between different species within the two classes of angiosperms. We present the results in terms of the estimated numbers of substitutions per site, both at synonymous sites (K_S) and at nonsynonymous sites (K_A). These estimated numbers are calculated from the observed differences between species, using the multiple-hits correction method of Li et al. (1985). Table 1 gives the results for 12 chloroplast and 6 nuclear genes compared between two monocots: maize and wheat (or its close relatives barley and rye). The synonymous substitution rate (K_S) in the nuclear data set varies considerably (by up to twofold) from gene to gene, whereas the chloroplast genes are more homogeneous with respect to K_S (the coefficient of variation among the nuclear genes is 23%, compared to 12% among the chloroplast genes, excluding the five short genes). The mean K_S seen in the nuclear genes is about four times that in the chloroplast genes.

Similar within-genome variation in synonymous substitution rate has been found in several other organisms and has been attributed either to selection among synonymous codons or to mutation rate variation. For example, in both enterobacteria and *Drosophila* the silent substitution rate in a gene is inversely related to its degree of codon usage bias (Sharp and Li 1987, 1989). This has been explained

Offprint requests to: K.H. Wolfe in Dublin

Table 1. Comparison between maize and wheat or barley genes

Gene	K_S[a]	L_S[b]	K_A[a]	L_A[b]	G+C[c]
Chloroplast genes:					
psbB	0.15 ± 0.02	343	0.01 ± 0.00	1178	28
atpA	0.15 ± 0.02	342	0.01 ± 0.00	1167	32
Five genes[d]	0.17 ± 0.03	242	0.01 ± 0.00	769	26
atpB	0.17 ± 0.02	343	0.02 ± 0.00	1148	30
atpE	0.19 ± 0.05	88	0.01 ± 0.01	321	35
psbC[e]	0.19 ± 0.04	153	0.01 ± 0.00	499	33
psbD	0.20 ± 0.03	240	0.01 ± 0.00	819	28
rbcL	0.20 ± 0.03	317	0.03 ± 0.01	1102	26
Total[f]	0.17 ± 0.01	2068	0.01 ± 0.00	7001	29
Nuclear genes:					
waxy	0.54 ± 0.06	354	0.07 ± 0.01	1233	90
adh2,3[g]	0.61 ± 0.05	251	0.07 ± 0.01	875	80
adh1	0.66 ± 0.05	253	0.03 ± 0.00	881	61
shrunken-1[e]	0.74 ± 0.12	130	0.04 ± 0.01	488	61
Chalcone synthase	0.86 ± 0.16	287	0.06 ± 0.01	907	95
gapC[e]	0.98 ± 0.15	201	0.09 ± 0.01	714	56
Total[f]	0.71 ± 0.04	1475	0.06 ± 0.00	5098	77

Rye sequences are used in place of wheat or barley for *psbB* and *orf43*. References to the original publications of the chloroplast DNA sequences are given in our recent compilation (Wolfe 1989). The nuclear sequences are from GenBank (release 58), Chojecki (1986), Brinkmann et al. (1987), Niesbach-Klösgen et al. (1987), Good et al. (1988), Maraña et al. (1988), Rohde et al. (1988), and Trick et al. (1988)

[a] K_S and K_A are the number of substitutions per synonymous and per nonsynonymous site, respectively (calculated by the method of Li et al. 1985)

[b] L_S and L_A are the numbers of synonymous and nonsynonympus sites compared

[c] Mean G+C content (%) at synonymous codon positions

[d] Sum of five genes of <100 codons: *atpH, psaC, psbH, orf43, orf62*

[e] Partial sequence

[f] The mean K_S and K_A values are weighted by the numbers of sites in each gene; the standard errors are theoretical values (on the assumption that all genes approximate the same substitution rate), calculated as in Wolfe et al. (1987)

[g] Mean of comparisons of maize *adh2* vs barley *adh2* and *adh3*

in terms of natural selection favoring certain translationally optimal codons; the intensity of selection is determined by the level of expression of the gene. On the other hand, although the synonymous substitution rate is also known to vary among mammalian nuclear genes (Li et al. 1987), there is no clear evidence of codon selection in mammals (Wolfe et al. 1989). Recently, a relationship has been found between the synonymous substitution rate in mammalian genes and their G+C content, and it may be possible to explain the rate variation in terms of differences in local mutation processes (Filipski 1988; Wolfe et al. 1989). Codon usage bias in plant nuclear genes has yet to be examined in detail, though large differences have been noted between monocots and dicots (Brinkmann et al. 1987; Salinas et al. 1988). The monocot nuclear genes in Table 1 vary considerably in G+C content, but this does not appear to be related to their synonymous substitution rates. Thus, the synonymous rate variation among plant nuclear genes remains to be explained.

We also have compared sequences between the dicot families Solanaceae and Brassicaceae (Table 2). As with the monocot nuclear data, K_S varies considerably from gene to gene (and is saturated for two genes), though the G+C contents of the dicot genes are uniform. Only four chloroplast genes can be compared between these two families, and their K_S values are rather variable (Table 2). Their mean is again about one-quarter of the mean nuclear K_S, though the nuclear genes are rather too close to saturation for the ratio to be estimated confidently. In our previous study (Wolfe et al. 1987), we suggested that the nuclear gene for plastocyanin had a very high rate of synonymous substitution. This was based on a comparison between spinach and *Silene* (white campion) sequences,[1] using a divergence date of 20–40 million years (Myr). However, we can now also compare a chloroplast gene (*rpoB*) across the same taxonomic distance [spinach vs *Saponaria* (soapwort)], and we find that the nuclear : chloroplast K_S ratio is still approximately 4 (1.19/0.28). Although it is possible that the molecular clock for both nuclear and chloroplast DNA has run more quickly in these species (order Caryophyllales) than in other angiosperms, it seems more likely that the diver-

[1] Minor corrections have been made to the *Silene* plastocyanin sequence since our previous analysis (Smeekens, personal communication)

Table 2. Comparison between genes from two dicot families: Solanaceae and Brassicaceae

Gene	K_S	L_S	K_A	L_A	G+C
Chloroplast genes:					
psbA	0.24 ± 0.04	231	0.00 ± 0.00	828	20
psbK	0.39 ± 0.12	38	0.08 ± 0.02	145	36
atpE	0.40 ± 0.08	88	0.06 ± 0.01	308	27
orfK	0.56 ± 0.06	305	0.20 ± 0.01	1192	27
Total	0.42 ± 0.03	662	0.11 ± 0.01	2473	25
Nuclear genes:					
gapC	1.29 ± 0.19	216	0.06 ± 0.01	759	41
gapA[a]	1.37 ± 0.20	167	0.06 ± 0.01	535	42
EPSPs	1.51 ± 0.17	309	0.10 ± 0.01	1021	32
Chalcone synthase	>2.50	271	0.10 ± 0.01	896	46
ALS	>2.50	405	0.10 ± 0.01	1353	38
Total[b]	1.91 ± 0.12	1368	0.09 ± 0.00	4564	39

The column headings are as explained in the footnotes to Table 1. Comparisons are between tobacco and mustard for all genes except: *atpE* (tobacco vs *Arabidopsis*), EPSPs (5-enolpyruvylshikimate-3-phosphate synthase; petunia and tomato vs *Arabidopsis*), chalcone synthase (petunia vs *Arabidopsis*), and ALS [acetolactate synthase; tobacco (2 loci) vs *Arabidopsis*]. Nuclear sequences are from GenBank, Klee et al. (1987), Mazur et al. (1987), and Lee et al. (1988). Chloroplast sequences are as compiled by Wolfe (1989), plus data from Chen et al. (1988) and Neuhaus (1989). *orfK* is the gene located within the intron of *trnK*$_{UUU}$; *gapC* and *gapA* are genes for cytosolic and plastid isozymes of glyceraldehyde-3-phosphate dehydrogenase, respectively

[a] Partial sequence

[b] Calculated by summation over genes before correction for multiple hits

gence date used is too recent, leading to an artificially high estimate of the rate.

Tables 1 and 2 show that the average synonymous substitution rate in plant nuclear genes is about four times that in chloroplast genes. In turn, we have shown previously that the synonymous rate in chloroplast genes is about three times higher than in plant mitochondrial genes. All the chloroplast genes used in the above comparisons are located in the unique regions of the chloroplast genome; there is evidence that genes located in the inverted repeat region have a synonymous substitution rate about four times lower than the rest of the chloroplast genome (Wolfe et al. 1987, unpublished results).

If we take the divergence time (T) between maize and wheat to be 50–70 Myr (Stebbins 1981; Chao et al. 1984), the nuclear data in Table 1 convert to an average synonymous rate of $K_S/2T = 5.1–7.1 \times 10^{-9}$ substitutions per site per year. This is similar to the absolute rates of synonymous substitution seen in mammalian genes: it falls between the rates calculated for primates and for rodents (2×10^{-9} and 8×10^{-9}, respectively; Li et al. 1987). Although the fossil data supporting the date of 50–70 Myr used above are rather limited, we have found using a molecular clock for chloroplast genes that this date is consistent with other fossil dates such as the divergence between angiosperms and bryophytes. The dicot data in Table 2 cannot be calibrated because there is no fossil record of the Solanaceae–Brassicaceae divergence. The ratio between mean K_S values for nuclear and chloroplast genes is the same in monocots and dicots, which suggests that the absolute rates may also be equal, though we cannot rule out the possibility that parallel rate changes have occurred in the nuclear and chloroplast genomes in one lineage.

We conclude that the synonymous substitution rates in plant mitochondrial, chloroplast, and nuclear genes are in the approximate ratio 1:3:12 and that the average absolute rates of synonymous nucleotide substitution in nuclear genes of plants and mammals are similar. Ochman and Wilson (1987) have proposed that there is a single rate of synonymous substitution in all cellular genomes (a "silent molecular clock"). The results from our laboratories have indicated considerable within-genome substitution rate heterogeneity in plants, enterobacteria, *Drosophila*, and mammals (this study; Sharp and Li 1987, 1989; Wolfe et al. 1989), and differences among mammalian lineages (Li et al. 1987). Nevertheless, it is perhaps remarkable that the mean synonymous substitution rates in each of these nuclear genomes differ by no more than a factor of 10.

References

Brinkmann H, Martinez P, Quigley F, Martin W, Cerff R (1987) Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Evol 26:320–328

Chao S, Sederoff R, Levings CS III (1984) Nucleotide sequence and evolution of the 18S ribosomal RNA gene in maize mitochondria. Nucleic Acids Res 12:6629–6644

Chen H-C, Wintz H, Weil J-H, Pillay DTN (1988) Nucleotide sequence of chloroplast CF1-ATPase ε-subunit and elongator tRNA^{Met} genes from *Arabidopsis thaliana*. Nucleic Acids Res 16:10372

Chojecki J (1986) Identification and characterisation of a cDNA clone for cytosolic glyceraldehyde-3-phosphate dehydrogenase in barley. Carlsberg Res Commun 51:203–210

Filipski J (1988) Why the rate of silent codon substitutions is variable within a vertebrate's genome. J Theor Biol 134:159–164

Good AG, Pelcher LE, Crosby WL (1988) Nucleotide sequence of a complete barley alcohol dehydrogenase 1 cDNA. Nucleic Acids Res 16:7182

Klee HJ, Muskopf YM, Gasser CS (1987) Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. Mol Gen Genet 210:437–442

Lee KY, Townsend J, Tepperman J, Black M, Chui CF, Mazur B, Dunsmuir P, Bedbrook J (1988) The molecular basis of sulfonylurea resistance in tobacco. EMBO J 7:1241–1248

Li W-H, Wu C-I, Luo C-C (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

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

Maraña C, García-Olmedo F, Carbonero P (1988) Linked sucrose synthase genes in group-7 chromosomes in hexaploid wheat (*Triticum aestivum* L.). Gene 63:253–260

Mazur BJ, Chui C-F, Smith JK (1987) Isolation and characterization of plant genes coding for acetolactate synthase, the target enzyme for two classes of herbicides. Plant Physiol 85: 1110–1117

Neuhaus H (1989) Nucleotide sequence of the chloroplast genes for tRNA^{Gln} and the 4 kD K polypeptide of photosystem II from mustard (*Sinapsis alba*). Nucleic Acids Res 17:444

Niesbach-Klösgen U, Barzen E, Bernhardt J, Rohde W, Schwarz-Sommer Zs, Reif HJ, Weinand U, Saedler H (1987) Chalcone synthase genes in plants: a tool to study evolutionary relationships. J Mol Evol 26:213–225

Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol 26:74–86

Rohde W, Becker D, Salamini F (1988) Structural analysis of the *waxy* locus from *Hordeum vulgare*. Nucleic Acids Res 16: 7185–7186

Salinas J, Matassi G, Montero LM, Bernardi G (1988) Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. Nucleic Acids Res 16:4269–4285

Sharp PM, Li W-H (1987) The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol Biol Evol 4:222–230

Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. J Mol Evol 28:398–402

Stebbins GL (1981) Coevolution of grasses and herbivores. Ann Mo Bot Gard 68:75–86

Trick M, Dennis ES, Edwards KJR, Peacock WJ (1988) Molecular analysis of the alcohol dehydrogenase gene family of barley. Plant Mol Biol 11:147–160

Wolfe KH (1989) Compilation of sequences of protein-coding genes in chloroplast DNA including cyanelle and cyanobacterial homologues. Plant Mol Biol Reporter 7:30–48

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Wolfe KH, Sharp PM, Li W-H (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285

Received December 19, 1988/Revised and accepted March 17, 1989