Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9315–9318, December 1990
Evolution

Convergent evolution of the red- and green-like visual pigment genes in fish, *Astyanax fasciatus*, and human

(color vision/genomic clones/DNA sequencing/adaptive evolution)

RUTH YOKOYAMA AND SHOZO YOKOYAMA

Department of Ecology, Ethology and Evolution, University of Illinois at Urbana-Champaign, 505 South Goodwin Avenue, Urbana, IL 61801

Communicated by James F. Crow, September 4, 1990 (received for review May 8, 1990)

ABSTRACT **We have isolated and sequenced genes from the blind cave fish, *Astyanax fasciatus*, that are homologous to the human red and green visual pigment genes. The data strongly suggest that, like human, these fish have one red-like visual pigment gene and multiple green-like visual pigment genes. By comparing the DNA sequences of the human and fish visual pigment genes and knowing their phylogenetic relationship, one can infer the direction of amino acid substitutions in the red and green visual pigments. The results indicate that the red pigments in human and fish evolved from the green pigment independently by identical amino acid substitutions in only a few key positions.**

In nature, a vast array of visual abilities exists, including phototaxis of bacteria, the ability to see at night for nocturnal animals, and complex color vision of humans. Thus, vision strongly influences not only the survival of organisms but also the processes of adaptation to different environments. With the first molecular characterization of the human color visual pigment genes (1), it became possible to study the molecular mechanisms for specific wavelength absorptions attributed to the different color visual pigments (2–6).

Absorption of different wavelengths has been accomplished by a series of gene duplications and accumulation of mutations (1). It is likely that the common ancestor of the human color visual pigment genes diverged first from that of the rhodopsin gene ≈800 million years before present (Mybp), that the long and short wavelength pigment genes diverged >500 Mybp, and that the two human long wavelength-sensitive pigments (red and green) diverged ≈30 Mybp (1, 7). Although fish and mammalian lineages diverged ≈400 Mybp, many fish are known to be similar to humans in having more than one long wavelength-sensitive visual pigment (e.g., see refs. 8–10). Thus, the molecular characterization of red and green visual pigment genes in fish is important in elucidating the evolutionary mechanisms of the long wavelength-sensitive pigments.

The sequence of one such visual pigment gene of the blind cave fish *Astyanax fasciatus* has already been characterized (11). In the present paper, we report two other long wavelength-sensitive pigment genes of *Astyanax*.* Data analysis indicates that the red visual pigments of human and fish evolved independently by identical amino acid substitutions in only a few key positions.

MATERIALS AND METHODS

Screening of a Genomic DNA Library. A genomic library was constructed by using the high molecular weight DNA made from one blind cave fish as described (11). In short, the genomic DNA was partially digested with *Sau*3AI and separated by size on an agarose gel. The DNA in the size range of 9–23 kilobases (kb) was electroeluted from the gel and ligated with λ EMBL3 vector DNA that had been double-digested with *Bam*HI and *Eco*RI. Approximately 2 × 10^6 recombinant plaques were screened with the human red cDNA clone, hs7, generously provided by J. Nathans (The Johns Hopkins University). After screening a genomic library of a blind cave fish with the human red cDNA clone hs7, 33 λ clones were obtained (11).

Table 1. Correspondence between hybridizing genomic bands and clones

	λ103	λ101	λ007
*Eco*RI*	5.5 kb, 1.9 kb	9.4 kb	5.5 kb
*Hin*dIII	8.8 kb	8.0 kb	4.5 kb

The genomic bands were detected by Southern hybridization of *Eco*RI- and *Hin*dIII-digested blind cave fish DNA to human red cDNA clone (hs7).
*A weakly hybridizing band of 4.0 kb is also present in some blots (see figure 1 in ref. 11).

DNA Sequencing. The coding regions and introns of two particular λ clones, 007 (designated R_{007}) and 101 (G_{101}), were sequenced by the dideoxynucleotide chain-termination method using double-stranded templates (12, 13) of subclones in Bluescript. The subclones were obtained either by isolation of specific restriction fragments and ligation with Bluescript vector or by deletions of some subclones using exonuclease III/mung bean nuclease following the protocol recommended by Stratagene.

Data Analyses. We have compared the DNA sequences of the two newly sequenced genes R_{007} and G_{101} and a green-like pigment gene (G_{103}; previously designated G_F in ref. 11) in *Astyanax*, red (R_H), green (G_H), and blue (B_H) pigment genes in human (1), and human (Rh_H; ref. 14), bovine (Rh_B; ref. 15), and chicken (Rh_C; ref. 16) rhodopsin genes.

Alignment of DNA sequences was initially performed by the method of Wilbur and Lipman (17) and refined visually. After the alignment, the proportion (p) of different nucleotides for each pair of genes was computed. From this proportion, the total number of nucleotide substitutions per site (d) was estimated by $d = -(3/4)\ln[1 - (4/3)p]$ (18). The phylogenetic tree was constructed by using the neighbor-joining method (19).

RESULTS AND DISCUSSION

Southern hybridization of *Eco*RI- and *Hin*dIII-digested blind cave fish DNA with the human red cDNA clone (hs7) showed three strongly hybridizing bands in each digest (Table 1). An

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: Mybp, million years before present; R_H, G_H, and B_H, red, green, and blue visual pigment genes in human; Rh_H, Rh_B, and Rh_C, rhodopsin genes in human, bovine, and chicken; R_{007}, G_{101}, and G_{103}, visual pigment genes in *Astyanax*.
*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M38619–M38630).

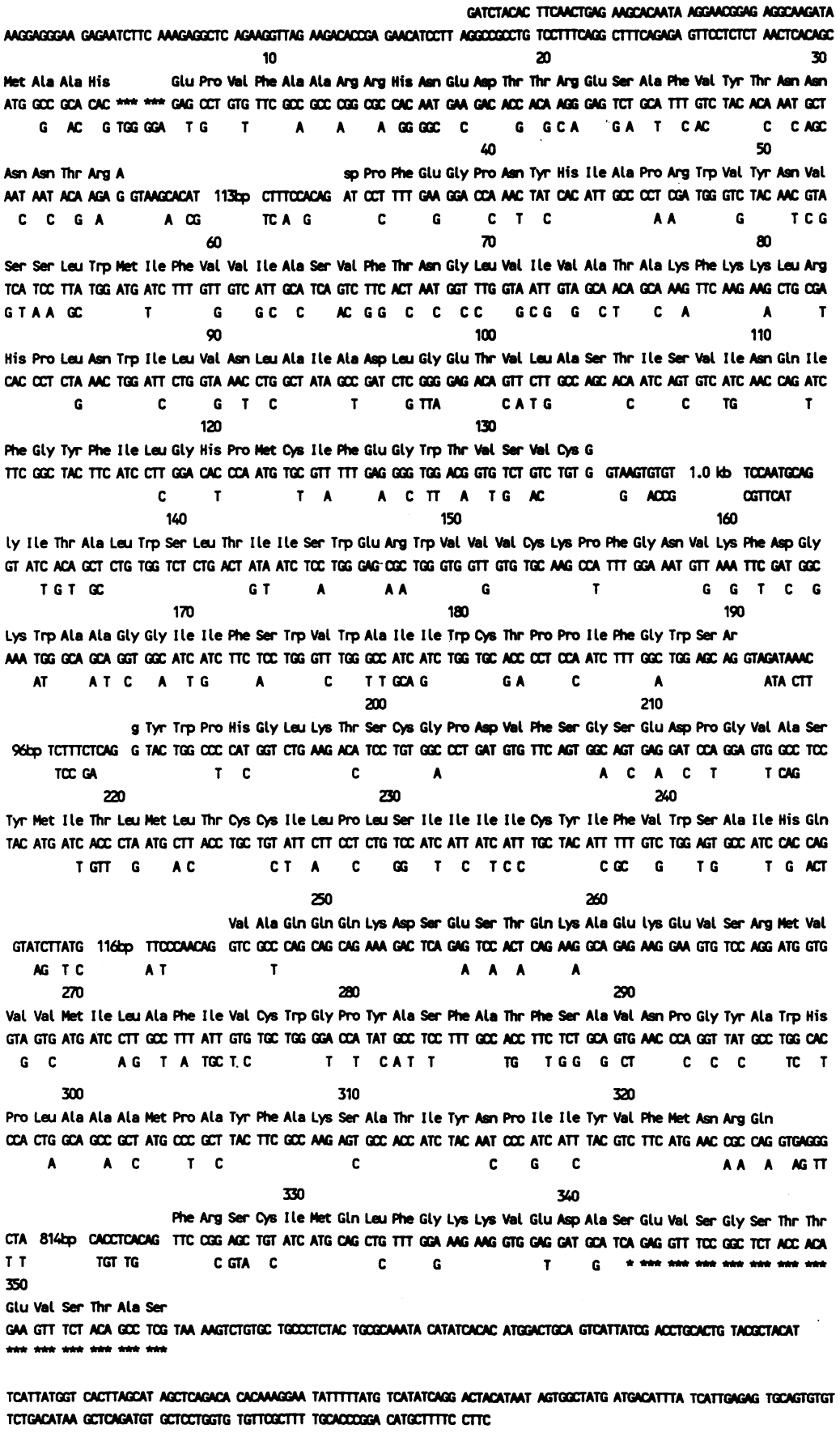

FIG. 1. Nucleotide sequences of G_{101} (first row) and R_{007} (second row) and deduced amino acid sequences of G_{101}. The sequence of R_{007} is shown only for nucleotides that differ from G_{101}, starting with the presumed first codon and ending 40 nucleotides prior to the presumed stop codon. Asterisks indicate gaps in the corresponding sequence. The first and last 10 nucleotides of each intron are also included. The first, second, third, fourth, and fifth introns of R_{007} consist of 120, 124, 84, 91, and 99 nucleotides, respectively.

Table 2. Number of nucleotides compared (n) and proportion of identical nucleotides ($1 - p$) between the coding regions of two genes

	R_H	G_H	B_H	Rh_H	Rh_B	Rh_C	G_{101}	G_{103}	R_{007}
R_H		1092	1035	1041	1041	1044	1059	1065	1025
G_H	0.98		1035	1041	1041	1044	1059	1065	1025
B_H	0.55	0.55		1035	1035	1044	1017	1017	977
Rh_H	0.54	0.55	0.54		1044	1044	1023	1023	983
Rh_B	0.55	0.56	0.56	0.89		1044	1023	1023	983
Rh_C	0.56	0.56	0.56	0.93	0.83		1023	1023	983
G_{101}	0.70	0.70	0.54	0.50	0.51	0.50		1059	1019
G_{103}	0.70	0.70	0.53	0.49	0.50	0.50	0.94		1025
R_{007}	0.74	0.74	0.53	0.53	0.51	0.53	0.76	0.75	

Values above the diagonal are n values; values below the diagonal are $1 - p$ values.

additional faint 4.0-kb *Eco*RI band can sometimes be detected (e.g., see figure 1 in ref. 11). Three blind cave fish genomic clones λ007, λ101, and λ103, which hybridized to the human red cDNA clone, represent segments from all three *Hin*dIII-hybridizing bands and the three strongest *Eco*RI-hybridizing bands (see Table 1). Thus, these clones seem to represent all the long wavelength pigment genes of the blind cave fish. Although both λ007 and λ103 hybridized to the same sized *Eco*RI band (5.5 kb), the molecular structures of the two clones are very different and the identical 5.5-kb size corresponds to nonhomologous regions (data not shown). λ103 (denoted G_{103}) was sequenced and found to be much more similar to the human green gene than to the human red gene (11). The DNA sequences of the two clones λ007 (designated R_{007}) and λ101 (G_{101}) and the deduced amino acid sequence of G_{101} are shown in Fig. 1. G_{101} is two codons shorter in the first exon than R_{007} (and G_{103}). 007 is incomplete at the 3′ region, missing the last 40 nucleotides of the coding region. The five intron positions have been completely conserved in R_{007}, G_{101}, G_{103}, R_H, and G_H.

Table 2 shows the proportion of identical nucleotides of the coding regions of the six visual pigment genes R_H, G_H, B_H, Rh_H, Rh_B, and Rh_C and the three fish genes R_{007}, G_{101}, and G_{103}. Clearly, G_{101} and G_{103} have 94% sequence similarity and are very closely related, whereas R_{007} and G_{101} (or G_{103}) have only ≈75% similarity. In fact, the latter value is about the same as the level of the sequence similarity between R_{007} and R_H (or G_H). Since nonsynonymous nucleotide substitutions, which cause amino acid changes, could affect the physicochemical properties of the protein, such changes often take place more slowly than synonymous substitutions. When only nonsynonymous sites are considered, the level of sequence similarity increases, but qualitatively the same evolutionary relationships are observed. For example, the proportion of identical nucleotides is 97% between G_{101} and G_{103}, whereas it is ≈87% between R_{007} and R_H (or G_H).

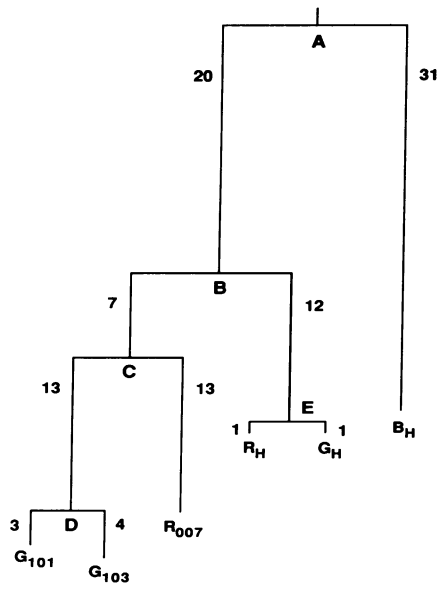

FIG. 2. Phylogenetic tree constructed for visual color pigment genes R_{007}, G_{101}, G_{103}, R_H, G_H, and B_H by using the number of nucleotide substitutions at all positions (d). The rooted tree was constructed by the neighbor-joining method [Saitou and Nei 1987 (19)], using Rh_H, Rh_B, and Rh_C as the outgroups.

Fig. 2 shows the phylogenetic tree based on the d values, where Rh_H, Rh_B, and Rh_C are taken as the outgroup. Clearly, G_{101}, G_{103}, and R_{007} in fish and R_H and G_H in human are clustered into two separate phylogenetic groups. In *Astyanax*, G_{101} and G_{103} were duplicated most recently and their common ancestor diverged from the ancestor of R_{007} before that. Even when only nonsynonymous nucleotide substitutions were considered, the same tree topology was obtained.

As already noted, the branch points A and E were estimated to be 500–600 and 30 Mybp, respectively. The time of divergence between fish and human (Fig. 2, B) is ≈400 Mybp. Knowing this divergence time and the relative branch lengths for all nucleotide substitutions, branch points C and D are estimated to be 240–290 and 50–70 Mybp, respectively. If we consider nonsynonymous substitutions, branch points C and D were estimated to be 190–320 Mybp and 40–70 Mybp, respectively.

Although the three genes R_{007}, G_{101}, and G_{103} were obtained from the blind cave fish, they are expected to be very similar to those of the eyed *Astyanax* because the cave isolation and eye loss occurred only within the last 1 million years (20, 21). That is, even at the evolutionary rate of nucleotide substitution in pseudogenes (4.8×10^{-9} per site per year; ref. 22), we would expect ≈5 nucleotide substitutions in the 1-kb coding region during the evolution of the blind cave fish. From all indications, the three genes still appear to be capable of encoding functional proteins (see also ref. 11). Thus, they may be regarded as typical long wavelength-sensitive visual pigment genes in *Astyanax*.

There are only 15 amino acid differences between the human red and green visual pigments (1) and we compared the amino acids between the fish and human visual pigments at these positions (Table 3; see also Table 4). From Table 3,

Table 3. Proportions of identical amino acids at the 15 polymorphic residues where the human red- and green-sensitive visual pigments differ

Pigments encoded by	Pigments encoded by	
	R_H	G_H
R_{007}	7/15	2/15
G_{101}	4/15	7/15
G_{103}	3/15	7/15

Residues considered were 65, 111, 116, 153, 180, 230, 233, 236, 274, 275, 277, 279, 285, 298, and 309.

it is clear that R_{007} is much more similar to R_H, whereas G_{101} and G_{103} are much more similar to G_H. Therefore, we regard R_{007} as a red-like visual pigment gene and both G_{101} and G_{103} as green-like visual pigment genes. Interestingly, like human, there appears to be one red gene and multiple green genes in *Astyanax*.

Fish and human lineages originally had the common long wavelength-absorbing visual pigment gene, and the duplications and evolutionary changes of the duplicate genes have occurred independently in the two lineages (see Fig. 2). Knowing the phylogenetic relationship of the long wavelength-absorbing genes of fish and human, we can infer the directions of amino acid substitutions. For that purpose, we compared the amino acid sequences deduced from the three fish genes R_{007}, G_{101}, and G_{103}; four human genes R_H, G_H, B_H, and Rh_H; and bovine and chicken rhodopsin genes (Rh_B and Rh_C) at the 15 polymorphic residues (Table 4). The three codon positions 180, 277, and 285 of R_H and G_H seem to be of particular importance.

At residue 277, only pigments encoded by R_{007} and R_H have tyrosine when the other pigments, including the three rhodopsins, have phenylalanine. Similarly, the pigments encoded by R_{007} and R_H have threonine at residue 285, while the others have alanine in that position. Clearly, in the red visual pigments, phenylalanine was replaced by tyrosine at residue 277, whereas alanine was replaced by threonine at residue 285. Because of the strong amino acid conservation in other pigments, these two amino acid substitutions are suspected to have had an adaptive significance in the development of the red visual pigment from the green visual pigment (see also ref. 11).

Similarly, alanine was replaced by serine at residue 180 of the pigments encoded by R_{007} and R_H (Table 4). However, at this residue the human blue pigment has the amino acid substitution from alanine to glycine (Table 4) and, therefore, the amino acid substitution at this residue may not have been as important as those at residues 277 and 285 in the development of the red visual pigment.

Thus, the present comparative molecular analysis of the fish visual pigment genes strongly suggests that the red visual pigment in human and fish evolved from the green visual pigment by identical amino acid substitutions at the two, or possibly three, residues.

Table 4. Amino acids deduced from different visual pigment genes

Residue	Visual pigments encoded by gene								
	R_{007}	G_{101}	G_{103}	R_H	G_H	B_H	Rh_H	Rh_B	Rh_C
65	V	I	I	T	I	L	V	M	L
111	V	V	V	I	V	F	L	L	M
116	F	F	F	S	Y	N	H	H	N
153	V	V	V	L	M	I	V	V	V
180	S	A	A	S	A	G	A	A	A
230	I	L	L	I	T	V	I	I	I
233	G	S	S	A	S	S	I	I	A
236	I	I	I	M	V	C	F	F	F
274	I	I	I	I	V	V	V	V	V
275	M	L	L	F	L	G	I	I	I
277	Y	F	F	Y	F	F	F	F	F
279	F	V	L	V	F	V	I	I	I
285	T	A	A	T	A	A	A	A	A
298	A	A	A	A	P	G	N	D	D
309	Y	Y	Y	Y	F	F	F	F	F

The 15 residues were selected because the human red- and green-sensitive visual pigments differ only at these sites.

This assertion is compatible with the results obtained in two other laboratories. First, from the comparative analysis of structural models, Kosower (23) observed that the difference between the red and green visual pigments is due to the net effect of residues 65, 180, 230, 233, 277, 285, and 309. Second, the Southern analysis of Nathans *et al.* (2) strongly suggests that the R_H and G_H sequence differences in exon 5 distinguish the spectral absorbance of the red and green wavelengths (see also ref. 6). Interestingly, both codons 277 and 285 are located in exon 5.

Molecular characteristics of the visual pigment genes in different species provide valuable information as to which residues are important for specific wavelength absorption. Fortunately, it will be possible to test such hypotheses of adaptive evolution by using site-directed mutagenesis at specific residues (such as 180, 277, and 285), expressing them in cultured cells and measuring their absorbance spectrum as has been done by Khorana and his colleagues (24, 25) and Nathans and his colleagues (26, 27).

We thank J. Nathans for providing hs7 and other cDNA clones. This research was supported by grants from the National Institutes of Health and the National Science Foundation.

1. Nathans, J., Thomas, D. & Hogness, D. S. (1986) *Science* **232**, 193–202.
2. Nathans, J., Piantanida, T. P., Eddy, R. L., Show, T. B. & Hogness, D. S. (1986) *Science* **232**, 203–210.
3. Nathans, J., Davenport, C. M., Maumenee, I. H., Lewis, R. A., Hejtmancik, J. F., Litt, M., Lovrien, E., Weleber, R., Bachynski, B., Zwas, F., Klingaman, R. & Fishman, G. (1989) *Science* **245**, 831–838.
4. Drummond-Borg, M., Deeb, S. & Motulsky, A. G. (1988) *Am. J. Hum. Genet.* **43**, 675–683.
5. Drummond-Borg, M., Deeb, S. & Motulsky, A. G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 983–987.
6. Neitz, J., Neitz, M. & Jacobs, G. H. (1989) *Nature (London)* **342**, 679–682.
7. Yokoyama, S. & Yokoyama, R. (1989) *Mol. Biol. Evol.* **6**, 186–197.
8. Levine, J. S. & MacNichol, F. (1982) *Sci. Am.* **246**, 140–149.
9. Loew, E. R. & Lythgoe, J. N. (1978) *Vision Res.* **18**, 715–722.
10. Archer, S. N. & Lythgoe, J. N. (1990) *Vision Res.* **30**, 225–233.
11. Yokoyama, R. & Yokoyama, S. (1990) *Vision Res.* **30**, 807–816.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
13. Hattori, M., Hidaka, S. & Sakaki, Y. (1985) *Nucleic Acids Res.* **13**, 7813–7827.
14. Nathans, J. & Hogness, D. S. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4851–4855.
15. Nathans, J. & Hogness, D. S. (1983) *Cell* **34**, 807–814.
16. Takao, M., Yasui, A. & Tokunaga, F. (1988) *Vision Res.* **28**, 471–480.
17. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 726–730.
18. Jukes, T. H. & Cantor, C. H. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–123.
19. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
20. Avise, J. C. & Selander, R. K. (1972) *Evolution* **26**, 1–19.
21. Chakraborty, R. & Nei, M. (1974) *Theor. Popul. Biol.* **5**, 460–469.
22. Li, W.-H., Luo, C.-C. & Wu, C.-I. (1985) in *Molecular Evolutionary Genetics*, ed. MacIntyre, R. J. (Plenum, New York), pp. 1–94.
23. Kosower, E. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1076–1080.
24. Karnik, S. S., Sakmar, T. P., Chen, H.-B. & Khorana, H. B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8459–8463.
25. Oprian, D. D., Molday, R. S., Kaufman, R. J. & Khorana, H. B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8874–8878.
26. Nathans, J., Weitz, C. J., Agarwal, N., Nir, I. & Papermaster, D. S. (1989) *Vision Res.* **29**, 907–914.
27. Nathans, J. (1990) *Biochemistry* **29**, 937–942.