

Assessment of concordance among genealogical reconstructions from various mtDNA segments in three species of Pacific salmon (genus *Oncorhynchus*)

D. CHURIKOV,*† M. MATSUOKA,* X. LUAN,* A. K. GRAY,* VL. A. BRYKOV† and A. J. GHARRETT*

*Fisheries Division, School of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 11120 Glacier Highway, Juneau, Alaska 99801, USA, †Institute of Marine Biology, Far East Branch, Russian Academy of Sciences, Vladivostok 690041, Russia

Abstract

Seven segments of mitochondrial DNA (mtDNA), comprising 97% of the mitochondrial genome, were amplified by polymerase chain reaction (PCR) and examined for restriction site variation using 13 restriction endonucleases in three species of Pacific salmon: pink (*Oncorhynchus gorbuscha*), chum (*O. keta*) and sockeye (*O. nerka*) salmon. The distribution of variability across the seven mtDNA segments differed substantially among species. Little similarity in the distribution of variable restriction sites was found even between the mitochondrial genomes of the even- and odd-year broodlines of pink salmon. Significantly different levels of nucleotide diversity were detected among three groups of genes: six NADH-dehydrogenase genes had the highest; two rRNA genes had the lowest; and a group that included genes for ATPase and cytochrome oxidase subunits, the cytochrome *b* gene, and the control region had intermediate levels of nucleotide diversity. Genealogies of mtDNA haplotypes were reconstructed for each species, based on the variation in all mtDNA segments. The contributions of variation within different segments to resolution of the genealogical trees were compared within each species. With the exception of sockeye salmon, restriction site data from different genome segments tended to produce rather different trees (and hence rather different genealogies). In the majority of cases, genealogical information in different segments of mitochondrial genome was additive rather than congruent. This finding has a relevance to phylogeographic studies of other organisms and emphasizes the importance of not relying on a limited segment of the mtDNA genome to derive a phylogeographic structure.

Keywords: gene genealogy, haplotype and nucleotide diversities, mitochondrial genome, Pacific salmon, PCR–RFLP analysis

Received 21 December 2000; revision received 16 May 2001; accepted 16 May 2001

Introduction

Widespread distribution of polymerase chain reaction (PCR) technology and automation of DNA sequencing have resulted in a noticeable shift from restriction site to nucleotide sequence analysis in population genetics studies; and even when restriction fragment length polymorphism (RFLP) is the method of choice, it is often

conducted on PCR-amplified genome segments. Although both methods allow increased scrutiny of genetic variation, they focus on small portions of the genome. Consequently, the question of the choice of the gene or genome segment for analysis arises. For population-level studies the most variable genes or segments of the genome are usually preferred, since they harbour potentially larger numbers of haplotypes providing a basis for finer resolution of intraspecific diversification. Mitochondrial DNA (mtDNA) and its genes have been used extensively in population genetic studies (Avisé 1998). However, although the whole mtDNA molecule may evolve five to 10 times faster than single-copy nuclear DNA (Brown *et al.* 1979), the

Correspondence: D. Churikov, Bodega Marine Laboratory, University of California, 2099 Westside Road/PO Box 247, Bodega Bay, CA 94923, USA. Fax: (707) 875-2009; E-mail: dychurikov@ucdavis.edu

evolutionary rates of the mitochondrial genes are heterogeneous (Meyer 1993).

Sequence comparisons among vertebrate species have shown that the levels of variability in mitochondrial genes are influenced by different selective constraints on amino acid composition and function of the gene product (Kocher *et al.* 1989; Saccone *et al.* 1991), and their position relative to the origin of heavy-strand replication (Bielawski & Gold 1996). Hence, each mitochondrial gene may evolve at its own rate (Meyer 1994). At the intraspecific level, much attention has been devoted to the noncoding control region containing the D-loop, because lack of the constraints on a protein product may result in elevated variability. An increased nucleotide substitution rate of the control region, compared to that in the rest of the mitochondrial genome, was indeed detected in humans (Aquadro & Greenberg 1983; Horai & Hayasaka 1990), but it is not necessarily the case for other species (Hoelzel *et al.* 1991; Bernatchez *et al.* 1992; Brown *et al.* 1993; Park *et al.* 1993).

Clonal haploid inheritance and absence of recombination are other properties of mtDNA that have made it a molecule of choice in the studies of intraspecific phylogeography (Avise *et al.* 1987; Avise 1998). Although different PCR-amplified mtDNA segments are used for haplotype genealogy reconstruction and making phylogeographic inferences, little is known about the extent to which variation in one mtDNA segment (even the most variable one) represents variation in the entire molecule. The question of the congruence between haplotype genealogies reconstructed from variation in a segment and in the whole mtDNA (or in different segments of the genome) has not yet been well explored; and only fragmentary data are interspersed in the literature (Edwards & Wilson 1990; Bernatchez & Danzmann 1993; Giuffra *et al.* 1994; Walker *et al.* 1995; Encalada *et al.* 1996; Fry & Zink 1998; Grant *et al.* 1998). Since mtDNA gene genealogies provide a basis for phylogeographic inferences, assessment of the concordance in genealogical partitions across mitochondrial genes or genome segments is warranted.

In this paper, we present the results of a systematic survey of variation in seven PCR-amplified segments of mitochondrial genomes of three Pacific salmon species. Sequence variation within each segment was surveyed using the same 13 restriction endonucleases. Levels of variability in each segment were quantified using diversity indices, and the distributions of variability across the mitochondrial genome were compared for chum, sockeye and even- and odd-year broodlines of pink salmon. The mtDNA haplotype genealogies were reconstructed for each species using the entire set of variable sites, and the contributions of variation within each mtDNA segment were compared to the genealogies resolved.

Pacific salmon exhibit some of the lowest mtDNA nucleotide diversities among fish species (Billington & Hebert 1991; Bernatchez & Wilson 1998). Several mtDNA-based studies of *Oncorhynchus* species (Wilson *et al.* 1987; Cronin *et al.* 1993; Bickham *et al.* 1995), including those based on the sequencing of the presumably hypervariable segments of the D-loop (Park *et al.* 1993; Nielsen *et al.* 1994), encountered low resolution among mtDNA lineages, which hindered comprehensive phylogeographic analysis. Therefore, achieving the maximum resolution of mtDNA genealogies was also an objective. However, in this paper, we focus on the analysis of concordance of genealogical information across different mtDNA segments. The species-specific topologies of the mtDNA trees will be discussed elsewhere in the framework of the phylogeographic surveys conducted for each species.

Materials and methods

Sample collections used in this study included: (i) 40 individuals from the even-year broodline of pink salmon (*Oncorhynchus gorbuscha*), which included 10 each from Port Caldera on Prince of Wales Island in Southeast Alaska, from both Duck Creek in Prince William Sound and Little Susitna River in Cook Inlet (both in Southcentral Alaska), and from Kushiro River on Hokkaido Island, Japan; (ii) 40 odd-year broodline pink salmon, 10 each from Karta River on Prince of Wales Island and Herman Creek on Behm Canal in Southeast Alaska, from Duck Creek, and from Kushiro River; (iii) 50 chum salmon (*Oncorhynchus keta*), 10 each from Karta River, from Wilson River on Behm Canal, from Little Susitna River, from Heilong (Amur) River at Fuyuan in northeastern Heilongjiang province of China, and from Suifen River at Dongning Hatchery in southeastern Heilongjiang province; and (iv) 40 sockeye salmon (*Oncorhynchus nerka*), 10 each from Hugh Smith Lake near southern Behm Canal, from Karta River, from Yentna River, a tributary of the Susitna River in Southcentral Alaska, and from Kamchatka River in eastern Kamchatka.

Total DNA was isolated from ~50 µg of frozen heart tissues using a proteinase K lysis and ammonium acetate protein precipitation method (Puregene™ DNA isolation kits, Gentra Systems Inc.). Seven segments of the mtDNA molecule (referred to subsequently as 12S/16S rRNA, ND1/ND2, COI/COII/A8, A8/A6/COIII/ND3, ND3/ND4, ND5/ND6, and Cyt *b*/D-loop regions) were amplified separately using primers described by Gharrett *et al.* (2001). These primers produce PCR products ranging in size from 2115 to 2689 base pairs (bp) (according to the *O. mykiss* mtDNA sequence, Zardoya *et al.* 1995). Positions of the amplified segments relative to the mitochondrial gene map are shown in Fig. 1, and their gene contents are given in Table 1. These seven PCR-amplified segments cover 97% of the mitochondrial genome. The exact locations of

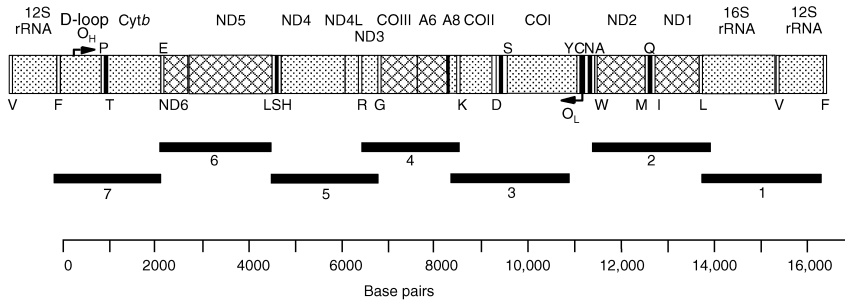


Fig. 1 Position of the seven PCR-amplified segments (black horizontal bars) relative to the linearized gene map of vertebrate mtDNA. The origins of H- and L-strand replication are indicated and all genes are marked next to the strand on which their coding sequences are found. The tRNA genes are depicted by their single letter amino acid codes. Exact locations of PCR primers are given in Gharrett *et al.* (2001).

priming sites and primer sequences are given in Gharrett *et al.* (2001).

PCR was carried out in 50- or 100- μ L volumes using 250–500 ng of total DNA preparation as template. The reaction mixture contained 10 mM Tris-HCl, pH 8.6; 50 mM KCl; 1.5 or 2.0 mM MgCl₂; 0.5 mM dNTPs; 0.25 μ M of each primer; and 0.5–1 units of *Taq* DNA polymerase (Perkin Elmer). Amplification was carried out in a Perkin Elmer 4800 thermal cycler as follows: preheating at 95 °C for 5 min, followed by 30–35 cycles of denaturing at 94 °C for 1 min, annealing at 52–56 °C for 1 min, and elongation at 72 °C for 2.5 min; amplification was completed with a final extension at 68 °C for 5 min. Each PCR-amplified mtDNA segment was digested in separate reactions by 13 restriction endonucleases (*AseI*, *BanII*, *BstNI*, *BstUI*, *DdeI*, *HhaI*, *HinfI*, *MboI*, *MspI*, *RsaI*, *Sau96 I*, *StyI* and *TaqI*). Reactions consisted of 9 μ L subsample of PCR product, 1 μ L of 10 \times restriction buffer, and 2–5 U of restriction enzyme. Restriction reactions were incubated at appropriate temperatures for at least 2 h and loaded onto 1.5% gels prepared from two parts Synergel™ (Diversified Biotech Inc.) and one part Ultra Pure™ agarose (Gibco BRL) in 0.5 \times TBE buffer (pH 8.3; 0.045 M Tris-borate and 0.001 M ethylenediaminetetraacetic acid) buffer. After electrophoresis, gels were stained with ethidium bromide (Sambrook *et al.* 1989) and photographed on a UV transilluminator. A mixture of 25 and 100 bp DNA ladders (Gibco BRL) was used as molecular weight standards. When it was necessary to resolve fragments smaller than 100 bp, 8–10% acrylamide gels and SYBR™ Green staining (Molecular Probes Inc.) was used. Restriction fragment sizes for each enzyme in each of the seven PCR-amplified mtDNA segments are given in Churikov (2000).

Presence or absence of restriction sites in all seven mtDNA segments was inferred for each of the 13 enzymes from series of restriction fragment patterns that differed by a single site. The site codes across the seven mtDNA segments were concatenated for each restriction enzyme; and each fish was assigned a 13-letter code that described its composite, multienzyme haplotype. A binary character state matrix consisting of presence or absence of all restriction sites in composite haplotypes was produced using the GENERATE program in REAP (McElroy *et al.* 1992). The

numbers of site differences between all pairs of haplotypes were deduced from this matrix and used to construct a mutational network among haplotypes. Haplotype ('nucleon', *h*) and nucleotide (π) diversities for each mtDNA segment and sampling variances for haplotype diversities were estimated according to Nei & Tajima (1981) and Nei (1987) using REAP.

Note that two pairs of amplified segments had a non-negligible overlap (Gharrett *et al.* 2001). This resulted in duplication of some sites used to calculate overall nucleotide diversity and puts extra weight on some character state changes or similarities. Since the total overlap between amplified segments constituted less than 5% of the genome, the impact on the overall nucleotide diversity estimate is small.

Correlation analysis (Sokal & Rohlf 1995) was conducted to determine whether there is a similarity among the species and between the species pairs in the variability levels over the mtDNA segments. For this analysis, segment-specific nucleotide diversity estimates were ranked within each species and sums of ranks were computed for each mtDNA segment across the species. Kendall's coefficient of concordance (*W*) was computed to obtain an overall measure of agreement in the rankings of the segment-specific variabilities among the species; and Kendall's coefficients of rank correlation (τ) were computed for pairwise species comparisons. Boxplots were used to show the differences in location and dispersion of ranked nucleotide diversities among mtDNA segments.

mtDNA haplotype genealogies were constructed as minimum spanning networks. Each branch in the genealogies was marked by the mtDNA segment(s) where restriction site change(s) took place, so that topologies of the networks that would result from restriction site variation in any single mtDNA segment could be easily deduced and compared to each other. Topological congruence or similarity between 'segment-specific' and whole-genome trees and between all pairs of 'segment-specific' trees was quantified as one minus standardized Symmetric Distance of Robinson & Foulds (1981), which is based on the number of branches that are not shared between trees.

The reliability of the network topologies (or the limits of parsimony) was evaluated using the method developed by

Table 1 Summary of the survey of restriction site variation in seven segments of salmon mitochondrial genomes

Seg. no.	Gene contents*	Size† (bp)	Base pairs surveyed/haplotype	No. variable sites	No. haplotypes	Haplotype diversity (± SE)	Nucleotide diversity (π)	Bootstrap 95% CI for π
Pink salmon, even-broodyear line (40 fish)								
1	12S/16S rRNA	2402	260.66	2	3	0.1885 ± 0.07967	0.0392%	
2	ND1/ND2	2689	260.83	3	4	0.1910 ± 0.08197	0.0553%	
3	COI/COII/A8	2471	242.00	3	3	0.0987 ± 0.06381	0.0320%	
4	A8/A6/COIII/ND3	2115	179.16	3	4	0.3103 ± 0.08816	0.0939%	
5	ND3/ND4L/ND4	2331	188.01	2	3	0.3833 ± 0.08792	0.1098%	
6	ND5/ND6	2488	184.00	4	5	0.6654 ± 0.05310	0.2528%	
7	Cyt b/D-loop	2599	287.62	6	7	0.6115 ± 0.07798	0.1518%	
	Whole mtDNA	16600	1597.46	23	16	0.8731 ± 0.04127	0.0971%	0.0782–0.1104%
Pink salmon, odd-broodyear line (40 fish)								
1	12S/16S rRNA	2402	256.66	5	6	0.2795 ± 0.09242	0.0583%	
2	ND1/ND2	2689	260.89	8	9	0.7744 ± 0.04299	0.2974%	
3	COI/COII/A8	2471	235.62	6	7	0.6897 ± 0.05453	0.1920%	
4	A8/A6/COIII/	2115	178.00	5	5	0.2359 ± 0.08803	0.0867%	
5	ND3/ND4L/ND4	2331	187.07	5	5	0.5744 ± 0.05776	0.2005%	
6	ND5/ND6	2488	187.83	3	4	0.6577 ± 0.05823	0.2729%	
7	Cyt b/D-loop	2599	285.33	4	5	0.3154 ± 0.09144	0.1036%	
	Whole mtDNA	16600	1594.03	36	24	0.9410 ± 0.02647	0.1667%	0.1467–0.1803%
Pink salmon (80 fish)								
1	12S/16S rRNA	2402	258.08	7	8	0.2345 ± 0.06310	0.0377%	
2	ND1/ND2	2689	260.60	9	10	0.5671 ± 0.05931	0.2244%	
3	COI/COII/A8	2471	237.56	8	9	0.4725 ± 0.06371	0.1232%	
4	A8/A6/COIII/	2115	178.50	7	8	0.2756 ± 0.06514	0.0923%	
5	ND3/ND4L/ND4	2331	187.44	6	6	0.5006 ± 0.05592	0.1663%	
6	ND5/ND6	2488	186.00	5	7	0.6687 ± 0.04290	0.2672%	
7	Cyt b/D-loop	2599	287.11	8	9	0.4804 ± 0.06608	0.1300%	
	Whole mtDNA	16600	1595.75	51	38	0.9519 ± 0.01280	0.1439%	0.1286–0.1551%
Chum salmon (50 fish)								
1	12S/16S rRNA	2402	220.67	7	7	0.4776 ± 0.07439	0.1498%	
2	ND1/ND2	2689	213.46	4	5	0.3461 ± 0.07940	0.1828%	
3	COI/COII/A8	2471	210.44	8	9	0.7282 ± 0.04502	0.5625%	
4	A8/A6/COIII/ND3	2115	167.09	10	8	0.4751 ± 0.08467	0.3264%	
5	ND3/ND4L/ND4	2331	184.53	11	10	0.7265 ± 0.05533	0.4831%	
6	ND5/ND6	2488	196.78	8	6	0.4408 ± 0.08266	0.3357%	
7	Cyt b/D-loop	2599	240.66	4	4	0.4022 ± 0.07593	0.1066%	
	Whole mtDNA	16600	1427.59	52	25	0.9224 ± 0.02213	0.2816%	0.2257–0.3359%
Sockeye salmon (40 fish)								
1	12S/16S rRNA	2402	226.66	2	3	0.5372 ± 0.02904	0.1258%	
2	ND1/ND2	2689	269.50	12	12	0.7500 ± 0.06432	0.5210%	
3	COI/COII/A8	2471	192.89	2	3	0.5372 ± 0.02904	0.1483%	
4	A8/A6/COIII/ND3	2115	173.33	1	2	0.5128 ± 0.01790	0.1493%	
5	ND3/ND4L/ND4	2331	207.34	2	4	0.6859 ± 0.04025	0.2499%	
6	ND5/ND6	2488	175.22	6	6	0.6077 ± 0.04351	0.2267%	
7	Cyt b/D-loop	2599	224.33	6	4	0.5987 ± 0.04344	0.5165%	
	Whole mtDNA	16600	1465.06	31	17	0.8692 ± 0.04141	0.2890%	0.2613–0.2964%

Seg. no., segment number.

*For tRNA gene contents see exact location of PCR primers in Gharrett *et al.* (2001) and Table 1 in Zardoya *et al.* (1995).

†Sizes are given according to *O. mykiss* mtDNA sequence (Zardoya *et al.* 1995).

Templeton *et al.* (1992). The parsimony probability for RFLP data was calculated using the program PARSROB v1.1 (http://bioag.byu.edu/zoology/crandall_lab/programs.htm) for each pair of RFLP haplotypes.

Results

Restriction enzyme coverage, numbers of variable sites and haplotypes, and diversity indices for each mtDNA

segment and the entire genome are given for each species in Table 1. Variable restriction sites were distributed unevenly among the seven mtDNA segments surveyed in the three salmon species. Both haplotype and nucleotide diversity estimates varied among genome segments within each species. Nucleotide diversities for a number of genome segments within species differed several-fold. The species-specific and overall levels of rank-ordered nucleotide diversities for seven mtDNA segments are shown in Fig. 2(b). Notably, three segments encompassing genes for NADH-dehydrogenase (ND) subunits show the three highest levels of the overall nucleotide diversity, indicating an overall tendency of these genes to have elevated variability. The Cyt *b*/D-loop region shows the next highest, two segments encompassing genes for the subunits of cytochrome *c* oxidase and ATPase show lower levels, and

as expected, the 12S/16S region shows the lowest overall nucleotide diversity level.

Variability of the genome segments differed among species, except for the ND3/ND4 and 12S/16S regions, which showed consistently high and low diversity, respectively (Fig. 2a). Correlation analysis showed low similarity of the levels of variability over the seven mtDNA segments across species. Kendall's coefficient of concordance (*W*), which ranges from 0 to 1, was 0.42, and it was not significant ($\chi^2 = 10.18$, 6 d.f., $P = 0.12$). However, the low *P*-value observed could be suggestive of a significant correlation, given the stochastic noise in the data. All pairwise correlations of the variability levels in mtDNA segments between species were also low (0–0.52) and insignificant ($0.1 < P < 0.9$). A significant correlation of the variability levels across species was observed only after the pooling of seven mtDNA segments into three groups: (i) ND1/ND2, ND3/ND4L/ND4, ND5/ND6; (ii) COI/COII/A8, A8/A6/COIII/ND3, Cyt *b*/D-loop; and (iii) 12S/16S rRNA. These three groups of genes/genome segments show high correlation ($W = 1$) across the species and significantly different variability levels (Friedman ANOVA $\chi^2 = 8$, $N = 4$, 2 d.f., $P < 0.02$). The group containing three segments that encompassed all ND genes had the highest [sum of ranks (SR) = 21], ribosomal RNA region had the lowest (SR = 6), and the group of the remaining genome segments had an intermediate (SR = 14.3) level of overall variability.

Mitochondrial haplotype genealogies for each species are presented in Fig. 3 in the form of minimum spanning haplotype trees/networks, where resolved haplotypes are connected by the shortest possible branches representing mutational changes of the restriction site states. The 95% confidence limits estimated for parsimonious connections between haplotype pairs (Templeton *et al.* 1992) were eight mutational differences for even- and odd-year pink, and seven mutational differences for chum and sockeye salmon. Only the longest branches, which link the most divergent groups of haplotypes in chum and sockeye salmon networks exceeded the 95% limit of parsimony. Those branches were drawn with dotted lines (Fig. 3c,d). The rest of the chum and sockeye salmon haplotype networks and the entire networks for even- and odd-year pink salmon represent true haplotype genealogies with high probability, although a few loops show uncertainty due to site homoplasy within the limits of parsimony.

Segments of mtDNA where each mutational change took place are indicated along the branches, so that the contribution of variation within any particular genome segment to the resolution of haplotypes and their genealogical groups can be seen. Typically, the segments with the highest nucleotide diversities contained variable sites that produced the major genealogical subdivisions. Most of the variable sites in the segments with low nucleotide diversities defined unique haplotypes at the periphery of

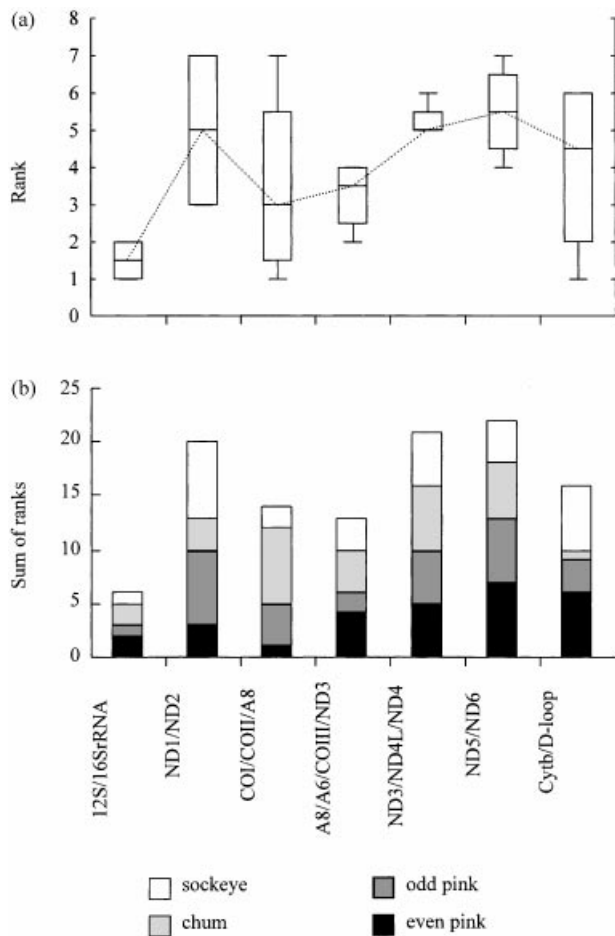


Fig. 2 Extent of variability in the seven segments of the mitochondrial genome of the three species of Pacific salmon. (a) The boxplots showing the medians (middle lines), interquartile ranges (boxes), and maximum and minimum values (upper and low whiskers) of the rank-ordered nucleotide diversities. (b) Stacked histogram illustrating overall and species-specific levels of variability in the different mtDNA segments.

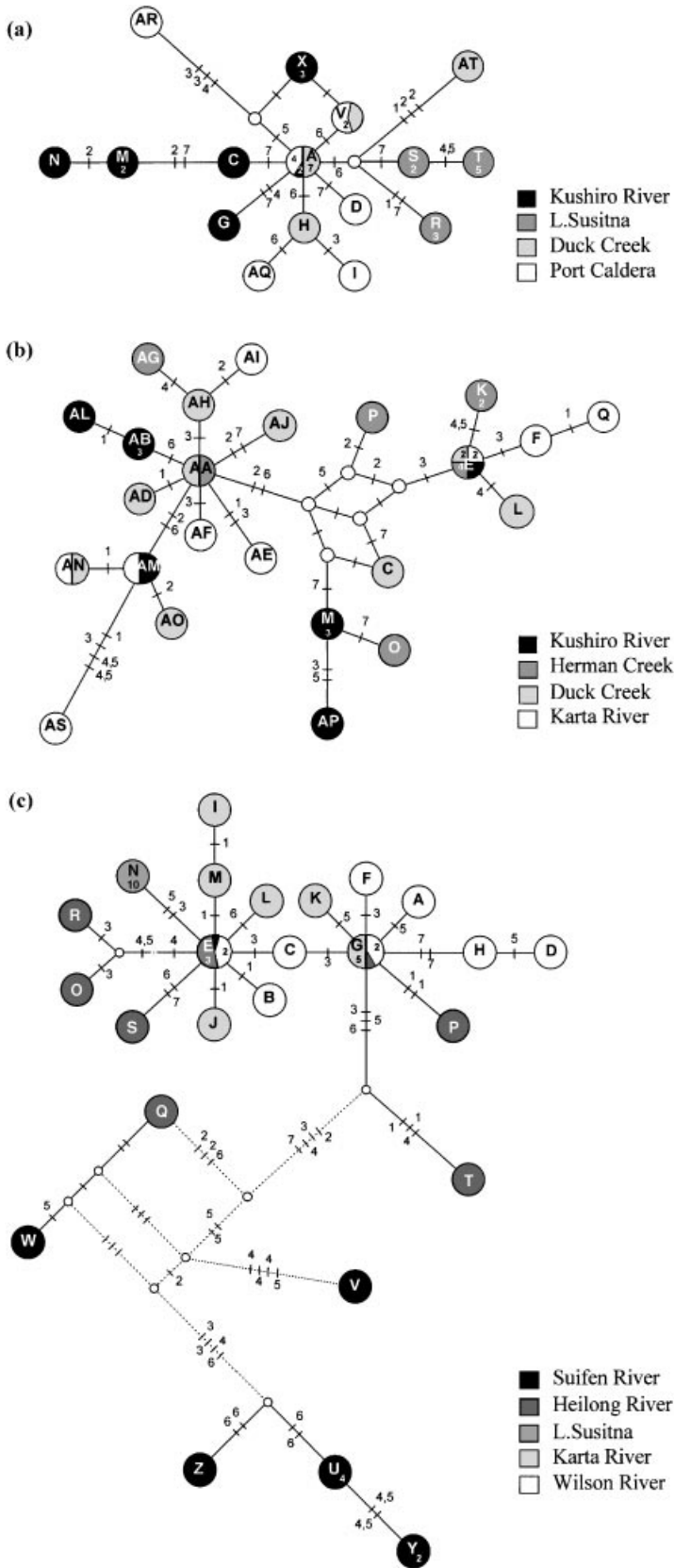


Fig. 3 Minimum spanning haplotype networks of even-brood-year (a) and odd-brood-year (b) pink, chum (c), and sockeye (d) salmon. The circles represent haplotypes and are not scaled according to haplotype abundance, but numbers of fish represented by each haplotype (if more than one) are shown within the circles. Small empty circles in the nodes of the networks indicate haplotypes inferred from the network, but not observed in the samples. Tick marks on the branches indicate mutational changes and nearby numbers refer to the mtDNA segment (Table 1) where mutational change took place. Branches drawn with dotted lines exceed 95% limits of parsimony.

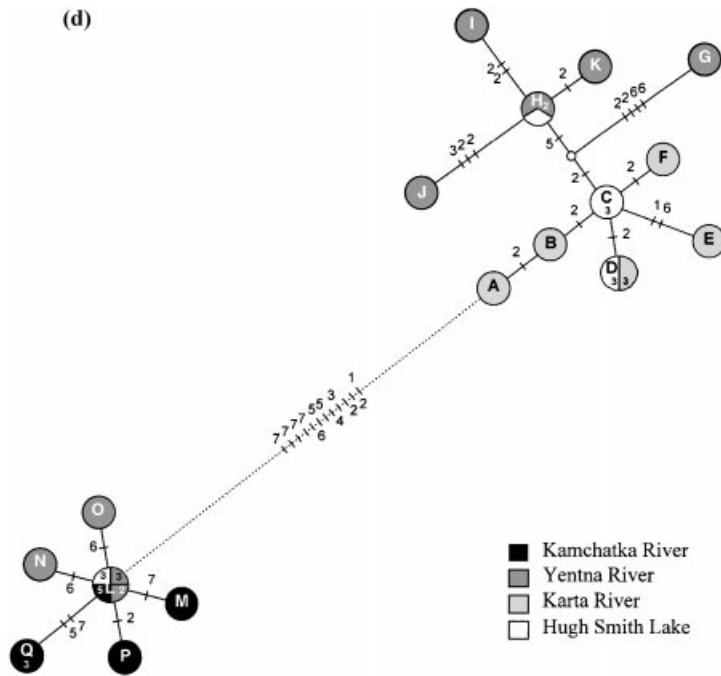


Fig. 3 Continued

the genealogical trees and had little influence on tree topology. Topological congruence between the pairs of 'segment-specific' trees (those based on variation in mtDNA regions) and between 'segment-specific' and whole-genome trees is summarized in Table 2. Generally, very little similarity was observed among haplotype partitions based on variation in different mtDNA segments, except for the partition of the sockeye salmon haplotypes into two groups that was imprinted in variation within every mtDNA segment.

Because the distributions of variability across mtDNA segments observed within each species are important for practical applications, they will be briefly characterized along with the similarities among haplotype partitions based on variation in different segments of mitochondrial genome.

Pink salmon, even-year broodline

The highest level of variation (both haplotype and nucleotide diversities) was observed in the ND5/ND6, and the lowest in the COI/COII/A8 region. The Cyt *b*/D-loop harboured the largest number of variable sites, but since the number of sites surveyed in this segment was also large, the nucleotide diversity estimate for the Cyt *b*/D-loop was only half that of ND5/ND6. Mutational changes in these two segments defined internal branches in the genealogical tree (Fig. 3a). Mutational changes in ND5/ND6 and Cyt *b*/D-loop do not, however, exhibit identical patterns across the individual fish analysed (as could be expected in the absence of recombination between mtDNA

lineages) and do not result in the congruent genealogies when considered separately. The rest of the mtDNA segments contained only a small amount of additional genealogical information, which added little to the topology of the tree defined by the two most variable segments.

Pink salmon, odd-year broodline

The ND1/ND2 region harboured the largest number of variable sites and had the highest diversity indices. The lowest nucleotide diversity was estimated in the 12S/16S region. The 12S/16S region was not more diverse at the nucleotide level in the odd- than in the even-year broodline, in spite of having nearly twice as many variable sites. This is due to the very low level of polymorphism of variable sites in the 12S/16S region of the odd-year broodline. However, overall in the odd-year broodline, there were more variable sites, and many of them were more polymorphic. Consequently, the mtDNA genealogy is more structured (Fig. 3b). The majority of the highly polymorphic sites (which determined topology of the genealogical tree) occurred in the ND1/ND2, ND5/ND6 and COI/COII/A8 regions. Few of the variable sites had identical patterns of polymorphism across individuals; hence, breaks in the genealogy are small. Genealogies reconstructed from segment-specific data have little (if any) congruence.

Chum salmon

The decisive variation for haplotype genealogy in this species occurred in the COI/COII/A8 and ND3/ND4L/

Table 2 Topological congruence* among haplotype trees reconstructed from seven mtDNA segments and between the 'whole-genome' and each of the 'segment-specific' trees

Seg. no.	Genes included	12S/16S rRNA	ND1/ND2	COI/COII/A8	A8/A6/COIII/ND3	ND3/ND4L/ND4	ND5/ND6	Cyt <i>b</i> /D-loop
Even-brood-year pink salmon								
1	12S/16S rRNA							
2	ND1/ND2	0.25						
3	COI/COII/A8	0.00	0.00					
4	A8/A6/COIII/ND3	0.00	0.00	0.25				
5	ND3/ND4L/ND4	0.00	0.00	0.00	0.25			
6	ND5/ND6	0.00	0.00	0.00	0.00	0.00		
7	Cyt <i>b</i> /D-loop	0.14	0.13	0.00	0.13	0.00	0.00	
Whole mtDNA		0.13	0.19	0.13	0.19	0.13	0.25	0.38
Odd-brood-year pink salmon								
1	12S/16S rRNA							
2	ND1/ND2	0.00						
3	COI/COII/A8	0.18	0.00					
4	A8/A6/COIII/ND3	0.11	0.11	0.10				
5	ND3/ND4L/ND4	0.11	0.00	0.22	0.33			
6	ND5/ND6	0.00	0.25	0.00	0.00	0.00		
7	Cyt <i>b</i> /D-loop	0.00	0.10	0.00	0.00	0.00	0.00	
Whole mtDNA		0.23	0.27	0.27	0.15	0.15	0.12	0.15
Chum salmon								
1	12S/16S rRNA							
2	ND1/ND2	0.00						
3	COI/COII/A8	0.00	0.09					
4	A8/A6/COIII/ND3	0.09	0.13	0.15				
5	ND3/ND4L/ND4	0.00	0.00	0.12	0.23			
6	ND5/ND6	0.00	0.11	0.14	0.08	0.06		
7	Cyt <i>b</i> /D-loop	0.00	0.20	0.09	0.13	0.00	0.11	
Whole mtDNA		0.20	0.10	0.30	0.20	0.33	0.23	0.10
Sockeye salmon								
1	12S/16S rRNA							
2	ND1/ND2	0.08						
3	COI/COII/A8	0.33	0.18					
4	A8/A6/COIII/ND3	0.50	0.09	0.50				
5	ND3/ND4L/ND4	0.25	0.08	0.25	0.33			
6	ND5/ND6	0.40	0.14	0.17	0.20	0.14		
7	Cyt <i>b</i> /D-loop	0.25	0.08	0.25	0.33	0.50	0.14	
Whole mtDNA		0.12	0.65	0.12	0.06	0.18	0.29	0.18

Seg. no., segment number.

* Topological congruence is $1 - D_s/n$, where D_s is the symmetric distance of Robinson & Foulds (1981), and n is a total number of distinct branches in two trees.

ND4 regions. These two mtDNA segments with highest nucleotide diversities contained the most polymorphic sites, which defined the major haplotype assemblages. The lowest diversity at the nucleotide level was observed in the Cyt *b*/D-loop region. The lowest haplotype diversity was observed in the ND1/ND2 region, and there were no polymorphic ND1/ND2 sites in Alaskan individuals. The existence of two genealogical assemblages in the collection of haplotypes from Alaska was reflected by variation in the COI/COII/A8 region alone (Fig. 3c). Analysis of data from this segment could parallel, to some extent, the analysis

based on the entire genome, and would eventually result in similar phylogeographic inferences. Variation in the rest of the segments poorly resolved the genealogy and, when analysed separately, resulted in noncongruent outcomes. Variation in the 12S/16S region, for instance, does not differentiate the most divergent haplotypes from Asia, and does not even resolve them as a separate group. Overall, higher nucleotide diversity was estimated in the sample of 50 chum than in the sample of 80 pink salmon. This was attributable to the elevated variation in mtDNA of Asian specimens.

Sockeye salmon

Substantial variability (as compared to other segments) was detected in ND1/ND2. This mtDNA segment harboured more than one-third of the variable restriction sites detected in the sockeye mitochondrial genome. Variation at ND1/ND2 sites resolved more than two-thirds of the haplotypes identified in the species. The ND1/ND2 region exhibited the highest diversity indices both at the haplotype and nucleotide levels (Table 1). A remarkable feature of sockeye salmon variation was that a common pattern of polymorphism was observed among 12 sites. At least one site from each genome segment exhibited this pattern. Together, they defined a major break in mtDNA genealogy of the species, which separated two haplotype assemblages (Fig. 3d). However, no similarity in the pattern of variation was observed beyond this point. Analysis of the variation in the Cyt *b*/D-loop region alone would be misleading in population studies. The variation in the Cyt *b*/D-loop region indicates a higher level of diversity in the haplotype assemblage, which includes fish from Asia, whereas this cluster is much less diverse than the other one (Fig. 3d) when variation across the entire mitochondrial genome is considered. The 12S/16S, COI/COII/A8 and A8/A6/COIII/ND3 regions all exhibited little variation and contributed little information to genealogical structure. Overall, the level of nucleotide diversity in sockeye was comparable to that in chum salmon, whereas haplotype diversity was lower than in either chum or pink salmon.

Discussion

The generalized pattern of distribution of variability in the mitochondrial genomes of three salmon species (Fig. 1) corresponds to those found earlier in the mouse and human mitochondrial genomes using high-resolution restriction site mapping (Ferris *et al.* 1983; Cann *et al.* 1984). The elevated variability of ND genes, as compared to other mitochondrial genes, was also detected in population genetic studies of fish closely related to salmon. Thus, higher variability was detected in the ND5 and ND6 genes than in the ATPase subunit 6 gene in rainbow smelt (Pigeon *et al.* 1998) or the mtDNA control region in brown trout (Apostolidis *et al.* 1997). A survey of variability in the same (as in this study) seven segments of coho salmon mtDNA, but conducted using a somewhat different set of restriction enzymes (Gharrett *et al.* 2001) also revealed the most variation in the ND5/ND6 and ND1/ND2 regions. Incorporating coho salmon mtDNA variation data into our analysis did not alter the overall pattern of distribution of variable sites across seven segments of salmon mitochondrial genome (Fig. 2b). We have to acknowledge, however, that the levels of variability estimated in various mtDNA segments should have large stochastic errors due

to sparse sampling (about 10% of the nucleotides) of the mtDNA segments; and the absolute values of nucleotide diversities shown in Table 2 represent observations from a single empirical study.

The observation of elevated variability of ND genes in Pacific salmon is in good agreement with the results of interspecific comparisons of mitochondrial gene and amino acid sequences (Meyer 1993; Zardoya *et al.* 1995). Recently, the mitochondrial genome of a second salmonid species, Atlantic salmon (*Salmo salar*), was completely sequenced (Hurst *et al.* 1999), and its sequence was aligned with that of the rainbow trout. This alignment revealed only slight differences in the extent of sequence variation among mitochondrial genes. Although, on average, ND genes were also more variable than ATPase and CO genes, the variance of variability within each group of genes was large.

Our data also support the view that in salmonids, the mtDNA control region has not generally evolved at a higher rate than that of the remainder of their mtDNA genomes (Shedlock *et al.* 1992). Although we estimated variability in the joint Cyt *b*/D-loop region and did not consider variability of the D-loop alone, there is evidence that only a slight difference in variation exists between the cytochrome *b* gene and the mitochondrial control region in brown trout (Giuffra *et al.* 1994), a related genus *Salmo*, hence the estimate of diversity in the segment may be representative of both genes.

An important characteristic of the distribution of variability across the mitochondrial genome that we observed at the intraspecific level is the large among-species variance of variability levels within mtDNA segments. Reduced variability in the chum salmon Cyt *b*/D-loop region is the most prominent example. The large difference between distributions of variable sites that is apparent even between brood-year lines of pink salmon supports the idea that such differences result from the stochasticity of lineage sorting and mutation processes rather than from differences in structural and functional constraints between homologous mtDNA segments. The stochastic component in the distribution of mutations along the DNA molecule is especially large when divergence among sequence types is low (as in the case of mtDNA haplotypes in salmon species), hence, little congruence of the distribution patterns is expected among mitochondrial genomes of different species. Consequently, preliminary screening of the mitochondrial genome is warranted for each species before a particular segment is chosen for a phylogeographic survey.

The absence of congruence between genealogical trees reconstructed from restriction site variation in different segments of pink and chum mitochondrial genomes can be explained by the nature of sequence polymorphism. Apparently, most sites mutated during the recent evolution of these species, and new sequence variants have not

had sufficient time to achieve higher frequencies (or become extinct). These recently arisen polymorphisms dominate in mtDNA of salmon species, and each mtDNA segment contains a unique set of them. Such polymorphisms define young mitochondrial lineages at the periphery of the genealogical tree. Sequence variants with higher frequencies (which presumably arose earlier in evolution) are rare in the mitochondrial genomes of Alaskan pink and chum salmon, and few have identical patterns of variation across individuals. Polymorphisms of the latter type produce breaks in the interior of the genealogical tree. Consequently, genealogical breaks between haplotype assemblages in pink and Alaskan chum are shallow and are not yet observed in the variation of every mtDNA segment. In contrast, more than one-third of the variable sites in sockeye salmon have an identical pattern of polymorphism across 40 specimens producing a deep break between two haplotype assemblages in mtDNA genealogy. Since this break is detectable in the variation of every mtDNA segment, some congruence exists between genealogies reconstructed from variation in a single mtDNA segment.

Low overall nucleotide diversity in salmon species imposes a problem on finding cladistically informative sites. A large genome screening effort was required to reveal the structure of the mitochondrial haplotype genealogy. In our study, examination of 167–288 bp per haplotype within a single mtDNA segment did not provide sufficient resolution to uncover essential features of haplotype genealogy. Only analyses of pooled data, that included 1424–1597 bp per composite haplotype across the entire genome, revealed genealogical structures. Consequently, we suggest that targeting a selected set of genealogically informative sites across the entire mtDNA genome by PCR-RFLP analysis would be the most efficient strategy for phylogeographic surveys of species with low intraspecific nucleotide diversities. Results of pink salmon phylogeographic and population genetic analysis using this method will be reported elsewhere (Churikov & Gharrett, in preparation).

Because genealogical information in the segments of mitochondrial genome is additive rather than congruent, sequencing a comparable portion of mitochondrial gene(s) of at least 1500 bp (certainly a more laborious way) would not guarantee resolution of the genealogy comparable to that derived from a random sample of sites. Cummings *et al.* (1995) showed that blocks of contiguous sites are not representative of the entire genome. If sequencing is to be performed anyway, the most appropriate target regions are ND5/ND6 in even brood-year pink salmon, ND1/ND2 in odd brood-year pink salmon and sockeye, and COI/COII/A8 in chum salmon. Variable sites that were found in these segments defined many (although not all) genealogical breaks and would be most informative for phylogeographic analysis of Asian and Alaskan salmon.

Acknowledgements

We are grateful to Chuck Guthrie III, Sharon Hawkins, Chris Kondzela and John Pohl for providing the best quality tissue samples for DNA isolation, and to Milo Adkison, Alexander Pudovkin and Richard Wilmot for helpful discussions. This research was funded by a grant from the US Department of Commerce, NOAA Saltonstall-Kennedy Fund NA76FD0032 awarded to A.J.G. D.C. was supported by a graduate fellowship from the Rasmuson Fisheries Research Center, University of Alaska Fairbanks.

References

- Apostolidis AP, Triantaphyllidis C, Kouvatzi A, Economidis PS (1997) Mitochondrial DNA sequence variation and phylogeography among *Salmo trutta* L. (Greek brown trout) populations. *Molecular Ecology*, **6**, 531–542.
- Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics*, **103**, 287–312.
- Avisé JC (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, **7**, 371–379.
- Avisé JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.
- Bernatchez L, Danzmann RG (1993) Congruence in control-region and restriction-site variation in mitochondrial DNA, of brook charr (*Salvelinus fontinalis* Mitchell). *Molecular Biology and Evolution*, **10**, 1002–1014.
- Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**, 431–452.
- Bernatchez L, Guyomard R, Bonhomme F (1992) DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology*, **1**, 161–173.
- Bielawski JP, Gold JR (1996) Unequal synonymous substitution rates within and between two protein-coding mitochondrial genes. *Molecular Biology and Evolution*, **13**, 889–892.
- Bickham JW, Wood CC, Patton JC (1995) Biogeographic implications of cytochrome *b* sequences and allozymes in sockeye (*Onchorynchus nerka*). *Journal of Heredity*, **86**, 140–144.
- Billington N, Hebert PDN (1991) Mitochondrial DNA diversity in fishes and its implications for introductions. *Canadian Journal of Fisheries and Aquatic Sciences*, **48**, 80–94.
- Brown WM, George M, Jr Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA*, **76**, 1967–1971.
- Brown JR, Beckenbach AT, Smith MJ (1993) Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). *Molecular Biology and Evolution*, **10**, 326–341.
- Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics*, **106**, 479–499.
- Churikov DY (2000) *Mitochondrial DNA haplotype genealogies and population histories in the Late Pleistocene: contrasts of pink salmon broodyears*. MSc Thesis. University of Alaska, Fairbanks.
- Cronin MA, Spearman WJ, Wilmot RL, Patton JC, Bickham JW (1993) Mitochondrial DNA variation in chinook (*Onchorynchus tshawytscha*) and chum salmon (*O. keta*) detected by restriction

- enzyme analysis of polymerase chain reaction (PCR) products. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 708–715.
- Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology and Evolution*, **12**, 814–822.
- Edwards SV, Wilson AC (1990) Phylogenetically informative length polymorphism and sequence variability in mitochondrial DNA of Australian songbirds (*Pomatostomus*). *Genetics*, **126**, 695–711.
- Encalada SE, Lahanas PN, Bjørndal KA *et al.* (1996) Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*: a mitochondrial control region assessment. *Molecular Ecology*, **5**, 473–483.
- Ferris SD, Sage RD, Prager EM, Ritte U, Wilson AC (1983) Mitochondrial DNA evolution in mice. *Genetics*, **105**, 681–721.
- Fry AJ, Zink RM (1998) Geographic analysis of nucleotide diversity and song sparrow (*Aves*: *Emberizidae*) population history. *Molecular Ecology*, **7**, 1303–1313.
- Gharrett AJ, Gray AK, Brykov VA (2001) Mitochondrial DNA variation in Alaskan coho salmon, *Oncorhynchus kisutch*. *Fishery Bulletin*, **99**, in press.
- Giuffra E, Bernatchez L, Guyomard R (1994) Mitochondrial control region and protein coding genes sequence variation among phenotypic forms of brown trout *Salmo trutta* from northern Italy. *Molecular Ecology*, **3**, 161–171.
- Grant WS, Clark A-M, Bowen BW (1998) Why restriction site polymorphism analysis of mitochondrial DNA failed to resolve sardine (*Sardinops*) biogeography: insights from mitochondrial DNA cytochrome *b* sequences. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 2539–2547.
- Hoelzel AR, Hancock JM, Dover GA (1991) Evolution of the cetacean mitochondrial D-loop region. *Molecular Biology and Evolution*, **8**, 475–493.
- Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *American Journal of Human Genetics*, **46**, 828–842.
- Hurst CD, Bartlett SE, Davidson WS, Bruce IJ (1999) The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. *Gene*, **239**, 237–242.
- Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the USA*, **86**, 6196–6200.
- McElroy D, Moran P, Bermingham E, Kornfield I (1992) REAP: an integrated environment for the manipulation and phylogenetic analysis of restriction data. *Journal of Heredity*, **83**, 157–158.
- Meyer A (1993) Evolution of mitochondrial DNA in fishes. In: *Biochemistry and Molecular Biology of Fishes Vol. 2* (eds Hochachka PW, Mommsen P), pp. 1–38. Elsevier Press, Amsterdam, the Netherlands.
- Meyer A (1994) Molecular phylogenetic studies of fish. In: *Genetics and Evolution of Aquatic Organisms* (ed. Beaumont AR), pp. 219–249. Chapman & Hall, New York.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **97**, 145–163.
- Nielsen JL, Gan C, Thomas WK (1994) Differences in genetic diversity for mitochondrial DNA, between hatchery and wild populations of *Oncorhynchus*. *Canadian Journal of Fisheries and Aquatic Sciences*, **51** (Suppl. 1), 1301–1305.
- Park LK, Brainard MA, Dightman DA, Winans GA (1993) Low levels of intraspecific variation in the mitochondrial DNA of chum salmon (*Oncorhynchus keta*). *Molecular Marine Biology and Biotechnology*, **2**, 362–370.
- Pigeon D, Dodson JJ, Bernatchez L (1998) A mtDNA analysis of spatiotemporal distribution of two sympatric larval populations of rainbow smelt (*Osmerus mordax*) in the St. Lawrence River estuary, Quebec Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1739–1747.
- Robinson DF, Foulds LR (1981) Comparison of Phylogenetic Trees. *Mathematical Biosciences*, **53**, 131–147.
- Saccone C, Pesole G, Sbisà E (1991) The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. *Journal of Molecular Evolution*, **33**, 83–91.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
- Shedlock AM, Parker JD, Crispin DA, Pietsch TW, Burmer GC (1992) Evolution of the salmonid mitochondrial control region. *Molecular Phylogenetics and Evolution*, **1**, 179–192.
- Sokal RR, Rohlf FJ (1995) *Biometry: the Principles and Practice of Statistics in Biological Research*, 3rd edn. W.H. Freeman and Co., New York.
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.
- Walker D, Burke VJ, Barak I, Avise JC (1995) A comparison of mtDNA restriction sites vs. control region sequences in phylogeographic assessment of the musk turtle (*Sternotherus minor*). *Molecular Ecology*, **4**, 365–373.
- Wilson GM, Thomas WK, Beckenbach AT (1987) Mitochondrial DNA analysis of Pacific Northwest populations of *Oncorhynchus tshawytscha*. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**, 1301–1305.
- Zardoya R, Garrido-Pertierra A, Bautista JM (1995) The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. *Journal of Molecular Evolution*, **41**, 942–951.

This paper represents a collaborative effort of D. Churikov, a graduate student working with A. J. Gharrett, and colleagues and graduate students. D. Churikov is investigating the Pleistocene history of salmon populations in the northern Pacific. Dr Gharrett, a faculty member at the University of Alaska Fairbanks, has been engaged in Pacific salmon genetics studies in Alaska since 1978. M. Matsuoka and X. Luan are interested in salmonid phylogenetics. A. Gray is studying the population genetics of salmon and rockfish. Dr Brykov is studying salmon population structures in the Russian Far East.
