

Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution

(retroposon/PCR/phylogeny/orthologous locus/salmon)

SHIGENORI MURATA, NOBUYOSHI TAKASAKI, MASAKO SAITOH, AND NORIHIRO OKADA*

Institute of Biological Sciences, University of Tsukuba, Tsukuba 305, Ibaraki, Japan, and Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 227, Japan

Communicated by C. Clark Cockerham, March 31, 1993 (received for review November 20, 1992)

ABSTRACT Several subfamilies of the salmonid *Hpa* I short interspersed element (SINE) family were isolated from salmonid genomes and were sequenced. For each genomic locus that represented the subfamily, amplification by PCR of the orthologous loci in the 12 fish allowed us to determine the order of branching of the Pacific salmonid species. The deduced phylogeny suggests three evolutionary lines, namely, a line of chum salmon, pink salmon, and kokanee; a line of coho salmon and chinook salmon; and a line of steelhead trout. Our data also support a change in the phylogenetic assignment of steelhead trout from *Salmo* to *Oncorhynchus*. We present here an extensive phylogenetic tree constructed from an analysis of differential insertion of SINEs, and we propose that SINE insertion analysis is one of the best available methods for clarifying the order of divergence of closely related species.

A retroposon is defined as a nucleotide sequence, present initially as a cellular RNA transcript, that has been reincorporated into the genome, presumably via a cDNA intermediate. Retroposons constitute roughly 10% of the human genome and are similarly abundant in other mammalian genomes (1, 2). As a result, the remarkable fluidity of eukaryotic genomes reflects the contributions of retroposition (2) as well as mechanisms operating at the DNA level such as mutation and recombination (1–4). Retroposons can be unique to one species, a few species, a genus, or in some cases a family. Retroposition is therefore a specialized form of gene duplication, which is believed to be of major importance in the creation of genetic diversity during evolution (5).

Nonviral retroposons are classified into three main groups: processed retropseudogenes, LINEs (long interspersed elements), and SINEs (short interspersed elements) (6). Except for the rodent type 1 and human *Alu* families (7, 8), all of the SINE families examined to date have been shown to be derived from tRNAs (9–14). In contrast to DNA transposable elements, which are often capable of being excised precisely, SINEs appear to be inserted irreversibly and should therefore provide an ideal evolutionary and phylogenetic marker (4).

The Pacific salmon and trout (*Oncorhynchus*) are a group of closely related species with complex life histories and an interesting global distribution (reviewed in ref. 15). Previously, in an attempt to elucidate a possible role of SINEs in the genomic organization and speciation of salmonids, we characterized three families of tRNA-derived SINEs in salmonid genomes (16, 17). The salmon *Sma* I family is restricted to the genomes of chum salmon and pink salmon. The charr *Fok* I family is present only in species that belong to the genus *Salvelinus*. The third family, the salmonid *Hpa* I family, is present in all species in the family Salmonidae but

not in other species (16). These results suggest that these SINEs were amplified specifically within certain salmonid lineages during evolution.

Our data prompted us to attempt to construct a phylogenetic tree for the salmonid species by using SINE insertions as irreversible events that would serve as informative markers of evolution. In this report, we present a characterization of the four subfamilies of the *Hpa* I family.[†] These subfamilies were amplified in the four different ancestral species within the genus *Oncorhynchus*. Such characterization provides a highly reliable order of branching of the various species of *Oncorhynchus*.

MATERIALS AND METHODS

Experiments were performed by using standard techniques (18–21).

The fish species examined in this study and their geographic sources are listed in Table 1. The family Salmonidae consists mainly of four genera: *Oncorhynchus*, *Salmo*, *Salvelinus*, and *Hucho*. The genus *Oncorhynchus* includes eight species, of which six species were analyzed in this study. Genomic DNAs from chum salmon (*O. keta*), kokanee (*O. nerka adonis*), and coho salmon (*O. kisutch*) were used to construct three genomic libraries. Each genomic library was screened for phage clones that contained the salmonid *Hpa* I family and their sequences were determined by the chain-termination method (19). When a unit of the family appeared to be integrated into a unique region of the genome, we synthesized 5' and 3' 20-meric primers that flanked the unit. Then PCR was performed (20), using the DNAs from the 12 listed species as templates. Each locus was named after the number of the clone and the name of the species from which it was isolated. For example, when a phage clone was isolated from the genomic library of kokanee and the number of the clone was 345, the locus was named Hpa(ON)-345 (where ON stands for *O. nerka*). The orthologous loci of chum salmon (*O. keta*), pink salmon (*O. gorbuscha*), coho salmon (*O. kisutch*), and steelhead trout (*O. mykiss*), which could be detected by PCR, were named Hpa(OK)-345, Hpa(OG)-345, Hpa(OKi)-345, and Hpa(OMy)-345, respectively. To confirm the presence or absence of a SINE unit, Southern hybridization experiments were performed (21) and several sequences of products of PCR at the orthologous loci were determined. To distinguish different loci from one another, different numbering systems were adopted for the different genomic libraries.

Abbreviation: SINE, short interspersed element.

*To whom reprint requests should be sent at the present address: Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 227, Japan.

[†]The sequences reported in this paper have been deposited in the GenBank data base (accession nos. D16238–D16246).

Table 2. Sequences of primers used for detection of the four loci analyzed

Locus	5' flanking	3' flanking
Hpa-345	5'-GCACTGTTACCACATAGTTAG-3'	3'-TCACGACGAACTCACAAATA-5'
Hpa-19	5'-AAACACATGGTCACGGTGTG-3'	3'-CACCCATGTCCTTTGTATGT-5'
Hpa-51	5'-GAGTTGAACATTACCAGTACTACCATT-3'	3'-ATACCTTAGTACATCATTTGGTTTTTCC-5'
Hpa-341	5'-CGAGTGTATCTGAAGTGTCC-3'	3'-GTCACCTATGAAAGACGGTG-5'

the present-day Pacific salmon and trout, and the unit in this locus was fixed among the population, most likely by random genetic drift. We have isolated several other loci that indicate a clear distinction between the genomes of the Pacific and Atlantic groups. A compilation of these loci, together with other loci that specify other subfamilies, will be published elsewhere.

Chum Salmon, Pink Salmon, and Kokanee Form a Monophyletic Group. The Hpa-19 locus was isolated from a genomic library of chum salmon and its nucleotide sequence was determined. Primer DNAs were synthesized (Table 2) and PCR was performed. Judging from additional DNA products amplified by PCR (Fig. 3A), sequence polymorphisms are present in this locus among several salmonid species. However, it is evident that a unit of this *Hpa I* subfamily is integrated in the genomes of chum salmon, pink salmon, and kokanee (Fig. 3). These results suggest the existence of a common ancestral species for the three present-day species. The nucleotide sequences of the Hpa(ON)-19 locus in kokanee and the Hpa(OT)-19 locus in chinook salmon were determined and confirmed the validity of this inference (Fig. 1B).

Coho Salmon and Chinook Salmon Are Sister Species. The Hpa-51 locus was isolated from a genomic library of coho salmon and its nucleotide sequence was determined (Fig. 1C). As shown in Fig. 4, it is evident that a unit of the *Hpa I* subfamily at this locus was integrated only in the genomes of coho salmon and chinook salmon, although some of the

orthologous loci in species other than *Oncorhynchus* were lost during evolution (Fig. 4C). The sequence of the orthologous locus Hpa(ON)-51 indicates that the unit is present in the genome of coho salmon but not in that of kokanee (Fig. 1C).

The Five Pacific Salmon Form a Monophyletic Group. The Hpa-341 locus was isolated from a genomic library of kokanee. A unit of the *Hpa I* subfamily was demonstrated to be integrated into the orthologous loci of the five Pacific salmon (Fig. 5). Thus, chum salmon, pink salmon, kokanee, chinook salmon, and coho salmon form a monophyletic group. The band in the lane for chum salmon in Fig. 5A has higher mobility than the other four bands, but a hybridization signal was detected in this lane (Fig. 5B). The nucleotide sequence of the Hpa(OK)-341 locus in chum salmon indicated that a deletion of 160 bp of DNA has occurred specifically in this lineage, with only 42 nucleotides of the unit at the 5' end being present at this locus. Fig. 1D shows the sequences of the three orthologous loci of kokanee, chum salmon, and steelhead trout.

This locus does not provide information about the divergence of the former *Salmo* (steelhead trout) from the genus *Oncorhynchus* (see *Discussion*), because the unit was not integrated at the orthologous locus of cherry salmon (*O. masou*), which belongs to *Oncorhynchus* (data not shown). Therefore, the data indicate the existence of an ancestral species that was common to the above five salmon after the divergence of cherry salmon and steelhead trout. The relationship between cherry salmon and steelhead trout is currently unknown.

DISCUSSION

SINE Insertion Analysis Is an Excellent Method for Determining the Order of Divergence of Closely Related Species.

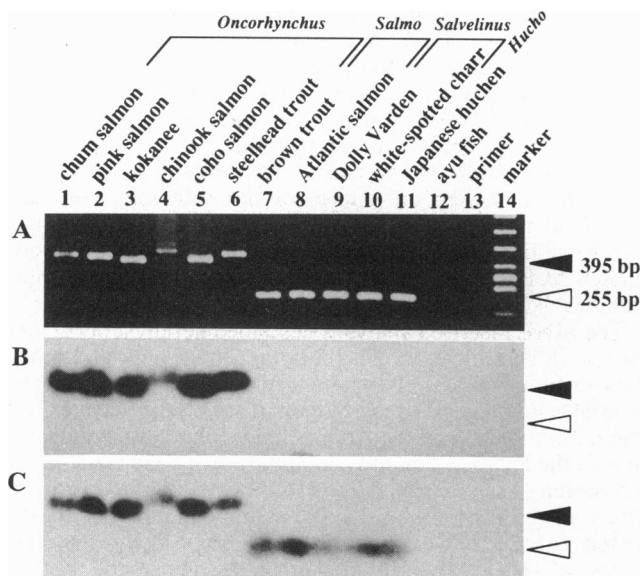


FIG. 2. The Pacific salmonids form a monophyletic group. (A) Analysis of the products of PCR of the orthologous loci of Hpa-345 by electrophoresis in an agarose gel. (B) Southern hybridization of a blot of the gel in A, using the unit sequence of the *Hpa I* family as a probe. (C) Southern hybridization of the same blot, using the Hpa(ON)-345 DNA as a probe. Hybridization was performed in buffer [50% (vol/vol) formamide/1% SDS/1 M NaCl] at 42°C for 16 hr. In the experiment in C, the blot was treated with an alkaline solution (0.4 M NaOH) at 42°C for 30 min before hybridization. Black and white arrowheads indicate positions of DNA with and without an integrated unit of the *Hpa I* subfamily, respectively.

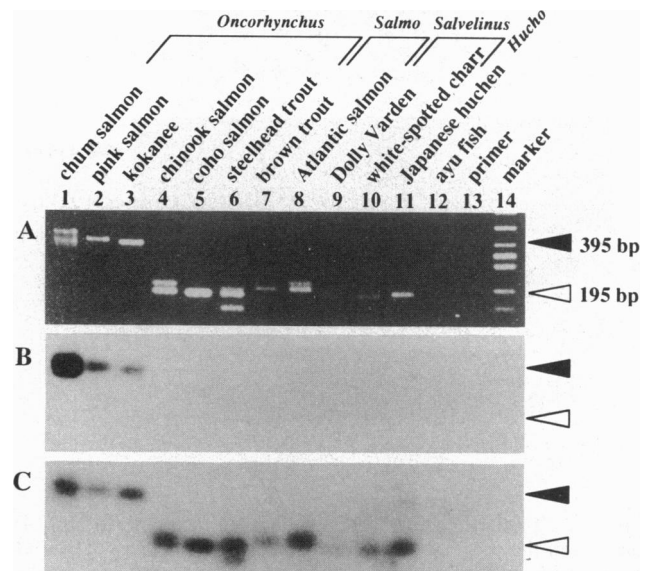


FIG. 3. Chum salmon, pink salmon, and kokanee form a monophyletic group. (A) Analysis of the products of PCR of the orthologous loci of Hpa-19 by electrophoresis in an agarose gel. (B) Southern hybridization of a blot of the gel in A, using the unit sequence of the *Hpa I* family as a probe. (C) Southern hybridization of the same blot, using the Hpa(OK)-19 DNA as a probe. The hybridization conditions were the same as those described in the legend to Fig. 2.

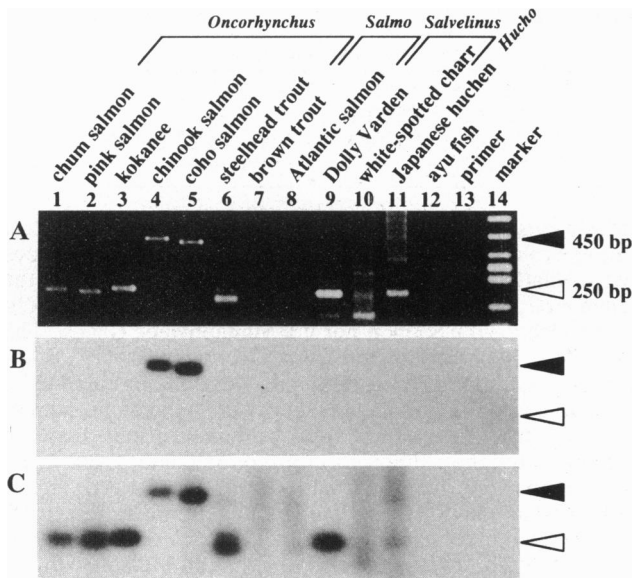


FIG. 4. Coho salmon and chinook salmon are sister species. (A) Analysis of the products of PCR of the orthologous loci of Hpa-51 by electrophoresis in an agarose gel. (B) Southern hybridization of a blot of the gel in A, using the unit sequence of the *Hpa* I family as a probe. (C) Southern hybridization of the same blot, using the Hpa(OKi)-51 DNA as a probe. The hybridization conditions were the same as those described in the legend to Fig. 2.

Several types of variation in DNA and protein have been used to infer phylogenetic relationships among species (22). The methods based on comparisons of sequence data among species must, however, always involve statistical errors and often require a large quantity of sequence data to ensure the validity of the branching orders. Moreover, the fact that the

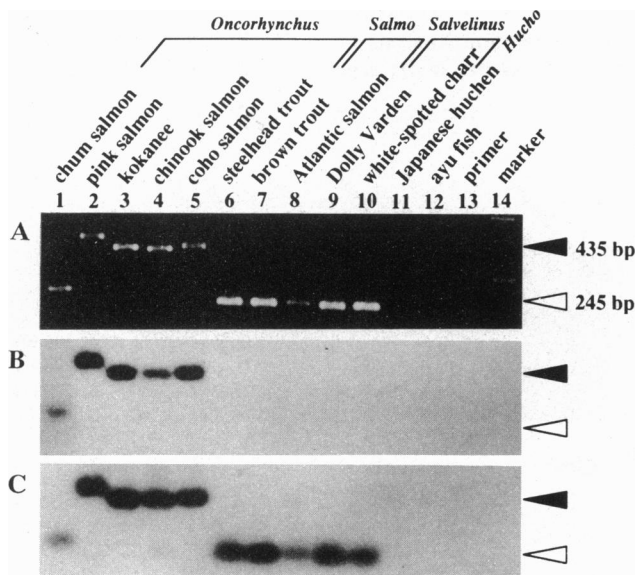


FIG. 5. The five Pacific salmon form a monophyletic group. (A) Analysis of the products of PCR of the orthologous loci of Hpa-341 by electrophoresis in an agarose gel. (B) Southern hybridization of a blot of the gel in A, using the unit sequence of the *Hpa* I family as a probe. (C) Southern hybridization of the same blot using the Hpa(ON)-341 DNA as a probe. The hybridization conditions in the experiment of C were the same as those described in the legend to Fig. 2. In the experiment of B, hybridization was performed in buffer (1% SDS/0.9 M NaCl/0.09 M sodium citrate, pH 7.0/0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/*Escherichia coli* DNA at 5 mg/ml) at 37°C for 16 hr.

rate of molecular evolution is sometimes not constant makes it very difficult to generate an accurate phylogeny.

We propose that SINE insertion analysis, as presented here, is one of the best methods for determining the branching orders of closely related species. Previously, several laboratories demonstrated that the human *Alu* family can be divided into several subfamilies, with each being inserted into the host genome at a different time during evolution (23–26). Britten *et al.* (24) and Jurka and Smith (25) clearly showed that the human *Alu* repeats that arose at earlier times shared correlated blocks of nucleotides that were different from the current consensus sequence at diagnostic positions. On the basis of these findings, Ryan and Dugaiczky (27) suggested the possibility that the distribution of *Alu* elements might be used to resolve the branching order in the evolution of primates. In this case, however, only the loci specific to the human genome were analyzed, and no extensive classification using this strategy has been attempted to date. We believe that our phylogeny is the first example of the full application of this methodology. We have isolated and characterized several loci per branching point, greatly reducing the probability of phylogenetic incongruence due to polymorphisms in the ancestral species (28).

With regard to the four loci described in the present study, it is not precisely determined whether each inserted SINE is fixed or polymorphic in the lineages of salmonid species. Recently we isolated one locus which contains a member of the *Hpa* I subfamily amplified specifically in the chum salmon, and we showed that the inserted SINE in this locus is fixed in all the populations of chum salmon from various places in the Pacific Ocean (unpublished results). This result strongly suggests that the SINE members in the four loci described above may be fixed in all the populations of salmonid species, because they are much older than the member of the species-specific SINE in the chum salmon.

Phylogeny of Pacific Salmonids. Western American trout consist of various species, which include steelhead trout (rainbow trout), cutthroat trout, and allied subspecies. The phylogenetic position of these species, represented by the steelhead trout in this study, has long been disputed because of their close relatedness both to Pacific salmon (*Oncorhynchus*) and to Atlantic salmon (*Salmo*) (reviewed in ref. 15). Recently, these species, which were previously classified as *Salmo*, were officially renamed *Oncorhynchus* (29). The decision was based on studies of the osteology (30) and biochemistry (31) of trout and salmon, which showed that steelhead trout and their close relatives are more closely related to Pacific salmon than to brown trout and Atlantic salmon.

The SINE insertion analyses described herein support this change in classification. The data presented here and other data to be published elsewhere demonstrate that there are several genomic loci that show clear distinctions between the Pacific salmonids and the Atlantic ones. Our results indicate that all the Pacific salmonids, including steelhead trout, have a common ancestral species and form a monophyletic group. Since all the present-day Pacific salmonids contain the inserted SINE in the Hpa-345 locus, it is likely that the ancestral species lived as a unique species in the Pacific Ocean during a long enough period of time for several variations in the DNA to be fixed among the population.

Chum salmon, pink salmon, and kokanee were shown to have a common ancestral species and to form a monophyletic group (Fig. 6). With respect to the relationship among these three species, we suggest that the kokanee lineage diverged before the divergence of the remaining two species, because the salmon *Sma* I family is present only in the genomes of chum salmon and pink salmon (16). The salmon *Sma* I family may have been amplified in a common ancestral species of these two species, suggesting that kokanee is distinct from

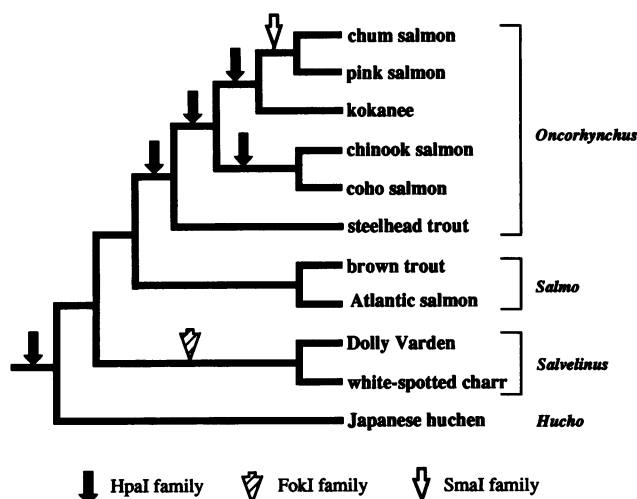


FIG. 6. Phylogenetic tree of the salmonid species as deduced from SINE insertion analyses. Three different kinds of tRNA-derived SINEs, namely, the *Sma* I family, the *Fok* I family, and the *Hpa* I family, were amplified at specific stages of evolution of salmonid species (16). Multiple filled arrows show the *Hpa* I sub-families described in the text and ref. 16.

that pair of species and might occupy an intermediate evolutionary position (16). The phylogenetic relationship of these three species in Fig. 6 is well correlated with that of Thomas *et al.* (32), obtained from analysis of mitochondrial DNA, but not with that obtained from allozyme variations by Utter *et al.* (33) or with that derived from morphological comparisons (29).

On the basis of ecological, morphological, behavioral, and biochemical data, there is general agreement that coho salmon and chinook salmon are the most closely related species (29, 30, 32–34). Our SINE insertion data support this conclusion.

With respect to the phylogenetic position of steelhead trout, our phylogeny (Fig. 6) indicates that steelhead trout diverged before the divergence of the other five Pacific salmon. This conclusion is in accordance with that from morphological comparisons (29) and with that from allozyme variations (33, 34) but not with that of Thomas *et al.* (32), who showed, by analysis of mitochondrial DNA, that rainbow trout (steelhead trout), coho salmon, and chinook salmon form a monophyletic group. When a root of the phylogenetic tree of Thomas *et al.* (32) is moved to the evolutionary line of rainbow trout, their phylogeny becomes the same as ours. Therefore, we suggest that the rate of mitochondrial molecular evolution altered in the lineages of rainbow and cutthroat trout. With respect to the order of branching, we believe that our phylogeny (Fig. 6) is the most reliable of those presented to date.

We thank Drs. Susumu Nishimura and Shozo Osawa for encouragement. We are grateful to Drs. Masami Hasegawa, Hidenori

Tachida, Toshio Okazaki, Ruth Phillips, Axel Meyer, Linda Park, and Alan Weiner for critical reading of the manuscript. We also thank Profs. Mineo Saneyoshi and Takeshi Matsunaga for giving us salmonid samples and Mr. Y. Hamada for help in isolation of phage clones.

- Rogers, J. (1985) *Int. Rev. Cytol.* **93**, 187–279.
- Weiner, A. M., Deininger, P. L. & Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55**, 631–661.
- Okada, N. (1991) *Trends Ecol. Evol.* **6**, 358–361.
- Okada, N. (1991) *Curr. Opin. Genet. Dev.* **1**, 498–504.
- Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, Heidelberg).
- Singer, M. F. (1982) *Cell* **28**, 433–434.
- Weiner, A. M. (1980) *Cell* **22**, 209–218.
- Ullu, E. & Tschudi, C. (1984) *Nature (London)* **312**, 171–172.
- Lawrence, C. B., McDonnell, D. P. & Ramsey, W. J. (1985) *Nucleic Acids Res.* **13**, 4239–4252.
- Daniels, G. R. & Deininger, P. L. (1985) *Nature (London)* **317**, 819–822.
- Okada, N., Endoh, H., Sakamoto, K. & Matsumoto, K. (1985) *Proc. Jpn. Acad. Ser. B* **61**, 363–367.
- Sakamoto, K. & Okada, N. (1985) *J. Mol. Evol.* **22**, 134–140.
- Endoh, H. & Okada, N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 251–255.
- Matsumoto, K., Murakami, K. & Okada, N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3156–3160.
- Phillips, R. B. & Pleyte, K. A. (1991) *J. Fish Biol.* **39**, Suppl. A, 259–275.
- Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M. & Okada, N. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2326–2330.
- Koishi, R. & Okada, N. (1991) *J. Mol. Evol.* **32**, 43–52.
- Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
- Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
- Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H., Jr., & Deininger, P. L. (1987) *Mol. Biol. Evol.* **4**, 19–29.
- Britten, R. J., Baron, W. F., Stout, D. B. & Davidson, E. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4770–4774.
- Jurka, J. & Smith, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4775–4778.
- Willard, C., Nguyen, H. T. & Schmid, C. W. (1987) *J. Mol. Evol.* **26**, 180–186.
- Ryan, S. C. & Dugaiczky, A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9360–9364.
- Tachida, H. & Iizuka, M. (1993) *Genetics* **133**, 1023–1030.
- Smith, G. R. & Stearley, R. F. (1989) *Fisheries* **14**, 4–10.
- Behnke, R. J. (1965) Dissertation (Univ. of California, Berkeley).
- Okazaki, T. (1984) *Jpn. Ichthyol.* **31**, 297–311.
- Thomas, W. K., Withler, R. E. & Beckenbach, A. T. (1986) *Can. J. Zool.* **64**, 1058–1064.
- Utter, F. M., Allendorf, F. W. & Hodgins, H. O. (1973) *Syst. Zool.* **22**, 257–270.
- Ferguson, A. & Fleming, C. C. (1983) in *Protein Polymorphism: Adaptive and Taxonomic Significance*, eds. Oxford, G. S. & Rollinson, D. (Academic, New York), Vol. 24, pp. 85–99.