Congruence in Control-Region Sequence and Restriction-Site Variation in Mitochondrial DNA of Brook Charr (*Salvelinus fontinalis* Mitchill)[1]

Louis Bernatchez and Roy G. Danzmann
Department of Zoology, University of Guelph

We compared the congruence in genetic variation and phylogenetic relationships among mitochondrial DNA (mtDNA) haplotypes detected with restriction-fragment-length polymorphisms (RFLPs) and sequence analysis in the brook charr, *Salvelinus fontinalis* Mitchill. This was accomplished by analyzing variation both at 172 restriction sites subsampling 903 bp over the entire mitochondrial genome and in the sequence of a 300-bp segment of the control region among the same 33 individuals. The sequence and RFLP analyses were highly congruent both in terms of mtDNA variants detected per number of nucleotides sampled and in assessing phylogenetic relationships among haplotypes. Therefore, variation observed in a small segment of the control region was representative of that existing over the entire genome. The major difference between both approaches was in the number of mutational changes observed per nucleotide in the sequence analysis. This number was approximately twice as high as that in the RFLP analysis. Nevertheless, the level of variation observed in the control region of charr was less than expected a priori on the basis of previous observations in other vertebrates. Lower mutation rates may be related to an unusually low transition:transversion ratio (8:3) for mtDNA intraspecific variation in vertebrates. The combined restriction-site and sequence data resolved at least three distinct phylogenetic groupings in *S. fontinalis*. However, and despite the high-resolution level of the study (>1,200 bp screened, or 7% of the mitochondrial genome), some major branching patterns in gene tree topology remained unresolved, which stresses the importance of recognizing large standard errors associated with sequence variability when phylogeny is reconstructed.

Introduction

The analysis of mitochondrial DNA (mtDNA) variation has been widely used to address questions in population biology. Until recently, mtDNA variation was almost exclusively assessed by digestion of the entire mitochondrial genome by using restriction endonucleases and by separation of the resulting fragments, according to their molecular size, by gel electrophoresis (RFLP analysis). Automation of the polymerase chain reaction (PCR; Mullis et al. 1986) provides sequence information on specific sections of the genome in large-scale population surveys. Sequencing via PCR procedures is now widely used for macroevolutionary studies, and its use for studying intraspecific genetic variation is increasing rapidly (Vigilant et al. 1989, 1991; Meyer et al. 1990; Thomas et al. 1990; Carr and Marshall 1991; McVeigh et al. 1991; Smith and Patton 1991; Fajen and Breden 1992; Patton and Smith 1992).

1. Key words: mitochondrial DNA, control region, RFLP, brook charr.

Address for correspondence and reprints: Louis Bernatchez, INRS-Eau, 2800, rue Einstein, Suite 105, C.P. 7500, Sainte-Foy (Québec) G1V 4C7, Canada.

Mol. Biol. Evol. 10(5):1002–1014. 1993.

0737-4038/93/1005-0005$02.00

The potential technical and analytical superiority of PCR sequencing versus RFLP analysis has been addressed by several authors (e.g., see Kocher et al. 1989; Swofford and Olsen 1990; Carr and Marshall 1991). However, there are few data available that allow empirical comparisons of the efficiency of both approaches in detecting genetic variation and assessing phylogenetic relationships. These two aspects are of concern in intraspecific mtDNA studies (Avise et al. 1987). Previous studies have not consistently supported the higher performance of sequence analysis in detecting variation. For example, Edwards and Wilson (1990) showed that sequence analysis of the cytochrome *b* gene and RFLP analyses were congruent both in the level of variation detected and in reconstructing phylogenies in Australian songbirds (*Pomatostomus*). In fishes, indirect comparisons of variation detected by mtDNA RFLPs and cytochrome *b* sequence analysis produced conflicting results. For instance, Carr and Marshall (1991) argued that higher variation was detected via sequencing in the Atlantic cod (*Gadus morhua*), while comparable low levels of variation were detected by both methods in Atlantic salmon, *Salmo salar* (Birt et al. 1986; McVeigh et al. 1991). In a comparative study performed on the same individuals, Beckenbach et al. (1990) showed that sequence divergence estimates obtained by sequencing 2,214 bp of mtDNA coding regions were approximatively half those measured by RFLP analysis in the rainbow trout (*Oncorhynchus mykiss*). The authors concluded that the lower variation observed in sequence analysis resulted from not surveying the most variable regions of the genome.

Studies of mammalian mtDNA variation have revealed that the most variable segments of the mitochondrial genome are located in the noncoding, control region (D-loop) (Cann et al. 1984). Consequently, it has been shown that sequencing of these "hypervariable" segments reveals higher variation, relative to RFLP analysis (e.g., see Desmarais 1989; Vigilant et al. 1989). There is little information available to confirm whether the higher variability of the control region also holds true in other vertebrates—namely, in fish. There are indications from the literature that the mitochondrial control region may not be as variable in some fish species as it is in mammals (Meyer et al. 1990; Bernatchez et al. 1992). Conversely, Brown et al. (1993) observed that the substitution rate of the white sturgeon (*Acipenser transmontanus*) D-loop region was comparable to rates for hypervariable sequences in the human D-loop region. These contrasting results suggest that differences in patterns of mutation and constraints may exist between the control region of teleost and chondrostean fish.

The objective of this paper was to assess the congruence in genetic variation and phylogenetic relationships among haplotypes detected with RFLPs and sequence analysis of the mitochondrial control region in a teleost fish. This was accomplished by performing DNA sequencing on selected brook charr (*Salvelinus fontinalis* Mitchill) identified as belonging to divergent RFLP haplotypes.

Material and Methods

RFLP Analysis

In a previous study of wild and hatchery brook charr from Ontario (Danzmann et al. 1991*a*, 1991*b*), 27 distinct mtDNA haplotypes were identified by using 46 hexameric and multihexameric restriction enzymes that surveyed 172 sites. This corresponds to 903 bp (5.4% of the molecule) after adjustment is made for the semi-isoschizomer sites. Details about technical procedures of mtDNA purification and analysis, as well as the complete list of endonucleases used, are provided elsewhere

(Danzmann et al. 1991*b*). A restriction site map of the brook charr mtDNA molecule was constructed for the polymorphic restriction enzymes by conducting a series of single and double digests (N. Billington, personal communication).

Sequence Analysis

Purified mtDNA template was available for 33 individuals representative of 21 of the 27 haplotypes identified by RFLP analysis in a larger number of individuals. mtDNA of one individual of each haplotype was sequenced, except for the most common haplotypes (1, 6, 18, and 21; table 1), for which four representatives were analyzed. PCR amplifications were performed as outlined in the paper by Bernatchez et al. (1992). The entire control region (~1 kb) was amplified by using primers LN20 (5′ACCACTAGCACCCAAAGCTA) and HN20 (5′GTGTTATGCTTTAGTTA-AGC), located, respectively, in the proline and phenylalanine tRNA genes. Double-stranded PCR product was purified and sequenced as described by Bernatchez et al. (1992), except that an internal primer (H2; 5′CGTTGGTCGGTTCTTAC) located 350 bp from LN20 was used for sequencing. Thus, we sequenced a total of 340 bp, including 40 bp of the proline tRNA gene and the adjacent 300 bp of the 5′ end of the control region. This segment of the control region is homologous to one of the hypervariable segments reported in mammals. The proline tRNA gene segment was not consistently resolved on all individuals and therefore was not considered in further analyses.

Nucleotide-Variation Analysis

The presence/absence matrix of restriction sites constituted the raw data for all RFLP analyses. The total number of nucleotide substitutions between each pair of mtDNA haplotypes (d_{seq}) was estimated by using the correction of Jukes and Cantor (1969) with DNADIST. This method does not make correction for different rates of substitution for different nucleotide pairs. However, this has little effect on estimates of d at low divergence levels (i.e., <0.15) (for a discussion on uses of different methods for estimating d, see Nei 1987, pp. 64–72). Divergence levels were much lower than 0.15 in this study. Thus, d estimates from both data sets were comparable. We then regressed corrected pairwise distance measures obtained from RFLP and sequence analyses.

We also estimated the efficiency of both methods of detecting mtDNA diversity, by use of the combinatorial approach (Hebert et al. 1988) as applied by Bernatchez et al. (1989) to mtDNA data. Thus, the relationship between the number of mtDNA haplotypes detected as a function of genome sampling was quantified by an incremental random choice of nucleotides. For RFLP data, results were obtained by an incremental random choice of two restriction enzymes. We then attributed the number of nucleotides to their respective restriction enzymes, which yielded a sum of nucleotides for any enzyme combination. For sequence data, we performed an incremental random choice of 25 nucleotides. In all cases, the random choice was repeated 20 times for each incremental step.

Phylogenetic Analysis

Both restriction-site and sequence data were subjected to character-based analyses using PHYLIP version 3.3 provided by Joe Felsenstein (Department of Genetics, University of Washington, Seattle). Phylogenetic trees were constructed according to a maximal parsimony criterion by using the MIX and DNAPARS programs. The

Table 1
Twenty-Seven mtDNA Haplotypes Identified in Ontario Brood Stocks and Wild Fish by RFLP Analysis

Haplotype	*Acc*I	*Asn*I	*Bam*HI	*Ban*I	*Ban*II	*Bst*XI	*Dra*I	*Hin*dIII	*Nco*I	*Nhe*I	*Nsi*I	*Pst*I	*Pvu*II	*Sph*I	*Nde*I
1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2	A	A	A	A	B	A	A	A	A	A	A	B	A	B	A
3	B	A	A	A	B	A	A	A	A	A	A	B	A	A	A
4	A	A	A	A	A	A	A	A	A	A	B	A	A	A	A
5	B	A	A	A	A	A	A	A	A	A	B	A	A	A	A
6	A	A	A	D	B	A	A	A	B	B	A	B	A	A	B
7	A	A	A	A	A	A	A	B	A	A	A	A	A	A	A
8	A	B	A	A	A	A	A	A	A	A	A	A	A	A	A
9	A	A	A	A	B	A	A	A	B	B	A	B	A	A	B
10	A	A	A	A	B	A	A	A	B	B	A	A	A	A	B
11	A	A	A	A	B	A	A	A	B	B	A	B	A	A	C
12	A	A	A	D	B	A	A	A	B	A	A	B	A	A	B
13	A	A	A	D	B	A	A	A	B	B	A	A	A	A	C
14	A	A	A	D	B	A	A	A	B	B	A	B	A	A	C
15	A	A	A	D	B	A	A	A	B	B	A	A	A	A	B
16	A	A	A	A	A	A	A	A	A	A	A	B	A	A	A
17	A	A	A	D	A	A	A	A	A	A	B	A	A	A	B
18	C	A	A	D	B	A	A	A	B	B	A	B	A	A	B
19	A	A	A	D	B	A	A	A	A	B	A	B	A	A	B
20	C	A	A	C	B	A	A	A	B	B	A	B	A	A	B
21	A	C	B	A	B	B	B	A	A	A	A	B	B	A	A
22	D	C	B	A	B	B	B	A	A	A	A	B	B	A	A
23	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A
24	A	A	A	D	A	A	A	A	B	B	A	B	A	A	B
25	C	A	A	A	B	A	A	A	B	B	A	B	A	A	B
26	A	A	A	A	B	A	A	B	A	A	A	B	A	A	A
27	A	A	A	A	B	A	A	A	A	A	A	B	A	A	A

NOTE.—Capital letters refer to different fragment patterns detailed by Danzmann et al. (1991*a*). DNA template for haplotypes 10, 11, 17, and 25–27 was not available for sequence analysis.

majority-rule consensus tree and confidence statements on branches were estimated by running 100 replicates of the bootstrapping procedure using BOOT and DNABOOT programs. As there was no overlap of mutational sites detected by RFLP and sequence analyses (N. Billington, personal communication), both data sets were pooled to construct the overall phylogenetic tree.

Results

Nucleotide Variation

The 27 haplotypes observed in the RFLP study were generated from variation at 20 polymorphic sites (=903 bp) surveyed (table 1). Pairwise estimates of sequence divergence among haplotypes were 0.0011–0.0132 [mean 0.0062, standard deviation (SD) 0.0034], implying 1–12 (mean 5.6, SD 3.0) observed mutations.

Of 340 bp sequenced, 300 bp of the control region were surveyed for all individuals (fig. 1). Eleven polymorphic positions were detected; they were due to eight transitions, three transversions, and a simple deletion. Nine sequence haplotypes were identified among the 33 individuals analyzed (table 2). Pairwise sequence divergence estimates (i.e., d_{seq}) among the 21 individuals sequenced who corresponded to 21 distinct RFLP haplotypes were 0.000–0.023 (mean 0.010, SD 0.0067), implying 0–7 mutational steps (mean 3.1, SD 2.0). While many distinct RFLP haplotypes collapsed into the same sequence haplotype, there was only one instance of a different sequence haplotype that was undetected with RFLP analysis (6d; table 2).

Fewer variable sites were detected in the sequence analysis; however, the observed number of mutations corrected for the number of nucleotides surveyed was higher in sequence analysis, except at low levels of divergence (fig. 2). There was a strong positive linear relationship between estimates based on both methods. The slope value (1.86) of the best-fit linear model indicated that the number of mutations per nucleotide that were detected with sequence analysis was approximately twice that of the number detected with RFLP analysis at $d > 0.0025$ (i.e., the intercept with the one-to-one curve in fig. 2).

```
Proline tRNA                                                        1
-------------------------------------------                         *
GCACCCAAAG CTAGGATTCT AAGTTAAACT ACCCTCTGAC CCGCCCTAAT ATGTACAGCA      60
                 2                3
                 *                *
ATAAATGTTT GTACCTCACA AATTAGGGTT ATAATACATC TATGTATAAT ATTACATATT     120
                   4       5
                   *       *
GTGTATTTAC CCATATACAA AATATCTCAA TGGTGAGTAG TACATCATAT GTATTATCAA     180
  6                              7
  *                              *
CATAAGTGAA TTTAAGCCCT CATATATCAG CATATACCCA AGGTTTACAT AAGCTAAACA     240
                   8                    9     10         11
                   *                    *     *          *
CGTGATAATA ACCAACTAGG TTGTTTTAAC CTAGGCAATT GCAACATTAA CAAAACTCCA     300

ACTAACACGG GCTCCGTCTT TACCCACCAA CTTTCAGCAT
```

FIG. 1.—Sequence for the 5′ end segment of the mtDNA control region from haplotype 1 in *Salvelinus fontinalis*. Dashes indicate part of the proline tRNA gene. The proline tRNA gene segment was not consistently resolved on all individuals and therefore was not considered in further analyses. Asterisks and numbers above them indicate the 11 variable positions among the 12 haplotypes resolved.

Table 2
Variable Nucleotide Positions of Nine Haplotypes Resolved among 33 *Salvelinus fontinalis* mtDNAs Sequenced

	Nucleotide at Variable Site[b]										
Haplotype(s)[a]	1	2	3	4	5	6	7	8	9	10	11
1a–1d,16,23	G	T	A	A	C	T	A	G	C	C	C
2,3	—	—	—	—	—	C	—	—	—	—	—
4,5,7	—	—	—	—	—	—	T	—	—	—	—
6a–6c,9,12,15,19	—	—	G	—	—	C	—	—	T	—	—
8	—	—	—	—	—	—	—	—	—	—	T
13,14,6d	—	—	G	—	—	C	—	—	T	A	—
18a–18d,20	—	—	G	Del	—	C	—	—	T	—	—
21a–21d,22	A	C	—	—	G	C	—	A	—	—	—
24	—	—	G	—	—	—	—	—	T	—	—

[a] Nos. refer to distinct haplotypes revealed by RFLP analysis.

[b] Nos. refer to sequence position in fig. 1. The nucleotide at each position is for the first haplotype; for the other haplotypes, the nucleotides are given when different, while identity is indicated by dashes. Del = a deletion.

The relationships between the numbers of haplotypes detected as a function of the number of nucleotides sampled indicated an almost identical efficiency of sequence and RFLP analyses in detecting mtDNA variants (fig. 3). In both cases, this relationship was best defined by a linear regression (fig. 3). When the regression equations were applied, an average of nine haplotypes were detected in sequence analysis, as opposed to eight detected by RFLP when 300 nucleotides were screened. At 900 nucleotides screened, 21 haplotypes were resolved by RFLPs, while 24 would have been found by sequence analysis, under the assumption of an extrapolation of the linear model beyond 300 bp for sequence data.

Phylogenetic Trees

High levels of phylogenetic congruence were obtained between the RFLP and DNA sequence results (fig. 4). The most parsimonious tree obtained for RFLP data required a minimum of 27 mutational steps for 20 mutated sites (consistency index 0.74). Two main clusters of haplotypes (A and C) were supported at 79% and 100% bootstrapping levels, respectively, while other haplotypes (cluster B) formed an unresolved grouping.

The consensus tree obtained from sequence analysis required a minimum of 12 mutational steps for 11 mutated sites (consistency index 0.91). All haplotypes forming clusters A, B, and C in the RFLP tree also clustered distinctively in sequence analysis, with very comparable confidence statements. Elements of congruence were also observed at lower branching levels. For instance, pairs of haplotypes 18 and 20, 13 and 14, and 4 and 5 clustered together in both analyses. Conversely, minor discrepancies were observed. Haplotype 6d was identical to haplotypes 13 and 14 in sequence analysis. Haplotype 7 clustered with haplotypes 4 and 5 in sequence analysis but not in the RFLP analysis.

The overall consensus tree obtained by pooling both RFLP and sequence data sets required 39 required mutational steps for 31 mutational sites (consistency index 0.80) (fig. 5). The combination of both data sets added a single additional haplotype (6d) to the 21 identified by RFLPs alone. The haplotypes clustered into three groups

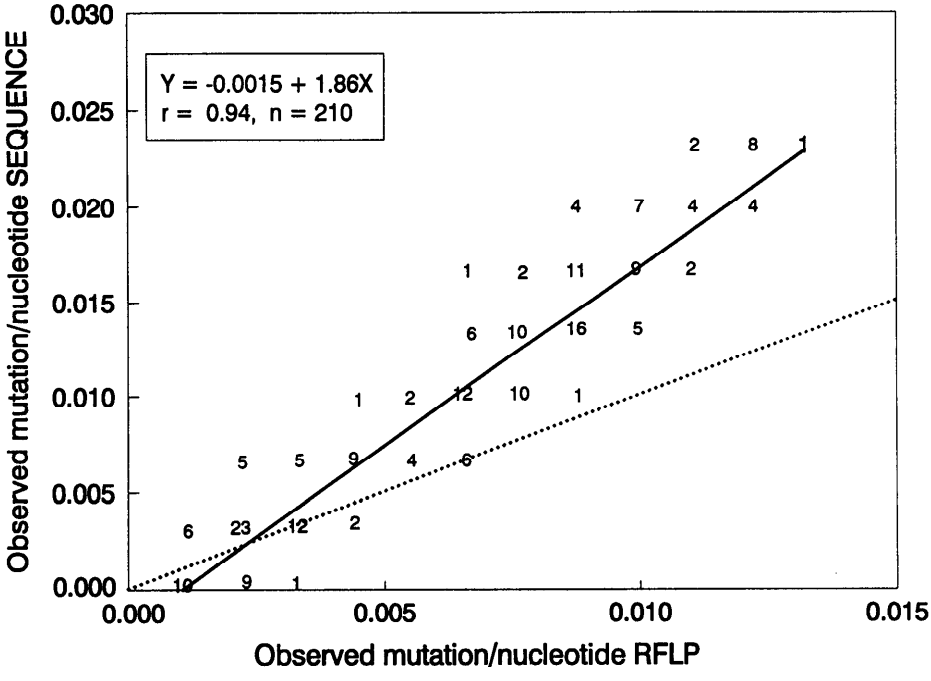

FIG. 2.—Pairwise relationships between observed number of mutations per nucleotide sampled, as detected by sequence and RFLP analyses for the 33 brook charr used in both studies. Values were estimated by dividing the observed number of mutations between each pair of individuals by the number of base pairs surveyed, for sequence (300 bp) and RFLP (903 bp) analyses. Numbers indicate the number of overlapping observations. The one-to-one relationship is illustrated by a dotted line.

identical to those described above. All bootstrapping values on major branches were increased by pooling of both data sets, which provided an additional indication of their congruence. However, group B still received low bootstrapping support, as did the branching pattern among the three groups that represent an unresolved trichotomy.

To further assess the statistical significance of similarity in phylogenies resolved from RFLP and sequence analyses, we conducted an analysis of correlation along branches of both trees. To do so, we constrained the RFLP and sequence data separately to fit the topology of the combined RFLP and sequence tree. We then correlated the lengths of each branch for both data sets. As expected from the observed similarity in the general topology and bootstrapping support of both RFLP and sequence trees (fig. 4), their branch lengths were positively correlated ($r = 0.78$, $n = 23$). This comparison gave additional support to the hypothesis that RFLP and sequence data sets produced similar trees.

Discussion

The present study indicates that both a sequence analysis of the control region and a random RFLP study over the entire mitochondrial genome were highly congruent in terms of both detecting the number of mtDNA variants per number of nucleotides sampled and depicting phylogenetic relationships among haplotypes. Therefore, variation observed in a 300-bp segment of the control region appeared to be representative of that existing over the entire genome.

The major difference between the two approaches is in the number of mutational

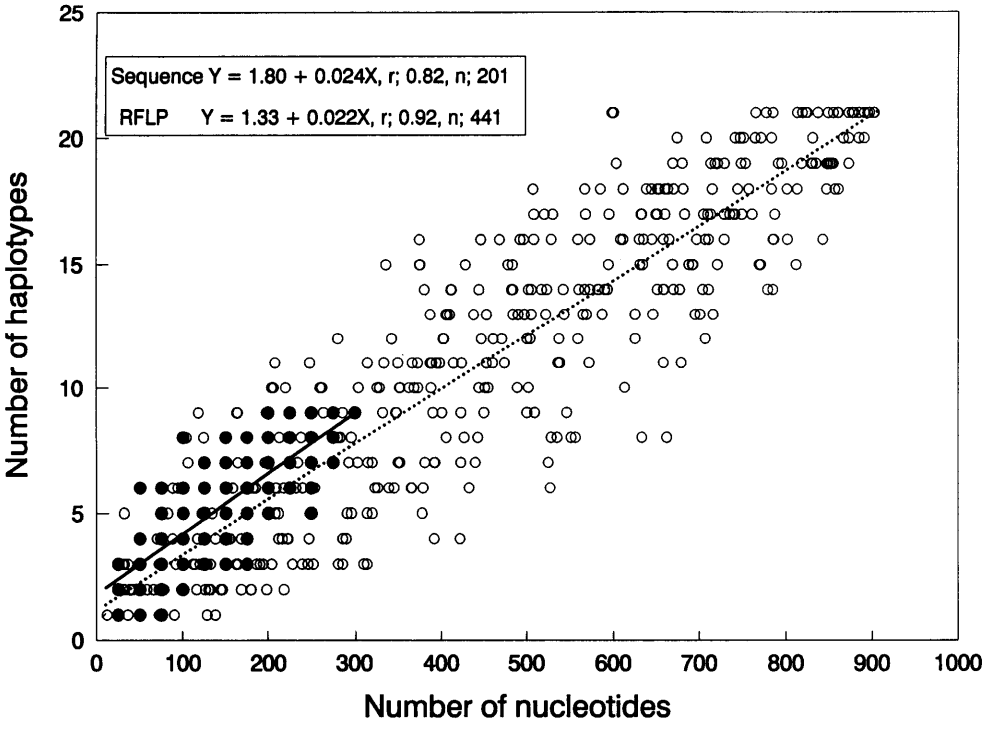

FIG. 3.—Number of mtDNA haplotypes depicted as a function of random nucleotide sampling for sequence (blackened circles) and RFLP (unblackened circles) analyses. Solid and dotted lines correspond to fitted linear regression curves for sequence and RFLP data, respectively.

changes observed per nucleotide sampled among the mtDNA variants found with sequence analysis, compared with RFLP analysis. At low divergence ($d < 0.0025$), higher estimates were observed with RFLP. This usually implied a situation where mtDNA variants differing by a single mutational change in the RFLP analysis collapsed into a single haplotype in sequence analysis. At $d < 0.0025$, higher divergence estimates in sequence analysis could be related either to a lower efficiency of RFLPs in detecting variation or to a higher mutational rate of the control region. The detection of mtDNA variants by RFLPs is only limited to variation found in palindromic sequences and may lead to an underestimation of the actual variation (see Carr and Marshall 1991). Furthermore, independent mutational changes occurring within the same restriction site may be undetected, thus underestimating divergence estimates or causing artifactual homoplasy. Mutational changes occurring in proximity may also go undetected because of limits in the resolution of fragment mobility detected by gel electrophoresis.

Despite these potential limitations, lower divergence estimates were observed in a sequence analysis of a 2,214-bp region encompassing ATPase 6, CO III, ND 3, ND 4L, $tRNA^{GLY}$, and $tRNA^{ARG}$ genes in rainbow trout, *Oncorhynchus mykiss,* where the same individuals were used for both sequencing and an RFLP study (Beckenbach et al. 1990). They found that divergence estimates based on RFLPs were approximately twice as high as those estimated from sequence analysis. Similarly, in a study of Australian songbirds (*Pomatostomus*), Edwards and Wilson (1990) reported lower or comparable divergence estimates in sequence analysis of the cytochrome *b* gene, versus RFLP analysis performed over the entire molecule. It thus appears more likely that

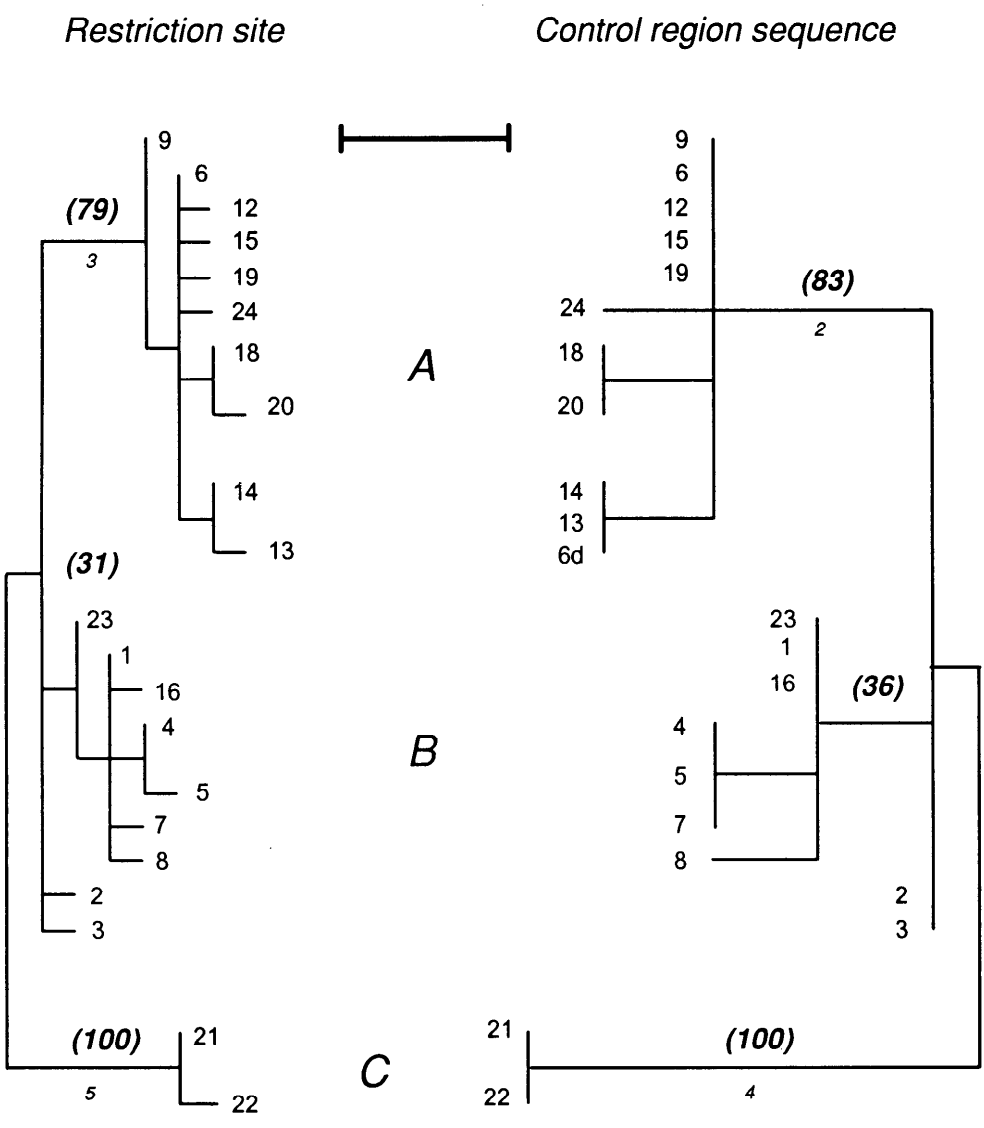

FIG. 4.—Majority-rule consensus trees relating haplotypes depicted by restriction-site (*left*) and control-region sequence (*right*) analyses among 33 brook charrs. Capital letters (A–C) identify the three haplotype groupings referred to in the text. Numbers at tip of branches refer to haplotype designation. Confidence statements (in percent) estimated from 100 bootstrapping replicates are given in brackets above branches for clades with ⩾31% support. Numbers of mutational events are given below branches (in small italics). Branches without numbers imply a single mutation. The scale bar represents 0.005 nucleotide substitutions/site.

the detection of higher divergence estimates in sequence analysis in the present study was more related to a higher mutational rate of the control region in *Salvelinus fontinalis,* compared with the rest of the molecule.

Our results corroborate the observations of a higher rate of mutation in the mitochondrial control region relative to coding regions, as generally reported in mammals as well as in chondrostean fishes (Brown et al. 1993). However, we observed major differences both in the level and the types of mutation occurring in brook charr. It

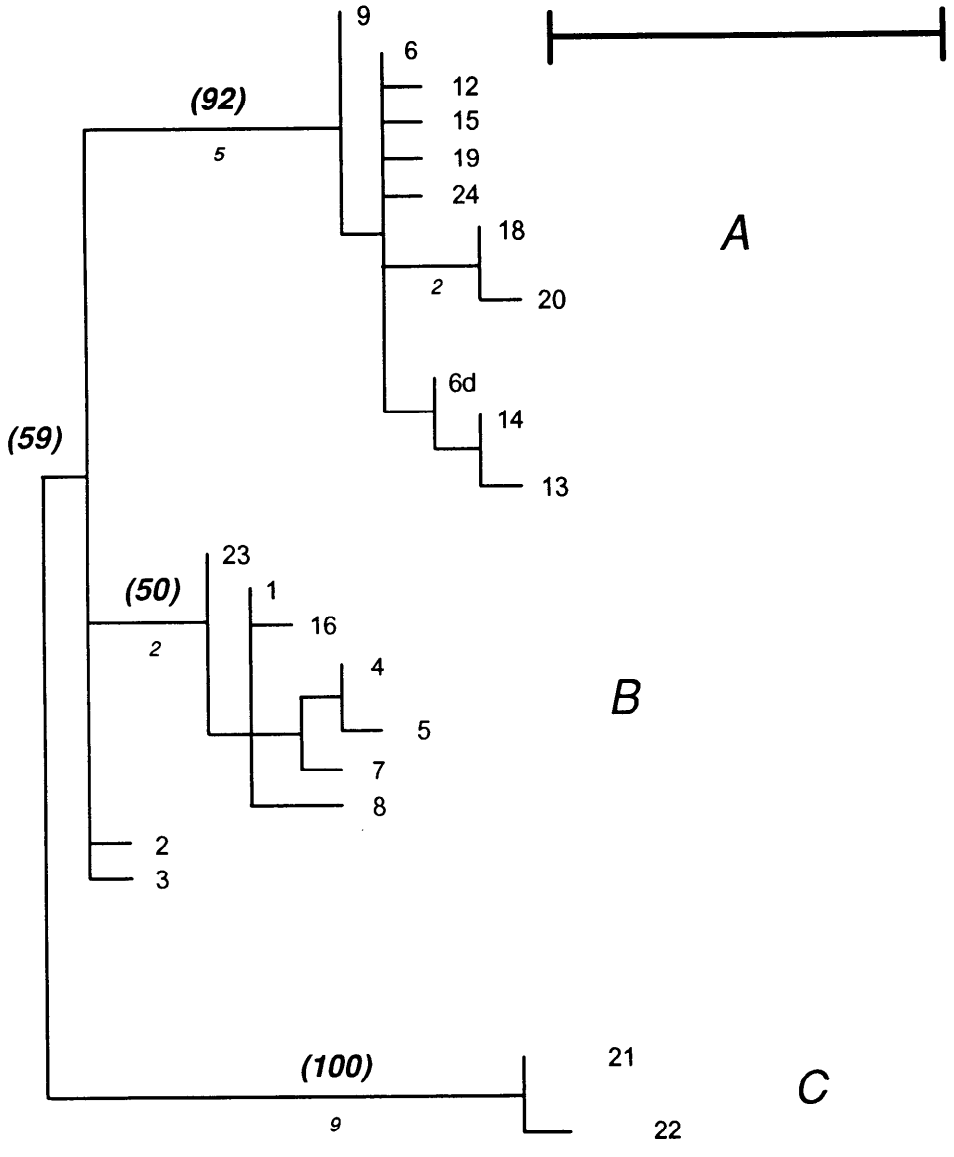

FIG. 5.—Majority-rule consensus tree among 22 brook charr mtDNA haplotypes, resolved by the combination of RFLP and control-region sequence data. Capital letters (A–C) identify the three haplotype groupings referred to in the text. Numbers at the tip of branches refer to haplotype designations. A single new haplotype (6d) was added to those resolved by RFLP analysis alone. Confidence statements (in percent) estimated from 100 bootstrapping replicates are given above branches (in brackets). Numbers of mutational events are given below branches (in small italics). Branches without numbers imply a single mutation. The scale bar represents 0.005 nucleotide substitutions/site.

has been frequently reported in mammals that segments of the control region evolve 5–15 times faster than the rest of the mtDNA genome (Aquadro and Greenberg 1983; Cann et al. 1984; Desmarais 1989; Vigilant et al. 1989). Brown et al. (1993) recently reported for white sturgeon a substitution rate in the D-loop region similar to that reported for humans. At best, we found in brook charr a ratio of 1.9:1 for a segment

of the control region homologous to a hypervariable segment reported in mammals and sturgeon.

The apparently lower mutation rate of the brook charr control region may be related to differences in the type of mutations. It has been widely documented in vertebrates that, at the intraspecific level, there is a strong mutational bias toward transitional events outnumbering transversions by a 10:1–32:1 ratio, in both protein-coding genes and noncoding regions of birds (e.g., see Edwards and Wilson 1990) and mammals (reviewed in Moritz et al. 1987), as well as in coding regions of fishes (Carr and Marshall 1991; McVeigh et al. 1991). Brown et al. (1993) also observed in the D-loop region of sturgeon a transitional bias (27:1) similar to that reported for humans. The 8:3 ratio observed in the present study contrasts with those observations but is comparable to the situation reported in the brown trout, *Salmo trutta,* where 17 transitions were observed for every 6 transversions (Bernatchez et al. 1992). This ratio also contrasts with a transition-based ratio (14:0) observed in an ongoing analysis of sequence variation in the coding cytochrome *b* and ATPase subunit VI genes of brown trout (E. Giuffra, L. Bernatchez, and R. Guyomard, unpublished results). This suggests that transitional mutations may be more restricted (or selected against) in the control region of these salmonids, thus leading to the decrease in the detectable mutation rate. The recent observation of a low transition:transversion ratio (approximately 3:1) in the D-loop region of a distantly related fish species (*Poecilia reticulata*) indicates that this phenomenon is not unique to salmonids but may be widespread in teleosts (Fajen and Breden 1992).

At present, we can only speculate on the causes of low transition/transversion mutational events observed in the control region of these salmonids. One explanation may be that the lower probability of transitions is related to structural or functional needs of maintaining a nonrandom nucleotide composition in the control region. For instance, it is well documented in *Drosophila* that the noncoding region containing the origin of replication for the mitochondrial genome (and therefore functionally homologous to the control region in vertebrates) is very A+T rich, i.e., 90%–100% (e.g., see Fauron and Wolstenholme 1980). These authors related such a disproportionately high A+T:G+C ratio to structural constraints linked to mechanisms of DNA replication in *Drosophila.* Consequently, and in opposition to the prevailing situation in vertebrates, transversions equaled or outnumbered transitions within the A+T-rich region of *Drosophila* species (e.g., see Clary and Wolstenholme 1987). Furthermore, the rate of mutation in this region was not significantly higher than that in the coding regions (Clary and Wolstenholme 1987).

The control regions of brook charr and brown trout are also biased toward a higher A+T:G+C ratio (66:34 and 68:32, respectively), compared with mammals. Conversely, the A+T:G+C ratio in the rest of the mtDNA molecule of *Salvelinus fontinalis* and *Salmo trutta* does not appear biased. On the basis of the A+T:G+C composition of the recognition sequences of the restriction enzymes used in this study, a 47:53 recognition ratio was expected in the restriction sites detected when random cuts to the genome and an equal A+T:G+C composition are assumed. The observed ratio was 49:51, which suggests that the frequency of A+T is not disproportionate in the overall mitochondrial genome. For brown trout the ongoing analysis of sequence variation in the coding cytochrome *b* and ATPase subunit VI genes (E. Giuffra, L. Bernatchez, and R. Guyomard, unpublished results) revealed an A+T:G+C ratio of 52:48. Structural or functional needs to maintain a given A+T composition in the control region of salmonids could therefore lead to a lower probability of transitions

(either A↔G or C↔T), because the frequency of different kinds of point mutations is expected to be a function of base composition (e.g., see Aquadro and Greenberg 1983) and because any transitional events will change the A+T:G+C content of the control region. Similarly, transversions that would not affect the A+T:G+C composition (i.e., A→T or G→C) should be more probable than other types. Although the present results tend to support this hypothesis, too few mutations were observed to quantify the frequency of each type of transversion.

Acknowledgments

We thank Drs. T. J. Crease and M. M. Ferguson for helpful discussion and suggestions. This study was financially supported by the Ontario Ministry of Natural Resources through the Ontario Renewable Resources Research Grant Program and by NSERC operating grant A46571. L.B. was supported by an NSERC postdoctorate fellowship.

LITERATURE CITED

AQUADRO, C. F., and B. D. GREENBERG. 1983. Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics **103**:287–312.

AVISE, J. C., J. ARNOLD, R. M. BALL, E. BERMINGHAM, T. LAMB, J. E. NEIGEL, C. A. REEB, and N. C. SAUNDERS. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. **18**:489–522.

BECKENBACH, A. T. 1991. Rapid mtDNA sequence analysis of fish populations using the polymerase chain reaction (PCR). Can. J. Fisheries Aquatic Sci. **48** Suppl. 1:95–98.

BECKENBACH, A. T., W. K. THOMAS, and H. SOHRABI. 1990. Intraspecific sequence variation in the mitochondrial genome of rainbow trout (*Oncorhynchus mykiss*). Genome **33**:13–15.

BERNATCHEZ, L., J. J. DODSON, and S. BOIVIN. 1989. Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. J. Fish. Biol. **35** Suppl. A:233–244.

BERNATCHEZ, L., R. GYUOMARD, and F. BONHOMME. 1992. Sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout (*Salmo trutta,* L.) populations. Mol. Ecol. **1**:161–173.

BIRT, P., J. M. GREEN, and W. S. DAVIDSON. 1986. Analysis of mitochondrial DNA in allopatric anadromous and nonanadromous Atlantic salmon, *Salmo salar.* Can. J. Zool. **64**:118–120.

BROWN, J. R., A. T. BECKENBACH, and M. J. SMITH. 1993. Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). Mol. Biol. Evol. **10**:326–341.

CANN, R. L., W. M. BROWN, and A. C. WILSON. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics **106**:479–499.

CARR, S. M., and H. D. MARSHALL. 1991. Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gadus morhua*) by the polymerase chain reaction. Can. J. Fisheries Aquatic Sci. **48**:48–52.

CLARY, D. O., and D. R. WOLSTENHOLME. 1987. *Drosophila* mitochondrial DNA: conserved sequences in the A + T rich region and supporting evidence for a secondary structure model of the small ribosomal DNA. Mol. Evol. **27**:116–125.

DANZMANN, R. G., P. E. IHSSEN, and P. D. N. HEBERT. 1991*a*. Genetic discrimination of wild and hatchery populations of brook charr (*Salvelinus fontinalis* Mitchill) in Ontario using mitochondrial DNA analysis. J. Fish Biol. **39A**:69–77.

DANZMANN, R. G., A. P. LISKAUSKAS, M. GERRITS, R. BARRETTE, and P. D. N. HEBERT. 1991*b*. An analysis of mitochondrial DNA variability in Ontario hatchery strains of brook charr (*Salvelinus fontinalis*). OMNR Tech. Rep. Ontario Ministry of Natural Resources, Ottawa.

DESMARAIS, E. 1989. Phylogénies intraspécifiques et histoire évolutive des populations de souris *Mus spretus* Lataste: analyse des lignées matriarcales par séquençage nucléotidique de l'ADN mitochondrial. Doctoral thesis, Université de Montpellier II, Montpellier, France.

EDWARDS, S. V., and A. C. WILSON. 1990. Phylogenetically informative length polymorphism and sequence variability in mitochondrial DNA of Australian songbirds (*Pomatostomus*). Genetics **126**:695–711.

FAJEN, A., and F. BREDEN. 1992. Mitochondrial DNA sequence variation among natural populations of the Trinidad guppy, *Poecilia reticulata.* Evolution **46**:1457–1465.

FAURON, C. M.-R., and D. R. WOLSTENHOLME. 1980. Intraspecific diversity of nucleotide sequences within the adenine + thymine rich region of mitochondrial DNA molecules of *Drosophila mauritiana, melanogaster,* and *simulans.* Nucleic Acids Res. **8**:5391–5410.

HEBERT, P. D. N., R. D. WARD, and L. J. WEIDER. 1988. Clonal-diversity patterns and breeding system variation in *Daphnia pulex,* an asexual-sexual complex. Evolution **42**:147–159.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–123 *in* N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KOCHER, T. D., W. K. THOMAS, A. MEYER, S. V. EDWARDS, S. PÄÄBO, F. X. VILLABLANCA, and A. C. WILSON. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86**:6196–6200.

MCVEIGH, H. P., S. E. BARTLETT, and W. S. DAVIDSON. 1991. Polymerase chain reaction/direct sequence analysis of the cytochrome *b* gene in *Salmo salar.* Aquaculture **95**:225–233.

MEYER, A., T. D. KOCHER, P. BASASIBWAKI, and A. C. WILSON. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature **347**:550–553.

MORITZ, C., T. E. DOWLING, and W. M. BROWN. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18**:269–292.

MULLIS, K. B., F. FALOONA, J. SCHARF, R. K. SAIKI, G. T. HORN, and H. A. ERLICH. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. **51**:263–273.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

PATTON, J. L., and M. F. SMITH. 1992. mtDNA phylogeny of Andean mice: a test of diversification across ecological gradients. Evolution **46**:174–183.

SMITH, M. F., and J. L. PATTON. 1991. Variation in mitochondrial cytochrome *b* sequence in natural populations of South American akodontine rodents (Muridae: Sigmodontinae). Mol. Biol. Evol. **8**:85–103.

SWOFFORD, D. L., and G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–500 *in* D. M. HILLIS and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

THOMAS, W. K., S. PÄÄBO, F. X. VILLABLANCA, and A. C. WILSON. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. **31**:101–112.

VIGILANT, L., R. PENNINGTON, H. HARPENDING, T. D. KOCHER, and A. C. WILSON. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc. Natl. Acad. Sci. USA **86**:9350–9354.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES, and A. C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.

WESLEY M. BROWN, reviewing editor

Received October 14, 1992; revision received March 27, 1993

Accepted March 27, 1993