Complex Evolution of a Salmonid Microsatellite Locus and Its Consequences in Inferring Allelic Divergence from Size Information

Bernard Angers† and Louis Bernatchez†*

*Département des sciences de l'environnement, Université du Québec à Montréal; and †GIROQ, Département de biologie, Université Laval, Québec

Sequence analysis of 78 alleles of a microsatellite locus was resolved at various taxonomic levels among salmonid fishes to test for congruence between size and sequence information. Allelic variability of this locus involved changes in number of dinucleotide repeats as expected for microsatellite loci. However, additional mutational events involving large indels and base substitutions also occurred frequently among and within species. These caused the evolution of different imperfect microsatellite variants, from an ancestral perfect one, which was undetectable from allelic size information only. Alleles of the same length resulting from different mutational events were also observed among and within species. Altogether, these results revealed the incongruence between size and number of mutation events, implying that it may be imprecise to interpret mutational rates and relationships on the basis of size information alone in interspecific comparisons and, to a lesser extent, within species.

Introduction

Microsatellites are short loci characterized by motifs of one to five base pairs (bp) repeated in tandem various times. Most likely due to the high slippage propensity of repeat arrays (Levinson and Gutman 1987; Schlötterer and Tautz 1992), microsatellites are amongst the fastest-evolving DNA sequences with mutational rates estimated to vary between 2.5×10^{-5} and 10^{-2} mutation per generation (Weber and Wong 1993). This results in an extensive polymorphism that makes microsatellite loci useful markers for genetic studies within species (Bowcock et al. 1994; Forbes et al. 1995; Goldstein et al. 1995b; Shriver et al. 1995; Nei and Takezaki 1996). Microsatellite loci are also generally conserved among closely related species (Moore et al. 1991; Schlötterer, Amos, and Tautz 1991; Stallings et al. 1991; Blanquer-Maumont and Crouau-Roy 1995; FitzSimmons, Moritz, and Moore 1995; Meyer et al. 1995; Pépin et al. 1995; Angers and Bernatchez 1996; Rico, Rico, and Hewitt 1996). This has raised the possibility that they could also be used in studies of interspecific relationships (Bowcock et al. 1994; Blanquer-Maumont and Crouau-Roy 1995; Meyer et al. 1995).

Genetic divergence among populations or species may be quantified from microsatellites without considering mutational events among alleles, for instance, by estimating the proportion of shared alleles (Bowcock et al. 1994; Estoup et al. 1995*a*) or computing standard genetic distance (Estoup et al. 1995*a*; Nei and Takezaki 1996). Slatkin (1995) and Goldstein et al. (1995*a*, 1995*b*) have suggested that classical divergence estimates for neutral loci under infinite allele model (IAM: Kimura and Crow 1964) were not appropriate for microsatellites because of high mutational rates and mutation processes which do not erase memory of the ancestral allelic state. Specific measures for microsatellites were proposed, using information of coalescence time between alleles based on strict stepwise mutation model (SMM: Ohta and Kimura 1973) (Goldstein et al 1995*a*; Shriver et al. 1995) or more generalized SMM processes (Goldstein et al. 1995*b*; Slatkin 1995) to take into account deviations of allele size variance from theoretical expectations of the SMM model as reported in several studies (Shriver et al. 1993; Di Rienzo et al. 1994; Estoup et al. 1995*b*; Garza, Slatkin, and Freimer 1995).

However, differences among alleles may involve additional types of mutations, such as insertions or deletions of nonrepeat sequences as well as single-base substitutions (Blanquer-Maumont and Crouau-Roy 1995; Estoup et al. 1995b; FitzSimmons, Moritz, and Moore 1995; Garza, Slatkin, and Freimer 1995; Meyer et al. 1995), which cannot be resolved by allele size estimation alone and potentially bias estimates of mutation rate. In such circumstances, sequence analysis may represent a valuable approach to document the actual mutational patterns driving allelic size variance at microsatellite loci.

In this study, we performed a sequence analysis of putative homologous alleles of a microsatellite locus (SFO-12) first identified in the brook charr, *Salvelinus fontinalis* (Angers et al. 1995; Angers and Bernatchez 1996), to test for congruence between size and sequence information. This was done by documenting mutational patterns of 32 alleles from intraspecific to intergeneric levels among twelve species representing four salmonid genera. We propose a model for the evolution of this locus based on mutational events driving allelic diversity, and we also describe the importance of mutational homoplastic events and allele size homoplasy.

Abbreviations: IAM, infinite allele model; SMM, stepwise mutation model; SSM, strand slippage mispairing; ITS-2, second internal transcripted spacer.

Key words: Salmonidae, fish, microsatellite, evolution, sequence, mutation models.

Address for correspondence and reprints: Louis Bernatchez, GIROQ, Département de biologie, Université Laval, Sainte-Foy Québec G1K 7P4, Canada. E-mail: louis.bernatchez@bio.ulaval.ca.

Mol. Biol. Evol. 14(3):230–238. 1997
© 1997 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Material and Methods

Screening of Allelic Diversity

Screening of allelic diversity was performed in one to four populations of the following species: huchen *(Hucho perryi),* brown trout *(Salmo trutta),* Atlantic salmon *(S. salar),* rainbow trout *(Oncorhynchus mykiss),*

Table 1
Name, Localization, Sample Size for Screening of Variation in Allele Sizes (*N*), and Alleles (in bp) Detected Within Each Population

Species	Populations	*N*	Alleles
Hucho perryi	Hokkaido Fish Hatchery, Japan	2	191, 197
Salmo trutta	Abant Lake, Turkey	2	185
	Tohmas River, Turkey	2	185
	Apshak River, Russia	2	185
Salmo salar	Ashuapmuchuane River, Québec, Canada	2	187
	Ste-Marguerite River, Québec, Canada	2	187
Oncorhynchus mykiss	Kispiox River, British Columbia, Canada	2	197
	Babine River, British Columbia, Canada	2	197
Oncorhynchus kisutsh	Unknown	1	195
Salvelinus leucomaenis	Belaya River, Sakhalin (Okhotsk)	4	225, 235
	Sokol'nikovka River, Japan	4	203, 225
	Hokkaido Fish Hatchery, Japan	1	203, 235
Salvelinus malma	Noatak River, Alaska, U.S.A.	4	223
	Auke Creek, Alaska, U.S.A.	3	223, 225
	Kamchatka River, Russia	4	223
Salvelinus albus	Kamchatka River, Russia	3	223
Salvelinus alpinus	Nayauk Lake, Northwest Territories, Canada	4	225
	Français Lake, Québec, Canada	4	225
	Thun Lake, Switzerland	30	225
	Dolly Varden Lake, Alaska, U.S.A.	3	225
Salvelinus confluentus	Upper Arrow Lakes, British Columbia, Canada	4	229
	Metiolus River, Oregon, U.S.A.	4	229
	Pinto Lake, Alberta, Canada	4	229
Salvelinus fontinalis	Grande-Baleine River, Québec, Canada	8	247, 269, 271, 273
	Avalon Lake, Québec, Canada	30	269
	Benoit Lake, Québec, Canada	30	197, 269
	Dauphinais Lake, Québec, Canada	30	267
	Du Fou Lake, Québec, Canada	30	263, 269, 271
	Waber Lake, Québec, Canada	30	247a, 251, 267, 271
	Ross Lake, Maine, U.S.A.	30	251, 269, 273
	Kennebago Lake, Maine, U.S.A.	30	251, 271, 273, 277
	Bluevale Creek, Ontario, Canada	8	247, 251, 265b, 271, 273
	Soper Creek, Wisconsin, U.S.A.	8	197, 275
	Timber Run, Maryland, U.S.A.	10	251, 255, 269, 271, 273
	Mill Creek, Maryland, U.S.A.	4	263, 265, 273, 281
	Shavers Run, West Virginia, U.S.A.	8	245, 263
	Walderman Run, West Virginia, U.S.A.	8	263, 265, 271
	Indian Camp Creek, Tennessee, U.S.A.	8	263
Salvelinus namaycush	Haliburton Lake, Ontario, Canada	4	257, 259
	Marquette stock, Michigan, U.S.A.	4	259
	Dezasdeash Lake, Yukon, Canada	4	283
	Des Chasseurs Lake, Québec, Canada	40	259, 263

coho salmon *(O. kisutsh),* white-spotted charr *(Salvelinus leucomaenis),* lake trout *(S. namaycush),* Arctic charr *(S. alpinus),* Dolly varden (*S. malma),* bull trout (*S. confluentus),* and white charr *(S. albus).* A more extensive survey of intraspecific variation was performed in *S. fontinalis,* for which 15 populations throughout its native range were screened. Origins of populations analyzed and sample sizes for screening of variation in size of microsatellite alleles appear in table 1.

Screening of allelic diversity was performed by autoradiography using one primer end-labeled with [γ ^{32}P] ATP (Angers et al. 1995), or by DNA sequencer (ABI 373A), as detailed in Ziegle et al. (1992), using one primer labeled to the 5′ end with a fluorescent dye (FAM). In both cases, PCR amplifications were performed in a programmable thermocycler (Perkin Elmer 480 or 9600), using 15-μl reaction volume and approximately 100 ng of DNA template as described in Angers et al. (1995). The PCR program consisted of an initial denaturating step of 3 min at 95°C, followed by 30 cycles of the following profile: denaturation at 94°C, 30 s at 60°C, and 1 s at 72°C. Denaturation time was 60 s for the first five cycles and was then reduced to 30 s.

Sequence Analysis

One representative of each size allele from each population and species was sequenced, for a total of 78 alleles (table 1). PCR amplifications of alleles to be sequenced were performed in 60-μl reaction volumes using the PCR profile described above. Double-stranded PCR products were isolated on 5% agarose gels and purified using the Gene Clean kit (Bio 101 Inc.) following the supplier's protocol. DNA sequencing was performed by using either the Sequenase 2.0 kit (USB) for the radioactive detection, or the ABI prism dye terminator cycle sequencing ready reaction kit (Perkin-Elmer) for DNA sequencer.

Across species the microsatellite region was composed of repeat and nonrepeat arrays (figs. 1 and 2). These arrays were divided into seven parts to clarify

```
Hucho:          CCCGTTTCAC AATCAGAGAA GAGATCTCAC AAATTAAGGA
Salmo:                             .. .......... ..........
Oncorhynchus:                      .. .......... ..........
Salvelinus:                        .. .......... ..........

Hucho:          GGCAGTTGCA TACCTAAACT CAC--GTGAA AAATAATGAG
Salmo:          .......... .......... ..G--..... ..........
Oncorhynchus:   .......... .......... ...AA..... ..........
Salvelinus:     .......... .......... ...AA..... ..........

Hucho:          TTTGAGAAAT AA         [microsatellite region]
Salmo:          ......G... ..         [microsatellite region]
Oncorhynchus:   .......... ..         [microsatellite region]
Salvelinus:     .......... ..         [microsatellite region]

Hucho:          ACTGCTGAGA GTGCAGGATA ATAATGAAAA CCTGCCTGTA
Salmo:          .......... .......... .......... ..........
Oncorhynchus:   .........T .C........ .......... ..........
Salvelinus:     .......... .......... .......... ....A.....

Hucho:          AAGTAAG
Salmo:          .......
Oncorhynchus:   .......
Salvelinus:     .......
```

FIG. 1.—Sequences of the flanking regions observed at locus SFO-12 among four salmonid genera. The primer used for sequencing is underlined. Sequences of the microsatellite region are detailed in figure 2. Dots indicate identity with the *Hucho* sequence, whereas dashes represent indels.

sequence variation among alleles and to visually align sequences based on similarity. Because parts 4 and 5 resulted from the division of part 3 by CC interruptions (see below), these interruptions were included within these parts.

Results

Locus Evolution

Sequence of alleles revealed only five base substitutions across species among the 121 nucleotides forming the two flanking regions (fig. 1). Considering the *H. perryi* sequence as the hypothetical ancestral state, two single-base substitutions were detected in *Salmo* and *Oncorhynchus* and one was detected in *Salvelinus*. An indel of two contiguous A nucleotides in *Salmo* and *Hucho perryi* represented a unique shared mutation within the flanking region compared to *Oncorhynchus* and *Salvelinus* species. No mutations were detected at or below genus level. Globally, sequence divergence estimates (*d*) among genera varied between 0.016 and 0.041, whereas the absence of variation among 121 nucleotides places the maximum divergence estimate among species within genera at 0.008.

The microsatellite region was much more variable, with 32 different allelic variants (fig. 2). A basic structure composed of a dinucleotide TG array, a nonrepeat sequence (CAYGYGCC where Y = C or T), and a short dinucleotide repeat, $(TG)_2$, were observed in all species. These sequence elements corresponded to parts 3, 6, and 7 (fig. 2).

Complex variation of this basic structure characterized allelic variation among genera (figs. 2 and 3). *H. perryi* alleles comprised only one pure uninterrupted TG array in part 3. Consequently, SFO-12 may be considered as a perfect microsatellite for that species. Based on the parsimony criterion of minimizing the overall number of mutational steps involved in the evolution of the locus among species, we hypothesized that this perfect locus represents the ancestral state. Different nucleotide substitutions interrupted the repeat array in other species, creating different imperfect microsatellite patterns. Within the genus *Salmo,* the basic repeat array (part 3) was interrupted by a C nucleotide. A cytosine doublet interrupted the basic TG array in *Oncorhynchus* and *Salvelinus* genera, dividing the repeat array into two parts, which we distinguished as parts 3 and 5 (fig. 2). Finally, a large insertion distinguished as part 1 and part 2 was detected in the genus *Salvelinus,* which produced an abrupt shift in allelic size for this genus. This insertion showed a high homology with *Oncorhynchus* alleles composed of parts 3, 5, and 6, both in organization and sequence, which suggests that it arose from duplication of these regions (fig. 3).

Mutational Diversity

The overall allelic diversity resulted from various combinations of three distinct types of mutational events; change in number of dinucleotide repeats, base substitution, and indels of repeat and nonrepeat segments. Most sequence differences observed among alleles involved multiple mutational types, as observed in nearly 100% (429/430) of allelic pairwise comparisons among species.

Within the genus *Salvelinus,* the number of dinucleotide repeats was highly variable across the different repeat arrays. However, large indels also represented an important source of allele size variation among species (figs. 2 and 3). Part 1 was missing in *S. leucomaenis,* as was part 2 in all *S. alpinus, S. malma, S. albus,* and *S. confluentus* allelic variants. Presence of an additional segment (part 4) composed of a guanine doublet and TG array was observed in *S. alpinus, S. malma, S. albus,* and *S. confluentus* alleles, while a cytosine doublet and a TG array were observed in *S. namaycush* alleles and in all but four alleles in *S. fontinalis* (fig. 2). The guanine doublet likely resulted from base substitutions, (T to G) while the cytosine doublet likely resulted from incorporation of CC in the indel mechanism of repeat motif as observed by Estoup et al. (1995*b*). Single-base substitutions not affecting allelic size were also observed among *Salvelinus* species. These occurred in repeat arrays as well as in nonrepeat sequences of the microsatellite region (fig. 2).

Variation in the number of dinucleotide repeats was a more important source of allelic diversity at the intraspecific level, implied in 55 out of the 131 (39%) com-

Allele	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7
Hucho perryi							
191	------	------	$(TG)_{11}$	------	------	CACGCGCC	$(TG)_2$
197	------	------	$(TG)_{14}$	------	------	CACGCGCC	$(TG)_2$
Salmo. trutta							
185	------	------	TGCG$(TG)_2$CG$(TG)_2$	------	------	CATGCGCC	$(TG)_2$
Salmo salar							
187	------	------	TGCG$(TG)_6$	------	------	CATGCGCC	$(TG)_2$
Oncorhynchus mykiss							
197	------	------	$(TG)_7$	------	CC $(TG)_5$	CACGCGCC	$(TG)_2$
Oncorhynchus kisutsh							
195	------	------	$(TG)_6$	------	CC $(TG)_5$	CACGCGCC	$(TG)_2$
Salvelinus leucomaenis							
203	------	------	$(TG)_8$	------	CC $(TG)_7$	CACGCGCC	$(TG)_2$
225	------	$(TG)_7$ TACGCGCC	$(TG)_8$	------	CC $(TG)_7$	CACGCGCC	$(TG)_2$
235	------	$(TG)_7$ TACGCGCC	$(TG)_{13}$	------	CC $(TG)_7$	CACGCGCC	$(TG)_2$
Salvelinus malma							
223	$(TG)_5$CG CC	------	$(TG)_5$	GG $(TG)_6$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
225	$(TG)_5$CG CC	------	$(TG)_4$	GG $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
Salvelinus albus							
223	$(TG)_5$CG CC	------	$(TG)_5$	GG $(TG)_6$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
Salvelinus alpinus							
225	$(TG)_5$CG CC	------	$(TG)_4$	GG $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
Salvelinus confluentus							
229	$(TG)_5$CG CC	------	$(TG)_5$	GG $(TG)_9$	GA CG$(TG)_5$	CACGTGCC	$(TG)_2$
Salvelinus fontinalis							
197	$(TG)_5$GG CC	------	------	------	$(TG)_6$	CACGCGCC	$(TG)_2$
245	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{13}$	------	CC $(TG)_6$	CACGCGCC	$(TG)_2$
247	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{14}$	------	CC $(TG)_6$	CACGCGCC	$(TG)_2$
247a	$(TG)_5$GG CC	------	$(TG)_{14}$	CC $(TG)_9$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
251	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{16}$	------	CC $(TG)_6$	CACGCGCC	$(TG)_2$
255	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{18}$	------	CC $(TG)_6$	CACGCGCC	$(TG)_2$
263	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{13}$	CC $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
265	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{14}$	CC $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
265b	$(TG)_5$GG CC	TGCG $(TG)_4$ CACGCGCC	$(TG)_{14}$	CC $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
267	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{14}$	CC $(TG)_9$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
269	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{16}$	CC $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
271	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{16}$	CC $(TG)_9$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
273	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{18}$	CC $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
275	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{18}$	CC $(TG)_9$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
277	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{20}$	CC $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
281	$(TG)_5$GG CC	$(TG)_6$ CACGCGCC	$(TG)_{22}$	CC $(TG)_8$	CC $(TG)_6$	CACGCGCC	$(TG)_2$
Salvelinus namaycush							
257	$(TG)_5$CG CC	$(TG)_7$ CACGCGCC	$(TG)_{11}$	CC $(TG)_5$	CC $(TG)_7$	CACGCGCC	$(TG)_2$
259	$(TG)_5$CG CC	$(TG)_7$ CACGCGCC	$(TG)_{12}$	CC $(TG)_5$	CC $(TG)_7$	CACGCGCC	$(TG)_2$
263	$(TG)_5$CG CC	$(TG)_7$ CACGCGCC	$(TG)_{14}$	CC $(TG)_5$	CC $(TG)_7$	CACGCGCC	$(TG)_2$
283	$(TG)_5$CG CC	$(TG)_7$ CACGCGCC	$(TG)_9$	CC $(TG)_{19}$	CC $(TG)_8$	CACGCGCC	$(TG)_2$

FIG. 2.—Sequence alignments of 32 microsatellite allelic variants resolved at locus SFO-12 among 12 species representing four salmonid genera. The sequences are divided into seven parts as described in the *Results* section. Gaps indicate the absence of the different parts in given alleles.

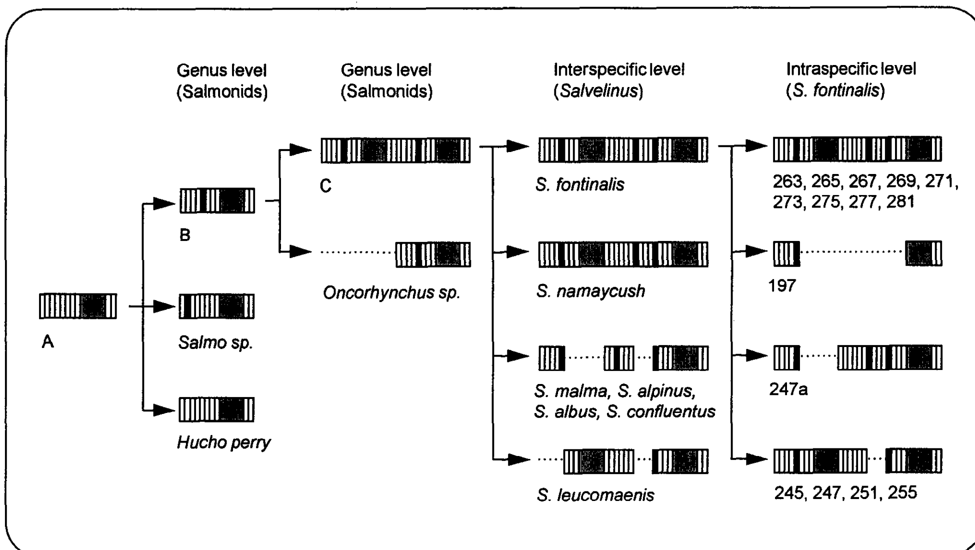

FIG. 3.—Schematic hypothesis of mutational patterns observed within locus SFO-12 microsatellite region. From a hypothetically ancestral perfect microsatellite (*A*), conserved in *Hucho,* single-base substitutions produced two different interrupted arrays; one in *Salmo,* and the other in the hypothetical ancestral state of *Oncorhynchus* and *Salvelinus* alleles (*B*), which was retained in *Oncorhynchus.* A large insertion resulting from duplication of the microsatellite region of the allele B produced the hypothetical ancestral pattern of *Salvelinus* alleles (*C*). Within *Salvelinus spp.* different deletions occurred within *S. leucomaenis* and the group composed of *S. malma, S. alpinus, S. albus* and *S. confluentus.* An additional (TG) array occurred in *S. namaycush* and *S. fontinalis,* which otherwise retained the ancestral pattern. Within *S. fontinalis,* allele 197 resulted from the deletion of several parts while deletion of various segments generated two groups of alleles with overlaping sizes (247a and 245–255). An additional size class (263–281) retained the ancestral pattern. White boxes: dinucleotide repeat units, gray box: nonrepeat sequences; black boxes: single-nucleotide substitutions (detailed in fig. 2); gaps: deletion of entire parts (detailed in text).

parisons. Nevertheless, only 22% of all intraspecific pairwise allelic comparisons solely involved changes of number of repeats within a single array. Changes in the number of repeats within more than one part occurred in 17% of all pairwise comparisons. For instance, differences between allele 283 and other alleles in *S. namaycush* involved differences in number of repeats within three different arrays (parts 3, 4, and 5). Finally, 18% of all intraspecific allelic pairwise comparisons involved large deletions. For instance, allele 197 in *S. fontinalis* was 48 bp shorter than the second shortest allele, and resulted from deletion of parts 2, 3, and 4, by either single or multiple mutational events from any specific alleles.

Table 2
Potential Relationships Among Alleles Detected in *S. fontinalis* According to the Number of Repeats of Part 3 (Vertical) and Part 4 (Horizontal) Observed at Locus SFO-12

	DEL	$(TG)_8$	$(TG)_9$
$(TG)_{13}$	245	263	—
$(TG)_{14}$	247	265	267
$(TG)_{16}$	251	269	271
$(TG)_{18}$	255	273	275
$(TG)_{20}$	—	277	—
$(TG)_{22}$	—	281	—

NOTE.—DEL refers to deletion of part 4 and dashes indicate that such alleles were not observed in this survey. Differences among alleles within a given column or line involve change in only one array, and otherwise involve the two different arrays. Alleles 247a and 265b are derived from alleles 267 and 265, respectively, and did not involve mutations in the parts represented here, while absence of both parts 3 and 4 makes the relationships to allele 197 ambiguous.

Mutational Homoplasic Events

In the absence of a consensus phylogeny within *Salvelinus* genus (Phillips, Sajdak, and Domanico 1995 and references therein), it is not possible to strictly quantify homoplasic events that occurred at locus SFO-12 during the family radiation. Nevertheless, sequence data clearly provided evidence for such events across species comparisons (fig. 2). For example, the absence of part 2 in all alleles of the *S. alpinus, S. malma, S. albus,* and *S. confluentus* species group was also observed within species (allele 203 in *S. leucomaenis,* allele 247a in *S. fontinalis*) for which this part was present for most alleles. This implies that homologous parts across species were independently subject to the same mutational events.

Within *S. fontinalis,* variation in the number of repeats in either part 3 or 4 generated different allelic groups as reflected in table 2, which represents a parsimonious network of relationships among allelic variants (excluding alleles 197, 247a and 265b) within *S. fontinalis.* From the 12 mutational steps inferred to explain allelic diversity under a parsimony criterion, five steps involved the parallel changes of same character states, suggesting that independent mutational events occurred frequently at SFO-12 for that species. For instance, relationships within columns involve two changes from $(TG)_{13}$ to $(TG)_{14}$, three from $(TG)_{14}$ to $(TG)_{16}$, and three from $(TG)_{16}$ to $(TG)_{18}$ and resulted in identical numbers of repeats within part 3 among different allele lineages.

Size Homoplasy

Among representatives of the 26 alleles resolved by size information, six additional alleles were revealed

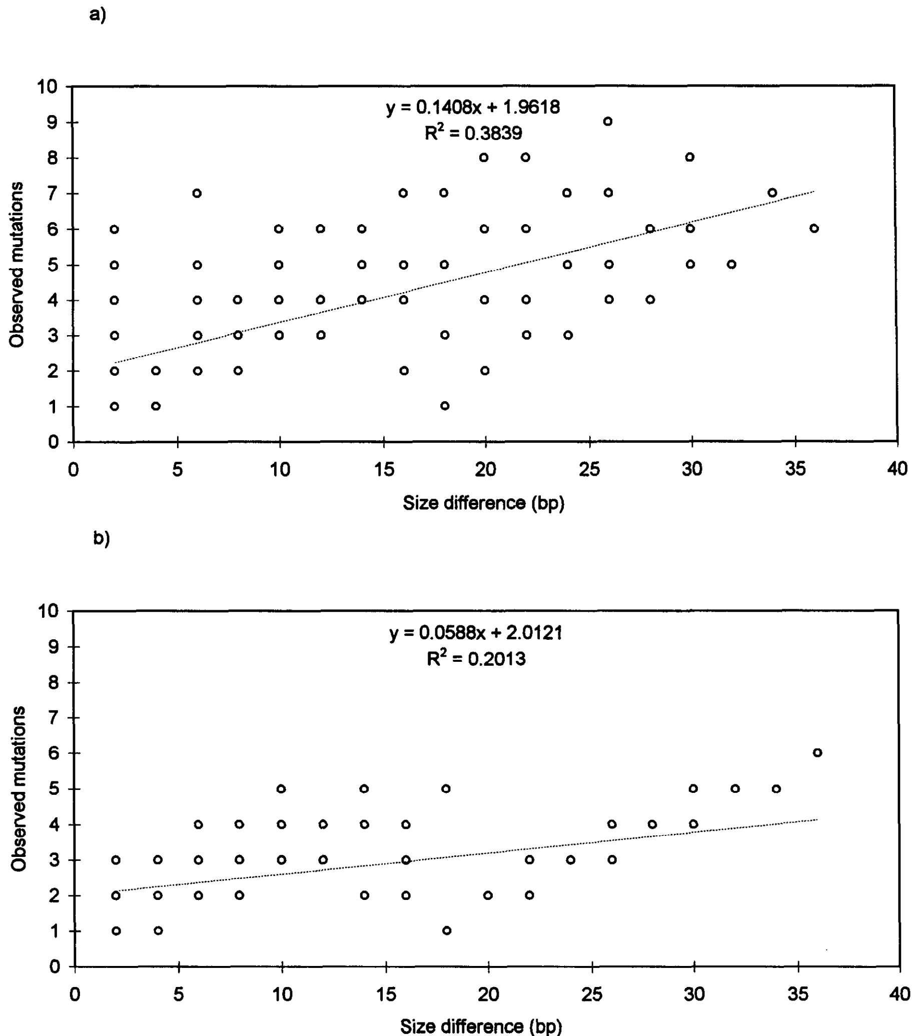

FIG. 4.—Relationships between observed number of mutational events as detected by sequence analysis (refering to table 2) and difference in size (bp) among 78 pairwise allelic comparisons (excluding alleles 197, 247a, and 267b) within *S. fontinalis*. *a*, by column; *b*, by line. In both cases, allele 263 was considered ancestral.

by the sequence analysis. Size homoplasy occurred at all taxonomic levels. Among genera, the allele 197 observed in *Oncorhynchus, Hucho,* and *Salvelinus* was clearly not identical by descent, differing substantially both in the flanking and microsatellite regions. Among *Salvelinus* species, allele 225 observed in *S. leucomaenis* differed by at least five mutations from the same electromorph detected in *S. malma* and *S. alpinus,* involving deletion of different sequences (parts 1 or 2 respectively), changes in number of repeats in two arrays (parts 3 and 5), and single-base substitutions (part 3). Allele 263 observed in *S. namaycush* and *S. fontinalis* differed by a minimum of five mutational steps, involving change in number of repeats in four different arrays (parts 2, 3, 4, and 5) and one nucleotide substitution (part 1). At the intraspecific level, allele 265b in *S. fontinalis,* most likely originating from allele 265 (fig. 2), differed by a single-base substitution in one TG array (part 2). The implication of size homoplasy may be more pronounced, as exemplified by alleles 247a and 247 in *S. fontinalis,* which differed by the deletion of two different sequences (parts 2 and 4 respectively).

Incongruence Between Size Information and Observed Mutational Events

The strong departure from a basic stepwise mutational pattern led to major discrepancies between size difference and the number of mutation events detected among alleles by sequence analysis. Globally, the number of mutational events among alleles (excluding alleles 197, 247a, and 265b) assumed from size information was poorly correlated with that quantified from parsimonious network resolved from sequence analysis (relationships by column or line in table 2), explaining only 38% of mutational events (R^2 = 0.38, $P < 0.0001$) by column and even less by line (R^2 = 0.20, $P < 0.0001$) (fig. 4). For instance, a set of four pairwise comparisons, suggesting numbers of mutational steps varying from two to 18 bp, all involved a single mutation as resolved by sequencing. Adding the alleles 197, 247a,

and 265b resulted in an almost complete absence of relationships between sequence and allelic size information ($R^2 = 0.05$, $P < 0.0171$).

Discussion

The conservation of a basic structure revealed by sequence analysis of SFO-12 among all species analyzed confirmed the homology of this locus within the salmonid family. Sequencing also revealed contrasting mutation rates between the flanking and microsatellite regions. In comparisons with sequence divergence estimates observed among salmonids from mtDNA (Grewe, Billington, and Hebert 1990) or nuclear ribosomal DNA (Phillips, Pleyte, and Brown 1992), mutation rate was approximately one order of magnitude lower within the flanking regions of SFO-12. Relative conservation of microsatellite flanking regions have previously been reported in other animal groups (Moore et al. 1991; Schlötterer, Amos, and Tautz 1991; Stallings et al. 1991; Blanquer-Maumont and Crouau-Roy 1995; FitzSimmons, Moritz, and Moore 1995; Meyer et al. 1995; Pépin et al. 1995; Rico, Rico, and Hewitt 1996). At present, explanations for the sequence conservation in these DNA segments are only hypothetical, and mostly involve selective effects (Rico, Rico, and Hewitt 1996 and references therein).

In contrast, high polymorphism was observed within the microsatellite region. This involved multiple mutational processes within both repeat arrays and nonrepeat sequences which clearly did not conform to SMM expectations. Several homologous loci sequenced across species in previous studies revealed mutational events in addition to changes of number of repeats (Moore et al. 1991; Schlötterer, Amos, and Tautz 1991; Stallings et al. 1991; Blanquer-Maumont and Crouau-Roy 1995; FitzSimmons, Moritz, and Moore 1995; Garza, Slatkin, and Freimer 1995; Meyer et al. 1995; Pépin et al. 1995). Within species, complex mutational events have also been documented among alleles detected in distantly related subspecies, while mutational patterns close to SMM expectations were observed among closely related subspecies or within populations (Estoup et al. 1995*a*, 1995*b*).

Globally, our results were congruent with the general observation that the resolving power of microsatellite markers under SMM assumptions decreases with evolutionary divergence, as illustrated by the higher proportion of stepwise mutations observed at the intraspecific compared to the interspecific level. However, the present study also revealed that complex mutational processes could be involved in closely related populations and even among alleles found within the same population. The most dramatic changes were observed in the genus *Salvelinus* following duplication of the basic microsatellite structure that characterized the other genera. While allelic divergence involved changes in simple number of dinucleotide repeats, these mutations did not reflect most of the differences as expected from size information. Mutational events involving large indels, as well as base substitutions, occurred among and within species. These resulted in numerous homoplastic events which independently generated alleles of similar sizes among distantly related groups or, conversely, alleles of disjunct sizes within a given group. Within *S. fontinalis,* this translated to a poor correlation between difference in size and number of mutations estimated from sequence analysis. The major consequence of these complex mutational patterns is that allelic coalescence cannot easily be inferred from size information only. This could potentially generate strong biases in quantifying relationships and the extent of divergence among species and populations on the basis of allelic size variance without knowledge of mutational events.

It may conceivably be argued that complex mutational processes described in the present study represent an exception rather than the rule for microsatellite loci and, at worst, only concern imperfect microsatellite loci. Alternatively, these observations could be of importance for several reasons. First, imperfect microsatellites are relatively abundant within animal genomes (Weber 1990; Ostrander, Sprague, and Rine 1992; Wintero, Fredholm, and Thomsen 1992; Estoup et al. 1993*b*), and particularly in fishes (Brooker et al. 1994; Estoup et al. 1993*a*; Slettan, Olsaker, and Lie 1993), and are routinely included in microsatellite studies. Second, our results illustrate that the dichotomy between imperfect and perfect microsatellite loci is not straightforward. They indicate that the state of a homologous microsatellite locus may evolve from a perfect to different imperfect stages, which cannot be detected by allelic size information only. Finally, mutational events differing from changes in stepwise number of repeats was previously inferred in perfect microsatellites in order to explain allelic distribution within populations (Shriver et al. 1993; Di Renzo et al. 1994). One reason why the occurence of non-SMM mutational patterns has often been ignored is that such departure cannot be detected by sequence analysis within perfect microsatellites. Instead, fit to SMM expectations must be inferred from predictions of allelic size distribution within a population, which rely on considerations, such as sufficient number of alleles detected within population, infinite population size, and random mating, that are rarely met.

Base substitutions are the type of mutations most likely to drive the evolution of imperfect microsatellites from a perfect one. Such mutations generally reduce the number of contiguous repeats and the propensity of slippage (Chung et al. 1993; Blanquer-Maumont and Crouau-Roy 1995; Pépin et al. 1995). This may reduce variability, since a minimal number of repeats is apparently necessary to generate microsatellite polymorphism (Weber 1990). In this study, base substitutions resulted in a reduction of mutation rate in *Salmo* and *Oncorhynchus* species, as well as within the *S. alpinus, S. malma, S. confluentus,* and *S. albus* species group. Large deletions that include repeat arrays may also reduce polymorphism, as in the case of allele 197 of *S. fontinalis,* for which no alleles of similar sizes were observed. This implies that levels of polymorphism may show important variability and do not appear to be strictly associated to the degree of evolution as reported in earlier

studies (Meyer et al. 1995; Rubinsztein et al 1995). In SFO-12, the complete deletion or duplication of microsatellite segments may have resulted from large strand slippage misparing (SSM) favored by the homology between repeat arrays or nonrepeat sequences of different parts of the microsatellite region (Levinson and Gutmann 1987). If so, this has caused single-step large mutations to occur frequently and in an unpredictable manner.

Despite the complex mutational patterns of SFO-12, which involved numerous homoplastic events, several nonhomoplastic characters were also identified which may be used as additional information to infer relationships among salmonid genera and species. Thus, the sharing of an adenine doublet in the flanking region and of a cytosine doublet that generated the presence of two repeat arrays in *Oncorhynchus* and *Salvelinus* as opposed to a single one in *Hucho* and *Salmo* are suggestive of their closer phylogenetic affinity. This is in contrast with phylogenetic hypotheses suggested by mtDNA (Grewe, Bilington, and Hebert 1990), nuclear rDNA (Phillips, Pleyte, and Brown 1992) or morphological (Behnke 1992) data, stating that *Salmo* and *Oncorhynchus* are more recently derived. While apparently controversial, it is nevertheless noteworthy that the *Oncorhynchus-Salvelinus* and *Hucho-Salmo* pairs are respectively more diversified in the Nearctic and Palearctic regions, and consequently may be presumed to have radiated on the different continents.

Additional potentially informative characters were observed within the genus *Salvelinus*. The occurence of a unique guanine doublet within part 3, the complete absence of parts 2 and 4 within the *S. malma, S. alpinus, S. albus,* and *S. confluentus* species group compared to their presence within both *S. fontinalis* and *S. namaycush,* and the occurence of numerous mutations unique to *S. leucomaenis* are indicative of three distinct evolutionary lineages within the genus. These relationships are congruent with the phylogeny suggested by sequence analysis of the second internal transcripted spacer (ITS-2) of nuclear rDNA (Phillips, Manley, and Daniels 1994). This example, along with their common occurence within the genome, suggests that sequences of imperfect microsatellites may represent an additional class of markers to be considered in phylogenetic inferences.

Acknowledgments

We are indebted to R. G. Danzmann, R. B. Phillips, A. Osinov, S. Pollard, J. Schoffman, P. C. Brunner, P. A. Crane, M. Plante, and N. Tessier for kindly providing DNA samples. The constructive comments of C. Moritz and two anonymous reviewers substantially helped to improve this article. This study was supported by the Canadian Parks Service and by NSERC (Canada) and FCAR (Québec) research grants to L.B.

LITERATURE CITED

ANGERS, B., and L. BERNATCHEZ. 1996. Usefulness of heterologous microsatellites obtained from brook charr, *Salvelinus fontinalis* Mitchill, in other *Salvelinus* species. Mol. Ecol. **5**:317–319.

ANGERS, B., L. BERNATCHEZ, A. ANGERS, and L. DESGROSEILLERS. 1995. Specific microsatellite loci for brook charr (*Salvelinus fontinalis* Mitchill) reveal strong population subdivision on a microgeographic scale. J. Fish Biol. **47**(Suppl. A):177–185.

BEHNKE, R. J. 1992. Native trout of western North America. American Fisheries Society Monograph **6**, Bethesda, Md.

BLANQUER-MAUMONT, A., and B. CROUAU-ROY. 1995. Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species. J. Mol. Evol. **41**:492–497.

BOWCOCK, A. M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, J. R. KIDD, and L. L. CAVALLI-SFORZA. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368**:455–457.

BROOKER, A. L., D. COOK, P. BENTZEN, J. M. WRIGHT, and R. W. DOYLE. 1994. Organization of microsatellites differs between mammals and cold-water teleost fishes. Can. J. Fish. Aquat. Sci. **51**:1959–1966.

CHUNG, M. Y., M. P. W. RANUM, L. A. DUVICK, A. SERVADIO, H. Y. ZOGHBI, and H. T. ORR. 1993. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat. Genet. **5**:254–258.

DI RIENZO, A., A. C. PETERSON, J. C. GARZA, A. M. VALDES, M. SLATKIN, and N. B. FREIMER. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA **91**:3166–3170.

ESTOUP, A., L. GARNERY, M. SOLIGNAC, and J.-M. CORNUET. 1995*a*. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hieracrchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics **140**:679–695.

ESTOUP, A., P. PRESA, F. KRIEG, D. VAIMAN, and R. GUYOMARD. 1993*a*. (CT)n and (GT)n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity **71**:488–496.

ESTOUP, A., M. SOLIGNAC, M. HARRY, and J.-M. CORNUET. 1993*b*. Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. Nucleic Acids Res. **21**:1427–1431.

ESTOUP, A., C. TAILLIEZ, J.-M. CORNUET, and M. SOLIGNAC. 1995*b*. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol. Biol. Evol. **12**:1074–1084.

FITZSIMMONS, N. N., C. MORITZ, and S. S. MOORE. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol. Biol. Evol. **12**: 432–440.

FORBES, S. H., J. T. HOGG, F. C. BUCHANAN, A. M. CRAWFORD, and F. W. ALLENDORF. 1995. Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. Mol. Biol. Evol. **12**:1106–1113.

GARZA, J. C., M. SLATKIN, and N. B. FREIMER. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol. Biol. Evol. **12**:594–603.

GOLDSTEIN, D. B., A. R. LINARES, L. L. CAVALLI-SFORZA, and M. W. FELDMAN. 1995*a*. An evaluation of genetic distances for use with microsatellite loci. Genetics **139**:463–471.

———. 1995*b*. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA **92**:6723–6727.

GREWE, P. M., N. BILLINGTON, and P. D. N. HEBERT. 1990. Phylogenetic relationships among members of *Salvelinus*

inferred from mitochondrial divergence. Can. J. Fish. Aquat. Sci. **47**:984–991.

KIMURA, M., and J. F. CROW. 1964. The number of alleles that can be maintained in a finite population. Genetics **49**:725–738.

LEVINSON, G., and G. A. GUTMANN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

MEYER, E., P. WIEGAND, S. P. RAND, D. KUHLMANN, M. BRACK, and B. BRINKMANN. 1995. Microsatellite polymorphisms reveal phylogenetic relationships in primates. J. Mol. Evol. **41**:10–14.

MOORE, S. S., L. L. SARGEANT, T. J. KING, J. S. MATTICK, M. GEORGES, and D. J. S. HETZEL. 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics **10**:654–660.

NEI, M., and N. TAKEZAKI. 1996. The root of the phylogenetic tree of human populations. Mol. Biol. Evol. **13**:170–177.

OHTA, T., and M. KIMURA. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22**:201–204.

OSTRANDER, E. A., J. F. SPRAGUE JR., and J. RINE. 1993. Identification and characterization of dinucleotide repeat (CA)n markers for genetic mapping in dog. Genomics **16**:207–213.

PÉPIN, L., Y. AMIGUES, A. LÉPINGLE, J.-L. BERTHIER, A. BENSAID, and D. VAIMAN. 1995. Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity **74**:53–61.

PHILLIPS, R. B., S. A. MANLEY, and T. J. DANIELS. 1994. Systematics of the salmonid genus *Salvelinus* inferred from ribosomal DNA sequences. Can. J. Fish. Aquat. Sci. **51**:198–204.

PHILLIPS, R. B., K. A. PLEYTE, and M. R. BROWN. 1992. Salmonid phylogeny inferred from ribosomal DNA restricion maps. Can. J. Fish. Aquat. Sci. **49**:2345–2353.

PHILLIPS, R. B., S. L. SAJDAK, and M. J. DOMANICO. 1995. Relationships among charrs based on DNA sequences. Nordic J. Freshw. Res. **71**:378–391.

RICO, C., I. RICO, and G. HEWITT. 1996. 470 million years of conservation of microsatellite loci among fish species. Proc. R. Soc. Lond. B **263**:549–557.

RUBINSZTEIN, D. C., W. AMOS, J. LEGGO, S. GOODBURN, S. JAIN, S.-H. LI, R. L. MARGOLIS, C. A. ROSS, and M. A. FERGUSON-SMITH. 1995. Microsatellite evolution–evidence for directionality and variation in rate between species. Nat. Genet. **10**:337–343.

SCHLÖTTERER, C., B. AMOS, and D. TAUTZ. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature **354**:63–65.

SCHLÖTTERER, C., and D. TAUTZ. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20**:211–215.

SHRIVER, M. D., L. JIN, E. BOERWINKLE, R. DEKA, R. E. FERRELL, and R. CHAKRABORTY. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol. Biol. Evol. **12**:914–920.

SHRIVER, M. D., L. JIN, R. CHAKRABORTY, and E. BOERWINKLE. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulations approach. Genetics **134**:983–993.

SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**:457–462.

SLETTAN, A., I. OLSAKER, and Ø. LIE. 1993. Isolation and characterization of variable (GT)n repetitive sequences from Atlantic salmon, *Salmo salar* L. Animal Genet. **24**:195–197.

STALLINGS, R. L., A. F. FORD, D. NELSON, D. C. TORNEY, C. E. HILDEBRAND, and R. K. MOYZIS. 1991. Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics **10**:807–815.

WEBER, J. L. 1990. Informativeness of human (dC-dA)n.(dG-dT)n polymorphism. Genomics **7**:524–530.

WEBER, J. L., and C. WONG. 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. **2**:1123–1128.

WINTERO, A. K., M. FREDHOLM, and P. D. THOMSEN. 1992. Variable sequences in the porcine genome. Genomics **12**:281–288.

ZIEGLE, J. S., Y. SU, K. P. CORCORAN, L. NIE, P. E. MAYRAND, L. B. HOFF, L. J. MCBRIDE, M. N. KRONICK, and S. R. DIEHL. 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics **14**:1026–1031.

CRAIG MORITZ, reviewing editor

Accepted November 19, 1996