DEVELOPMENT OF MOLECULAR GENETIC MARKERS IN ATLANTIC SALMON *(SALMO SALAR)* AND AN ILLUSTRATION OF THEIR APPLICATION TO AQUACULTURE AND FISHERIES

by

Patrick T. O'Reilly

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

September 19, 1997

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Development of Molecular Genetic Markers in Atlantic Salmon (*Salmo salar*) and an Illustration of their Potential Application to Fisheries and Aquaculture"

by Patrick O'Reilly

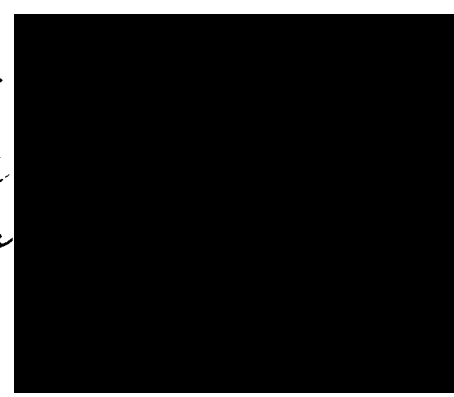
in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: September 15, 1997

External Examiner

Research Supervisor

Examining Committee



DALHOUSIE UNIVERSITY

DATE: September 19, 1997

Author: Patrick T. O'Reilly

TITLE: Development of molecular genetic markers in Atlantic salmon *(Salmo salar)*, and an illustration of their application to aquaculture and fisheries

DEGREE: Ph. D. Convocation: May Year: 1997

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above titled upon the request of individulas or institutions.

[redacted]

Signature of Author

For my mom:

TABLE OF CONTENTS

LIST OF FIGURES viii
LIST OF TABLES x
ABSTRACT xii
ABBREVIATIONS xiii
PUBLICATIONS xiv
ACKNOWLEDGEMENTS xv

CHAPTER 1: GENERAL INTRODUCTION 1
Multilocus fingerprinting 3
Single locus fingerprinting 9
Minisatellite Variant Repeat mapping 21
Microsatellites 22
Properties of microsatellites 23
Cloning of microsatellites 29
Application of microsatellites to aquaculture and fisheries 32
Disadvantages and limitations of microsatellites as molecular markers 34
Future developments 38

CHAPTER 2: ISOLATION AND CHARACTERIZATION OF VNTR LOCI, AND DEVELOPMENT OF PCR BASED ASSAYS OF VARIABILITY 44
INTRODUCTION 44
METHODS 45
Initial characterization of minisatellite loci 45
RESULTS AND DISCUSSION 49
Characteristics of original EMBL3 minisatellite loci 49
Minisatellite Variant Repeat mapping in Atlantic salmon 50

CHAPTER 3: ISOLATION AND CHARACTERIZATION OF MICROSATELLITES 61
INTRODUCTION 61
METHODS 63
Library construction and microsatellite screening procedures 63
Analysis of microsatellite flanking sequence 65
RESULTS 67
Abundance of GT(n) microsatellites 67
Microsatellite array type and length 67
Locus characteristics and suitability for PCR based assays of variation 72
Isolation of tetranucleotide microsatellites 82
DISCUSSION 82

Abundance of GT(n) microsatellites..... 82
Locus characteristics and suitability for PCR based assays of variation..... 85
Abundance and characteristics of tetranucleotide microsatellites..... 87

CHAPTER 4: RAPID ANALYSIS OF GENETIC VARIATION IN ATLANTIC SALMON *(SALMO SALAR)* BY PCR MULTIPLEXING OF DINUCLEOTIDE AND TETRANUCLEOTIDE MICROSATELLITES..... 90
INTRODUCTION..... 90
METHODS..... 95
Multiplexing loci..... 95
Allelic ladder development..... 95
PCR analysis..... 97
Patterns of Inheritance of microsatellite loci..... 97
Population analyses..... 98
RESULTS AND DISCUSSION..... 98
Characteristics of individual microsatellites..... 98
Microsatellite multiplex..... 102
Allelic Ladders and typing loci..... 105
Population differentiation..... 107

CHAPTER 5 AN ANALYSIS MICROSATELLITE-BASED PARENTAGE DETERMINATION IN ATLANTIC SALMON 110
INTRODUCTION..... 110
MATERIALS AND METHODS..... 118
Fish rearing and DNA typing..... 118
Measures of variability and information content..... 120
Parentage determination using compatibility analysis..... 120
Analysis of parent-offspring mismatches..... 123
Probability of detecting scoring errors and mutations..... 124
Classification of scoring errors..... 126
RESULTS AND DISCUSSION..... 127
Information content and power of exclusion..... 127
Determining parentage using compatibility analysis..... 130
Types and rates of mutations..... 133
Types and rates of scoring errors observed..... 137
Implications of mutations and scoring errors on determinations of parentage..... 142
Experimental design recommendations for monitoring and managing the impacts of errors on parentage determination analyses..... 145
Early environmental influence on growth rates of Atlantic salmon..... 147

Summary..... 148
CONCLUDING REMARKS..... 150
REFERENCES..... 152

LIST OF FIGURES

Figure 1.1.	Schematic demonstrating the molecular basis of multilocus fingerprinting	4
Figure 1.2	Multilocus DNA fingerprint of tilapia	6
Figure 1.3	Single locus DNA fingerprint of tilapia	10
Figure 1.4	Schematic demonstrating PCR amplification	12
Figure 1.5	Schematic demonstrating the basis of allele binning	19
Figure 1.6	Co-amplification of di- and tetranucleotide microsatellites from Atlantic salmon	35
Figure 2.1	Partial characterization of the Atlantic salmon minisatellite bearing *Ssa* 1 clone	46
Figure 2.2	Initial PCR amplification of the microsatellite locus *Ssa* 197	51
Figure 2.3	DNA sequence of the entire Ssa 197 clone, relative positioning of micro- and minisatellite arrays, and location of primers used in assaying variation at both types of repeated DNA	53
Figure 2.4	Initial MVR-PCR trials at the Atlantic salmon locus *Ssa* 197	56
Figure 2.5	Schematic demonstrating MVR-PCR at the Atlantic salmon locus Ssa 197 using tagged primers	58
Figure 3.1	DNA sequencing autoradiograph of 18 microsatellite 'positive' clones isolated from Atlantic salmon	68
Figure 3.2	Frequency histogram of array size classes of microsatellites isolated from Atlantic salmon	70
Figure 3.3	Microsatellite development pyramid	74
Figure 3.4	Alignment of four *Ssa* clones isolated here with the Atlantic salmon *Ssa*1 TC1-like transposone clone	76
Figure 3.5	Allele signatures of representative dinucleotide microsatellite loci developed from Atlantic salmon	78

Figure 4.1 Multiplex PCR of one dinucleotide (*Ssa* 85) and three tetranucleotide (*Ssa* 171, *Ssa* 197 and *Ssa* 202) isolated from Atlantic salmon..... 100

LIST OF TABLES

Table 1.1.	Multilocus minisatellite probes used to generate fingerprint profiles in fish	8
Table 1.2	Polymorphism at single-locus minisatellites detected by DNA probes from fish	15
Table 1.3	Levels of polymorphism observed at microsatellite loci from several fish species	25
Table 3.1	Sequence similarity of Atlantic salmon microsatellite clones with repetitive DNA submitted to Genebank	73
Table 3.2	Primer sequences and PCR amplification conditions of all dinucleotide microsatellite loci developed here	80
Table 3.3	Locus characteristics of all dinucleotide microsatellite loci developed here	81
Table 3.4	Genebank submission numbers of microsatellite loci potentially suitable for PCR based assays of genetic variation	83
Table 4.1	Nucleotide sequence of Multiplex microsatellites and primers (a and b) used for PCR amplification	96
Table 4.2	Numbers of alleles and heterozygosity at microsatellite loci from three Atlantic salmon populations in Nova Scotia	99
Table 4.3	Allele sizes for dinucleotide and tetranucleotide microsatellites in fish and mammals	106
Table 4.4	Pairwise test for heterogeneity between all pairs of three Nova Scotia Atlantic salmon populations for all four multiplex microsatellite loci	108
Table 5.1	Microsatellite locus characteristics and measures of information content	129
Table 5.2	Percentage of offspring resolved to one, two, three or greater than three families	131
Table 5.3	Primer sequence and PCR conditions for additional loci used to confirm status of putative mutant alleles	135

Table 5.4 Results of microsatellite scoring error and mutation detection analysis..... 139

Table 5.5 Types and rates of microsatellite allele scoring errors observed in parentage analysis..... 141

ABSTRACT

DNA fingerpinting technology has changed considerably over the last decade. With the advent of PCR-based methods, new avenues of inquiry are becoming accessible. However, our understanding of the processes responsible for creating variation at mini- and microsatellite loci is poorly understood. Also, little effort has been made to evaluate the accuracy of these technologies, or the implications of mutations or scoring errors on the interpretation of results. Here, I have characterized several minisatellite loci from Atlantic salmon *(Salmo salar),* and have developed MVR-PCR technology at one. I have also isolated and sequenced over 180 microsatellites from this species, and have developed a primarily tetranucleotide microsatellite-based multiplex system for efficient and accurate analysis of genetic variation. Due to the high variability of these loci (>87 % heterozygosity at all three tetranucleotide loci), fewer than two in 30,000 individuals are expected to exhibit identical composite genotypes. To evaluate the utility of this system for assessment of population differentiation, genetic variation was surveyed in 3 rivers from Nova Scotia, Canada. Significant differences in allele frequencies were observed between all Atlantic salmon populations surveyed. In the second segment of this research, approximately 800 communally reared offspring, and their 12 possible sets of parents, were typed at four multiplex loci. The goals of this research were: (1) to analyze the resolution and accuracy of the multiplex system in determining parentage, and (2) to assess rates of mutation at these loci. Over 99.6 % of offspring could be unambiguously matched to one set of parents in the original 12 X 1 cross (each of 12 males uniquely crossed to one of 12 females) and in the simulated 36 X 1 cross; and over 80 % in a 12 X 12 cross (involving multiple half-sib matings). Of the approximately 6,400 parent-offspring transfers of alleles screened in this parentage study, only two mutations were observed. The overall combined mutation at these four loci was estimated to be 3.4 X10-4 per gamete.

ABBREVIATIONS

A.L.-	allelic ladder
bp -	base pairs
CEPH-	Centre d'Etudes du Polymorphisme Humain
Het. -	heterozygosity
kb –	kilobase pairs
MVR-	minisatellite variant repeat
MVR-PCR-	minisatellite variant vepeat via Polymerase Chain Reaction
μM-	micro molar
mM-	milli molar
n/a-	not available
PCR-	Polymerase Chain Reaction
SSC-	NaCl & Na Citrate
SSR-	simple sequence repeat
STR-	short tandem repeat
TAE-	tris; acedic acid; EDTA
TBE-	tris; boric acid; EDTA

PUBLICATION LIST

Research presented here has appeared, at least in part, in the following published papers.

McConnell, S. K., O'Reilly, P., Hamilton, L., Wright, J.M. & Bentzen, P. 1995. Polymorphic microsatellite loci from Atlantic salmon *(Salmo salar)*: Genetic differentiation of North American and European populations. Can. J. Fish Aqu. Sciences **52**, 1863-1872.

O'Reilly, P., and Wright, J. M. 1995. The evolving technology of DNA fingerprinting and its application to fisheries and aquacluture. J. Fish Biol. **47A**, 29-55.

O'Reilly, P. T., Hamilton, L. C., McConnell, S. K., and Wright, J. M. 1996. Rapid analysis of genetic variation in Atlantic salmon *(Salmo salar)* by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Can. J. Fish. Aquat. Science **53**, 2292-2298.

McConnell, S. K., Ruzzante, D. E., O'Reilly, P. T., Hamilton, L. and Wright, J. M. 1997. Microsatellite loci reveal highly significant genetic differentiation between Atlantic salmon *(Salmo salar)* populations from the East Coast of Canada. Mol. Ecol. 6, 282-296.

ACKNOWLEDGEMENTS

Many individuals have assisted in one way or another over the far too many years that have taken to complete this thesis. Often, these people have helped me, and a great many other graduate students, with little or no formal credit, and, sometimes, little general acknowledgement. First, I would like to thank D. Cook for enduring more questions from myself, and nearly every other student on the sixth floor over the last 5 or so years, than the Dalhousie switchboard. It is amazing that he also found time to keep the M.P.G.L. lab running, given the great many little known, and even less appreciated, chores and tasks involved.

I would also like to thank M. Dillon for all the work she has done over the years in maintaining the lab, and training myself and far too many other students; we have all taken far more than we gave. L. Hamilton has similarly helped both myself and countless others with laboratory and computer problems, to the extent that many of us should be, well, embarrassed; I am quite sure that we have added a year to the time it will take her to complete her degree. To all three, I do appreciate your situation at the end of every August; in my case, the hazards of owning a van at moving time, in yours, the hazards of knowing what your doing. My next vehicle will be a motorcycle, I don't know what you all are going to do.

I thank J. Wright for supervising my thesis, and for far too much time spent writing grants. I also appreciate the freedom, at the end of my first year, to follow my own ideas and pursue my own areas of interest. Thanks also for putting me

back on the track of developing an MVR-PCR locus in Atlantic salmon. That was a lot of fun.

I am also greatly indebted to C. Herbinger for teaching me much about determining relatedness using genetic information, for helping with other analytical questions, and for coming to my rescue with a do-able research project. Thanks also for helping supervise much of my research, and for coming up with several original and insightful approaches to problems. P. Bentzen also helped supervise early on in my thesis. R. Doyle, L. Zouros, F. Doolittle and J. Hutchings were either formally on my committee, or attended one or more examinations, and read one or more dry reports. Thank you all. P. Amiro from D.F.O. helped track down samples, and provided background information on several Atlantic salmon populations from Nova Scotia. M. Jones and T. McParland graciously provided genotype information from the Upper salmon river, thanks a lot guys. S. McConnell, the great diplomat, helped establish collaborations with D.F.O., which eventually lead to some interesting research. M. O'Connell, another great diplomat, answered many of my questions, and helped with some analyses.

I would also like to thank all of my lab mates over the years, J. Frank, A. Brooker, D. Coltman, M. Dillon, L. Hamilton, S. Lang, M. O'Connell, S. McConnell, B. Pitmann, D. Paquet, and S. Mockford for helping create a very interesting and enjoyable environment in which to do research. Much thanks to Mary Dillon and Jennifer Crystal for time spent in the sometimes thankless task of helping to administer the Partick Lett fund. Finally, I am very grateful to Dr.

Lett himself for establishing the rather significant Lett fund for needy Graduate students- a great many students, including myself, have benefited immensely from this contribution.

This work was supported by an NSERC strategic grant, a Department of Fisheries and Oceans NSERC subvention grant, an American Fishing Tackle Manufactures Fellowship, and by Patrick Lett.

CHAPTER 1: GENERAL INTRODUCTION

Dispersed throughout the genomes of most, if not all, eukaryotic organisms are tandemly repeated blocks of DNA of identical or similar sequence. The length of these blocks may range from two to thousands of base pairs. DNA consisting of long repeat units (hundreds to thousands of base pairs) was originally termed satellite DNA because of the manner in which it was first identified. Due to the abundance of the repeats and their slightly different base pair composition compared to bulk genomic DNA, this class formed a separate, "satellite" band distinct from the main genomic band in equilibrium density centrifugation. A second class of tandemly repeated DNA, consisting of shorter repeat units (10-64 base pairs) was inadvertently discovered during DNA sequence analyses of the human insulin gene (Bell *et al.*, 1982). The reduced length of the repeat unit, compared to satellite DNA inspired the term "minisatellite" DNA (Jeffreys *et al.*, 1985). Loci with repeat unit lengths of between 10 and 64 base pairs are also commonly referred to as **V**ariable **N**umber of **T**andemly **R**epeat (VNTR) DNA (Nakamura *et al.*, 1987). A third class of tandemly-arrayed sequences with repeat unit lengths of only one to four base pairs (Litt and Luty, 1989) has been termed, not surprisingly, microsatellite DNA, but is also called simple sequence (Tautz, 1989) or **S**hort **T**andem **R**epeat (STR) DNA (Edwards *et al.*, 1991). Length variation in tandemly-arrayed repetitive DNA, particularly in the latter two classes, is usually due to increases or decreases in repeat unit copy number (Jeffreys *et al.*, 1988a; Weber, 1990).

These differences in repeat numbers represent the basis for most DNA profiling technologies used today.

Since the discovery of DNA fingerprinting by Alec Jeffreys almost a decade ago, there has been ongoing progression in the technological development and breadth of application of genetic typing methodology. Initially, genetic variation was surveyed at VNTR loci using multilocus DNA fingerprinting techniques (Jeffreys, 1985b), primarily for forensic and paternity investigations in humans. Shortly thereafter, the technology was applied to a variety of biological problems in numerous taxa, from paternity assessment in birds (Burke and Bruford, 1987) to strain identification in plants (Dallas, 1988). Several modifications to the basic approach, including single-locus fingerprinting, quickly followed. However, it was the advent of the **P**olymerase **C**hain **R**eaction (PCR) that made feasible major technological innovations, including Minisatellite Variant Repeat (MVR) mapping (Jeffreys *et al.*, 1991), and the assessment of variation at microsatellite loci (Weber and May, 1989). These new approaches surveyed greater amounts of genetic variation in the genome than was previously possible, often with much less effort and expense, and with increased resolution. Furthermore, the ability of PCR based techniques to detect genetic variation using extremely small amounts of DNA, and partially degraded DNA, permitted inquiry into issues in fisheries, aquaculture, management, and conservation that were not possible using earlier methods.

Multilocus fingerprinting

In multilocus DNA fingerprinting, length variation is surveyed at many VNTR loci simultaneously (see Figures 1.1 and 1.2). Due to the large number of loci examined and the extremely variable nature of this particular class of repeated DNA, each profile of bands (the so called "fingerprint") is usually highly informative and individual specific.

Numerous probes are available that hybridize to different VNTR loci possessing similar repeat unit sequences (Table 1.1). In some of these studies, applications involve resolution of first order relationships (parent-offspring-sib) in a limited number of individuals, permitting side by side comparisons of samples within the same gel. Although employed with varying degrees of success in several population level applications (Table 1.1), multilocus fingerprinting is not usually the method of choice for these uses. DNA profiles are often very complex, and it is usually not possible to identify both members of allelic pairs at individual loci. This makes it almost impossible to estimate allele frequencies, necessary in many population analyses. Also, multilocus fingerprinting protocols, although rigorously adhered to, often do not generate reproducible results. The intensity of some bands in the multilocus fingerprint, and the presence of others, can vary between gels containing the same samples, even under carefully controlled conditions. A second disadvantage of multilocus fingerprinting is the methodology-- Southern blot and hybridization. This method is much less sensitive (requiring more DNA), and more time consuming, than PCR based approaches.

Figure 1.1. Molecular basis of multilocus DNA fingerprinting. **Top Panel:** A1 and A2, and B1 and B2, show organization of three minisatellite loci on homologous chromosomes in individuals A and B, respectively. Genomic DNA from two parents (A and B) is digested with the restriction enzyme *Hae* III. **Bottom Panel:** the digested DNA is size fractionated by agarose gel electrophoresis. Using the Southern blotting technique, multiple VNTR arrays are detected simultaneously by hybridization to a radiolabelled probe with sequence similarity to minisatellite loci. Offspring inherit one allele from each of their parents, at all three loci.

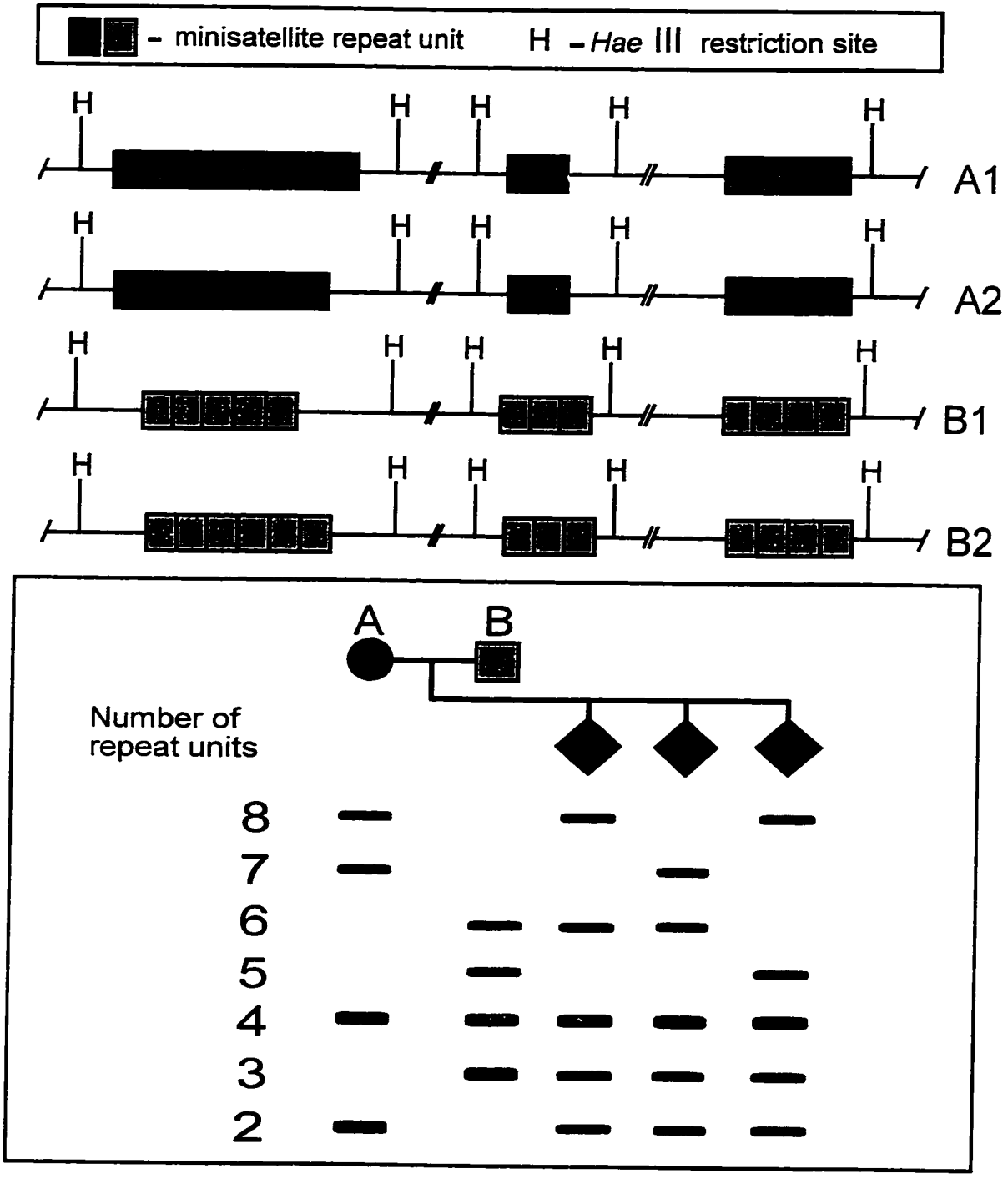

– minisatellite repeat unit
H – Hae III restriction site
H
A1
A2
B1
B2
A
B
Number of repeat units
8
7
6
5
4
3
2

Figure 1.1.

Figure 1.2. Fish multilocus DNA fingerprint from two parents and three offspring, showing Mendelian inheritance of minisatellite loci in tilapia. Approximate fragment lengths were determined using molecular size standards (in kilobase pairs at right of panel). (Harris, 1995).

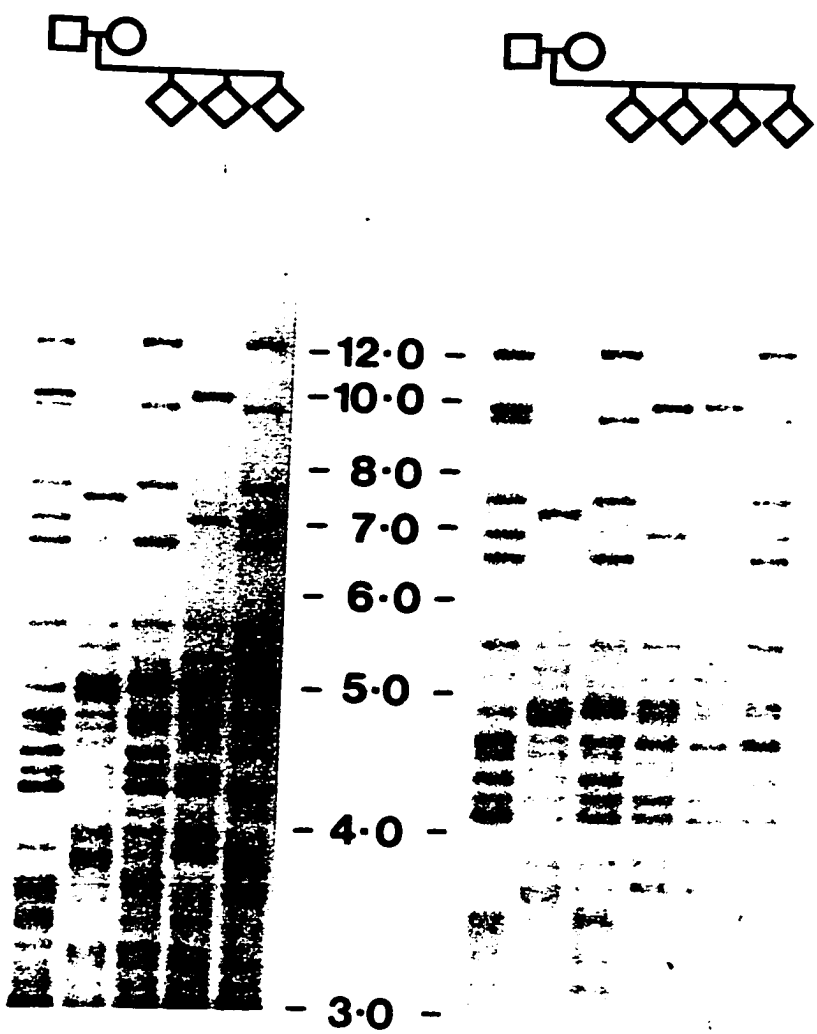


Figure 1.2

Table 1.1. Multilocus minisatellite probes used to generate fingerprint profiles in fish.

Probe(s)	Species assayed that hybridize to probe	Reference	Application
$GACA_{(4)}$	Sailfin molly *(Poecilia latipinna)*	Laughlin and Turner (1994)	Analysis of population variation
33.6	Brown trout *(Salmo trutta)*	Prodohl *et al.* (1992)	Analysis of population variation
33.6 33.15	Tilapia *(Oreochromis niloticus)*	Harris *et al.* (1991)	Initial survey of utility for studies of quantitative genetics and assessments of individual identity
Pspt18.15 pSPT19.6	Tilapia	Carter *et al.* (1991)	Analysis of gynogenesis
33.15 M13 3'HVR	Orange roughy *(Hoplostethus atlanticus)*	Baker *et al.* (1992)	Analysis of population variation
B2-2	Chinook salmon *(Oncorhynchus tshawytscha)*	Stevens *et al.* (1993)	Initial survey of utility for identifying family structure, populations, and relatedness issues in hatchery and aquaculture broodstock
pYNZ132	Stickleback *(Gasterosteus aculeatus)*	Rico *et al.* (1991)	Paternity assessment

A major practical advantage of multilocus fingerprinting, however, is the availability of probes that cross-hybridize in distant taxa. As such, genetic variation can be surveyed in just about any species with a minimum of development time and cost. Before using this approach, however, the researcher should ensure that (1) the fingerprint profiles are reproducible, (2) the diagnostic bands are inherited in a Mendelian fashion, (3) variable loci exhibit sufficient stability in the germ line and somatic tissue, and (4) that the restriction enzymes used are not methylation sensitive (Wright, 1993).

Single locus VNTR profiling

In single locus profiling, allelic variation is surveyed at individual VNTR loci, using one of two approaches. The first method involves restriction endonuclease digestion of genomic DNA, separation of fragments by electrophoresis through agarose gels, and blotting (Southern, 1975) onto DNA binding membranes (Jeffreys *et al.*, 1988a; Bentzen and Wright, 1993; Taylor *et al.*, 1994) (see Figure 1.3). Membranes are then probed with denatured, labelled DNA from a single VNTR locus, preferably the unique flanking region. The second method is to PCR amplify the locus using primers flanking the array (Jeffreys *et al.*, 1988b; Jeffreys *et al.*, 1994; Galvin *et al.*, 1995a,b) (see also Figure 1.4). PCR products are separated by standard gel electrophoresis, and can be visualized by a variety of methods, including, under certain conditions, post-separation staining with ethidium bromide.

Figure 1.3. Single locus fingerprinting of tilapia. Lanes a-e demonstrate Mendelian inheritance, where a homozygous male (a) was crossed with homozygous female (b). Offspring (c-e) show one allele from each parent. Individuals f-v are unrelated fish. Approximate lengths were determined using molecular weight standards (in kilobase pairs at left of panel). (Harris, 1995).

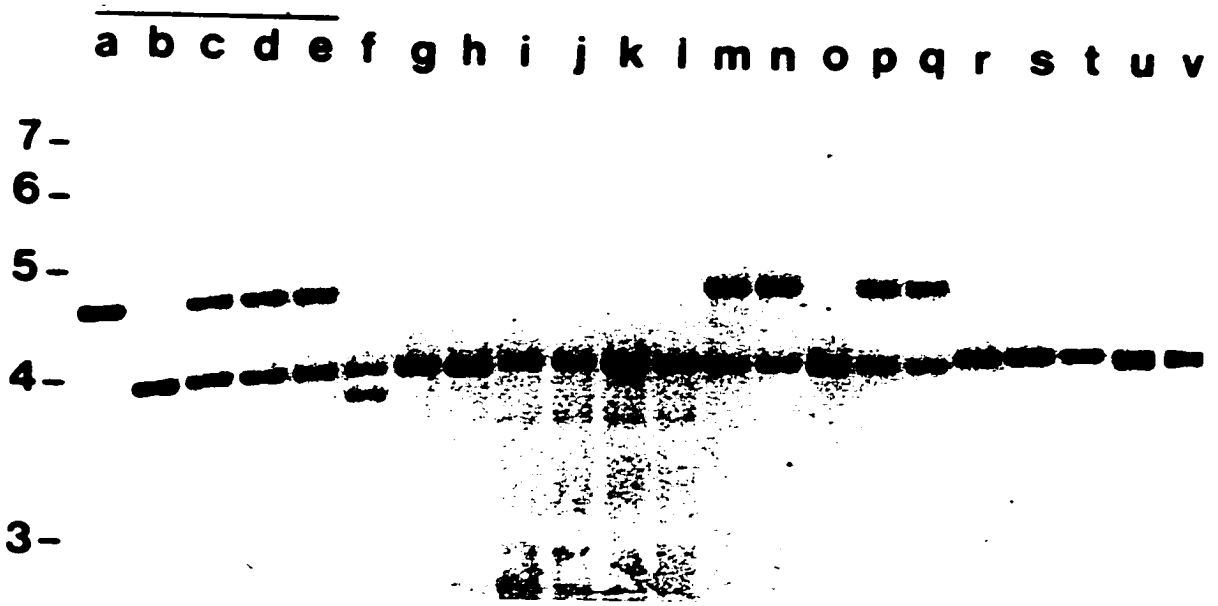


Figure 1.3

Figure 1.4. Mini- and microsatellite loci can be amplified via PCR. Using primers complementary to DNA on both sides of the array, loci are amplified many thousand fold by repeated denaturing, annealing, and extension cycles. In this illustration, multiple copies of two alleles from a heterozygous individual are produced. Alleles from minisatellite arrays would be assayed for length variation using agarose gel electrophoresis; microsatellite products would typically be resolved by electrophoresis in denaturing polyacrylamide gels.

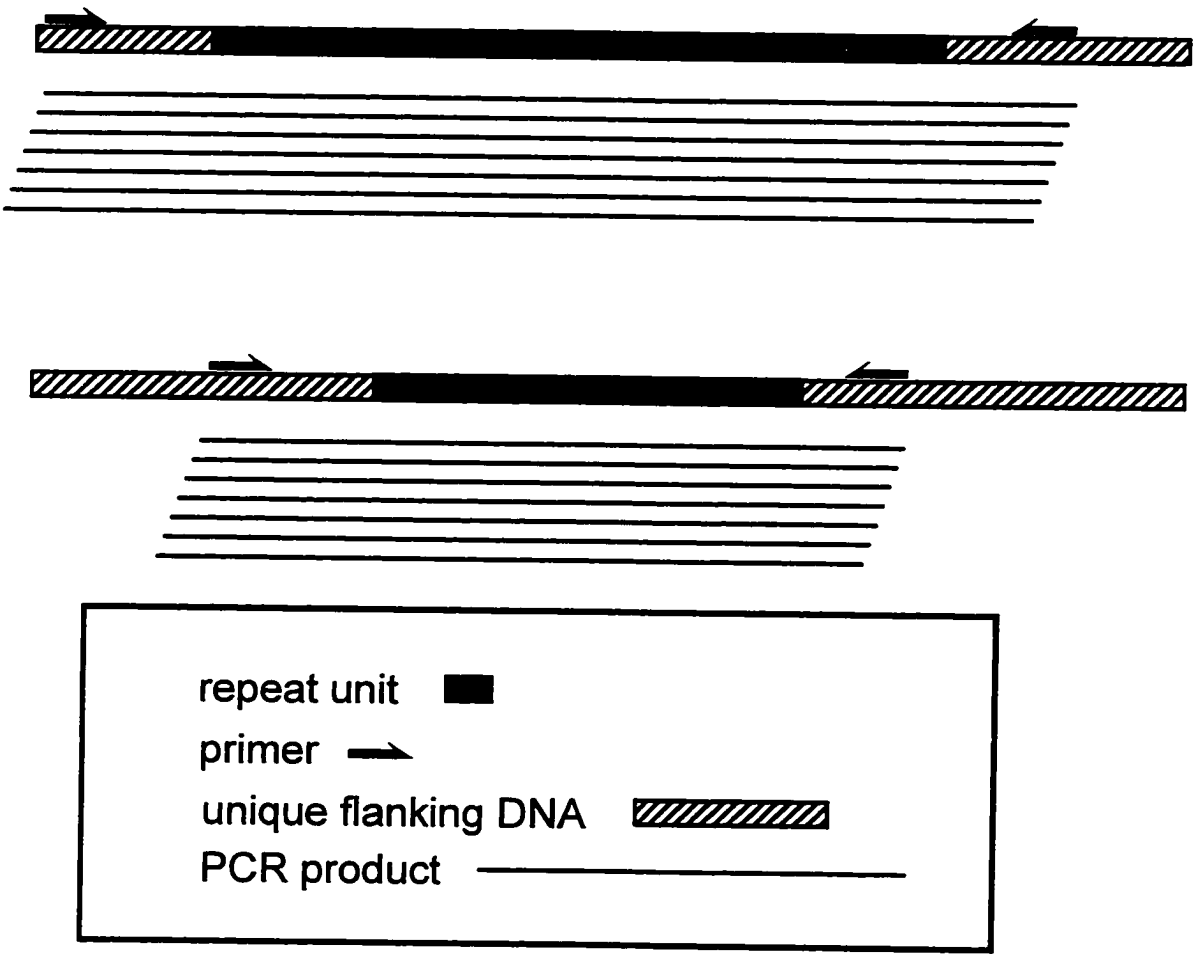


Figure 1.4.

Development of either approach requires the isolation and cloning of single VNTR loci. In the second method, DNA sequence is needed from both sides of the array, which, in some instances, may require considerable additional effort. Once developed, PCR- based VNTR systems are much less labour intensive than methods using Southern blotting, and may be conducted in a standard molecular biology laboratory with a minimum of expensive equipment. There is, however, one serious draw-back. In conventional PCR, fragments much larger than four or five kilobases are not readily amplified and detected. Many VNTR loci surveyed in humans (Jeffreys *et al.*, 1988a) and fish (Taylor *et al.*, 1994) range from less than 1000 base pairs (bp) to over 10 kilobase pairs (kb). For some applications, for example, maternity/paternity assessment of offspring, where the expected product sizes can be readily determined by genotyping the parents, this is not an insurmountable problem; by surveying, ideally, only a few VNTR loci, markers that are both informative and that consistently amplify may be identified. In population applications, however, it is much more difficult to estimate the possible range of allele sizes likely to be encountered.

To date, most surveys of VNTR loci in fish involve Southern blotting and hybridization (Stevens *et al.*,1993; Taggart and Ferguson, 1990; Taylor *et al.*, 1994; Bentzen and Wright, 1993). Species-specific probe development and their application has been limited to gadoids, salmonids and tilapia (Table 1.2). Recently, Galvin *et al.* (1995a) has used a PCR based method to assess variation at a single minisatellite locus in several gadoids. Amongst the

Table 1.2. Polymorphism at single-locus minisatellites detected by DNA probes from fish.

Source	Probe	Species assayed that hybridize to probe	Number of alleles	Heterozygosity
Atlantic salmon[1] *(Salmo salar)*	3.15.33	Atlantic salmon[1]	7	72
	3.15.34	Atlantic salmon[1]	5	64
		Chinook salmon[2] *(Oncorhynchus tshawytscha)*	8-20	n/a
	3.15.45	Atlantic salmon[1]	5	66
	3.15.60	Atlantic[1] salmon[1]	6	59
Atlantic salmon[3]	S*sa*1	Atlantic salmon[3]	9	36-77
		rainbow trout[3] *(O. mykiss)*	n/a	n/a
		brook trout[3] *(Salvelinus fontinalis)*	n/a	n/a
		sockeye salmon[3] *(O. nerka)*	n/a	n/a
		chum salmon[3] *(O. keta)*	n/a	n/a
Whiting[4] *(Merlangius merlangus)*	Mmer-AMP2	whiting[4]	24	93-95
		Atlantic cod[5] *(Gadus Morhua)**	22	76-90
Tilapia[6] *(Oreochromis niloticus)*	OniMS17	tilapia	4	n/a
	OniMS34	tilapia	6	n/a

* single locus profile generated via PCR amplification
n/a not available

References:

1. Taggart and Ferguson (1990)
2. Stevens *et al.* (1993)
3. Bentzen and Wright (1993)
4. Galvin *et al.* (1995a)
5. Galvin *et al.* (1995b)
6. Harris and Wright (1995)

salmonids, levels of polymorphism are generally moderate, with most loci exhibiting 5 to 10 alleles, and heterozygosity values typically ranging from 36% to 77%.

VNTR probes have been used to assess population structuring and identification. The Ssa1 probe developed for Atlantic Salmon *(Salmo salar)* produces highly variable and informative profiles in chum salmon *(Oncorhynchus keta)* (Taylor *et al.*, 1994). Using discriminant function and neural network analyses, Taylor *et al.* (1994) classified fish to their region of origin with a high degree of success. Stevens *et al.* (1993) used single locus markers (in concert with multilocus fingerprinting) to assess population structure in chinook salmon *(Oncorhynchus tshawytscha)* on a much finer scale. Despite the increased polymorphism of the single locus 3.15.34 probe over the multilocus B2-2 probe also used, the latter proved more valuable in resolving family structure in chinook salmon, presumably due to the increased number of informative independent loci detected. Galvin *et al.* (1995a) surveyed variation at the Atlantic whiting Mmer-AMP2 minisatellite locus. Average heterozygosity values (94%) were among the highest cited for any fish species. Significant differences in allele frequencies between regional populations of whiting off the west coast of Europe were reported. Using the Mmer-AMP2 probe isolated from the whiting, Galvin *et al.* (1995b) also assessed geographic variation in Atlantic cod. Significant allele frequency differences amongst regional populations from the north-east Atlantic, and between east and west Atlantic cod, were also found. Several observations from Galvin *et al.* (1995b) suggest that the presence and detection of larger

alleles was not a serious problem in this study: 1) allele size ranges were quite small, between 460 and 1870 bp; 2) the longer products, though fainter than the smaller products, were still readily visible; 3) two alleles were observed in 94 % of the individuals surveyed; and 4) an excess of heterozygotes, as opposed to homozyogotes, was detected.

Single locus approaches to DNA fingerprinting obviate several major problems associated with multilocus methods. First, band profiles are much simpler (Figure 1.3), and co-migration of alleles from non-homologous loci is unlikely to occur. Second, since bands can be ascribed to a given locus, allele frequencies (or more realistically, allele bin frequencies, discussed below), can be estimated. Third, with the inclusion of proper allelic standards, comparisons are possible across many gels.

Some of the limitations of single locus minisatellite approaches are as follows. From high resolution mapping of the interspersion pattern of minisatellite repeat unit types (Jeffreys *et al.*, 1994), it is clear that the mechanisms of mutation that generate variation at some minisatellite loci are very complex, involving intra-allelic duplication or deletion, but also inter-allelic recombination and gene conversion. Hence, devising realistic mutation models for such loci will be exceedingly difficult. Also, inter-allelic recombination and gene-conversion will serve to scramble linear assemblages of repeat units between homologous arrays, and therefore, many VNTRs will not evolve along haploid chromosomal

lineage's as once thought (Jeffreys *et al.*, 1990; Wright, 1993). Finally, minisatellite alleles identified in surveys often do not differ from one another in length by discrete, integral increases or decreases in the number of repeat units (Jeffreys *et al.*, 1988a; Taylor *et al.*, 1994). For example, two alleles may differ by 15 bp, half the standard 30 base pair repeat length at a hypothetical locus. This observation is due to one or more of following: (1) as revealed by DNA sequence analysis, the frequent observation of repeat units with non-standard lengths (Bentzen & Wright, 1993; S. McConnell, unpublished data, P. O'Reilly, unpublished data); (2) the frequent occurrence of dinucleotide and tetranucleotide loci immediately flanking many of the VNTR arrays sequenced in fish (Bentzen & Wright, 1993; P. O'Reilly, unpublished data); and (3) the limited resolving power of agarose gel electrophoresis and detection of alleles via Southern blotting. Given the length of some VNTR arrays compared to the repeat unit size, the percent size difference between many alleles can be unmanageably small (e.g., 0.2% for neighbouring alleles approximately 10 kb in length, and with repeat unit lengths of 20 base pairs). Combined, these factors make comparisons of allele sizes between gels difficult, and have necessitated the grouping or binning of alleles into defined size classes (Figure 1.5) (Budowle *et al.*, 1991; Taylor *et al.*, 1994). In addition to underestimating genetic variation, binning also makes it difficult to determine if populations are in Hardy-Weinberg equilibrium as heterozygous individuals with alleles differing only slightly in length may be scored as homozygotes (Devlin *et al.*, 1991).

Figure 1.5. Schematic showing the basis of allele binning, a method frequently employed in large scale surveys of length variation at individual minisatellite loci. In this example, bin intervals are arbitrarily defined at 1 kilobase intervals. Alleles within delineated boundaries are binned, and various population parameters are estimated for allele groups instead of specific alleles. A common criticism of this practice is that alleles of similar size are often grouped together, thus contributing to the excess of homozygotes, and underestimating genetic diversity. For example, the individual in lane 5 is clearly heterozygous; however, owing to binning, it would be classified as homozygous.

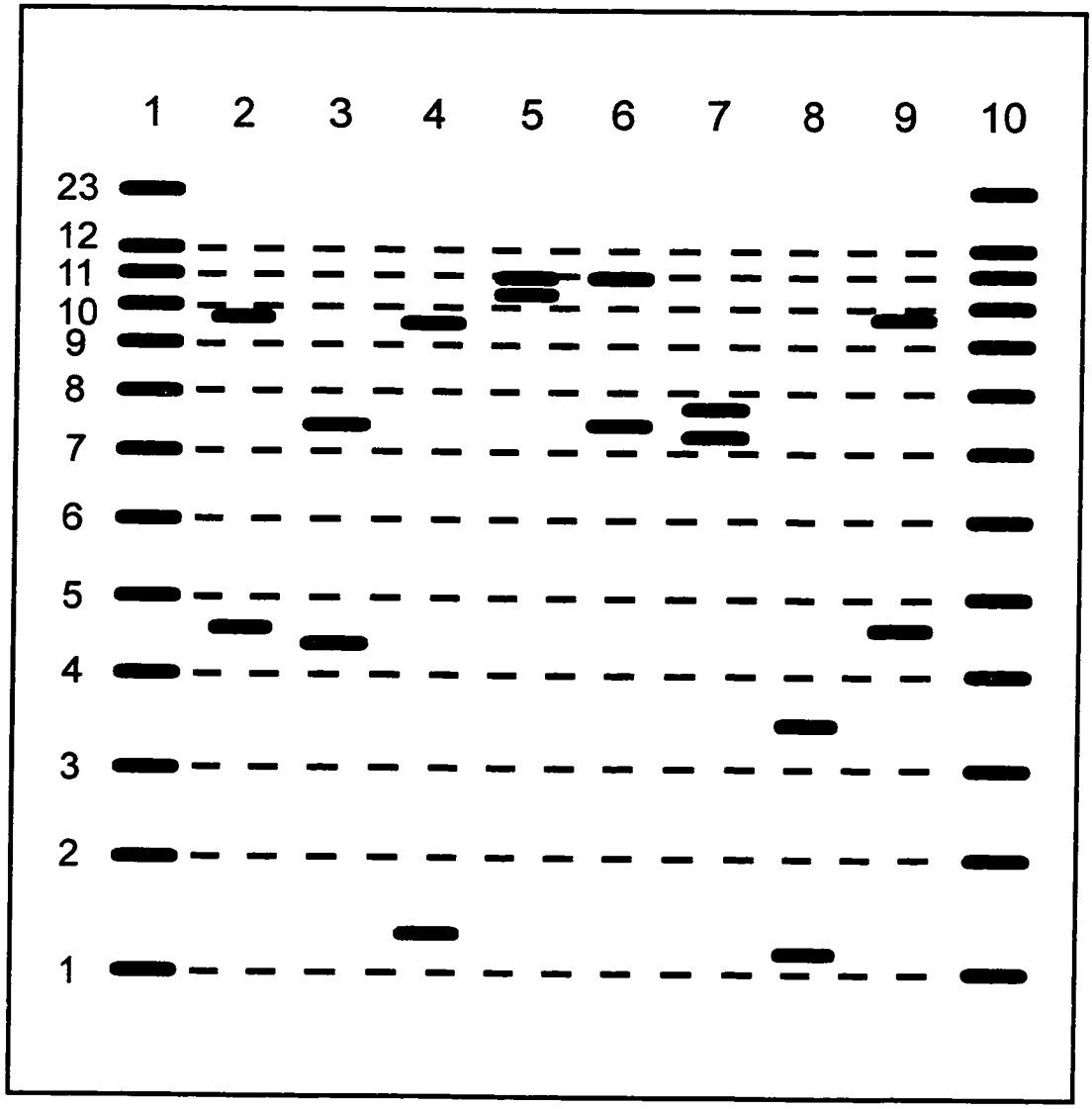


Figure 1.5.

Minisatellite variant repeat mapping

The extremely high levels of allelic variation seen at some minisatellites using conventional size separation methods may reflect only a portion of existing variation at these loci. Minisatellite alleles can differ not only in the number of copies of repeats, but also in the interspersion pattern of different types of repeats comprising the array. Jeffreys *et al.* (1990) first assayed repeat unit sequence variation within single (haploid) molecules by using restriction enzymes which recognize and cleave specific repeat unit types. Although this method revealed variation never before seen, it was also laborious, technically demanding, and was limited to short alleles readily amplified by PCR. Shortly thereafter, Jeffreys *et al.* (1991) developed a more expedient PCR based method, termed **M**inisatellite **V**ariant **R**epeat-PCR (MVR-PCR) (also referred to as digital DNA fingerprinting). In this strategy, the sequence of repeat units in both (diploid) arrays is determined simultaneously by PCR using primers complementary to the different repeat unit types, and a primer in nearby unique flanking DNA. During successive rounds of PCR, repeat unit primers anneal to internal sites, causing a progressive shortening of PCR products. This results in the rapid attenuation of signal strength of bands representing increasingly distal repeat positions. Jeffreys *et al.* (1991) were able to circumvent this problem by using 'tagged' repeat unit primers. In this strategy, initial PCR amplifications were performed using a flanking primer, and low concentrations of repeat unit specific primers with approximately 20 base pair extensions. After several cycles, a series of PCR products differing in length by integral numbers of repeat

units are generated. However, these products now terminate in single copy tag or extension sequence. By including high concentrations of extension oligonucleotide in the PCR reaction mix, subsequent amplifications involved mainly flanking primer (also present in high concentrations) and extension primer. Since the complement of the extension sequence occurs, at most, only once in each PCR product, internal priming and product length reduction is avoided.

Repeat unit positions along the array can be assigned numerical values representing the state (homozygous type A, homozygous type T, heterozygous, etc.) of sequential repeats, producing an individual-specific diploid code, ideal for computer databasing. Of the 334 unrelated individuals surveyed by Jeffreys *et al.* (1991), no two diploid MVR codes were the same. Whereas standard assays of length variation are susceptible to band shifting and gel distortion problems, digital mapping provides its own internal standards and is largely immune to electrophoresis-related complications.

To date, accounts of only three MVR-PCR amenable loci have been published, all developed in humans by Jeffreys and colleagues (Jeffreys *et al.* 1991; Jeffreys *et al.* 1994). In Chapter 2, attempts to develop MVR-PCR in Atlantic salmon are discussed.

Microsatellites

Polymorphism at microsatellite loci was first demonstrated by Tautz (1989) and Weber and May (1989). Within a year or two, microsatellites were

being employed as genetic markers in medical, gene mapping and forensics studies. Their use continued to expand in the early to mid 1990's (in 1994 alone, there were over 200 publications with the term Microsatellite in the title or abstract). The rapid use and acceptance of microsatellites as genetic markers is due primarily to their comparative ease of assay via PCR, and accuracy of scoring allele types. Outlined below are several specific characteristics of microsatellites that make these ideal markers for a range of applications in aquaculture and fisheries biology.

Properties of Microsatellites

Array Length: Microsatellite arrays are generally quite short, 20 to 300 bp. This size range is well within the capabilities of conventional PCR, and "allele drop out", the tendency for large alleles to amplify poorly to the point of being difficult to detect visually, is rarely a problem. The short length of the PCR products amplified from microsatellite loci allows alleles to be size-fractionated on denaturing polyacrylamide gels, providing considerably more resolving power than standard agarose gel electrophoresis. With the inclusion of allelic size standards, single base pair differences between even the largest alleles (300 bp) can be distinguished. This has important implications for the scoring of alleles, but also for the databasing or management of information (discussed in detail in section on *"future developments"*).

Repeat Unit Simplicity: From DNA sequence information from Atlantic cod *(Gadus morhua)* (Brooker *et al.*, 1994); Atlantic salmon (Slettan *et al.*, 1993;

McConnell *et al.*, 1995a; and brown trout (Estoup *et al.*, 1993), many microsatellites consist of perfect uninterrupted stretches of identical repeat units. At such loci, most or all differences between alleles seem to be due to integral increases or decreases in the number of repeats of identical length. This greatly facilitates the scoring of alleles, but also offers the promise of a much simpler, less expensive means of surveying variation at some loci (discussed below). By contrast, nearly all of the minisatellites analyzed from fish to date exhibit various levels of sequence and length heterogeneity amongst repeat units comprising the arrays (Bentzen and Wright, 1993; Harris and Wright, 1995; McConnell unpublished data, O'Reilly, unpublished data). As mentioned earlier, the latter type of variation contributes to the discontinuous size distribution of VNTR alleles at a given locus.

Levels of Polymorphism: A very high proportion of microsatellite loci surveyed in fish are polymorphic (Table 1.3). In all these species, the extent of polymorphism varies greatly amongst loci, with some exhibiting only a few alleles, and others dozens. The range of variability makes it possible to select loci for specific applications. For example, markers with only a few alleles are well suited for population genetic studies (Carvalho and Hauser, 1994; O'Connell and Wright, 1997). The more variable microsatellite loci, however, are ideal for gene mapping, determination of parentage in aquaculture genetics, and pedigree analysis.

Table 1.3. Levels of polymorphism observed at microsatellite loci from several fish species.

Species	No. of loci*	No. of variable loci	No. of alleles per locus	Het.** (Percent)	Reference
Zebra fish *(Brachydanio rerio)*	17	16	2-5	n/a	Goff *et al.* (1992)
Atlantic cod *(Gadus morhua)*	7	6	8-46	14-92	Brooker *et al.* (1994)
Atlantic salmon *(Salmo salar)*	4	4	4-52	28-91	McConnell *et al.* (1995a)
Rainbow trout *(Oncorhynchus mykiss)*	2	2	8-9	72-78	Morris (1993) Morris *et al.* (1996)
Brown trout *(S. trutta)*	3	3	5-6	18-74	Estoup *et al.* (1993)
Bluegill sunfish *(Lepomis macrochirus)*	17	7	2-8	27-63	Colbourne *et al.* (1996)
Gulf pipefish *(Syngnathus scovelli)*	4	4	19-28	89-95	Jones and Avise (1997)

* number of loci surveyed for variation
** Het. = Heterozygosity
n/a = not available

gene mapping, determination of parentage in aquaculture genetics, and pedigree analysis.

Jeffreys *et al.* (1988a) found that extremely polymorphic minisatellite loci with heterozygosity levels much in excess of 97% exhibited mutation rates on the order of 0.005 to 0.05 per gamete per generation (measured by direct counts from large human pedigrees). These values matched remarkably well with predicted mutation rates based on heterozygosity levels, assuming a neutral mutation/random drift hypothesis. Extremely high mutation rates such as these can limit the effectiveness of markers in paternity assessment (Jeffreys *et al.*, 1988a; Jeffreys *et al.*, 1991; Farr and Goodfellow, 1991) by leading to false exclusions. With the exception of several highly unstable trinucleotide microsatellites associated with various human diseases, mutation rates thus far estimated for dozens of human microsatellites (using several different methods) are typically below 0.005 per gamete per generation (Weber and Wong, 1993; Hastbacka *et al.*, 1992; Edwards *et al.*, 1992). The two most variable microsatellite loci reported have germline mutation rates of approximately 0.008 (Mahtani and Willard, 1993; Weber and Wong, 1993). In fish, no direct assessments of microsatellite mutation rates have been attempted. However, given the similar upper heterozygosity levels to those reported in humans (approximately 90%), microsatellites should provide the maximum amount of information for gene mapping, parentage determination, and resolving pedigrees, without significant risk of mutation to confound analyses.

Nature and mechanisms of mutations: Understanding mutational processes at microsatellite loci is important in developing statistical procedures for inter-population comparisons (Di Rienzo *et al.*, 1994). Most observed mutations at microsatellite loci involve single repeat unit increases or decreases in copy number (Kwiatkowski *et al.*, 1992; Sirugo *et al.*, 1992; Oudet *et al.*, 1993; Weber and Wong, 1993; Mahtani and Willard, 1993; Hastbacka *et al.*, 1992; Henderson and Petes, 1992), with occasional changes involving several repeats (Weber and Wong, 1993). However, at many trinucleotide microsatellite loci associated with various diseases in humans, increases of dozens or even hundreds of repeat units have been reported (reviewed in Nelson and Warren , 1993). Di Rienzo *et al.* (1994) found that of the three models of mutation tested, (1) the one step, (2) geometric, and (3) two-phase (involving mostly single-step changes but also occasional larger additions/deletions), the third model best reflected allele frequency distributions at ten microsatellite loci in a human Sardinian population tested.

The molecular mechanism thought to be responsible for most mutations at microsatellite loci is slipped-strand mispairing during DNA replication (Levinson and Gutman, 1987; Litt and Luty, 1989; Tautz, 1989). *In vitro* experiments using synthetic oligonucleotides have demonstrated that arrays can both arise and expand due slipped strand mispairing alone (Behn-Krappa and Doerfler, 1994; Schlotterer and Tautz, 1992). Also, an increasing number of studies (Mahtani and Willard, 1993; Weber and Wong, 1993; Morral *et al.* 1991, Kwiatowski *et al.*, 1992; La Spada *et al.*, 1992; Shelbourne *et al*, 1992), involving several lines of

evidence, have reported findings consistent with either slipped-strand mispairing or unequal exchange between sister chromatids, and not unequal crossing over between homologous chromosomes seen at some minisatellite loci (Jeffreys *et al.*, 1994).

Abundance in fish genomes: Based on screening of size-selected genomic libraries from several species, microsatellites appear to be abundant in the genomes of teleost fish. Arrays consisting of $(GT)_n$ repeats were reported to occur, on average, every 7 kb in Atlantic cod (Brooker *et al.*, 1994), 12 kb in zebra fish (Goff *et al.*, 1992), 11-56 kb in Atlantic Salmon (McConnell *et al.*, 1995a), and 24 kb in brown trout (Estoup *et al.*, 1993). Estoup *et al.* (1993) also screened for $(CT)_n$ microsatellites, and found that they were about 1/3 as abundant as $(GT)_n$ arrays. The proportion of $(GT)_n$ to $(CT)_n$ microsatellites is surprisingly similar to that reported for rats and humans (3:1) (Beckmann and Weber, 1992). Although these values are rough approximations, similarly derived estimates in humans (Tautz and Renz, 1984; Litt and Luty, 1989; Stallings *et al.*, 1991) were found to be in agreement with actual numbers obtained from human sequence databases (Beckmann and Weber, 1992).

In addition to lessening cloning and development efforts for polymorphic markers, the abundance of microsatellites in the teleost genome has several other important implications. First, given their frequency and dispersion throughout the genome, and high levels of variability (Brooker *et al.*, 1994; McConnell *et al.*, 1995a,b; Nielsen *et al.*, 1994, 1995; Estoup *et al.*, 1993), these

sequences promise to be ideal for genome mapping in fish (Wright, 1993). Second, the almost unlimited number of potential markers makes it possible to accumulate assemblages of compatible loci well-suited for various general applications. For example, in humans several sets of three to six loci have been developed (Edwards *et al.*, 1991; Urquhart *et al.*, 1995; Kimpton *et al.*, 1993) that are unlinked, highly variable, and co-amplify or multiplex (discussed below). These marker systems permit rapid, inexpensive, and highly accurate determinants of individual identity for paternity and forensics. Similar marker systems in fish comprised of several highly variable loci could be developed for pedigree, parentage assessment, and forensic use. Also, systems employing suites of less variable loci could be developed for population applications, where increased numbers of alleles complicate analyses, and necessitate larger sample sizes (Carvalho and Hauser, 1994; S. McConnell, pers. comm.).

Cloning of microsatellites

Methods describing the cloning and characterization of dinucleotide microsatellites are abundant in the literature (e.g. Weber and May, 1989; Weber, 1990; Estoup *et al.*, 1993; Weissenbach *et al.*, 1992; Rico *et al.*, 1994). Briefly, partial or total genomic libraries are constructed using, for example, pUC18 or M13 cloning vectors. Insert size is typically kept small (300-600 bp) so that DNA flanking both sides of the microsatellite array can be readily sequenced (if libraries with larger inserts are necessary, see Baron *et al.* (1992)). Clones containing microsatellite arrays of a given sequence, (GT) for example, can be

identified by hybridizing labelled oligonucleotides of the complementary sequence (AC_{15}) to membrane lifts of plates containing the original colonies or plaques. Inserts containing microsatellites from positive colonies are then sequenced and flanking PCR primers designed.

Tetranucleotide microsatellites are gradually replacing dinucleotide loci as the preferred genetic marker for certain applications (discussed below), despite reports that they are more difficult to isolate and clone than traditional GT or CT tracts (Paetkau and Strobeck, 1994). The primary factor contributing to the challenge of finding these loci is the scarcity of specific array types. Using hybridization techniques, Edwards *et al.* (1991) estimated the frequency of each of the $(AATC)_n$, $(AATG)_n$, $(ACAG)_n$, and $(AGAT)_n$ loci on the human X chromosome at about 1 in every 150-500 kb. Similar values were obtained by Beckmann and Weber (1992) in a survey of all possible mono-,di-,tri-,and tetranucleotide microsatellites from 745 kb of human genomic DNA. $(AAAN)_n$ microsatellites were by far the most common found in the latter study, and were often associated with the poly A tail of *Alu* repetitive elements. The abundance of these loci in other taxa lacking this element, therefore, is expected to be markedly lower. Most of the remaining motifs were observed once or not at all in the 745 kb of human DNA surveyed. Overall, tetranucleotide microsatellites were similarly abundant in rats. In both humans and rats, they represent 23% of the total mono-,di-,tri-, and tetranucleotide microsatellites observed. Although the relative abundance of the different repeat motifs varied between the two species, $(AAGG)_n$ and $(AGAT)_n$ microsatellites were relatively common in both.

Given the generally low abundance of specific types of tetranucleotide compared to GT and CT arrays, and the marked variation seen in the frequency of the different tetranucleotide motifs, careful probe selection is important in isolating loci by traditional hybridization techniques. Unfortunately, no reports on the relative occurrence of the different tetranucleotide motif any fish have been published. Screening for $(AAGG)_n$ and $(AGAT)_n$ microsatellite loci, given their abundance in humans and rats, is a promising option for isolating these markers. Another approach to cloning tetranucleotide microsatellites is to screen libraries containing large numbers (10^5 - 10^6) of colonies or plaques, or libraries containing large sized inserts (Edwards *et al.*, 1991; Li *et al.*, 1993). Alternatively, methods that enrich for specific classes of loci may be employed (Armour *et al.*, 1994).

A further caution when screening for microsatellites is that they are frequently observed to occur within or near large dispersed repetitive elements in humans (Beckmann and Weber, 1992), artiodactyls (Kaukinen and Varvio, 1992; Buchanan *et al.*, 1993; Wilke *et al.*, 1994), seals (Coltman and Wright, 1994), wolverine (Duffy *et al.* 1996), and Atlantic salmon (W. Davidson, pers. comm.). This can be problematic if one or both of the primer sites is located within the element. PCR primers targeted at microsatellites embedded in repetitive elements may either produce multiple accessory bands or low product yields. I recommend that during the design of PCR primers for amplifying microsatellite loci, flanking sequence be first checked for sequence similarity to any of the published repetitive elements.

Application of Microsatellites to Aquaculture and Stock Analyses

Only recently has cloning and characterization of microsatellite loci in teleost fishes (e.g., Atlantic cod, rainbow trout, and Atlantic salmon) been reported (Estoup *et al.*, 1993; Brooker *et al.*, 1994; Morris, 1993; Morris *et al.*, 1995; Slettan *et al.*, 1993; Sakamoto *et al.*, 1994; McConnell *et al.*, 1995b), and studies applying these new markers to problems in aquaculture and population biology are just starting to appear. Herbinger *et al.* (1995a) have demonstrated the feasibility of determining pedigrees in a mixed family rainbow trout population. Using only five loci, 91% of the 873 offspring analyzed could be traced to one or two parental couples out of a possible 100 couples. Significant differences in the growth and survival of progeny from different sires and dams, and preliminary findings of reduced performance of inbred progeny, were also reported. In an initial study describing the use of microsatellite DNA profiles to analyze genetic correlations between hatchery and natural fitness, Doyle *et al.* (1995) assessed the extent of domestication selection (selection for improved hatchery performance, usually associated with reduced fitness in the wild) in laboratory populations of Atlantic cod. Doyle and Herbinger (1995) have developed within-family selection procedures using microsatellite markers. The technique is designed to greatly increase selection intensities while decreasing levels of inbreeding in commercial aquaculture operations. Using recently developed likelihood ratio methods, Herbinger *et al.* (1997) have looked for evidence of family structuring in a natural cohort of cod larvae on the western

Bank of the Scotian Shelf. The population sampled seemed to be comprised of a mixture of unrelated individuals derived from a fairly large genetic pool.

Several recent studies have observed various levels of regional population structuring of allele frequencies at microsatellite loci. For example, Nielson *et al.* (1994, 1995) found significant differences in allele frequencies between Pacific steelhead from several regions within California. Their findings have important implications for the conservation of these rapidly declining stocks: by cataloging geographic structuring of microsatellite allele frequencies, brood stock collection sites for proposed hatchery-based enhancement efforts can be chosen so as to minimize the loss of genetic variation in these populations. In a study of microsatellite variation between Atlantic salmon populations from different drainage basins in Nova Scotia, McConnell *et al.* (1997) report increased genetic distance (at four loci surveyed) with increasing geographic distance.

The limited length of microsatellites and resulting amenability to PCR also makes possible assays of microsatellites from sources of DNA that were not previously feasible, thus increasing the range of potential applications. For example, DNA can be amplified from minute amounts of tissue, including scales (Whitmore *et al.*, 1992; Park & Morgan, 1994), fin clips or small amounts of blood, permitting the non-destructive sampling of individuals. Moreover, microsatellite variation has been assayed using single sperm (single molecules) (Hubert *et al.*, 1992). PCR amplification of microsatellite loci has also been achieved from DNA extracted from otoliths (D. Cook, unpublished data) and from scales air-dried and stored for 10 or more years. This has permitted

retrospective analyses of allele frequencies. For example, I have been able to assess long-term (over 40 years) genetic changes in hatchery-enhanced LaHave River stocks of Atlantic salmon. In this study, genetic variation was surveyed at 5 microsatellite loci in a minimum of 80 individuals from multiple periods dating back to 1953 (manuscript in prep.).

Disadvantages and Limitations of Microsatellites as Molecular Markers

Development and Costs: A major criticism that may be leveled against the use of microsatellites as genetic markers, compared to allozymes and multilocus fingerprinting, is high development costs. I have found, however, that several reliable and polymorphic microsatellite loci can be developed in two to three months. Furthermore, many loci cross-amplify in a range of related species. For example, microsatellites developed in rainbow trout and Atlantic salmon PCR amplify (and are variable) in many *Salmo* and *Onchorynchus* species, respectively (Morris *et al.*, 1995; McConnell *et al.*, 1995b). Furthermore, microsatellite markers developed for an one application (e.g. resolving pedigrees), may prove equally useful for other applications (e.g. gene mapping or forensic identification).

The appearance of shadow or stutter bands: Most dinucleotide microsatellite alleles are visible as a ladder of bands, rather than a single discrete product (Figure 1.6). "stutter" bands are generally thought to be due to slipped-strand mispairing during PCR (Litt and Luty, 1989; Tautz, 1989; Luty *et al.*,

Figure 1.6. Multiplex PCR of several unrelated individuals showing characteristics of different classes of microsatellites isolated from Atlantic salmon. *Ssa* 171 and *Ssa* 197 are tetranucleotide loci, exhibiting minimal PCR stuttering compared to *Ssa* 85, a typical dinucleotide locus. Whereas *Ssa 197* alleles are consistently separated by 4 base pairs, non standard 2 base pair differences (microheterogeneity) are often observed at *Ssa* 171. Lengths are determined by comparisons with allelic ladder standards (A.L.) and M13 DNA sequence, shown on the far left and right.

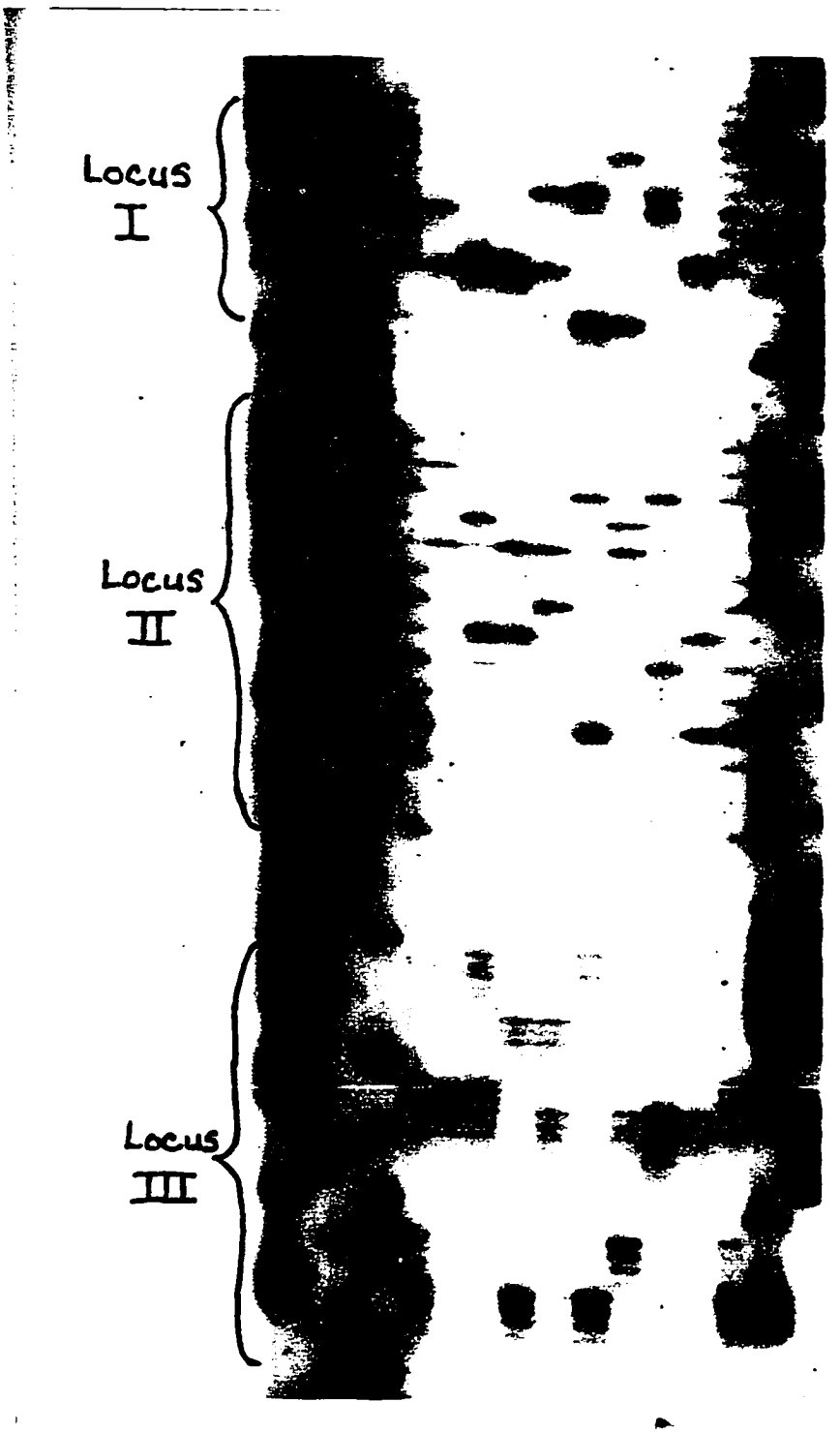


Figure 1.6

1990), but alternative explanations include *in vivo* slipped-stand mispairing and incomplete denaturation of amplification products, and "out of register" annealing of truncated products (Hauge and Litt, 1993). By sequencing PCR products directly, Hauge and Litt (1993) demonstrated that the shadow bands observed at the human D11S527 locus were indeed due to slipped strand mispairing during PCR.

The practical outcome of PCR "stutter" is that it may cause problems scoring alleles. Often, the pattern of "stutter" (number of base pairs separating bands and the number of bands) is constant amongst individuals, such that well-separated alleles can be unambiguously identified. It should be noted though that it is sometimes difficult to distinguish homozygotes from heterozygous individuals when alleles are separated by only two base pairs. In practice, the correct allelic state can usually be determined by analyzing the number and relative intensity of bands in the "stutter" ladder. Mis-identification of allelic states, however, usually leads to an excess of homozygotes and artificial departures from Hardy-Weinberg equilibrium in population studies.

Null alleles: Existing alleles that are not observed using standard assays (null alleles) are well-known at allozyme and VNTR loci. Null alleles occur at microsatellite loci as well. Unidentified polymorphism (base substitutions or deletions) at primer sites may cause some microsatellite alleles to amplify poorly or not at all. In some instances, the presence of null alleles may be detected as departures from classical Mendelian inheritance, or heterozygote deficiency as predicted from Hardy-Weinberg expectations. Once realized, the problem may

be resolved by re-designing the primers from flanking sequence. Null alleles, however, probably go undetected in most studies, especially in population surveys, as information on the segregation and transmission of alleles is usually not available. Furthermore, departures from Hardy-Weinberg equilibrium may also be caused by other factors, such as PCR "stuttering" and population substructuring. Therefore, homozygote excess may not always be obviously attributable to the presence of non-amplifying alleles.

Null alleles have been identified in deer (Pemberton *et al.*, 1995), Scallops (Herbinger, pers. comm.), human microsatellite loci (Phillips *et al.*, 1991; Callen *et al.*, 1993), and, relatively recently, several fish species (reviewed in O'Connell and Wright, 1997). In studies involving parentage determinations, an effective strategy for dealing with null alleles is to use only those loci that are typed as heterozyous in a given offspring (Pemberton *et al.*, 1995).

Future developments

Digital recording, data management, and automation: The short length of microsatellite loci and the ability to readily separate microsatellite alleles on denaturing polyacrylamide gels, allows for very accurate size determination. Unknown alleles can be assigned a precise length in bases arrived at by comparisons with allelic ladder standards (Puers *et al.*, 1994) and DNA sequence derived from a known source, such as M13 (see Figure 1.6). These integral values may then be used for digital storage and comparisons of large numbers of individuals.

The reproducibility and accuracy with which microsatellite alleles can be routinely separated has made possible automated detection and scoring (Edwards *et al.*, 1991). Two types of systems are available for microsatellite assays: Automated DNA sequencers (for example, ABI's 373A and Licor's Infrared systems), and fluorescent imaging devices (for examples, Molecular Dynamic's FluorImager™, Biorad's Fluor-S™ multiimager, and Hitachi's FMBIO™). All of these systems offer many of the following benefits over conventional methods of assaying microsatellite variation: (1) non-radioisotopic visualization of alleles, (2) greater throughput (discussed below) (3) reduced labour and laboratory consumable costs, (4) capability of dual or multiple wavelengths and resulting ability to run internal size standards, (5) increased ability to multiplex loci, and (6) automated scoring of allele sizes. The ABI system has been on the market for several years, and has been successfully used in numerous surveys of microsatellite length variation (Edwards *et al.*, 1991; Urquhart *et al.*, 1995), whereas other systems have been only recently developed.

DNA typing with tetranucloetide repeats: In addition to the higher mutation rates and levels of polymorphism at tetranucleotide microsatellite loci compared to dinucleotide and trinucleotide microsatellite loci (Weber and Wong, 1993; Armour *et al.*, 1994), there are several reasons why tetranucleotide microsatellites are becoming increasingly popular markers in paternity and forensic investigations in humans. First, PCR "stutter" at microsatellite loci tends to decrease with increasing repeat length (Edwards *et al.*, 1991; Figure 1.6).

This all but eliminates the possibility of mis-identifying alleles separated by single repeat units as homozygotes. Also, the greatly reduced "stutter" can facilitate automated scoring of alleles. Currently, some automated systems cannot accurately assign alleles at many dinucleotide loci because of the presence of "stutter" bands. Second, the minimum four base pair difference typically observed between alleles at tetranucleotide microsatellite loci makes comparisons across multiple lanes easier (see Figure 1.6), and requires fewer allelic size standards for estimating allele size.

For many applications, tetranucleotide microsatellites offer the opportunity to survey variation using much simpler, less labour intensive methods. The larger difference between alleles, and near absence of shadow bands, has permitted the resolution of length polymorphism in native polyacrylamide gels (Santos *et al.*, 1993). A still simpler, more efficient approach would be to size fractionate alleles using newly developed metaphore agarose (from FMC BioProducts) which can resolve 4 base pair differences between products 200-300 base pairs in length. Occasionally, PCR yields derived from amplification of microsatellite loci are too low to be detected by standard ethidium bromide staining. Several alternative strategies have been devised to circumvent this problem. Valdes *et al.* (1993) and Hubert *et al.* (1992) have modified PCR procedures to yield sufficient product for standard detection by ethidium bromide staining. Santos *et al.* (1993) have employed relatively simple, highly-sensitive silver staining techniques to detect alleles. Also promising is the recently available SYBR™ Green nucleic acid stain (Molecular Probes, Inc) which, under

appropriate conditions, is 25 times more sensitive than ethidium bromide. The amplification, separation, and visualization of alleles, using standard denaturing gel electrophoresis and radiolabelling techniques (Sambrook *et al.*, 1989), takes two to three days. Under a system employing post-separation staining of double stranded DNA, allelic variation can be assayed within eight hours. Furthermore, this approach enables genetic surveys with a minimum amount of specialized equipment, without concerns over the handling and disposal of radioisotopes.

Due to the possibility of microheterogeneity (non-standard repeat unit differences), a system employing post-separation staining of double stranded products may not be ideal for population applications. For example, given the limited resolving capability of these methods, a two base pair insertion or deletion in a particular tetranucleotide microsatellite allele may go undetected. In some instances, such alleles could be inadvertently typed as either longer or shorter standard alleles. This approach, however, could be useful in parentage determination studies where all possible parents are available for DNA analysis. Genotypes of the parents could first be assayed using denaturing polyacrylamide electrophoresis to confirm the constancy of the four base pair differences, and to determine the potential pattern of alleles in the offspring. Only one of the three tetranucleotide loci surveyed in our laboratory has exhibited non-integral repeat unit differences.

As mentioned earlier, specific tetranucleotide microsatellites, $(AGCA)_n$ for instance, occur infrequently in the genome. However, there are 32 different four base pair motifs. Collectively, tetranucleotide loci are abundant, occurring in the

genomes of the species examined at nearly 1/2 the frequency of dinucleotide loci (Beckmann and Weber, 1992). Furthermore, procedures have been recently developed to enrich for specific tetranucleotide microsatellites during screening of genomic libraries (Armour *et al.*, 1994).

Multiplexing loci: In most studies to date, microsatellite loci have been PCR-amplified and the products size-fractionated individually. Marked increase in efficiency can be achieved, however, by multiplexing, or co-amplifying multiple microsatellite loci simultaneously in the same PCR reaction (Figure 1.6). By selecting microsatellite loci with alleles of non-overlapping size ranges, up to four loci can be co-amplified, and alleles resolved on a single denaturing polyacrylamide gel using standard labelling or detection systems (Huang *et al.*, 1992; P. O'Reilly, unpublished data). Studies utilizing automated systems have routinely resolved six or seven human microsatellite loci (Urquhart *et al.*, 1995). This is achieved by labelling overlapping microsatellites with different fluorescent tags.

In addition to non-overlapping allele sizes, loci amplified together must also have similar primer annealing temperatures. Furthermore, primer sets must be compatible, exhibiting a minimum of interaction (e.g. primer dimerization). Loci may either be chosen from a pool of available microsatellites or designed specifically to perform with existing compatible marker(s).

In humans, multiplex systems have been used in gene mapping (Huang *et al.*, 1992) and forensic investigations (Edwards *et al.*, 1991; Urquhart *et al.*, 1995). Hexaplex systems developed by Urquhart *et al.* (1995) have combined

matching probabilities of 10^{-8}, comparable to four hypervariable minisatellite loci (Gill *et al.*, 1991). Multiplexed systems should, therefore, provide more than sufficient information to resolve many issues in aquaculture (parentage determinations and pedigree constructions) with a minimum of effort and expense.

In Chapter 2, I discuss attempts to develop PCR based methods for surveying variation at minisatellite loci. Several minisatellites are characterized, and the development of MVR-PCR at one locus is presented. The isolation and characterization of over 180 microsatellite loci is discussed in Chapter 3. Also in this chapter, efforts to amplify 15 $(GT)_n$ dinucleotide, and three tetranucleotide loci via PCR are described, as is the interpretation of allele sizes, and reliability of these markers. In Chapter 4, the development of a primarily tetranucleotide microsatellite based multiplex is reviewed. The overall information content is determined, and the likely utility in population analyses assessed. In Chapter 5, approximately 800 communally reared offspring, and their 12 possible sets of parents, are typed using the four multiplex loci discussed in detail in chapter 4. Over 6,400 parent-offspring transfers of alleles are screened for possible mutations. In this chapter, I also investigate the resolution and accuracy of compatibility-based parentage determination using these loci.

CHAPTER 2: ISOLATION AND CHARACTERIZATION OF VNTR LOCI, AND DEVELOPMENT OF PCR BASED ASSAYS OF VARIABILITY.

INTRODUCTION

One of the early objectives of this research was to develop technology for surveying genetic variation at minisatellite loci in Atlantic salmon. Length variation at some minisatellite loci can be rapidly and efficiently surveyed by standard PCR and simple agarose gell-electrophoresis, followed by ethidium bromide staining (Galvin, *et al.* 1995a). First, individual loci must be isolated and flanking regions sequenced in order to develop primers on both sides of the array. Ideally, loci should possess minimal repeat unit length variation (Jeffreys' *et al.* 1991), and be free of microsatellites or other repetitive DNA's in regions flanking VNTR arrays. Loci not possessing these characteristics could exhibit a continuous size distribution of PCR amplified alleles, necessitating binning practices and complicating interpretation of genetic variation observed.

Development of PCR-based **M**inisatellite **V**ariant **R**epeat (MVR) mapping (Jeffreys *et al.* 1991) requires DNA sequence from only one flanking region, but also from many repeats units as well. Loci with minimal repeat unit length variation or excessive repeat unit sequence variation (at multiple bases) are not suitable for MVR-PCR. Length variation amongst repeats off-sets the "register" between repeat unit types along the ladder of bands of the diploid sequence (Jeffreys *et al.* 1991). Too much repeat unit sequence variation would result in

either excessive "null" positions along the array, or partial band signatures, both of which would complicate interpretation of repeat unit sequence variation.

Bentzen *et al.* (1991) had previously isolated 30 Atlantic salmon clones from an EMBL3 genomic library that hybridized to Jeffreys' 33.15 and 33.6 minisatellite probes. One clone, *Ssa*1, had been analyzed further by Bentzen and Wright (1993), and did indeed contain an approximately 2.9 kb minisatellite array (see Figure 2.1a, array 1), consisting of about 180 repeats of a 16 bp motif (Figure 2.1b). Because of considerable repeat unit length and sequence variation observed within the array, this minisatellite array was not suitable for MVR-PCR analysis. Supplementary work, however, suggested that one or more additional arrays might be present within the 5.6 kb *Sal* I fragment of the original *Ssa* I clone (P. Bentzen pers. comm).

METHODS

Initial Characterization of Minisatellite loci

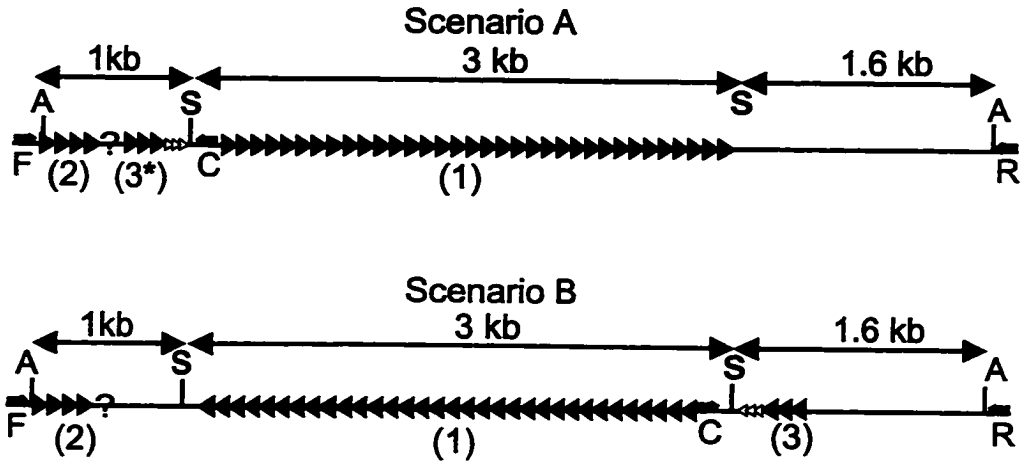
Four of the original 30 clones isolated from an Atlantic salmon EMBL3 genomic library (*Ssa* 1, *Ssa* 7, *Ssa* 10 and *Ssa* 11) were chosen here for further study, in an attempt to identify a locus that was both highly variable, and that exhibited minimal repeat unit sequence and length variation. Given the increased suitability of bacterial as opposed to viral vectors for the kinds of analyses intended here, insert DNA from the original EMBL3 clones was excised

Figure 2.1a. The original 5.6 kb *Sal* 1 minisatellite bearing subfragment identified by Bentzen et al. (1993) was subcloned into pUC 18, and sequenced using forward (F), reverse (R) and internal (C) primers. Two orientations of the 3 kb Ssa1 minisatellite bearing *Sau*3a fragment are possible. Scenario A is suggested by sequence similarity observed between repeat unit of array 2 and repeat units from array 3* (determinied using primer C). Under scenario B, however, the CT rich repeat units of array 3 may be from a third minisatellite from this clone.

Figure 2.1b. DNA sequence dissimilarity between repeat units from array 1 and array 2 suggest that these are two distinct, though closely linked minisatellites. Two repeat units [2(1) and 2(2)] from array 2 are given to demonstrate the considerable repeat unit length and sequence differences within this minisatellite, and to demonstrate the CT rich nature of these repeat units.

Figure 2.1c. Alternative orientations of the 5.6 kb minisatellite bearing *Sal* I fragment within the original EMBL3 *Ssa*1 clone. Scenario A: unique DNA is available on both flanks of array 2 for further sequence analysis and potential development of single locus PCR analyses. Scenario B: unique DNA is available from only one flank of array 2.

a)



b) Repeat unit sequence (concensus)

1) Array 1 CAGGGCTGGTTCTATA (Bentzen and Wright, 1993)
2) Array 2(1) CAGCCTCTCCTCTTCTCTGCCTCTCCTCTTCT
Array 2(2) CAGCCTCTGTTCCTCCTCTCAGACTCTGTTCCTCCTC

c)

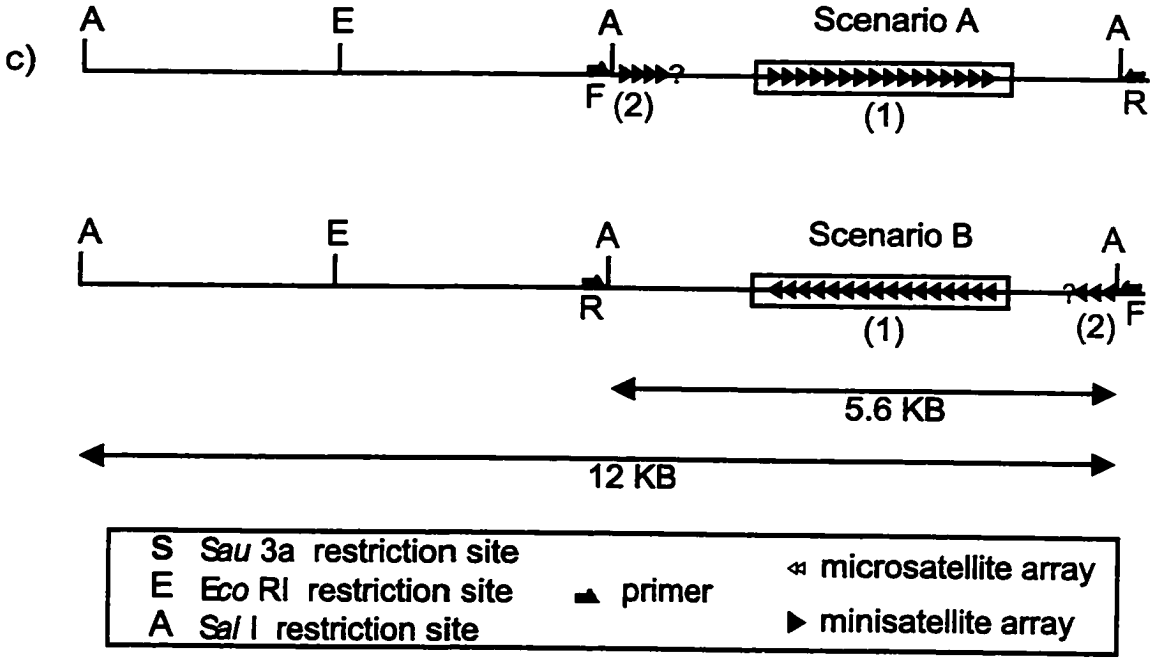


Figure 2.1.

using *Sal* I, and inserted into pUC18, a bacterial plasmid. To increase the ease of handling these sequences further, insert DNA from *Ssa* 7, *Ssa* 10 and *Ssa* 11 was cut with various restriction enzymes, and the resulting fragments "shot gun" cloned into pUC18. Insert DNA, which now ranged from one to four kb, was restricted with additional enzymes, separated by agarose gel electrophoresis, transferred to nylon membranes, and hybridized to Jeffreys' 33.15 and 33.6 minisatellite probes. Multiple fragments from all 3 clones hybridized to both probes, suggesting either multiple arrays ***or*** hybridization to vector sequence. In order to distinguish between these two possibilities, an aliquot of pUC18 cut once by digestion with *Hind*III alone was size fractionated on an agarose gel adjacent to pUC18 treated with *Hind*III and *Eco*RI, to remove the entire polylinker site. DNA was then transferred to a nylon membrane, and hybridized to the two Jeffreys' probes. Only vector with polylinker intact was strongly positive, suggesting contaminating polylinker sequence in the commercially prepared 33.6 and 33.15 probe (CellMark diagnostics). All subsequent experiments required multiple pre-digestion steps to remove polylinker DNA prior to digestion analysis the final desired enzymes. Following this procedure, one fragment from each of the two clones 7(1) and 10(3) still hybridized to Jeffreys' 33.6 and 33.15 probes.

These two fragments were then subjected to ExoIII/Mung bean deletion mapping procedures following Sambrook *et al.* (1989). This permitted DNA sequencing of the both the 1kb 10(3) and the 3kb 7(1) segments. Insert DNA from the original EMBL3 clone *Ssa*1 was first digested with *Sal* I, and the minisatellite bearing 5.6 kb fragment sub-cloned into pUC18 (Figure 2.1a). Large

portions of non-repetative regions of this clone were sequenced using forward, reverse and internal primers.

RESULTS AND DISCUSSION

Characteristics of initial EMBL3 minisatellite loci

Sub clone 7(1) contained an array of at least 30 repeat units of a GGAGGAGGNCT like motif. Considerable length (11-15 bp) and sequence variation was also observed at this locus, thereforestandard PCR ***or*** MVR-PCR assays were not attempted. No minisatellite arrays were observed in subclone 10(3); however, $(AC)_{37}$ and $(AT)_{10}(AT)_{22}$ microsatellites, both with suitable flanking sequence, were identified.

DNA sequencing of *Ssa* 1 using the forward (universal) primer revealed an additional minisatellite array number 2 consisting of primarily two repeat unit types (see Figure 2.1a and 2.1b). Using an internal sequencing primer (C, Figure 2.1a), developed from DNA sequence information from Bentzen and Wright, (1993), additional (CT) rich minisatellite repeat units were also observed (array 3^{*}, Figure 2.1a). This minisatellite was several hundred base pairs from primer C, and was positioned at the end of an $(CT)_5(CCT)_2(CCCT)_4$ microsatellite (Figure 2.1a); due to the distance from primer C and the intervening microsatellite, accurate DNA sequence of the minisatellite repeat units could not be readily obtained. However, the CT- rich nature of repeat units observed using both forward and C primers suggests these two sets of repeat units may represent a single minisatellite array (see Figure 2.1a, scenario A), possibly

having arisen through expansion of the original $(CT)_5$ microsatellite. Given the alternative alignment of the 3 kb *Sau*3A fragment, however, these repeats may be from a third minisatellite locus, positioned within the 1.6 kb *Sau*3A- *Sal* I fragment (Figure 2.1a, scenario B). Considerable length and sequence variation was observed amongst repeat units within all arrays from clone *Ssa*1. Also, the alignment of the 5.6 kb *Ssa*I sub-fragment in the original EMBL3 clone could be such that additional DNA sequence flanking array 2 (and 3?) would not be available without substantially more cloning and screening efforts (see Figure 2.1c, scenario B). For these reasons, this locus was not considered for MVR-PCR, or conventional single locus PCR, and was not studied further.

Minisatellite Variant Repeat Mapping

One minisatellite locus (*Ssa* 197) was serendipitously discovered while sequencing microsatellites isolated using a $(GT)_n$ oligoncleotide probe. Initial DNA sequence of the minisatellite repeat array, and information from microsatellite primer trials (see Figure 2.2) suggested that this locus might be highly suitable for MVR-PCR. Further sequence efforts using PCR primer 197A, revealed very minimal repeat unit length, and suitable repeat unit sequence variation (a C to T transition, equally distributed amongst repeat units within the array; Figure 2.3). Initial MVR-PCR experiments were carried out using a fixed primer flanking the minisatellite array (primer 197F), and primers complementary to the two repeat unit types observed (minisatellite primers G and A). Details of

Figure 2.2. Initial PCR amplification of locus *Ssa* 197 from 8 individuals. The ladder of bands of increasing size are due to unintentional annealing of one of the original microsatellite primers (B) (see figure 2.3) to increasingly distal repeat units.

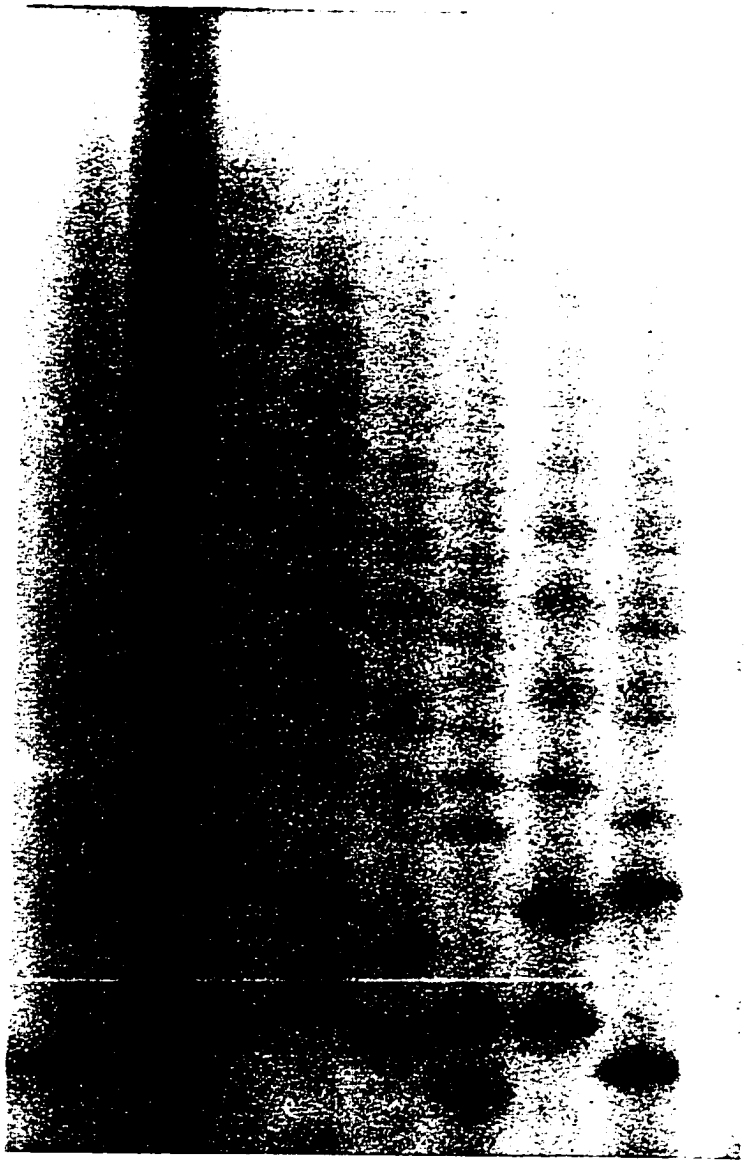


Figure 2.2

Figure 2.3. DNA sequence of the entire *Ssa* 197 clone. Microsatellite and minisatellite repeat units are indicated. Primers 197A and 197 B are the original microsatellite primers developed at this locus. Additional DNA sequence obtained using 197 A provided further information from which microsatellite primer 197 D could be designed. Primers 197 A and 197 D are currently being used in multiplex PCR discussed in Chapter 4. Primers 197 F, minisat primer A, and minisat primer G were used in initial MVR-PCR analyses. Later MVR-PCR trials were attempted using tagged sequences added to minisat primers A and G (see fig 2.5).

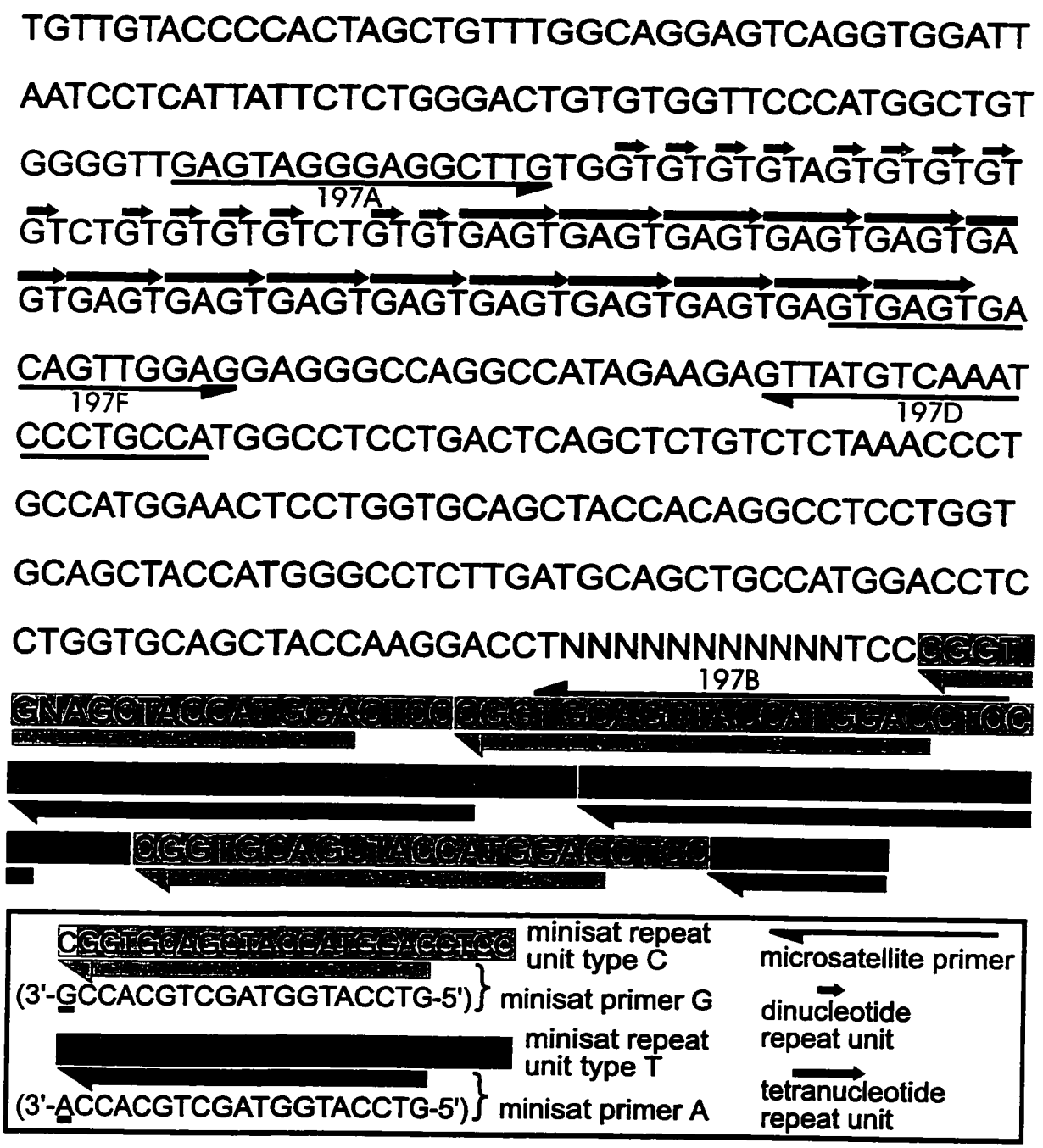

TGTTGTACCCCACTAGCTGTTTGGCAGGAGTCAGGTGGATT
AATCCTCATTATTCTCTGGGACTGTGTGGTTCCCATGGCTGT
GGGGTTGAGTAGGGAGGCTTGTGGTGTGTGTAGTGTGTGT
197A
GTCTGTGTGTGTCTGTGTGAGTGAGTGAGTGAGTGAGTGA
GTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGA
CAGTTGGAGGAGGGCCAGGCCATAGAAGAGTTATGTCAAAT
197F
197D
CCCTGCCATGGCCTCCTGACTCAGCTCTGTCTCTAAACCCT
GCCATGGAACTCCTGGTGCAGCTACCACAGGCCTCCTGGT
GCAGCTACCATGGGCCTCTTGATGCAGCTGCCATGGACCTC
CTGGTGCAGCTACCAAGGACCTNNNNNNNNNNTCCCGGT
197B
minisat repeat unit type C
(3'-GCCACGTCGATGGTACCTG-5') minisat primer G
minisat repeat unit type T
(3'-ACCACGTCGATGGTACCTG-5') minisat primer A
microsatellite primer
dinucleotide repeat unit
tetranucleotide repeat unit

Figure 2.3.

the experimental procedure, a schematic of the MVR-PCR process at this locus, and diploid minisatellite variant repeat maps from two individuals are shown in Figure 2.4. The repeat unit sequence of a single (haploid) molecule, as ascertained by MVR mapping, was confirmed through the first 10 repeats using DNA sequence information.

In MVR mapping, internal priming results in decreased PCR yields of larger products, corresponding to increasingly distal repeat positions (Jeffreys *et al.*, 1991). This is evident in Figure 2.4 where bands much in excess of ten repeat units in length appear increasingly faint. To circumvent this problem, primer tags were used to prevent internal priming (see Figure 2.5). The 20 bp tag extension was designed to minimize within and between primer dimerization. The use of tagged primers here, however, did not increase PCR yields of larger products. Several variations in experimental conditions were attempted (e.g., $MgCl_2$ concentrations, annealing temperatures, primer concentrations, etc.) but did not have a positive, noticeable effect. To date, only two different diploid MVR profiles at *Ssa* 197 have been observed in approximately 12 individuals from two populations surveyed. MVR mapping of the three human loci surveyed thus far has demonstrated that the vast majority of repeat unit sequence variation can be restricted to one end of an array (Jeffreys *et al.* 1994). Variation at this locus too may be concentrated at one (the opposite) end of the minisatellite terminus. Another attempt at redesigning the primer tag could greatly increase the efficiency of PCR for larger products, and the amount of information that might be obtained from this locus. Alternatively, MVR-PCR could be attempted from the

Figure 2.4. Minisatellite Variant Repeat (MVR) mapping of the Atlantic salmon minisatellite locus *Ssa*197. Panel **a** shows the methods used to determine the sequence of the two variant repeat unit types (C and T) at one end of the minisatellite array. Two PCR amplification reactions are performed in separate experiments. In the first reaction (panel **a** (i)), PCR is performed using primer f, located to the left in unique flanking DNA, and primer c (complementary to repeat unit type C). In the second reaction (panel **a** (ii)), PCR is performed using primer f and primer t (complementary to repeat unit type T). In both reactions, PCR products corresponding in length to the distance between the flanking primer and the repeat unit primer (c or t) are produced. Longer products are generated for increasingly distal repeat units. In panel **b**, reaction products from experiments **(i)** and **(ii)** are resolved in separate, adjacent lanes (C and T, respectively), using polyacrylamide electrophoresis. A single heavy band in lane C, position one (individual 1) is interpreted as a homozygous state for repeat unit type C (see also panel a (i)). The presence of equally intense bands in both C and T lanes at positions two through four, indicate heterozygous states (see also panel a (i) and (ii)). A diploid code, representing the state of each repeat unit position, may be read directly from the MVR profile in panel **b**. Positions homozygous for type C repeats are designated the diploid code 1, positions homozygous for type T repeats 2, and positions heterozygous for C and T repeat types, 3. The diploid code for repeats 1 through 10 of individual one is thus 1,3,3,3,3,1,3,3,1,1 (see also panel **a**), and individual two 1,2,2,2,2,1,2,2,1,1.

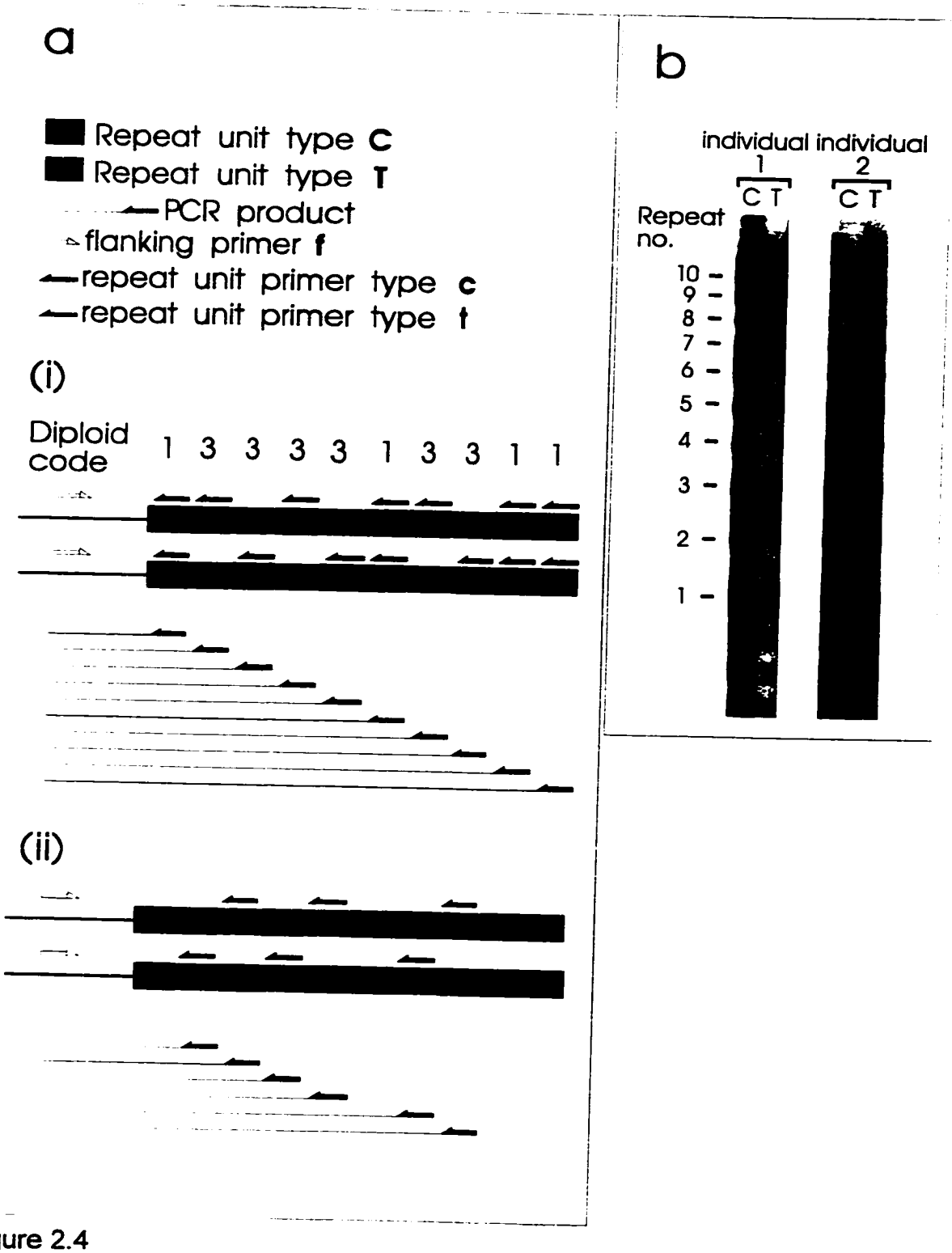

a
Repeat unit type C
Repeat unit type T
PCR product
flanking primer f
repeat unit primer type c
repeat unit primer type t
(i)
Diploid code
1 3 3 3 3 1 3 3 1 1
(ii)
b
individual 1
individual 2
C T
C T
Repeat no.
10
9
8
7
6
5
4
3
2
1

Figure 2.4

Figure 2.5. Simplified schematic of MVR-PCR using tagged primers to prevent internal annealing. Here, only a single (haploid) molecule is analyzed, at a single repeat unit type, (T). Initial PCR is carried out using flanking primer (F) and 'A' type tagged primers. After only several cycles, PCR products corresponding in length to each repeat unit position are generated. These products terminate in a unique 'tag' sequence. Subsequent cycles are carried out using flanking primer (f) and 'tag' primers. Since complementary sequence occurs only at the end of each molecule, internal priming (and product shortening) is prevented. This methodology was adapted from Jeffreys et al. (1991).

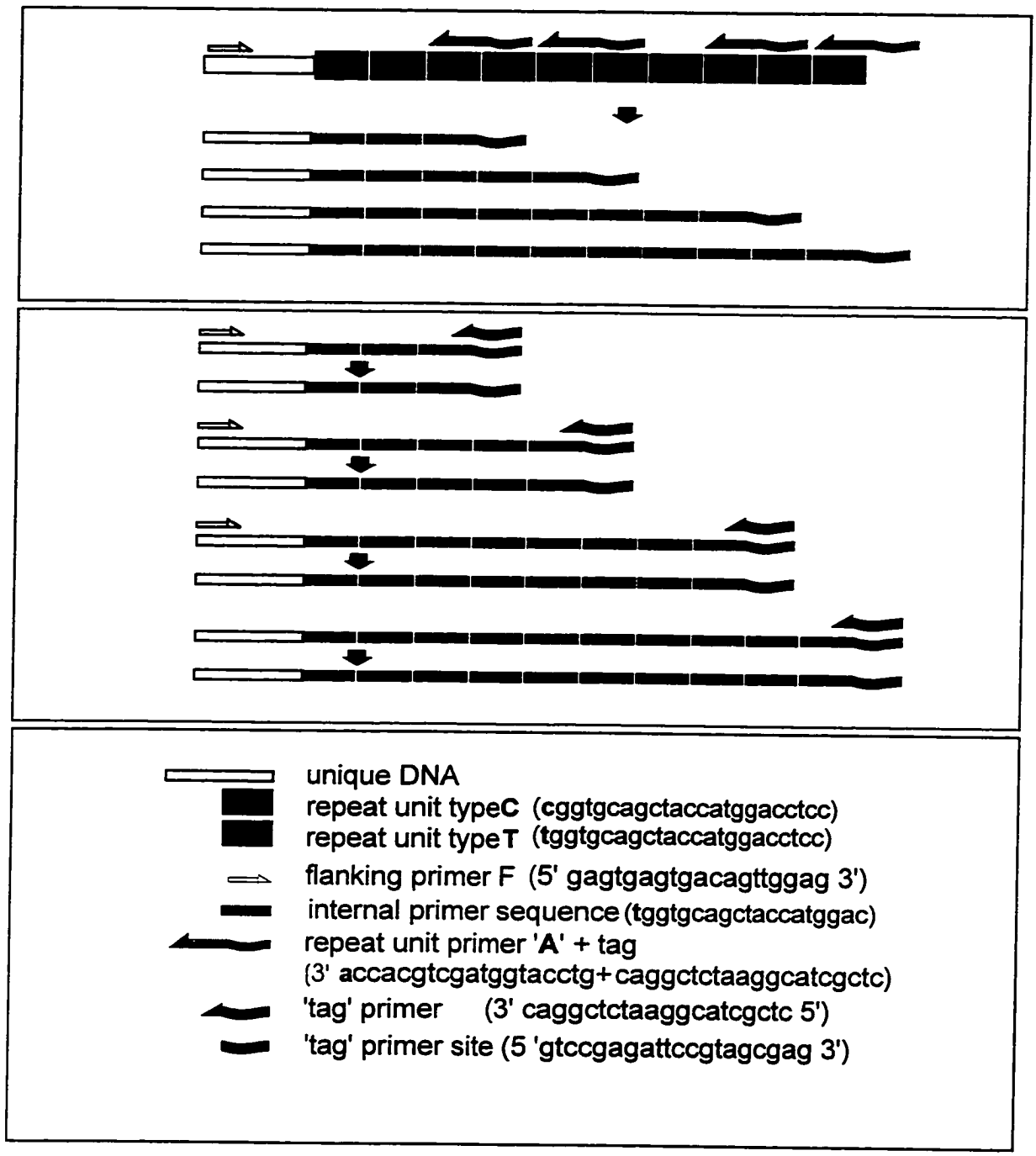

unique DNA
repeat unit typeC (cggtgcagctaccatggacctcc)
repeat unit typeT (tggtgcagctaccatggacctcc)
flanking primer F (5' gagtgagtgacagttggag 3')
internal primer sequence (tggtgcagctaccatggac)
repeat unit primer 'A' + tag
(3' accacgtcgatggtacctg+caggctctaaggcatcgctc)
'tag' primer (3' caggctctaaggcatcgctc 5')
'tag' primer site (5 'gtccgagattccgtagcgag 3')

Figure 2.5.

other end of the minisatellite array. This, however, would require considerable additional cloning and sequencing effort.

CHAPTER 3: ISOLATION AND CHARACTERIZATION OF MICROSATELLITES

INTRODUCTION

Given the growing interest and acknowledged utility of microsatellite genetic markers, and the amount of progress made isolating minisatellite loci suitable for PCR-based assays of variation, the decision was made at the end of my first year of study to focus on the isolation and application of microsatellites from Atlantic salmon. Large numbers of microsatellites can be isolated and characterized in a small fraction of the time required to develop several minisatellite loci. Typically, a genomic library consisting of 300-600 bp inserts is probed with multiple tandem copies of a desired repeat motif (i.e $(GT)_{12}$). Positive clones are then isolated and sequenced. Due to the limited size of the genomic DNA selected, most inserts can be completely sequenced using forward and reverse primers only, without the need of sub-cloning, deletion mapping, or other lengthy procedures. Primers flanking microsatellite arrays can designed, and allele length assayed by standard PCR followed by denaturing polyacrylamide electrophoresis, and autoradiographic detection.

By far the most commonly used class of microsatellites are those with a two base pair repeat motif, typically (GT) or (CT) (Jarne and Lagoda, 1996). The heavy dependence on dinucleotide microsatellites for gene mapping applications is in large part due to their great abundance and apparently even distribution in mammalian genomes surveyed to date (up to one every 30 kb, Beckmann and Weber, 1992). Because of their abundance and hence relative ease of isolation,

these loci have also been extensively employed in population analyses and relatedness determinations in many species.

One of the major limitations of most dinucleotide microsatellites is the difficulty in accurately and reliably scoring some alleles due to PCR induced complications (discussed in detail in chapter 5), and the tendency of upper alleles (visualized as bands on X-ray film following autoradiography) at loci with longer GT arrays to become diffuse. This second shortcoming is due to re-annealing of complementary strands, often during electrophoresis (see Litt *et al.* 1993). Both limitations are less of a concern in gene mapping studies, where researchers must distinguish between, at most, four possible alternate allelic states among offspring. In parentage analyses and population studies, however, there may be many (dozens) of possible allelic states. Also, in the latter case, pedigree information is rarely available to help resolve ambiguous allele size determinations. Hence, much care should be taken in selecting dinucleotide loci that are to be used for applications other than genome mapping. Alternatively, many researchers are using tri or tetranucleotide microsatellites, which exhibit considerably less PCR-related "stutter" (Strassmann *et al.* 1996, 1997; Edwards *et al.* 1991; Sprecher *et al.*, 1996). Also, the upper alleles at these loci do not typically become as diffuse compared to those of long tract $(GT)_n$ loci (personal observations). This is not surprising given the increased sequence complexity of tetranucleotide arrays, and the well-known relationship between annealing rates and sequence simplicity.

The objective of this research was to isolate and characterize a large number of microsatellite loci from Atlantic salmon for development of molecular genetic markers suited for (1) determining relatedness in aquaculture and studies of social behaviour, (2) addressing population issues important in fisheries management and conservation, and 3) studying rates and modes of mutation of microsatellites in Atlantic salmon.

METHODS

Library construction and microsatellite screening procedures

Two Atlantic salmon genomic libraries were constructed; the first consisted of approximately 1 X 10^4 colonies, and the second 4 X 10^4 colonies. In the construction of both libraries, 50 micrograms of Atlantic salmon genomic DNA (provided by earlier researchers at the M.G.P.L.) was digested with the restriction endonuclease *Sau*3A and size fractionated by electrophoresis through 1 % agarose and 1X TAE running buffer. Insert DNA (10-20 micrograms) in the range of 300-600 bp was excised, and subsequently extracted from agarose blocks by centrifugation through filter paper columns as described in Chuang and Blattner (1994). Size-selected genomic DNA was extracted twice with phenol, once with chloroform, and percipitated with ethanol following Sambrook *et al.* (1989). Purified DNA was then quantified by comparing various dilutions with standards at several concentrations size fractionated in adjacent lanes of an agarose gel, and visualized using ethidium bromide staining. *Sau*3A generated fragments were then ligated to dephosphorilated *Bam*HI pUC18 vector

(Pharmicia), using a vector to insert ratio of ends of 2:1, 1:1 and 0.5:1, in three separate experiments. The ligation mixture was then phenol extraced once, chloroform extracted once, and ethanol precipitated. Purified ligation mixture was then quantified as described above. Approximately 50 ng (5 ul total volume) of each of the three ligation mixtures was used to transform 200 ul of DH5 α super competent cells, following the suppliers' protocol (Gibco BRL). An aliquot of each of the 3 preparations of transformed cells was plated so as to achieve a fairly high colony density of approximately 500-1000 per 130 cm^2 plate. Colonies were lifted onto "Hybond N" nylon membranes as directed by the supplier, Dupont. Membranes were prehybridized in 50 ml of hybridization solution as described by Westneat *et al.* (1988) for 2-3 hours. In the first library, $(GT)_{12}$ probe was 5' end-labelled using Pharmacia's T4 polynucleotide kinase. Hybridization was carried out overnight in the same solution at 62 ° C. The following day, membranes were washed once for 15 minutes at 42 ° C in 2 X SSC (0.3M NaCl and 0.03 M sodium citrate)- 0.2% SDS, and twice at 30 ° C in 2 X SSC for 15 minutes each. By minimizing the stringency and duration of the washes, non-microsatellite bearing colonies exhibited faint background signal, while $(GT)_n$ bearing clones were strongly positive. The faint background on non-target colonies was very important in increasing the accuracy with which GT microsatellite clones were isolated. Once identified, positive colonies were picked using sterilized P10 pipet tips; tooth picks, inoculation loops or other similarly course instruments were not suitable at this higher density. Positive clones were then sequenced using forward and reverses primers. DNA sequence was size

fractionated on 8 % polyacrylamide wedge gels. Primers were designed by GSD oligos LTD., and by myself using GeneRunner software.

In the second larger library, membranes were screened using a $(GTCA)_8$ probe as described above, except that hybridization was carried out at 58 ° C. The (GTGA) motif was selected because this class of microsatellite had been observed in the 1st library screened, and the motif can be created by a single base transition of a T to an A in the very common $(GT)_n$ class of microsatellites. Colonies were screened at 58 ° C instead of 62 ° C in response to findings of earlier experiments, where plates containing colonies bearing (1) $(GTGA)_{17}$ microsatellites, (2) short tract $(GT)_n$ arrays, (3) long tract $(GT)_n$ arrays, (4) minisatellite arrays, and (5) no detectable repeat unit sequence, were hybridized to $(GTGA)_8$ probe under various conditions. At 62 ° C, long tract $(GT)_n$ colonies hybridized more strongly to $(GTGA)_8$ probe than did (GTGA) colonies. The best signal-to-noise ratio was achieved using, surprisingly, a less stringent hybridiztion temperature, 58 ° C. All subsequent steps were carried out as described for the first library.

Analysis of microsatellite flanking sequence

Prior to the selection of loci for consideration as candidates for PCR amplification, flanking regions were analyzed for the presence of one or more factors that may influence the reliability and robustness of PCR based assays, and interpretation of microsatellite allele sizes. Many $(GT)_n$ microsatellite loci are associated with repetitive elements in Atlantic salmon (W. Davidson, pers.

comm). The presence of one or both primers in a highly repeated, fairly conserved stretch of DNA could result in a multilocus type profile, and reduced amplification product of the target sequence (Edwards *et al.* 1991). I attempted to identify such occurrences by comparing all flanking sequences to the EMBL3 database, which included many types of repeated DNA from Atlantic salmon and other salmonids. In addition, multiple segments of each clone were compared to the remaining approximately 200 clones isolated here to identify microsatellite loci associated with repeat DNA not submitted to the EMBL database, and to identify clones that were inadvertently sequenced twice, but treated as unique.

Another concern is the presence of repeated and potentially cryptic satellite or minisatellite DNA immediately flanking microsatellite arrays. The occurrence of one of the flanking primers in a larger tandemly-repeated sequence can result in a ladder of bands above the main PCR product. In addition to using up primer and other PCR components in the targeted reaction, the smaller alleles from upper (more distal) minisatellite repeat units can obscure longer alleles from the intended, proximal repeat (personal observations). This was minimized by performing within clone checks for sequence duplication: every second 10 bp block of sequence was compared with the remaining sequence from a given clone.

RESULTS

Library I.

Abundance of GT(n) microsatellites

Of the approximately 10,000 colonies screened in the first library for $(GT)_n$ microsatellites, 203 positive clones were identified. Based on DNA sequence analysis, 180, or approximately 89%, contained microsatellite arrays with 5 or more GT repeats (Figure 3.1). Given an average insert size of 400 bp (deduced from DNA sequence of the 203 positive clones), and the total number of colonies screened, $(GT)_n$ microsatellites occur at an approximate frequency of 1 in 20 kb of DNA. The average insert size of clones screened was also estimated using a second approach. Here, several dozen of the original 10,000 colonies were randomly selected. Inserts were PCR amplified using forward and reverse primers, and sized using agarose gel electrophoresis. Average insert size was considerably less than previously determined, approximately 150 bp. The frequency of $(GT)_n$ microsatellites using this revised estimate of insert size is approximately 1 every 8 kb. Assuming a haploid genome size of approximately 2.0×10^9 (Hinegardner and Rosen, 1972) , there is an estimated 500,000 $(GT)_n$ microsatellites in the Atlantic salmon genome.

Microsatellites array type and length

Of the 180 microsatellites observed, 129 were perfect, 47 imperfect and 4 compound. Array length ranged from fewer than 10 to over 90 repeats (Figure 3.2). The shortest size class (10 or fewer repeats), was most common, followed

Figure 3.1. DNA sequencing autoradiograph of 18 positive clones, giving examples of compound, perfect, imperfect, short array (<12 repeats), and long array (>50 repeats) microsatellites. The bases A, C, G, and T were loaded left to right.

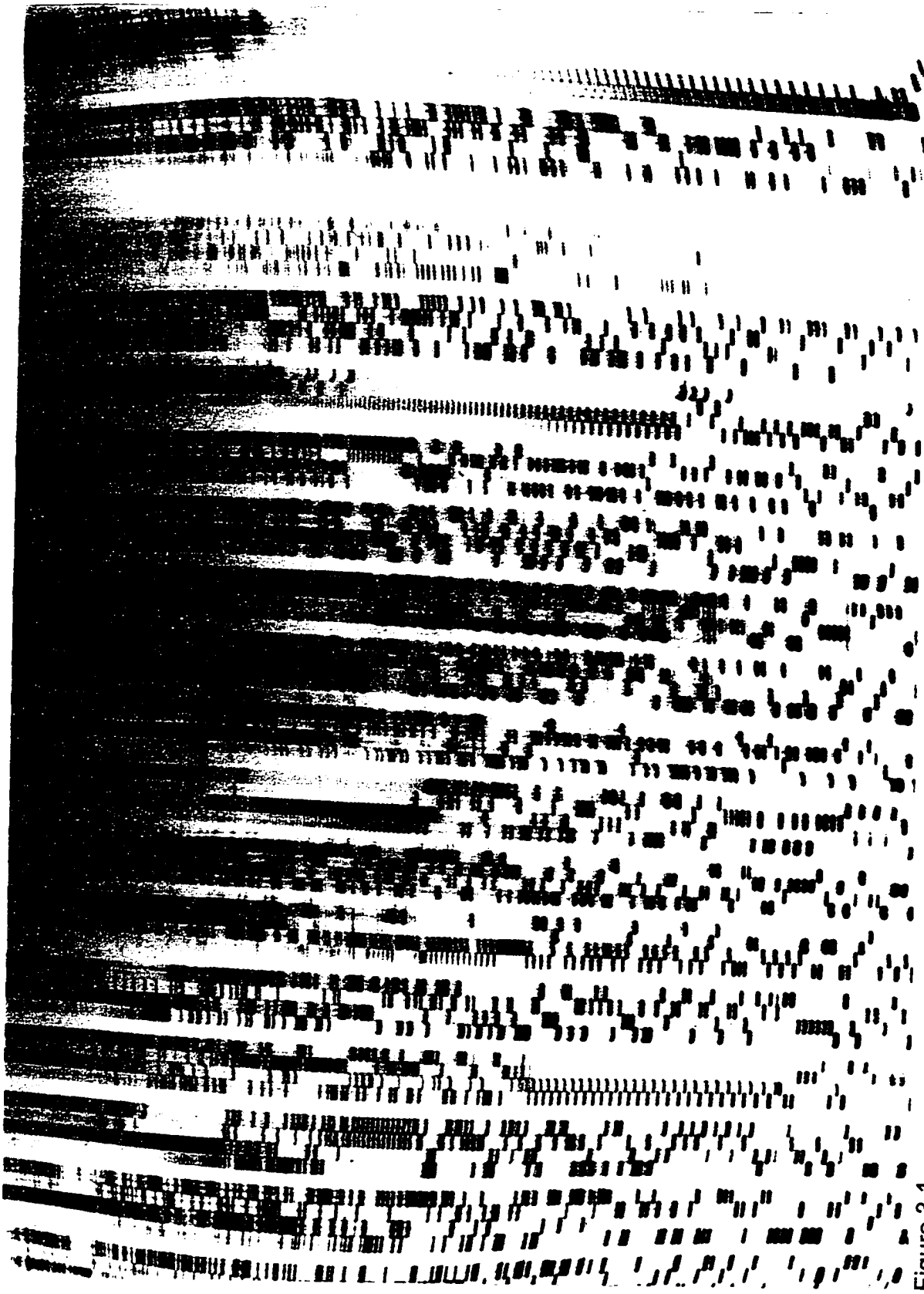


Figure 3.1

Figure 3.2. Frequency histogram of array size classes of microsatellites isolated from Atlantic salmon.

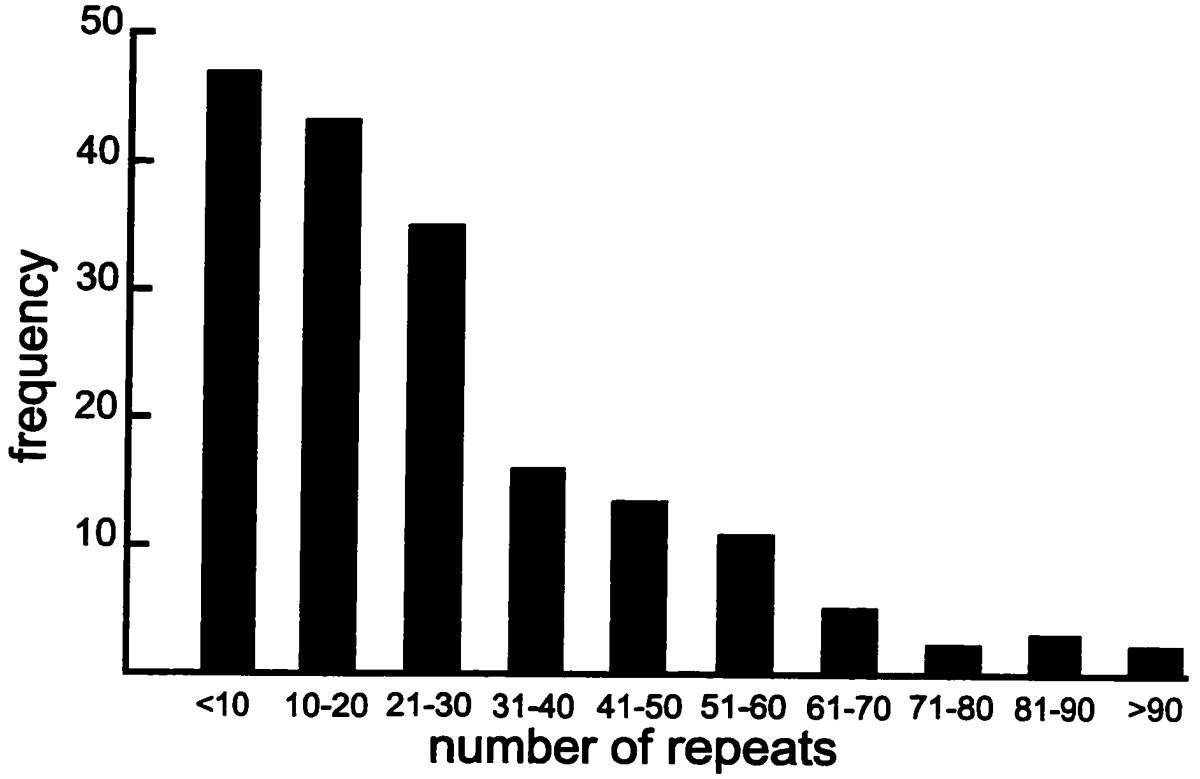

50
40
30
20
10
frequency
<10
10-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
>90
number of repeats

Figure 3.2.

closely by the second smallest class. Each subsequent class was less common than the previous for all but the longest arrays.

Locus characteristics and suitability for PCR based assays of variation

Not all dinucleotide microsatellite loci identified were suitable candidates for PCR based assays of variability (Figure 3.3). For forty seven loci isolated, arrays were comprised of too few repeat units (10 or less) to be expected to be variable. Suitable primer sequence was unavailable for a further 41 loci. Another 6 were likely embedded in segments of highly repeated DNA identified from Atlantic salmon or other salmonids (Table 3.1; Figure 3.4), and two sets of loci were either duplicates, or from highly conserved and not yet published repeat DNA. Also, one microsatellite clone contained a minisatellite array immediately flanking the GT(n) array. The remaining 81 microsatellites (~40% of the total number screened) were potentially informative and amenable to PCR amplification.

Primers were developed and tested for 15 dinucleotide microsatellite loci (Table 3.2). PCR products interpreted as amplification of the targeted microsatellite arrays were observed for 12 of the 15 loci tested (see Figure 3.5). Array sequence, number of loci amplified, and variability are given in Table 3.3. Diploid states were observed at 13 of the 15 loci surveyed; two sets of loci were amplified using primers derived from DNA sequence at *Ssa* 119 and *Ssa* 120. At only two microsatellite loci was no variation observed in the 8 individuals assayed. In total, 8 of the 15 dinucleotide loci appear to be suitable molecular

Table 3.1. Sequence similarity of Atlantic salmon microsatellite clones with repetitive DNA deposited in Genbank.

Clone	Database sequence identity	Percent sequence similarity
Ssa 2	Fok (SLL-1)[1] (white spotted charr)	79
Ssa 33	S.I.N.E.[2] (Coho salmon)	95
	S.I.N.E.[2] (kokanee salmon)	94
	Fok(SLL)-6 gene[1] (white spotted charr)	86
Ssa 52	TC1-like transposon[3] (Atlantic salmon)	89
Ssa 127	TC1-like transposon[3] (Atlantic salmon)	86
Ssa 151	TC1-like transposon[3] (Atlantic salmon)	85
Ssa 199	TC1-like transposon[3] (Atlantic salmon)	85

1 Murata et al. (1993)
2 Kido et al. (1991)
3 Goodier and Davidson (1994)

Figure 3.3. Schematic showing microsatellite locus development pyramid. The bottom tier represents the total number of positive (putative microsatellite bearing) clones sequenced. Reduction in size of increasingly upper levels represents reduction in the proportion of potentially useful microsatellite genetic markers. Percentages are calculated with respect to the bottom level, the total number of positives sequenced. Values given in levels above the dotted line are based on the 15 dinucleotide loci for which PCR was actually attempted; due to the cost of synthesizing primers, many potential dinucleotide microsatellites were not analyzed further.

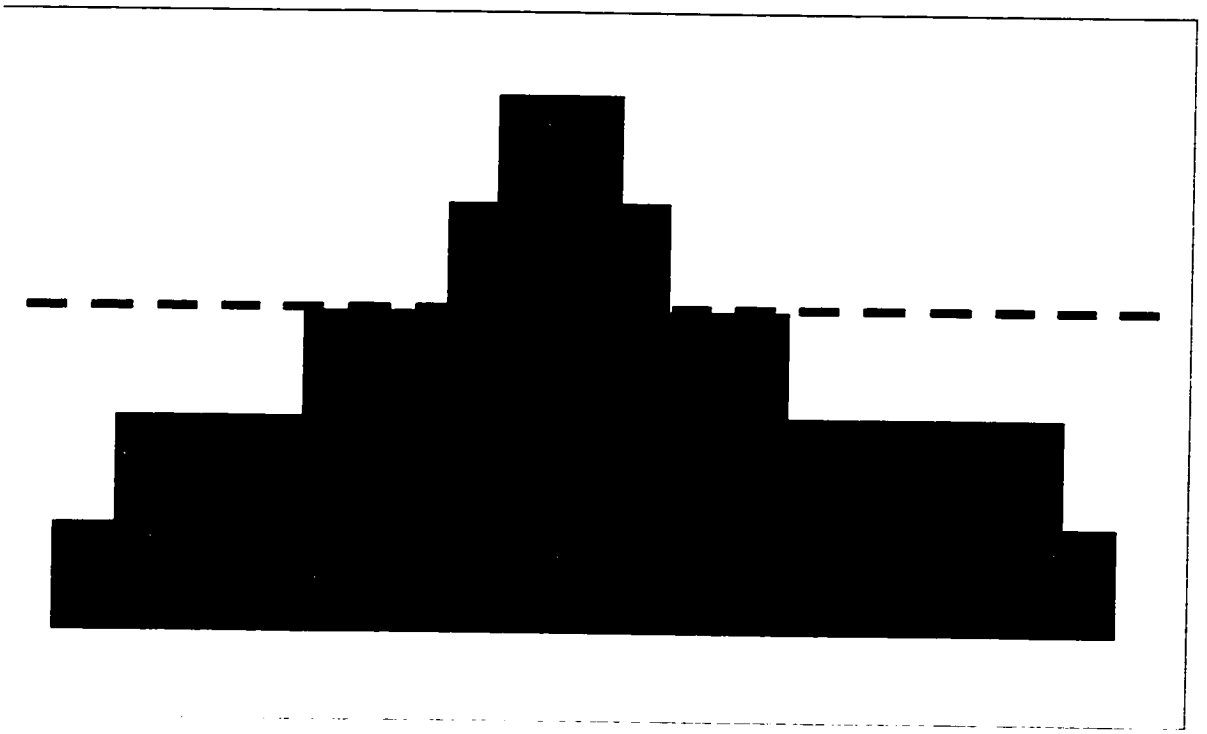


Figure 3.3.

Figure 3.4. Alignment of four *Ssa* clones islolated here with the Atlantic salmon *Ssa*1 TC1-like transposon clone (Goodier and Davidson, 1994). No microsatellite array was observed in clones *Ssa* 52 or *Ssa* 127. An approximately 200 bp segment of *Ssa* 127 was tandemly duplicated, though the second repeat was separated from the first by approximately 100 bp of unique DNA. All clones except *Ssa* 52 contain large blocks of unique DNA in addition to TC1 sequence. Positions of ITRs (35 bp inverted terminal repeats) and IPRs (27 bp inverted internal repeats) are given.

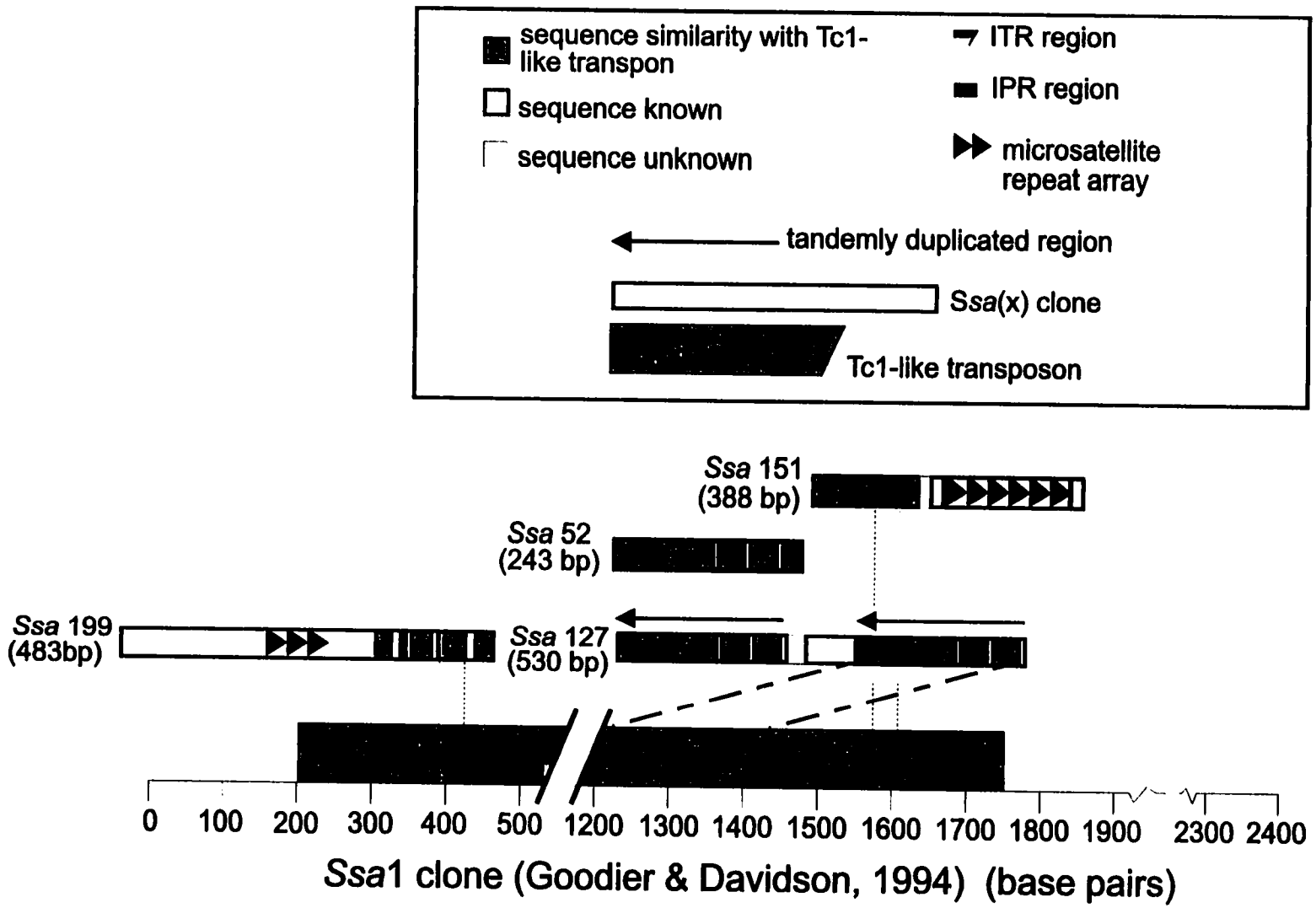

sequence similarity with Tc1-like transpon
sequence known
sequence unknown
ITR region
IPR region
microsatellite repeat array
tandemly duplicated region
Ssa(x) clone
Tc1-like transposon
Ssa 151 (388 bp)
Ssa 52 (243 bp)
Ssa 199 (483bp)
Ssa 127 (530 bp)
0 100 200 300 400 500 1200 1300 1400 1500 1600 1700 1800 1900 2300 2400
Ssa1 clone (Goodier & Davidson, 1994) (base pairs)

Figure 3.4.

Figure 3.5. Representative dinucleotide microsatellite loci developed from Atlantic salmon. *Ssa* 12 (a) is a short array microsatellite locus of low variability that exhibits minimal PCR stutter, and sharp well-resolved bands. The array length at *Ssa* 85 (b) is slightly longer. Stutter ranges from minimal to moderate (increasing with increasing allele length), and bands are generally sharp, and well resolved. Array lengths at *Ssa* 03 (c) and *Ssa* 76 (d) are considerably longer. The amount of stutter at these loci is moderate, and bands, especially for the larger alleles, are very blurry and not well resolved.

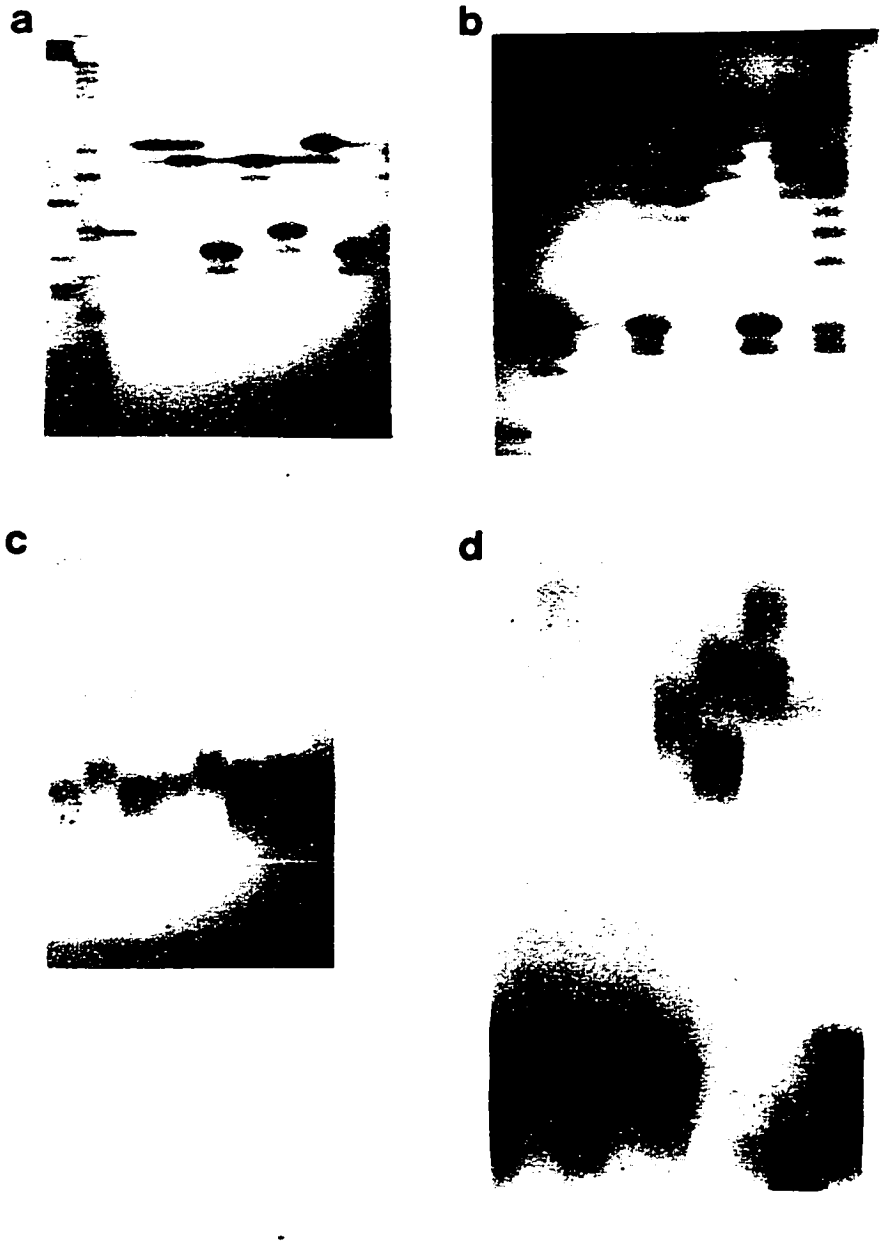


Figure 3.5

Table 3.2. Primer sequences and PCR amplification conditions of Atlantic salmon dinucleotide microsatellites.

Locus	Primers (5' to 3')	$MgCl_2$ (mM)	Annealing temp. (C)
Ssa 03	(a) aatttcttggaggtcactacagtc (b) acaacgagagacaaaccaacag*	1.5	59
Ssa 12	(a) gaatctctcctttgtctgc* (b) catccgacaacagcactg	1.5	55
Ssa 22	(a) tcacacccttctgctctctt (b)tcaatcaatagaaccataggct*	1.5	57
Ssa 49	(a) tcaatcaatagaaccataggct* (b) ttcaatctcagtttacctccct*	1.0	58
Ssa 58	(a) aatccctctcctggcttactt (b) cttcttcttctcttcccctcc*	1.0	55
Ssa 61	(a) gccacctactacctacaagcc (b) aggtaggggtcagaggcaa*	1.5	58
Ssa 66	(a) accaccataggacacaataatg (b) acggattacagatgtctacg*	1.5	55
Ssa 73	(a) atacattcagcagtccatt (b) tgacaggcttattatccac*	1.5	55
Ssa 76	(a) agagtcccagaaatacaggca (b) gactcatctcccaaccaactg*	1.0	60
Ssa 85	(a) ggtgggtcctccaagctac (b) acccgctcctcacttaatc	1.5	58
Ssa 86	(a) tagatgaagcgtgtgcgtg (b) tgaatgctactgatgccttg*	1.5	56
Ssa 119	(a) tgcactatccagggaacagg (b) ataggagattggacagtcatttc*	1.5	59
Ssa 120	(a) aaagggtatcggtagagaagtcaa* (b) acctcttgctttacctcttcctc	1.5	60
Ssa 126	(a) cttcatctctcccctctcc (b) tactgcctatgttggtctgag*	1.5	55
Ssa 129	(a) tgtattgtgacgttcagactgc (b) gaacagttccattcaatttgtaaa*	1.0	57

*labelled primer
Annealing Temp. (C) = Annealing temperature (Celsius)

Table 3.3. Locus characteristics of Atlantic salmon dinucleotide microsatellites.

Locus	Sequence (array)	Number of loci amplified	Variation (# alleles)*	Comments
Ssa 03	$(GT)_{38}C(TG)_4$ $TC(TG)_2$	1	> 5	alleles blurred
Ssa 12	$(GT)_{12}$	1	3	alleles distinct
Ssa 22	$(TG)_{18}$	1	3	alleles distinct
Ssa 49	$(GT)_{25}$	1	>5	alleles blurred
Ssa 58	$(TG)_{18}$	1	>5	alleles blurred
Ssa 61	$(TG)_{25}$	unknown	n/a	uninterpretable (multiple bands)
Ssa 66	$(TG)_{25}$	none	n/a	no visible PCR products
Ssa 73	$(GT)_{13}$	1	1	alleles distinct
Ssa 76	$(GT)_{28}$	1	>5	alleles blurred
Ssa 85	$(GT)_{14}$	1	4	alleles distinct
Ssa 86	$(GT)_{20}$	1	4	alleles distinct
Ssa 119	$(GT)_{19}CC(GT)_1TCC$ $(GT)_6CCA(TG)_3$	2	4	alleles distinct
Ssa 120	$(TG)_{25}$	2	4	alleles distinct
Ssa 126	$(GT)_{13}$	none	n/a	no visible PCR products
Ssa 129	$(GT)_{15}C(TG)_1TC(TG)_{10}$ $C(TG)_4C(TG)_{13}$	1	1	alleles distinct

* number of alleles based on assays of eight individuals from a single population.

genetic markers for certain applications. Only four dinucleotide microsatellites are suitable for most applications (variable, alleles distinct, and only a single locus amplified). DNA sequences of microsatellite loci potentially suitable for PCR assays of genetic variation have been deposited in Genbank (Table 3.4).

Library II.

Isolation of Tetranucleotide microsatellites

Of the approximately 4.0 X 10^4 colonies screened in the second library for $(GTGA)_n$ microsatellites, 12 positive colonies were identified. A single clone contained a $(GTGA)_n$ locus (*Ssa* 202) of 14 repeat units in length. The remaining positives contained either long tract $(GT)_n$ loci or no identifiable microsatellite whatsoever. The tetranucleotide loci *Ssa* 171 and *Ssa* 197 were serendipitously found while screening for $(GT)_n$ loci in the first library.

DISCUSSION

Abundance of GT(n) microsatellites

Assuming that the portion of the genome cloned in library I is representative of the entire Atlantic salmon genome, the two estimates of locus frequency calculated here are 1 every 20 kb, and 1 every 8 kb. The former estimate compares well with values reported by Estoup *et al.* (1993) of 1 every 24 kb in brown trout, but is considerably lower than the 1 every 90 kb given by Slettan *et al.* (1993) for Atlantic salmon, using a similar approach. All these

Table 3.4. Genebank submission numbers of microsatellite loci potentially suitable for PCR based assays of genetic variation.

Locus	Genebank accession number	Locus	Genebank accession number
Ssa 01	U58897	*Ssa* 118	U58890
Ssa 03	U58895	*Ssa* 119	U58892
Ssa 09	AF019197	*Ssa* 120	U58896
Ssa 10	U58891	*Ssa* 126	U58893
Ssa 11	U58894	*Ssa* 128	U58898
Ssa 12	U58900	*Ssa* 129	U58899
Ssa 13	U58903	*Ssa* 132	U58901
Ssa 15	AF019149	*Ssa* 138	U58902
Ssa 16	AF019153	*Ssa* 140	U58904
Ssa 17	AF019154	*Ssa* 141	U58905
Ssa 20	AF019162	*Ssa* 142	U58906
Ssa 30	AF019169	*Ssa* 143	U58907
Ssa 31	AF019170	*Ssa* 145	U58908
Ssa 38	AF019172	*Ssa* 147	U58909
Ssa 41	AF019173	*Ssa* 151	U58910
Ssa 42	AF019174	*Ssa* 162	AF019150
Ssa 47	AF019175	*Ssa* 163	AF019151
Ssa 48	AF019176	*Ssa* 164	AF019152
Ssa 50	AF019177	*Ssa* 171	U43693
Ssa 53	AF019178	*Ssa* 185	AF019155
Ssa 57	AF019179	*Ssa* 195	AF019156
Ssa 58	AF019180	*Ssa* 196	AF019157
Ssa 60	AF019181	*Ssa* 197	U43694
Ssa 61	AF019182	*Ssa* 199	AF019158
Ssa 64	AF019183	*Ssa* 201	AF019159
Ssa 65	AF019184	*Ssa* 202	U43695
Ssa 66	AF019185	*Ssa* 204	AF019160
Ssa 73	AF019186	*Ssa* 208	AF019161
Ssa 75	AF019187	*Ssa* 213	AF019163
Ssa 76	AF019188	*Ssa* 216	AF019164
Ssa 81	AF019189	*Ssa* 218	AF019165
Ssa 83	AF019190	*Ssa* 219	AF019166
Ssa 85	U43692	*Ssa* 223	AF019167
Ssa 86	AF019191	*Ssa* 224	AF019168
Ssa 87	AF019192	*Ssa* 333	AF019171
Ssa 91	AF019193	*Ssa* 999	AF019196
Ssa 94	AF019194		
Ssa 98	AF019195		

values are based on insert size determinations of clones selected by hybridization to $(GT)_n$ probe DNA. Since larger inserts are more likely to contain a particular sequence than smaller inserts, $(GT)_n$ bearing clones were probably not representative of the respective library averages. This conclusion is supported by the findings here of considerably smaller insert size of randomly chosen (non-$(GT)_n$ selected) clones. Therefore, the average frequency is probably more on the order of 1 every 8 kb.

Brooker *et al.* (1994) reported an average frequency of 1 every 8 kb in Atlantic Cod. Since calculations were based on non-adjusted insert size, the frequency may be even higher. In humans, the frequency of $(GT)_n$ microsatellites was surveyed in long stretches (hundreds of kilobases) of genomic sequences in EMBL databases (Beckmann and Weber, 1992), and were found, on average, every 30 kilobases. Differences between humans and fish described here may simply reflect the lack of precision in estimating the frequency of microsatellite loci using hybridization-based techniques. However, it is interesting to note that several researchers have reported increased array size in fish compared to that observed in mammals (Slettan *et al.* 1993; Brooker *et al.* 1994; McConnell *et al.* 1995b). Also, an intriguingly high percentage of loci isolated from fish exhibit very high heterozygosity levels (greater than 85%), compared to mammals. For example, 4 of 7 loci assayed in Atlantic cod by Brooker *et al.* (1994), 4 of 5 assayed here in Atlantic salmon (Chapter 4, unpublished data), and 4 of 4 assayed from Gulf pipefish (Jones and Avise, 1997). In humans, 4 of 24 (Armmour *et al.* 1994), and 92 of 810 (Weissenbach *et al.* 1992) exhibited

heterozygosities of greater than 85%. Increased replication slippage, proposed as a possible explanation for larger array lengths in Atlantic Cod (Brooker *et al.* 1994), could also contribute to increased heterozygosity (via mutation), and the increased creation of microsatellite loci in fish.

Locus characteristics and suitability for PCR based assays of variation

Of the 203 positives sequenced, only 81 (40%) were potentially suitable for PCR amplification. Of the 15 dinucleotide loci tested, eight (~50%) amplified a single locus and were variable. As part of a very large effort to construct a second-generation linkage map of the human genome, Weissenbach *et al.* (1992) isolated and sequenced 12,014 similarly isolated $(GT)_n$ positive microsatellite clones. In their study, fewer still (25%) were considered amenable to PCR amplification. Most were excluded for various reasons, including unsuitable primer sequence, too short array length, duplicate sequences, etc.. Of the 2995 assayed for variation using PCR, approximately 1750 (58 %) were variable and amplified a single locus.

Each of the eight dinucleotide loci developed here have unique allele signatures, varying in the number and intensity of "stutter" bands, band resolution, and background signal (see Figure 3.5). In general, two undesirable qualities increase with increasing array length at these loci: the number and intensity of "stutter" bands, and the smearing or blurring of bands. This is readily evident by comparing alleles within a locus, where increasing length is presumably due to additional repeats (see Figure 5a, 5b for examples of

increased "stutter", and 5d for increased blurring or smearing). Also, loci with shorter array lengths (*Ssa* 12, *Ssa* 85) tend to exhibit less PCR induced "stutter" overall. This is not unexpected given observations of increased mutation rates for longer alleles compared to shorter alleles (Talbot *et al.* 1995; Ellegren *et al.* 1996). On the other hand, *in vitro* experiments of slippage synthesis of simple sequence DNA found no relationship between array length and slippage rate (Schlotterer and Tautz, 1992).

A likely major contributing factor to the smearing observed is the increasing amount of simple DNA comprising the longer alleles, and hence greater tendency for re-association of complementary strands compared to shorter alleles. This latter difficulty can be partially remedied by using high concentrations of formamide in the preparation of denaturing acrylamide gels for size fractionation of microsatellite alleles (Litt *et al.* 1993). However, formamide is highly toxic, and additional precautions must be taken during electrophoresis. Because of these considerations, use of formamide denaturing gels is not widely adopted in managing the extent of smearing seen at many dinucleotide loci.

Overall, allele signatures at loci with shorter dinucleotide arrays may be clearer and more accurately typed compared to longer tract loci. However, locus variability is negatively correlated with decreasing number of repeat units (Weber 1990; Goldstein and Clark (1995); P. O'Reilly, personal observations). The lack of any relationship between array length and variability reported by Valdes *et al.* (1993), is likely due to several biases inherent in their experiment, such as the inclusion of mainly loci with large repetitive arrays (see Jarne and Lagoda, 1996

for details). Thus, few short array dinucleotide loci are likely to be suitable for parentage determination, where scoring accuracy and locus variability are both of paramount importance (discussed in detail in Chapter 5). The longer array loci may be acceptable for genome mapping purposes, where the number of possible alternate allelic states in the offspring is limited, and complete pedigree information is readily available to confirm allele designations. The short array loci are suitable for population applications, where large numbers of alleles can be problematic (although see discussion below). Furthermore, amplification success of very old, degraded DNA seems to be negatively correlated with PCR product length (primer to primer), and in part, array length (personal observations). Some of these loci, particularly *Ssa* 12, are likely to be very useful in retrospective studies of population variation using archival tissue samples that have not been well preserved (personal observations).

Abundance and characteristics of Tetranucleotide microsatellies

A single $(GTGA)_n$ tetranucleotide microsatellite locus was observed in the 40,000 colonies screened using in the second library. This either reflects the rarity of this type of locus in the Atlantic salmon genome, or is due to sub-optimal screening procedures, despite the inclusion of several controls not commonly employed in these types of studies. Given the current trend towards the use of tri- and tetranucleotide loci for parentage and populations applications, more information will become available on the relative abundance of the different motifs in related taxa. Moreover, Beckman and Weber (1992) have

demonstrated that although the relative ranking of the abundance of different motifs shifts somewhat between human and rats, approximate frequency of occurrence in one species may be a reasonable predictor of frequency in not too phylogenetically distant taxa.

All three tetranucleotide loci investigated here were similar in exhibiting minimal PCR "stutter". A single faint "stutter" band 4 base pairs below the primary band was observed at *Ssa* 197 and *Ssa* 202 (see Figure 4.1). Either no "stutter" band or a very faint "stutter" band two base pairs below the primary band was observed at *Ssa* 171 (see Figure 4.1). Interestingly, while only standard four base pair differences were observed amongst alleles at *Ssa* 197 and *Ssa* 202, non-standard 2 bp differences (microheterogeneity) were commonly observed at *Ssa* 171, a locus flanked by a perfect array of 7 dinucleotide repeats. Microheterogeneity was almost absent in several hundred European Atlantic salmon surveyed (unpublished data), and the "stutter" pattern resembled that described above for *Ssa* 197 and *Ssa* 202.

Blurring of alleles was never observed at *Ssa* 202 or *Ssa* 171. However, both upper and lower alleles were occasionally diffuse at *Ssa* 197. This was not due to re-annealing during electrophoresis, however, but rather during gel loading. Re-annealing and resultant blurring of the upper alleles was completely eliminated by simply maintaining sample-containing microtitre plates above 75° C while loading denaturing gels.

In summary, designating allele sizes at tetranucleotide microsatellites loci is generally much less problematic, and much more accurate, than at dinucleotide loci (see Chapter 5 also). For this reason alone, these loci are preferable for all applications, except perhaps genome mapping where markers are required at regular and frequent intervals. Tetranucleotide loci are also, on average, more variable, a definite advantage for forensic and parentage studies. The large number of alleles commonly observed in population surveys is becoming less of a problem. Programs for analyzing population variation now typically accommodate 50 or more alleles (Raymond and Rousset, 1995). Furthermore, although findings of large numbers of alleles necessitates large sample sizes to adequately represent all of the alleles present in the population, methods are beeing developed that, based on a limited data set, extrapolate to the number of alleles likely to be present in a particular population given complete sampling (C. Herbinger, pers. comm.) Lastly, Ruzzante (1997) has demonstrated that even for extremely variable loci with up to 60 allelic variants, 50 individuals is an adequate sample size for reasonable estimation of many commonly used genetic distance and structure metrics.

CHAPTER 4: RAPID ANALYSIS OF GENETIC VARIATION IN ATLANTIC SALMON *(SALMO SALAR)* BY PCR MULTIPLEXING OF DINUCLEOTIDE AND TETRANUCLEOTIDE MICROSATELLITES

INTRODUCTION

Identification and analysis of genetic variation has two primary applications in Atlantic salmon biology, resolution of proximal relationships for studies of reproductive strategies and implementation of breeding programs, and analysis of population structuring for fisheries management and conservation (i.e. mixed stock analysis and targeted enhancement of productive catchments or tributaries, Davidson *et al.* 1989; Galvin *et al.* 1995c; M. O'Connell, pers. comm.). Population variation in Atlantic salmon has been extensively surveyed, throughout much of the species home range, using allozymes markers (reviewed in Davidson *et al.* 1989). Significant differences among populations is often reported at these loci (Hovey et al. 1989; McElligott and Cross, 1991; O'Connell et al. 1994), although levels of variation detected are typically much lower than observed using recently developed methods (O'Connell and Wright, 1997). In fact, due to the limited levels of variation observed, these loci are rarely useful in determining relatedness amongst individuals. Many of these loci also appear to be acted on by selection (Jordan and Youngson, 1991), and are thus of limited use in estimating population parameters, which assume neutrality. Nevertheless, assays of variation at these loci can be considerably less time consuming and

expensive then currently available DNA-based methods, and are suitable for some applications.

More recently, mtDNA variation has also been used to differentiate between populations of Atlantic salmon (O'Connell *et al.* 1995). Because of the primarily uniparental (maternal) mode of inheritance, and resulting reduced Ne (N) compared to nuclear markers (4N), mtDNA is more sensitive to bottlenecks than allozyme or nuclear DNA markers (Wilson *et al.*, 1985). However, mtDNA is not useful in most analyses of relatedness, as it provides no information on paternity. A more serious general limitation of this marker is that, because of the molecules overall size (typically < 20 KB), and general lack of recombination between molecules in heteroplasmic individuals (Wilson *et al.*, 1985) all sites in the genome are effectively linked. Hence, this marker should be considered as a single locus, and alternate mtDNA types, alleles. MtDNA, however, is very useful in addressing population issues when used in conjunction with other (nuclear) markers.

Genetic variation in Atlantic salmon has also been surveyed at minisatellite or VNTR loci (Taggart and Ferguson, 1990; Bentzen and Wright, 1993). Methods thus far developed to assay variation at these loci, however, are based on blotting and hybridization procedures (Southern, 1975), and are therefore very laborious and time consuming. Furthermore, unlike current PCR-based methods, southern blotting procedures require non-degraded, high molecular weight DNA, and are thus not suitable for retrospective or historical analyses using scales or other poorly preserved sources of tissue.

Microsatellites are both abundant and highly polymorphic in the several fish species examined to date (Goff *et al.* 1992; Estoup *et al.* 1993; Sakamoto *et al.* 1994; Brooker *et al.* 1994; Wright and Bentzen 1994; McConnell *et al.* 1995b; Morris *et al.* 1996; Colbourne *et al.* 1996), including Atlantic salmon (Slettan et al. 1993; McConnell et al. 1995b). Due to the variability of these loci and ease of assay using the polymerase chain reaction (PCR), microsatellites are becoming increasingly popular as genetic markers for a range of applications in fisheries and aquaculture, such as parentage determination in mixed family groups (Herbinger *et al.* 1995; Doyle and Herbinger 1995), family relatedness in natural populations (Herbinger *et al.* 1997), and population differentiation (Nielsen *et al.* 1995; McConnell *et al.* 1995b).

Microsatellites isolated from fish thus far are comprised predominantly of two base pair repeat units, usually $(GT)_n$ or $(CT)_n$ motifs (Goff *et al.* 1992; Morris 1993; Estoup *et al.* 1993; Brooker *et al.* 1994; McConnell *et al.* 1995b; Colbourne *et al.* 1996). One of the primary disadvantages of many dinucleotide markers is that amplification of an allele via PCR often generates a ladder of bands (one or two base pairs apart) when resolved on standard denaturing polyacrylamide gels. These accessory bands (also known as "stutter" or shadow bands) are thought to be due to slipped-strand mispairing during PCR (Tautz 1989; Litt and Luty 1989; Hauge and Litt 1993; O'Reilly and Wright 1995), but alternative scenarios include out of register annealing of truncated products (Hauge and Litt 1993), or addition of a single nucleotide to the 3' end of some strands (Weber 1989). Regardless of the mechanism, the appearance of "stutter" or shadow bands may cause

problems identifying or scoring alleles. For example, "stutter" bands from a particular allele may overlap bands from adjacent alleles differing in length by a single repeat unit. In such instances, heterozygous individuals can be mis-identified as homozygotes, though the correct state can sometimes be resolved by comparing the intensity and number of "stutter" bands (O'Reilly and Wright 1995).

Trinucleotide and tetranucleotide microsatellites typically exhibit little or no "stuttering" (Edwards *et al.* 1992; Urquhart *et al.* 1995). Growing interest in these microsatellites stems partly from the role many trinucleotide arrays play in human diseases (reviewed by Nelson and Warren 1993), but also because of the ease and accuracy of scoring alleles, and the ability to survey variation using much simpler and more efficient separation and visualization techniques (Valdes *et al.* 1993; Santos *et al.* 1993). In mammalian genomes, tri- and tetranucleotide microsatellites occur at much lower frequency than dinucleotide repeats (Beckmann and Weber 1992) , but can be isolated by screening large genomic libraries (Edwards *et al.* 1991; Li *et al.* 1993), or by using recently developed enrichment procedures (Armour *et al.* 1994).

Variation at microsatellite loci can be surveyed much more efficiently by co-amplifying or multiplexing several loci in a single reaction. Several multiplex systems employing previously identified human microsatellites have been published (Huang *et al.* 1992; Edwards *et al.* 1992; Kimpton *et al.* 1993; Urquhart *et al.* 1995). In many of these studies, automated detection systems and

different fluorescent tags were used to help determine the identity (locus) of alleles (bands) in profiles containing multiple microsatellites.

Typically, microsatellite allele sizes are determined by comparisons with known sequence (M13 bacteriophage DNA) and one or two reference alleles (Weber 1990; Weber and May 1989; Estoup *et al.* 1993). More accurate and efficient typing of some microsatellites, however, can be achieved by the application of allelic ladder standards (Puers *et al.* 1994). By pooling PCR products from select individuals, a continuous ladder of alleles separated by four base pairs (the length of the standard tetranucleotide repeat unit), and spanning the normal range of alleles observed, may be generated.

In this chapter, I report the development of a system for surveying genetic variation in Atlantic salmon, based on one dinucleotide and three tetranucleotide loci. All four microsatellites can be amplified in a single reaction, and exhibit non-overlapping allele size distributions, permitting identification using standard autoradiographic detection methods. Allele sizes from all three tetranucleotide loci were determined using allelic ladder standards. Allele frequency variation was surveyed from three rivers (Country Harbour, 45° 15' N, 61° 50' longitude; LaHave, 44° 20' N, 64° 25' longitude; Stewiacke, 45° 9' N, 63° 23' longitude) from Nova Scotia, Canada, to evaluate the utility of the loci used here in differentiating between populations.

METHODS

Multiplexing loci

Three loci were identified from the first library (*Ssa* 85, *Ssa*171 and *Ssa*197; see Chapter 3) with alleles exhibiting primarily non-overlapping size ranges, similar primer annealing temperatures, high heterozygosity values, and that co-amplified in the same reaction (Table 4.1). One additional locus (*Ssa* 202), isolated in the second library, was later added to the multiplex. Problems such as overlapping allele size ranges, and co-amplification of two loci at *Ssa*202 (a possible result of earlier tetraploidization in this species) were solved by redesigning primers at this locus. Initially, signal intensity amongst loci within the multiplex system varied greatly. To detect some loci, others had to be overexposed, reducing the ease of scoring certain loci. A more uniform signal intensity was achieved by labeling alternate primers, and by adjusting primer concentrations.

Allelic Ladder Development

Allelic ladders for all three tetranucleotide loci were constructed by pooling multiplex PCR products from eight carefully selected individuals. Aliquots (4 μL) of this cocktail were added to every 10th lane, so that unknown samples were never more than 4 lanes from allelic ladder standards.

Table 4.1. Nucleotide sequence of Atlantic salmon multiplex microsatellites, and primers (**a** and **b**) used for PCR amplification.

Locus	Sequence (5' to 3')	Allele size range (bp)
*Ssa*85	$(GT)_{14}$ (**a**) AGG TGG GTC CTC CAA GCT AC (**b**) ACC CGC TCC TCA CTT AAT C*	110-138
*Ssa*197	$(GT)_5C(TG)_4TC(TG)_3A(GTGA)_{15}$ (**a**) GGG TTG AGT AGG GAG GCT TG (**b**) TGG CAG GGA TTT GAC ATA AC*	131-203
*Ssa*171	$(TGTA)_{14}(TG)_7$ (**a**) TTA TTA TCC AAA GGG GTC AAA A* (**b**) GAG GTC GCT GGG GTT TAC TAT	214-278
*Ssa*202	$(CA)_3(CTCA)_{17}$ (**a**) CTT GGA ATA TCT AGA ATA TGG (**b**) TTC ATG TGT TAA TGT TGC GTG*	268-320

* labeled primer in PCR reaction

PCR analysis

One member of each primer pair (Table 4.1) was labeled at 37° C for 1 hour in a 10 μL volume containing 250 K Bq [γ ^{32}P] ATP (Dupont), 1μM primer, 10 units of T4 polynucleotide kinase ,and *1X one-phor-all* reaction buffer (Pharmacia). Amplification reactions were carried out in a 10 μl volume containing 1-100 ng of template, 10mM Tris-HCl (pH 8.3), 50mM KCL, 1.5mM $MgCl_2$, 200 μM each dNTP, 0.1 % Tween, 0.01 % gelatin, 0.50 μM (85a & b), 0.225 μM (171a & b), 0.150 μM (197a & b), 0.50 μM (202a & b), and 0.5 units of *Taq* Polymerase. The molar ratio of labeled to non-labeled primers was 1:9 at all four loci. Thermal cycling conditions were: 5 cycles of 20 s at 94° C (denaturing), 20 s at 58° C (primer annealing), 20 s at 72° C (extension), and 35 cycles of 20 s at 90° C, 20 s at 58° C and 20 s at 72° C. Amplifications were performed in a PCT-100 MJ Research thermal cycler. Following PCR, an equivalent volume of stop dye was added to each reaction, and heated to 95° C for 15 minutes prior to loading. PCR products were separated by denaturing electrophoresis through standard wedge sequencing gels, composed of 8% acrylamide, 47 % urea, and 1XTBE. Gel electrophoresis was carried out for 5.5 hours at 60 mA, dried and exposed at -70° C to Kodac XR film for 1-3 days.

Patterns of Inheritance of microsatellite loci

A minimum of 20 offspring from three sets of parents were examined to ensure that alleles from all four loci were inherited in Mendelian fashion.

Population analysis

Heterozygosity values were estimated according to Nei (1978). Tests for conformity to Hardy-Weinberg expectations were performed following Guo and Thompson (1992). Pairwise tests of allele frequency heterogeneity were performed for all four loci following Raymond and Rousset (1995). All of the above calculations were performed using Genepop, version 1.2 (Raymond and Rousset 1995). Probability of match values (the probability that two unrelated individuals chosen at random would have the identical genotypes) were computed using the most common genotypes, following Edwards *et al.* (1992).

RESULTS AND DISCUSSION

Characteristics of individual microsatellites

Ssa 171 is the most variable locus examined (Table 4.2), and may be the most variable locus (VNTR or microsatellite) isolated from a salmonid to date (Taggart and Ferguson 1990; Bentzen and Wright 1993; Estoup *et al.* 1993; Slettan *et al.* 1993; Sakamoto *et al.* 1994; McConnell *et al.* 1995b). A total of 25 alleles were observed at this locus, and the average heterozygosity was 91% (Table 4.2). The number and intensity of "stutter" bands was greatly reduced at the three tetranucleotide loci, compared to the dinucleotide locus, *Ssa* 85 (Figure 4.1).

Table 4.2. Number of alleles and heterozygosity at microsatellite loci from three Atlantic salmon populations in Nova Scotia, Canada.

	Locus			
Population (sample size)	*Ssa* 85 No. of alleles Het.*	*Ssa* 197 No. of alleles Het.*	*Ssa* 171 No. of alleles Het.*	*Ssa* 202 No. of alleles Het.*
Stewiacke River (n=43)	12 (0.85)	14 (0.88)	17 (0.91)	11 (0.87)
Country Harbour River (n=30)	7 (0.74)	11 (0.89)	14 (0.90)	10 (0.84)
LaHave River (n=36)	9 (0.80)	9 (0.83)	18 (0.93)	13 (0.90)
Total no. of alleles	12	15	25	14
Average Het.	(0.80)	(0.87)	(0.91)	(0.87)

*Het. = Heterozygosity

Figure 4.1. Multiplex PCR of one dinucleotide (*Ssa*85) and three tetranucleotide (*Ssa*171, *Ssa*197 and *Ssa*202) microsatellites isolated from Atlantic salmon. M13 DNA sequence and allelic ladder (A L) standards were used to size alleles from several unrelated individuals. Size references in base pairs are given on the left.

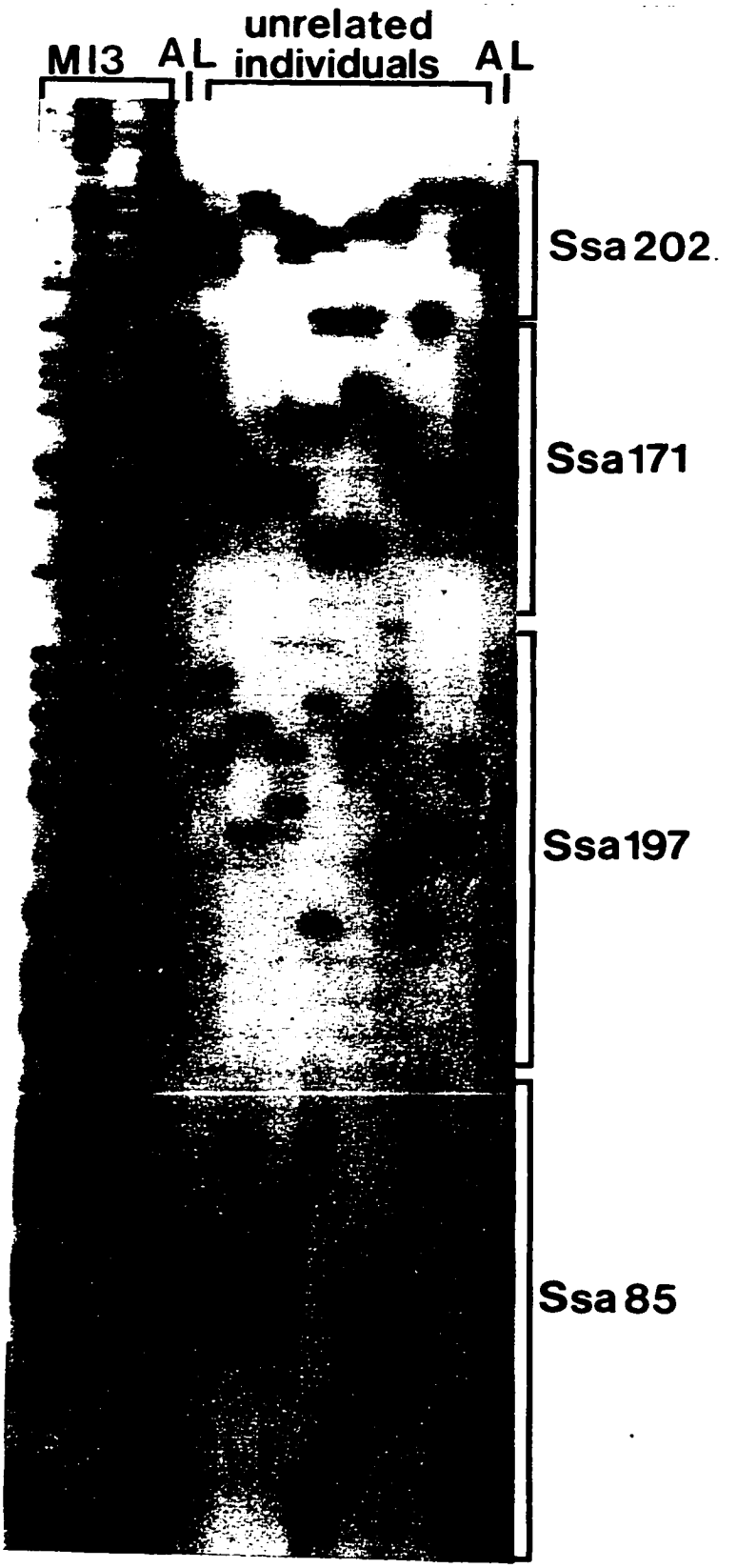

M13
AL
unrelated individuals
AL
Ssa202
Ssa171
Ssa197
Ssa85

Figure 4.1

Tetranucleotide microsatellites in humans were found to have higher mutation rates and levels of polymorphism compared to dinucleotide and trinucleotide loci (Weber and Wong 1993; Armour *et al.* 1994). In Atlantic salmon, I found that heterozygosity levels were higher at the three tetranucleotide loci than at the single dinucleotide locus reported here, as well as three other dinucleotide loci previously surveyed in two of these populations (McConnell *et al.* 1995a; McConnell *et al.* 1995b)

From analysis of parents and offspring of several progeny groups, all loci appear to exhibit classic Mendelian inheritance. Linkage disequilibrium analysis detected no genetic associations between any pair of loci presented here. No significant departures from Hardy-Weinberg expectations were observed at α=0.05 (with Bonferroni correction for 12 comparisons). The lack of an excess of homozygotes detected suggests that null alleles were not prevalent at any of the four loci. Also, no incidences of null alleles were observed in analyses of patterns of inheritance at these loci in 17 families comprised of individuals from several geographic regions (unpublished data).

Microsatellite multiplex

All four loci were amplified together in the same reaction, size fractionated using denaturing polyacrylamide electrophoresis, and visualized using standard autoradiographic techniques (Figure 4.1). The reliability of the system is dependent on the quality of the template DNA amplified: recently extracted DNA isolated from fresh or well preserved tissue amplify faithfully and consistently at

all four loci. All four loci could also be readily amplified from air-dried scales 1 to 2 years old. However, as sample age increases, failure of the larger alleles (primarily from *Ssa* 202) to amplify becomes increasingly common. In such instances, loci *Ssa* 171 and *Ssa* 202 should be amplified individually. For very old or degraded samples, *Ssa* 202 can be amplified individually using an alternate primer set internal to the original primers specified in the multiplex. This system is also sensitive to variation in between locus primer concentrations. Upon receipt of new primer allotments, primer concentrations often have to be re-optimized to achieve similar band intensity at the four loci; I have found that the primer concentrations given by several vendors are often inaccurate. Provided the optimal conditions specified above are adhered to, alleles at all four loci are readily typable. However, excess template DNA and unknown impurities in some samples will cause additional artifact bands that may complicate scoring of alleles at some loci.

Over 200 alleles from three river drainage's in Nova Scotia have been typed at all four loci. Alleles at loci *Ssa* 197 and *Ssa* 171 exhibited non-overlapping size distributions (Table 4.1). Less than 1 % of all observed alleles from *Ssa* 197 overlap with the largest alleles from *Ssa* 85. Given the marked differences in banding patterns of PCR amplified alleles from these two loci (heavy one base pair "stutter" versus minimal four base pair "stutter", Figure 4.1) determining the locus of individual alleles is not a concern. Less than 5 % of the observed alleles from *Ssa* 202 are below 280 base pairs in length, and less than 2 % of observed alleles from 171 are greater than 260 base pairs in length.

Instances where an allele from *Ssa* 171 and an allele from *Ssa* 202 overlap in the 260-280 base pair range (thus complicating the assignment of an allele to a specific locus) would be expected to occur in approximately 1 in 250 (~ 10% X 4%) Atlantic salmon surveyed from these locations. In such instances, loci should be amplified and size fractionated separately. The different 'signatures' of alleles from *Ssa*202 and *Ssa* 171 (a single faint shadow band 4 bp below the main band versus almost no "stutter", respectively) are also be useful in resolving occasional overlapping alleles.

Preliminary data from several additional populations from Nova Scotia indicate that overlap between alleles form these four loci is not a common occurrence. The extent to which alleles from different loci may overlap in populations from other geographical regions is not yet known. In studies where overlap is common, primers from potentially overlapping loci could be labeled with different fluorescent tags (Edwards *et al.* 1992; Urquhart *et al.* 1995), and detected using one of several automated systems currently available.

In this multiplex system, alleles from the four microsatellite loci span a range of over 200 base pairs, essentially the effective range of separation using standard denaturing acrylamide electrophoresis and autoradiography technology. It is unlikely, therefore, that additional loci can be added above, between or below existing loci. In humans, multiplex systems involving four loci span less than 100 base pairs (Huang *et al.* 1992). Allele length distributions for dinucleotide loci and possibly tetranucleotide loci also appear to be wider in fish than in mammals, although there is still little information on the latter class of

microsatellites in fish (Table 4.3). This is not surprising given that microsatellites are, on average, considerably longer in fish than in mammals (Slettan *et al.* 1993; Brooker *et al.* 1994). The implication of these findings is that, using conventional techniques, it will be more difficult to multiplex several microsatellites from fish than has been demonstrated for mammals. Fisheries biologists wishing to analyze microsatellite variation would especially stand to benefit from the use of different fluorescent tags and automated separation and detection systems.

Allelic ladders and typing loci

The 4 base pair allelic ladder developed here nearly spans the entire range of alleles observed at all three tetranucleotide loci. The occasional rung is missing at the top or bottom of each microsatellite distribution. Many of the absent alleles from *Ssa* 171 and *Ssa* 197 have only rarely been observed in the hundreds of individuals surveyed from 15 populations (see McConnell *et al.* 1997).

The allelic ladder permitted alleles to be accurately and rapidly typed at all three tetranucleotide loci. Alleles at locus *Ssa* 171 were relatively large (214-278bp), and many differed in length by non-standard, two base pair intervals. In practice, using this multiplex system, efficient and accurate typing of alleles at

Table 4.3. Allele sizes for dinucleotide and tetranucleotide microsatellites in fish and mammals.

Species	Average range (bp)*	Number of loci surveyed	Number of individuals sampled	Reference
Dinucleotide loci (fish)				
Atlantic Cod *(Gadus morhua)*	71	6	127	Brooker *et al.* 1994
Brown Trout *(Salmo trutta)*	14	3	40	Estoup *et al.* 1993
Rainbow Trout *(Oncorhynchus mykiss)*	40	2	614	Morris *et al.* 1996
Atlantic salmon *(Salmo salar)*	69	4	162	McConnell *et al.* 1995a
Atlantic salmon	21	3	146	McConnell *et al.* 1995b
Tetranucleotide loci (fish)				
Atlantic salmon	54	3	109	this study
Dinucleotide loci				
Humans	11	4	58	Huang *et al.* 1992
Pigs	21	357	104	Rohrer *et al.* 1994
Cattle	12	1	426	Machugh *et al.* 1994
Tetranucleotide loci				
Humans	38	17	>20	Armour *et al.* 1994
Humans	27	3	300	Edwards *et al.* 1992
Humans	36	7	150	Urquhart *et al.* 1995

* average range (largest size - smallest size) of microsatellite loci presented in study

Ssa 171 can only be achieved using multiple and frequent M13 sequence standards, or (preferably) allelic ladder standards.

Population differentiation

Significant differences in allele frequencies were observed between all pairs of populations at all four loci, except between Stewiacke and LaHave at loci *Ssa* 85 and *Ssa* 202 (Table 4.4). The lack of differentiation observed between these two populations for these two loci is somewhat surprising given the greater distance (in marine kilometers) separating the localities, compared to LaHave and Country Harbour sites. Also, adult returns to the Stewiacke are predominantly grilse that remain in the Bay of Fundy, whereas a large proportion of LaHave and Country Harbour populations are comprised of multi-sea-winter fish which migrate to feeding grounds off Greenland (Ritter 1989).

Admittedly, less variable loci are generally more suitable for population studies; for statistical analyses, increased numbers of alleles may impose notable constraints on sample sizes required (Carvalho and Hauser, 1994). However, in this study, highly significant differences in allele frequency were observed more often for the more variable loci used (*Ssa* 171 and *Ssa* 197, Table 4.4). I recommend, therefore, that the multiplex system presented here may be suitable in various population genetics contexts.

In conclusion, the microsatellite loci described here are ideal for assessing genetic variation in Atlantic salmon, and offer the following advantages over other markers developed for this species: (1) these four loci exhibit average

Table 4.4. Pairwise test for heterogeneity between all pairs of three Nova Scotia Atlantic salmon populations for all four microsatellite loci.

Locus	Populations	Probability
*Ssa*85	Stewiacke and Country Harbour	P<0.00001***
	Stewiacke and LaHave	P=0.04172 N.S.
	Country Harbour and LaHave	P=0.00104**
Ssa1 71	Stewiacke and Country Harbour	P<0.00001***
	Stewiacke and LaHave	P<0.00001***
	Country Harbour and LaHave	P<0.00171**
*Ssa*197	Stewiacke and Country Harbour	P<0.00001***
	Stewiacke and LaHave	P=0.00016***
	Country Harbour and LaHave	P<0.00001***
*Ssa*202	Stewiacke and Country Harbour	P<0.00001***
	Stewiacke and LaHave	P=0.28115 N.S.
	Country Harbour and LaHave	P<0.00001***
Overall pair-wise comparisons		
	Stewiacke and Country Harbour	P<0.00001***
	Stewiacke and LaHave	P<0.00001***
	Country Harbour and LaHave	P<0.00001***

* significant given Bonferroni correction for K number of tests at α =0.05
** significant given Bonferroni correction for K number of tests at α=0.01
*** significant given Bonferroni correction for K number of tests at α=0.001

heterozygosity levels between 80-91%; (2) all four loci can be multiplexed in a single reaction and resolved in a single lane on a standard denaturing polyacrylamide gel, greatly increasing the speed of assay; (3) three of these loci exhibit minimal "stuttering", and therefore can be typed much more accurately than dinucleotide loci, and as such are more amenable to automation; (4) three of these loci can be typed with the aid of a readily constructed allelic ladder, further increasing the ease and accuracy of scoring alleles; and (5) alleles from two loci presented here are consistently separated by standard four base pair differences, thus offering the potential for much simpler, less expensive methods of analyses.

Given the combined information content and discriminating ability of these loci (for example, in a randomly selected sample of unrelated individuals, I would expect to find two identical composite genotypes in 30,000 individuals surveyed), this system should be ideal for parentage determination in behavioral and artificial selection studies, and pedigree construction. Furthermore, these loci may be informative for population studies of Atlantic salmon.

CHAPTER 5: AN ANALYSIS OF PARENTAGE DETERMINATION IN ATLANTIC SALMON USING MICROSATELLITES

INTRODUCTION

Microsatellite markers have been used extensively in the reconstruction of proximal relationships between individuals for many and varied applications. Some of these uses include paternity assessment for litigation purposes in humans (Alford *et al.*, 1994; Maha *et al.*, 1995); parentage determination and breed identification in domestic livestock (Sakagami *et al.* 1995; Usha *et al.* 1995); identification of relatedness for inbreeding avoidance of captivated populations of endangered species (Breen *et al.* 1994); parentage determination for assessment of reproductive strategies in wild plants (Dow and Ashley, 1996), birds (Ellegren, 1992), seals (Coltman *et al.* 1996), insects (Paxtone *et al.* 1996), and fish (Rico *et al.* 1993; Colbourne *et al.* 1996); and implementation of safe and efficient artificial selection programs in aquaculture (Doyle and Herbinger, 1995). Despite the widespread utilization of these markers, there has been little discussion on types and rates of errors encountered, on the impact of errors (and mutations) on the resolution and accuracy of parentage determinations, and on the power or ability to resolve relatedness under variable conditions (but see Thomas *et al.* 1994, and Blouin *et al.* 1996). In many instances, these omissions may be due to one or more of the following. First, most studies involving parentage assessment are based on a relatively limited number of parents and offspring; estimates of error rates, and certainly of mutation rates, based on so

few genotypic determinations would not be very meaningful. Second, since studies often involve individuals from wild populations, many analyses are not well controlled. For example, due to phenomena such as egg dumping in birds, or sneak copulations by unexpected (non-genotyped) "sneaker" males in fish, birds, mammals and other groups, one cannot be certain that all, or even most, of the potential parents have been typed. As a result, it is difficult to determine whether offspring/parent incompatibilities (or false exclusions) are due to scoring errors (or mutations), versus mating involving non-genotyped individuals. Third, in many instances, only the minimal number of loci needed to resolve parentage to a desired level have been employed. As discussed below, additional information is often required to assess scoring error and mutation rates, and the impacts of these on parentage assessment.

Parentage determinations based on microsatellite assays are much more simple and reproducible than previous efforts using multilocus fingerprinting and southern blotting analysis (see O'Reilly *et al.* 1996). Still, there are many opportunities for errors to occur: mis-placing samples during DNA extraction, PCR setup, and gel loading; PCR-based errors including *in vitro* replication slippage and the generation of artifact bands; electrophoretic anomalies such as "band shifting"; and finally, allele designation and transcription mistakes. Analyses of rates of DNA sequence errors have been conducted (Kristensen *et al.* 1992), and were found to be surprisingly high (3-4 % per base submitted to the EMBL sequence database). Similar error rates, for example, in microsatellite assays (3-4% per allele scored) could have a serious impact on efforts to

determine parentage by leading to frequent incompatibilities between offspring and their true parents. In an analysis employing four microsatellite loci to resolve maternity and paternity of communally raised offspring , 24-32 % of all individuals assayed would be expected not to match their true parents due to experimental error alone.

Allele typing at microsatellite loci may be even more problematic than assigning bases in DNA sequence. In both DNA sequencing and Minisatellite Variant Repeat mapping using PCR (MVR-PCR) (Jeffreys *et al.* 1991), an internal DNA standard exists that greatly increases the accuracy of scoring. In these analyses, the position or size of a particular band is inferred by its location relative to bands above and below along a continuous sequence of bases (in DNA sequence) or repeat units (in MVR-PCR). Except in cases employing automated, multi-wavelength imaging equipment and different primer colour tags to separate and visualize standards and unknowns in the same lane, microsatellite allele sizes are determined by comparing mobility of unknown alleles to standards located elsewhere on the gel. Hence, sizing is subject to electrophoretic anomalies such as "gel smiling" or "frowning", and "band shifting". Also, the failure of PCR to amplify an allele so that it can be readily detected and scored (a result of a null allele or upper allele drop out) will result in incorrect genotype designations. Failure of polymerase mediated extension reactions in DNA sequencing should result in no or little product, not erroneous or incorrect sequence. Finally, in designating bases in DNA sequence, there are only 4

possible alternate states at any one position; there are often 10 or more alternate allelic states for the more variable microsatellite loci.

A mutation which occurs in the germ line of a parent, and which is subsequently passed on to offspring, will have the same effect as a scoring error in causing false exclusions or true parent/offspring mismatches. Mutations were thought to be of potential concern in multilocus fingerprinting-based parentage determination (see Maha *et al.*, 1995). These assays were based on variation at multiple VNTR or minisatellite loci, where mutation rates are often quite high, up to 5% per gamete (Jeffreys *et al.* 1988a). From an ever-expanding database of microsatellite mutation rates derived from (1) actual counts from CEPH reference families (Weissenbach *et al.* 1992; Bowcock *et al.* 1993; Kwiatowski *et al.* 1992; Petrukhin *et al.* 1993; Straub *et al.* 1993; Weber and Wong, 1993; Mahtani and Willard, 1993; Talbot *et al.* 1995; Beckmann *et al.* 1993), (2) linkage disequilibrium between microsatellites and other loci (Hastbacka *et al.* 1992; Crouau *et al.* 1996), and (3) heterozygosity measurements and simultaneous estimates of μN (Edwards *et al.* 1992), rates at non-disease causing microsatellites are generally on the order of 10^{-4} or lower (see also Goldstein and Clark, 1995). Of the many non-disease associated human loci assayed for mutations, however, a few tetranucleotide loci are very unstable, with germ line mutation rates approaching some VNTRs (7.5×10^{-3}, Mahtani and Willard, 1993; 1.4×10^{-2}, Talbot *et al.*, 1995). For most purposes, though, mutation rates roughly similar to those reported for the vast majority of human microsatellite loci should not be a concern in microsatellite based analyses of parentage.

Few empirical data, however, are available on mutation rates at microsatellite loci from non-mammalian species. Microsatellite arrays in fish are, on average, notably longer than those documented in mammals (Slettan 1993, Brooker 1994, and McConnell 1995a). This may be due to a relaxation of constraints on increased size, or perhaps some as yet unknown function favouring longer lengths in fish. Alternatively, the different characteristics may reflect different mutational processes, modes, or rates of mutation between the two groups. An overall average difference in mutation rates in fish is also not unexpected given the increased array length compared to mammals, and findings of greater microsatellite instability with increasing number of repeats in both disease-associated trinucleotide loci (Caskey *et al.* 1992; Richards *et al.* 1992; Imbert *et al.* 1990, see also Richards and Sutherland, 1994) and non-disease-associated tetranucleotide loci in humans (Talbot *et al.* 1995) and birds (Ellegren *et al.* 1996). Moreover, variation in mutation rates has been reported between less phylogenetically divergent taxa, rodents and humans (see Ellegren, 1995).

To date, basic measures of variability (heterozygosity and number of alleles) have been surveyed at thousands of human microsatellite loci (see Weissenbach *et al.* 1992 and Matise *et al.* 1994), whereas only a handful, perhaps several dozen, published reports of variation at fish microsatellite loci are available. Still, several loci from Atlantic Cod (D. Cook, pers. comm.), Gulf pipefish *(Syngnathus scovelli)* (Jones and Avise, 1997), Pacific herring (O'Connell *et al.* 1996b), and Rainbow trout (M. O'Connell, pers. comm.) have

been identified that exhibit levels of heterozygosity between 95 and 99%, comparable to the extremely rare hypervariable microsatellite loci from humans mentioned above.

The primary objective of this study was to determine parentage of approximately 800 Atlantic salmon, from 12 full-sib crosses, reared communally immediately following fertilization as part of experiment designed to investigate possible early environmental effects on growth performance; a detailed account of this research is given in Herbinger *et al.*, 1997). Briefly, pedigree information is essential in the implementation of any safe and effective breeding programs. In larger animals (mammals, for example), newly born offspring can be readily tagged to identify family of origin. For most fish, however, fry are too small (and often, too numerous) to permit physical tagging. Typically, to allow identification of cross of origin, offspring from different families are kept in separate tanks until large enough to tag, and then combined to minimize environmental influence on growth rates (and other characteristics of interest) of different families. One obvious limitation of this approach if that unknown environmental factors (tank density, water quality, food availability, competition, Herbinger *et al.*, 1982), may significantly influence rates of growth amongst the different families during early, pre-communal rearing periods. In fact, to the best of my knowledge, no attempt has yet been made to assess the presence (or extent) of confounding environmental influence on among family differences in growth rates, during this early pre-tagging stage.

Here, fertilized eggs from each of 12 full-sib crosses were divided into two numerically equal groups. One group from each family was combined with the remaining 11, and raised communally. The remaining allotment of eggs from each of the 12 crosses were raised in a separate tank until large enough to tag (after which whey were combined, and raised communally. Growth rates amongst the 12 families were compared between the communally and separately reared fish to determine the extent of environmental influence on early growth of Atlantic salmon.

A second objective was to determine the frequency of mutations at these four loci. This analysis was important in assessing the impact of mutations on parentage assessment, but also in contributing to the taxonomic base of a growing list of empirically-derived determinations of mutation rates at microsatellite loci. A third objective was to assess the types and rates of allele scoring errors at the one dinucleotide and three tetranucleotide loci employed here, and the impact of these on both the level of resolution of parentage achievable, and the accuracy of such analyses using microsatellite data. At the end of this chapter, I also make several recommendations on experimental design based on findings of type, rate and impact of scoring errors and mutations.

Several aspects of this study make it particularly suitable to assess types and rates of scoring errors and mutations in microsatellite-based analyses of parentage. First, in typing 800 offspring and all of their putative parents at four microsatellite loci, I have screened approximately 6,400 parent-offspring

transfers of alleles (opportunities for mutations to occur). Second, the experimental design was quite simple and well controlled: each of 12 males was crossed uniquely to one of 12 females, and offspring raised communally immediately after fertilization. Third, parentage of nearly all offspring could be determined using only 3 loci; an additional locus was included to help identify parent/offspring incompatibilities (errors), and to permit resolution of parentage in the event of allele scoring problems at a particular locus. Fourth, studies of microsatellite mutations in humans have been based on DNA extracted from CEPH lymphoblastoidal cell lines. Estimates from such analyses may be problematic, as many mutations appear to arise in somatic cells lines, rather than in the germ lines of the parents, hence overestimating actual mutation rates in most instances (Banchs *et al.* 1994). Here, DNA was obtained directly from blood or muscle tissue isolated from offspring and parents, therefore, all mutations must have occurred either in the germ line of the parents, or very early in the soma of offspring. Fifth, analyses were performed using real genotype information from actual parent-offspring groupings, not overly optimistic error-free data generated from population allele frequency information.

A final objective was to evaluate the degree of success in resolving parentage using varying amounts of information (number of loci), under actual and increasingly difficult simulated conditions (assuming considerably more crosses, and/or half-sib instead of full-sib crosses).

METHODS

Fish rearing and DNA typing

Twelve mature female and twelve mature male Atlantic salmon *(Salmo salar)* were obtained in November, 1989 from broodstock line two of the Salmon Genetic Research Program, St. Andrews, New Brunswick. This line was originally derived from the St. John River, New Brunswick. Each of the females was crossed to one of the twelve males in a simple full-sib 12 X 1 experimental design. Approximately 1 ml of blood was taken from each of the males and females for later DNA extraction and typing. Records of which male was crossed with which female were kept. After water hardening, approximately 260 eggs from each of eight lots of fertilized eggs were combined into two tanks and reared communally (~2100 eggs/tank). 525 eggs of the remaining four of twelve lots of fertilized eggs were combined into a second set of two tanks, and also reared communally. The remaining 2100 eggs of each of the 12 groups of eggs were transferred to individual tanks and reared separately until fry were large enough to tag, after which they were reared communally. This was part of a second study conducted to assess early environmental effects on growth performance of Atlantic salmon (see Herbinger *et al.*, 1997). After approximately 8 months, blood and/or muscle tissue was taken from the communally reared Atlantic salmon parr. Blood was preserved by combining with an equal volume of buffer A (Marine Gene Probe Laboratory), and freezing at –70. This preparation was originally intended for long term storage and later Southern blot-based

fingerprinting techniques. Several attempts were made to extract DNA from this solution using standard phenol-chloroform methods (Sambrook *et al.*1989). However, most of the DNA appeared to be tightly bound to a highly viscous core. Eventually, it was found that an approximately 1:10 dilution of the supernatant surrounding the viscous center was suitable for microsatellite based assays. DNA was also extracted from muscle tissue using standard phenol-chloroform methods. PCR failure rates, and the appearance of artifact bands, were notably higher in the blood-derived samples than either muscle or scale samples from this or other studies.

Microsatellite variation was surveyed at one dinucleotide (*Ssa* 85) and three tetranucleotide loci (*Ssa* 171, *Ssa* 197, *Ssa* 202) (see Chapters 3 and 4). When possible, the four loci were multiplexed or co-amplified two, three, or four at a time. However, due to poor template quality or contamination with unknown impurities, approximately 60% of all samples had to be amplified on an individual basis. Allelic ladder standards for the three tetranucleotide loci, and select alleles for the one dinucleotide locus, were positioned in the gel frequently and at regular intervals, every 9 lanes. Hence, no sample was more than 4 lanes from an allelic ladder, and in the case of the three tetranucleotide loci, nearly all alleles from the offspring were within one repeat unit of a standard allele of previously determined size. In addition to contributing to scoring accuracy, the repeated placement of allelic ladder standards also served to frequently "re-set" sample placement in the gel. Unintentional and unnoticed duplicated loading of a particular sample, for instance, could thus affect the identity of fewer samples

"downstream" of the mistake. All alleles were scored by comparing mobility (position) relative to standards without automated assistance of any sort.

Measures of Variability and Information content

Several measures of locus variability and information content were calculated for microsatellite data from the 24 parents in this study and from larger data sets obtained from a wild populations from the LaHave River, Nova Scotia (N=96). Number of alleles was simply the total number of unique or different alleles identified. Observed heterozygosity was the percent of all individuals assayed that were heterozygous at a particular locus. Expected heterozygosity was estimated following Nei (1978). As an indication of the usefulness of each of these four microsatellites in resolving parentage in this experiment, Power of Exclusion values were calculated. Here, the percent of all 11 non-parental families excluded for each offspring was determined for each of the four loci. These values were then averaged over all approximately 800 offspring.

Parentage Determination Using Compatibility Analyses

Parentage was determined by simply comparing sets of alleles at a given locus from each offspring with all four possible sets of alleles from each of the potential parental crosses. This was repeated for all four loci. Potential crosses with alleles incompatible with those of a particular offspring, at one or more loci, were excluded from the set of possible parental couples (except for one analysis based on '3 of 4' or '4 of 4' compatible loci, discussed below). Due to the large

number of necessary comparisons (for example, 800 offspring X 12 possible sets of parents X 4 loci), analyses were preformed on a computer using Microsoft Excel.

In order to assess the accuracy and resolving power of this system under a range of conditions, several parentage determination experiments were preformed. Preliminary analyses were conducted prior to extensive re-typing and re-analysis to assess the robustness of the assay to actual, observed rates of scoring errors or mutations. All subsequent analyses were based on genotype information corrected for scoring errors, as discussed below. In these experiments, parentage was determined using different numbers of loci, and under actual and simulated experimental designs, involving varying numbers and types (half versus full-sib) of crosses.

The few fish determined to be "jumpers" from nearby tanks, or for which genotype information from all four loci was unavailable, were excluded from all analyses, but will be discussed in the *Results and Discussion* section. Offspring exhibiting mutations were excluded from all but the initial, preliminary analysis. Information about whether a particular individual originated from the first communally reared group of eight families, or the second communally reared group of four families, was not utilized in determining parentage.

12X1 parental cross experimental design; "uncorrected" genotype information; preliminary analysis: After one or more attempts were made to type each of the approximately 800 offspring at all four microsatellite loci, those

individuals (N=674) with complete genotype information from all four microsatellite loci were subjected to compatibility analysis against the original 12 possible parental crosses. As many individuals were incompatible with any of the parental crosses at 4 of 4 loci (due to scoring errors and mutations), matches with potential parents based on either 4 of 4 ***or*** 3 of 4 compatible loci were accepted. The purpose of this analysis was to demonstrate the level of resolution and accuracy expected given observed rates of scoring errors or mutations, without extensive re-typing of incorrect loci, and without re-analysis of the data.

Following completion of this portion of the study, all individuals not matching one of the 12 sets of parents at 4 loci were analyzed further to identify the source of the incompatibility and to correct typing errors (discussed below).

12X1 parental cross experimental design; "corrected" genotype information; all four microsatellite loci: Using genotype information from 4 microsatellite loci, attempts were made to match each of approximately 800 offspring to one (or as few as possible) of the 12 original parental crosses.

12X1 parental cross experimental design; "corrected" genotype information; three (tetranucleotide) microsatellite loci: Using genotype information from only 3 tetranucleotide microsatellite loci, I again attempted to match each of 800 offspring to one (or as few as possible) of the 12 original parental crosses.

12X12 parental cross experimental design; "corrected" genotype information; four microsatellite loci: Here, parentage was determined using four microsatellite loci without knowledge of which male was crossed to which female. In other words, under a scenario whereby each of 12 males was crossed to each of 12 females in a simulated 12 by 12 experimental design, producing 12 actual and 132 simulated sets of parents.

36 X 1 parental cross experimental design; "corrected" genotype information: In this experiment, parentage was assessed under a scenario involving the addition of considerably more hypothetical parents. A further 24 males and 24 females were obtained from an existing database of genotype information from the Upper Salmon River, New Brunswick (courtesy of M. Jones and T. McParland). This population was the least geographically distant site from the St. John river for which information from *Ssa* 85, *Ssa* 171, *Ssa* 197 and *Ssa* 202 was already available (150 km North East). All approximately 800 offspring were compared to an expanded set of 36 parents: the original set of 12 actual parents, and 24 hypothetical sets of parents produced by uniquely grouping each of the 24 Upper Salmon River males with one of the 24 Upper Salmon River females.

Analysis of parent-offspring mismatches

All offspring that did not match any of the 12 sets of original parents, at all 4 loci, were analyzed further to identify the source of the mismatch or

incompatibility. First, relevant autoradiographs were re-examined to identify transcription errors (mistakes in documenting correctly scored alleles) and scoring errors (a detailed classification of which is given below). Next, unresolved incompatible samples were re-amplified and typed at the single problem locus. Still incompatible samples were then re-typed at the remaining 3 "compatible" loci in the event that of the 800 samples analyzed, a few may have been incorrectly typed at 3 loci, and correctly typed at a single locus. This could be explained, for instance, by a single sample placement error during PCR set up in which three of four loci were co-amplified or multiplexed in a single reaction.

Probability of Detecting Scoring Errors and Mutations

Two of the objectives of this part of the study were to determine the type and rates of scoring errors, and to determine rates of mutations. Both scoring errors and mutations were detected by searching for parent/offspring incompatibilities. However, some scoring errors and mutations may also produce new composite genotypes that (1) match the "true" set of parents, or (2) match one of the remaining 11 sets of "false" parents. With regards to the detection of scoring errors or mutations, it was important only to distinguish between whether the mutation or error is detected, or not. Distinguishing between the later two processes was important in assessing the impact of scoring errors or mutations on the accuracy of parentage determination analyses, and will be discussed below.

The likelihood of detecting an error or mutation was determined using two approaches. In the first, all "incorrect" or "mutant" composite genotypes possible in the offspring were generated and tested against the 12 sets of original parents using Microsoft Excel. This was done by first generating all composite genotypes possible in the offspring of each set of parents, assuming no errors or mutations. For each composite genotype, one allele at a time was varied, substituting a new, different allele from the total set of alleles present in the 24 actual parents. Alleles absent in the parents were not considered, since any non-parental allele would have been immediately detected in the compatibility analysis. This was repeated for both alleles at a given locus and for all four loci. The approximately 180,000 "mutant" genotypes were then subjected to the same compatibility analysis used in analyzing parentage in the 800 offspring discussed earlier. From this, cross specific detection rates for each locus were made. Overall error detection rates per locus were obtained by weighting each cross in accordance with its actual representation in the 800 offspring analyzed.

In the second approach, rather than generating all possible mutant genotypes , a subset of possible mutant composite genotypes were produced for each cross by randomly selecting new alleles from the set of potential alleles in the parents. This was done for all 4 loci. Computer generated "mutants" were then subjected to compatibility analysis using the Database program Borland from Paradox. Both approaches gave similar results (within, on average, 0.5%).

In an extension of this second approach, all artificially generated mutant composite genotypes that would not normally be detected as incompatibilities

were studied further to ascertain the impact of mutations and scoring errors on the accuracy of parentage assignments. Knowing the mutant genotype's actual cross of origin, and the cross implicated in the analysis, all "mutants" could be classed as having matched either the original "true" set of parents, or one of the 11 remaining "false" sets of parents. Occurrences of the latter would reflect the likelihood of an undetected error causing a false or incorrect offspring/parent match.

Classification of scoring errors

As discussed earlier, putative scoring errors were identified as incompatibilities between individual composite genotypes and any of the 12 sets of actual parents. After ruling out alternative sources of incompatibilities, such as mutations, extraneous origins of offspring ("jumpers"), and the incorrect typing of alternate loci, scoring errors were classed according to probable cause of the mistake. Alleles that amplified poorly relative to other alleles of approximate similar size were defined as "partial null alleles". "Upper allele drop out" was ascribed to instances were the largest alleles at a particular locus amplified markedly less well than the much shorter alleles observed. In instances of partial null alleles and upper allele drop out, individuals were initially identified as homozygous for the alternate, more strongly amplifying alleles. After compatibility analysis, and upon re-examination of relevant autoradiographs, the weakly amplifying allele could always be unambiguously identified. Errors due to difficulties distinguishing between homozygous allelic states and heterozygous

states with alleles one repeat unit apart, were defined as "stutter related homozygote/heterozygote scoring errors". Here, the direct cause of the incorrect identification of genotypic states was the occurrence of multiple, heavy "stutter" bands below the first primary band in the upper allele. These accessory bands could potentially overlap and mask a second smaller allele one repeat unit below the first. Hence, it was at times difficult to distinguish between homozgous from heterozygous states at a particular locus.

Instances where a single allele was identified incorrectly, but where overlap of "stutter" bands between adjacent alleles was not a factor, were defined as "allele designation errors". Typically, mistakes of this type were out by a single repeat unit, and were usually due to inaccurate estimates of size based on comparisons with standards, though locus specific contributing factors were also observed at *Ssa* 85. Errors involving both alleles at a single locus were classed as unexplained single locus errors, and those involving multiple alleles at two or more loci were classified as unexplained multiple locus errors. Errors due to mistakes in documenting allele sizes were termed transcription errors.

RESULTS AND DISCUSSION

Information content and power of exclusion

The dinucleotide locus *Ssa* 85, was the most variable and consistently interpretable of eight dinucleotide loci developed here for studies of Atlantic salmon. Nevertheless, expected heterozygosity, number of alleles, and power of exclusion values for this locus were lower than any of the three tetranucleotide

loci, *Ssa* 171, *Ssa* 197 and *Ssa* 202 (Table 5.1). Human and porcine tetranucleotide loci were similarly found to be, on average, more variable and informative than dinucleotide loci (Armour *et al.* 1992; Weber and Wong, 1993; Ellegren, 1995; but see also Chakraborty *et al.* 1997), and are being increasingly used in resolving forensic and paternity issues (Urquhart *et al.* 1995; Gill *et al.*1994; Hammond 1994). The locus *Ssa* 171 was the most informative (Table 5.1). This may be a result of non-standard 2 bp differences (microheterogeneity) between alleles, and hence a greater number of possible alternate allelic states, assuming some degree of upper and lower constraint on array length. Higher overall information content compared to other tetranucleotide loci, reflected in both number of alleles observed and probability of match values, were also reported in human tetranucleotide microsatellites exhibiting non-standard, two base pair differences (Urquhart *et al.* 1995). Given the minimal effect on scoring accuracy (discussed below), this type of tetranucleotide locus is demonstrably valuable in assessing relatedness.

Determining parentage in potentially inbred aquaculture stocks can be considerably more difficult than in non-inbred captive or wild populations due to the greater likelihood of allele sharing amongst potential parents. The expected heterozygosity of the St. Andrew's broodstock line two was, on average, less than 4% lower than the wild population from the LaHave River, Nova Scotia. This example, therefore, should be useful in assessing the power of these four loci in resolving parentage in wild or outbred stocks of Atlantic salmon.

Table 5.1. Microsatellite locus characteristics and measures of information content.

	Ssa 202	*Ssa* 171	*Ssa*197	*Ssa*85
Standard repeat unit motif (base pairs)	4	4	4	2
Number of alleles	11(12)	14(22)	10(10)	8(9)
Observed Heterozygosity	83.3(90.9)	91.7(90.8)	79.2(81.2)	83.3(85.1)
Expected Heterozygosity (Nei, 1978)	87.9(89.0)	86.1(92.7)	83.3(85.4)	78.1(82.1)
Power of Exclusion*	84.7	85.7	80.4	72.2

( ) Information for the wild LaHave population given in brackets
*percent of *'non-parental crosses'* excluded averaged over all offspring

Parentage determination using compatibility analysis

In a preliminary analysis involving the original 12 crosses, parentage was assessed using uncorrected data, based on 4 of 4, or 3 of 4 compatible loci. Of 674 offspring, 663 (97.32 %) could be assigned unambiguously to a single cross (Table 5.2). Three individuals (0.45%) were incorrectly assigned to a set of parents in this initial round of analyses. Using corrected genotype information from 4 loci, all (792) offspring were re-analyzed. All but one individual (99.9 %) could be unambiguously traced to a single set of parents. Three offspring could only be typed at either two or three loci, but were still resolved to a single cross. Compatibility analysis was also performed using only three (tetranucleotide) loci. Parentage of 773 of 792 offspring (97.6%) could be unambiguously resolved to a single set of parents. The remaining 19 individuals could be traced to two of 12 possible crosses.

Without knowledge of which female was mated with which male, resolving parentage becomes considerably more difficult, involving now 144 full and half-sib crosses. Using corrected genotype information from 4 loci, 81.6 % of offspring could be assigned to a single set of parents. A further 13.5 % could be traced to 2 or 3 of 144 potential crosses. In an additional experiment, parentage of the approximately 800 offspring was assessed while including a further 48 hypothetical parents, 24 additional crosses, for a total of 36 full-sib crosses. Using corrected genotype information, over 99.5 % offspring could be unambiguously traced to a single set of parents.

Table 5.2. Percentage of offspring resolved to one, two, three or greater than three families.

Cross information	Number of families compatible with genotype			
	One family	Two families	Three families	> Three families
-12X1 CROSS (12 families) -uncorrected data set -3 of 4 compatible loci **N=674**	97.92*** (663)	0.45 (3)	0.15 (1)	1.04 (7)
-12X1 CROSS (12 families) -corrected data set -4 loci **N=792**	99.87 (791)	0.13 (1)	0 (0)	0 (0)
-12X1 CROSS (12 families) -corrected data set -3 tetranucleotide loci **N=792**	97.6 (773)	2.4 (19)	0 (0)	0 (0)
-12X12 CROSS (144 families) -corrected data set -4 loci **N=792**	81.57 (646)	6.19 (49)	7.32 (58)	4.92 (39)
-36X1 CROSS (36 families) -corrected data set -4 loci **N=792**	99.62 (789)	0.38 (3)	0 (0)	0 (0)

***this includes 3 individuals (0.45%) that were incorrectly assigned to a single family

Given a moderate number of full-sib crosses, it is clear that this microsatellite system can accurately resolve parentage of nearly all (~99.9%) offspring. I would also argue that analysis based on either uncorrected, non-verified data from 4 loci, or corrected genotype information from the 3 tetranucleotide loci, would provide sufficient accuracy and resolution for most purposes. More important in contributing to the degree of difficulty in resolving parentage than the addition of even considerably more potential parents, is the inclusion of multiple half-sib crosses. In the 12 X 12 analysis, instances where offspring matched two or three potential crosses nearly always involved a common male or female parent. Although the level of resolution obtained in this experiment (~82% single cross assignments) would be suitable for many purposes, inclusion of a single additional locus as informative as any of the three tetranucleotide loci used here, would likely allow unambiguous identification of parentage in greater than 90 % of all offspring analyzed.

Other studies examining relatedness have used considerably more microsatellite markers than employed here (Morin *et al.* 1994; Blouin *et al.* 1996), and have reported similar levels of resolution or success. This is in part due to the highly polymorphic nature of the three tetranucleotide loci, *Ssa* 171, *Ssa* 197, and *Ssa* 202 (He between 83.3 and 87.9 %) . The significance of using more variable markers in reducing costs and labor is reflected in Blouin *et al*'s (1996) findings that nealy twice as many loci of He = 0.62 are required to obtain the same level of resolution as that achieved using loci of He=0.75.

Types and Rates of mutations

In determining parentage of communally reared Atlantic salmon, 794 offspring and 12 sets of potential parents were assayed at 4 microsatellite loci. This represents 3100 single locus genotypes, or 6400 parent-offspring transfers of alleles (opportunities for mutations to occur).

Compatibility analysis identified cases of scoring errors ***or*** mutations. After retyping and reanalyzing all incompatible offspring, and subsequent to the correction of scoring errors, 11 individuals still did not match any of the 12 parents at one or more loci. Three individuals were incompatible at only a single locus, involving a single allele. These fish could have been either "jumpers" or offspring that inherited a single mutation from one of the two parents. To distinguish between these two possible scenarios, these offspring, and their respective putative parents (based on earlier analyses using 3 of 4 or 4 of 4 compatible loci) were typed using an additional 3 variable microsatellite loci, *Ssa* 12, *Ssa* 49, and *Ssa* 76 (see Table 5.3). Two of these three fish matched their respective putative parents at all 3 loci. In both instances, the single incompatible allele at *Ssa* 197 and *Ssa* 202 (discussed in more detail below) was presumed to be a result of a single mutation, either in the germ line of one of their respective parents, or in somatic tissue of the offspring early in embryonic development. Multiple incompatibilities at *Ssa* 12, *Ssa* 49 and *Ssa* 76 were observed between the third fish and the most genotypically similar set of parents based on assays with the original set of four loci. This individual was considered to be a "jumper".

Eight of the 11 remaining incompatible fish did not match any of the 12 sets of parents at two or more of the original four loci; these were also considered to be "jumpers". It could be argued that individuals with 2 or more alleles not present in the most genotypically similar set of parents could also be the result of multiple mutations. However, given the frequency of individuals with confirmed single mutant alleles (1 in 400), the likelihood of even a single offspring with two mutations is quite small ($800(1/400)^2$), or 0.005. Even assuming an average per locus mutation rate of 1 in 500, one would expect fewer than 0.2 individuals with double mutations.

Several lines of evidence support the plausibility of an exogenous origin of some fish. First, even after re-typing, several individuals were incompatible at two or even three loci, involving 2-5 alleles. Second, one individual was incompatible with its putative parents at several additional loci (*Ssa* 12, *Ssa* 49 and *Ssa* 76). Third, tanks containing fish from other parents were close enough that occasional leaps between tanks were possible, and occasional fish were found nearby on the floor indicating jumps from tanks had occurred.

One of the mutations occurred at the microsatellite locus *Ssa* 197, and the second at *Ssa* 202. Given parental and offspring genotypes, the mutation at *Ssa* 197 was either a 4 or 16 base paiir increase. The mutation observed at *Ssa* 202 could only have been a 28 base pair insertion. The overall rate of mutation observed was approximately 1 in 3,200, or $3.1 X10^{-4}$. Given an error detection rate of about 90 % (Table 5.4), I applied a correction factor of 1.11 to give an adjusted rate of $3.4 X 10^{-4}$. Weber and Wong (1993) and Ellegren (1995) used a

Table 5.3. Primer sequence and PCR conditions for additional loci used to confirm status of putative mutant alleles.

Locus	GeneBank accession number	Primer sequence (5' to 3')	Annealing temperature (Celsius)
Ssa 12	U58903	(a) gaa tct ctc ctt tgt ctg c* (b) cat ccg aca aca gca ctg	55
Ssa 49	N/A	(a) gaa ggt cag agc aac aca gc (b) ttc aat ctc agt tta cct ccc t*	58
Ssa 76	AF019188	(a) gac tca tct ccc aac caa ctg (b) aga gtc cca gaa ata cag gca*	60

* labeled primer
N/A not available

correction factor of 1.18 to accommodate undetected mutations in their studies. The mutation rate reported here was similar to that reported in humans (cited above), and slightly higher than seen in pigs (Ellegren, 1995).

Understanding how microsatellite arrays evolve (whether mutations involve single or multiple changes in repeat unit number, and whether there exist any bias in increases versus decreases) is important in accurately estimating various parameters, including population size and population structuring (O'Connell and Wright, 1997). The two main models of microsatellite mutation proposed are 1) the infinite allele model, and 2) the stepwise mutation model (reviewed in Jarne and Lagoda, 1996). In the former, all mutations create new alleles not previously seen in the population. These may involve single, but also multiple repeat unit increases or decreases. In the latter, arrays change incrementally, involving single repeat unit increases or decreases. Recently, a two-phase model of mutations at microsatellites has been proposed (DiRienzo et al. 1994). Under this model, most new alleles involve single repeat unit increases, though some involve multiple repeat units.

Weber and Wong (1993) identified several dozen microsatellite mutations in human microsatellites; most changes involved single repeat unit increases or decreases, with occasional jumps involving multiple repeat units, reflecting most closely the two phase model. Too few new alleles were observed here to directly assess the likely mode of mutation at these loci. Insight into modes of mutation operating at microsatellite loci can also be obtained from analysis of patterns of

variation observed in natural populations. For example, Estoup et al. (1995) compared the number of alleles expected given observed levels of heterozygosity under the infinite allele model with expected numbers under the stepwise mutation model. However, all of the Atlantic salmon populations surveyed here have been heavily impacted by restocking practices; patterns of variation likely reflect both the natural history of these populations and the results of extensive stocking from often relatively few hatchery fish. Inferring the evolutionary history of microsatellite arrays from allele frequency information from such populations would be dubious at best.

Types and rates of scoring errors observed

The type and rate of scoring error observed varied considerably among the four microsatellite loci investigated. Allele designation errors (mistakes in assigning allele size, usually by a single repeat unit), were uncommon at the two teteranucleotide loci (*Ssa* 197 and *Ssa* 202) where alleles were consistently separated by standard 4 bp differences. Errors were only slightly more prevalent at *Ssa* 171, which exhibited non-standard 2 bp differences. Due to the size of PCR products at this locus (220 to 276 bp), many adjacent alleles were not well separated. Frequent placement of allelic ladder standards helped reduce incorrect size determinations at *Ssa* 171. Allele designation errors were highest at *Ssa* 85. Here, alleles were often incorrectly scored due to occasional variation in the pattern of "stutter" bands. Typically, "stutter" bands decrease in intensity from top to bottom, and the upper darkest band in the parent usually

corresponded to the upper, darkest band in the offspring. Occasionally, however, the upper band in some individuals was slightly less intense than the second band. In such instances, the upper band usually corresponded to the parental allele. Sometimes, however, the second slightly stronger band corresponded to the parental allele (this was deduced by re-typing relevant, conflicting offspring at *Ssa* 85, and from compatibility analyses based on the remaining three loci). The first band would appear to be an upper "stutter" artifact.

Another type of error caused by PCR-induced accessory bands, "stutter" related homozygous/heterozygous scoring errors, occurred at a high frequency at *Ssa* 85 (Table 5.5). This appears to be a very common source of error in parentage and population analyses, and is quite possibly a very major contributing factor in cases of homozygote or heterozygote excess often reported in population genetic surveys in fish (McConnell, pers. comm.; Nielsen *et al.*, 1994; O'Connell *et al.*,1996a, 1996b; Bentzen *et al.*, 1996; Scribner *et al.*,1996). The complete absence of this class of error at the tetranucleotide loci was due to the much reduced "stutter" at these microsatellites. In fact, distinguishing between homozygous and heterozygous states at these loci was always straightforward.

Upper allele "drop out", observed at *Ssa* 197, was the single most frequent cause of scoring error encountered. Interestingly, maximum PCR product lengths (and maximum array lengths) at this locus were less than those observed at the other two tetranucleotide loci. Overall product or band intensity, however,

Table 5.4. Results of microsatellite scoring error and mutation detection analysis.

	Locus				
	Ssa 202	*Ssa* 171	*Ssa* 197	*Ssa* 85	Average
Percentage of errors leading to Incompatibility (detection rate)	91.96	90.42	88.80	86.41	89.40
Percentage of errors leading to incorrect (non-paternal) cross	0.10	0.30	0.30	0.40	0.28

was lower for *Ssa* 197 compared to *Ssa* 85 or *Ssa* 171, due either to less efficient amplification or to less efficient primer labeling.

Partial null alleles, or alleles that amplify weakly compared to other alleles of approximate similar size, were also observed exclusively at *Ssa* 197, at a single allele 171 bp in length (Table 5.5). Null or partial null alleles are often due to DNA sequence variation (or deletions) in regions encompassing PCR primer sites (Pemberton *et al.* 1995; Callen *et al.* 1993; Koorey *et al.*, 1993). This is likely the case here. It is interesting to note, however, that the partial null 171 bp allele amplified strongly in the single parent in which it was found. After re-extracting and re-typing DNA from "affected" offspring, it was determined that the efficiency of amplification was largely dependent on the sample or template preparation. In other words, the "171 bp" allele derived from a single individual can amplify poorly or efficiently relative to the alternate alleles of similar size, depending on some unknown DNA extraction-related factor.

Unexplained single locus errors were observed at all loci. At any one locus this class of mistake was not overly common. Collectively, however, it represented approximately 25 % of all errors encountered. Possible explanations include band shifting during electrophoresis, and unknown PCR related errors. Another explanation is sample placement error which could have occurred during PCR procedures, or while loading samples into gels.

Unexplained multiple locus errors were also observed at all loci, and collectively represented about 19% of all errors. Possible explanations are the same as those stated for "unexplained" single locus errors.

Table 5.5. Types and rates of microsatellite allele scoring errors observed in parentage analyses.

Type of error	Locus			
	Ssa 202	*Ssa* 171	*Ssa* 197	*Ssa* 85
Transcription error	0	0	0	0
Allele designation Error	3	6	3	12
Stutter related Homozyggote/ Heterozygote Scoring error	0	0	0	11
Partial null allele Error	0	0	7	0
Upper allele "drop out" error	0	0	14	0
Unexplained single Single locus errors	7	2	5	9
Locus total	10	8	29	32
Total		79		
Unexplained multiple loci errors		18		

No transcription errors were observed. This is in part due to careful documentation of allele sizes, and the practice employed here of scoring and entering each allele size into computer databases twice.

In conclusion, 97 of the 674 offspring surveyed during preliminary analyses (~14 %) were incorrectly typed at one or more alleles. This translates to an approximately 2-3 % error rate per allele scored. It should be kept in mind that there are approximately half a dozen procedural steps involved in the overall assay (PCR set up, gel loading, etc.). An error at any single step could result in an incorrect assignment of alleles at one or more loci. As few as 1-1.5 errors (human and PCR) in 400 operations could account for the 14 % parent-offspring mismatch rate reported here.

Implications of mutations and scoring errors on determinations of parentage

All genetically based parentage determinations in sexually reproducing, diploid organisms are based on the premise that offspring receive one allele from each of their two parents at all autosomal loci. In compatibility analyses, potential parents of unknown progeny are often excluded because they do not possess either allele observed in a particular offspring. Both mutations and scoring errors result in "novel" alleles in offspring that were not present in the actual parents, and, hence, may lead to true parent-offspring incompatibilities, and possibly, incorrect assignments of parentage. When determining parentage in natural or artificially propagated populations, attention should be paid to (1) the frequency

of error rates (mutations and scoring errors), and (2) the impact of such errors on degree of resolution achievable, and on the accuracy of parentage determinations.

Error rates can be assessed, as done above, by identifying incompatible offspring, and analyzing sources of error. However, not all errors are detected as parent-offspring incompatibilities; some may generate new composite genotypes that match either 1) the true set of parents, or 2) one of the possible alternate, incorrect crosses. Hence, simply counting errors observed in incompatible offspring will underestimate the actual rate of mutations and scoring errors. In order to assess the true frequency of single allele errors, the probability of detecting an error must first be calculated (percentage of errors leading to incompatibilities, Table 5.4). Single allele detection rates were similar across the four loci used, averaging about 89%. The lowest rate of detection (86%) for locus *Ssa* 85 is not unexpected given the reduced number of alternate alleles observed at this locus. "Mutant" but compatible offspring were further analyzed to determine the impact of undetected scoring errors or mutations on the accuracy of parentage assignments. In 99.97 to 99.99 % of all cases (depending on the locus), incorrect genotype information due to a single error still implicated the single, true set of parents, despite the considerably greater number of potential alternate, incorrect possible crosses. The tendency of the new, incorrect composite genotype to match the true cross due both to the rarity of alternate crosses genotypically similar at the remaining 7 correct alleles, and to the small

likelihood that an error or mutation will be in the direction of the rare, genotypically similar cross.

The detection rate and impact of errors involving multiple alleles was not assessed. However, it is expected that the detection rate should roughly increase in proportion to the increased number of incorrect alleles. The more incorrect alleles involved, however, the greater the chance of implicating a false cross (if undetected), since the number of correct alleles relating offspring to true-parents necessarily decreases.

Several lines of evidence suggest that in this analysis errors involving single alleles, and quite likely those involving multiple alleles as well, had little effect on the accuracy of parentage determination. First, errors were detected in 97 of 674 individuals assayed using uncorrected data at 3 of 4 or 4 of 4 compatible loci. Given an error detection rate of approximately 90 %, I would expect that errors occurred in a further 9 of 674 offspring that were neither detected nor corrected. Of the total 792 individuals eventually assayed, errors that were neither detected nor corrected may exist in approximately 11 offspring. Since only 0.1 to 0.3% of all undetected errors lead to incorrect assignment of parentage, I would not expect ***any*** mismatched offspring. Second, of a total 674 offspring analyzed using uncorrected genotype information, only 3 were re-assigned based on re-typing and analysis. Third, all 792 offspring were analyzed without using information about which of 2 mixed sets of families they originated. All but a single fish were correctly assigned to their true group, and this individual could have jumped from an adjacent tank. At a parent-offspring mis-match rate

of only 1 percent, I would have expected to see 4 individuals placed into an incorrect group.

Another impact of scoring errors is to reduce the ability to resolve parentage. In the preliminary analysis based on 3 of 4 or 4 of 4 uncorrected loci, parentage was unambiguously assigned in 97 % of individuals surveyed, comparable to analyses based on corrected data from 3 variable tetranucleotide data, but slightly less than that achieved using corrected data from 4 loci. Although analyses based on uncorrected data at 4 loci and analyses based on corrected data from 3 loci are similarly informative, the later involves considerably more time, effort and expense.

The overall impact of mutations on the accuracy of parentage determinations was very minimal. First, mutation rates at these loci were quite low, between 10^{-3} and 10^{-4}. Second, ~90% of mutations should be detected using compatibility analyses, and hence, accommodated. Third, of those that are not detected, few (0.1-0.3%) will lead to incorrect assignments of parentage.

Experimental design suggestions for monitoring and managing the impacts of errors on parentage determination analyses

I have demonstrated that the frequency of undetected scoring errors leading to incorrect parentage assignments can be quite low. However, the importance of errors of this magnitude may vary considerably depending on the application and specific questions asked. For example, a parent/offspring mismatch rate of 1 in 200 is probably suitable for determining parentage in

aquaculture selection programs, or studies designed to assess the relative contribution of precocious parr versus adult males in contributing to a large pool of offspring. However, this rate of error may be unacceptably high in cases where investigators are looking for rare events (egg dumping), or in analyses of paternity for litigation purposes (but see Maha *et al* 1995).

Clearly, the first step in experimental design for any parentage determination analysis should be careful consideration of 1) the level of resolution required, and 2) the acceptable rate of scoring errors and parent-offspring mismatches. Next, potential parents and a small subset of offpring to be analyzed should be typed at several loci. This information should be used to assess 1) the information content of loci available given the particular genotypes of the parents, 2) the reliability of available loci, and 3) scoring error types and rates. A minimum set of reliable, easily interpretable loci should then be selected that allow resolution of parentage to the required level. One or more additional loci should then be added such that the same level of resolution can be achieved in the absence of one of the original loci (in the event of an incompatibility at a single locus). Analysis should then be extended to all offspring in the study. For some purposes, compatibility analysis based on uncorrected data at n ***or*** n-1 loci may be suitable. Increased accuracy (and resolution) can be obtained by re-typing the incompatible loci. Further accuracy can be achieved by adding still more loci. However, given the time and expense of doing so, I recommend that error and mismatch rates be monitored, and that

experimenters choose a level of effort appropriate to the level of resolution and accuracy dictated by the questions being asked.

Additional general recommendations aimed at minimizing effort and maximizing resolution and accuracy are as follows. First, several sources of errors are associated with PCR-induced "stutter", and can be avoided entirely by the exclusive use of tetranucleotide microsatellites. Second, when possible, select loci that produce strong, easily detectable products. Third, a common and potentially significant class of errors involves multiple alleles, and may be due to sample mix-up during PCR set up or gel loading. Researchers should be especially aware of this source of error, and should devise laboratory protocols to minimize sample mix-up during DNA extraction, PCR set-up and gel loading. Finally, careful consideration should be given to multiplexing all loci in a single PCR reaction, as errors due to sample placement during PCR set up and gel loading cannot be detected through compatibility analyses.

Early Environmental Influence on growth rates of Atlantic salmon

One of the objectives of this research was to identify the cross of origin of communally reared Atlantic salmon, as part of experiment designed by C. Herbinger to investigate early environmental influence on growth rates of Atlantic salmon. Details of findings of this research can be found in Herbinger *et al.* 1997, but are summarized below. Briefly, among family differences were observed in communally and separately reared Atlantic salmon. Early growth performance of families raised communally was poorly correlated with growth

rates of families raised separately. This lack of congruence between families reared separately and families reared communally must reflect early environmental influence of initial rates of growth. However, subsequent family growth performance (up to the smolt stage) was not well correlated with early family growth performance, suggesting that the early environmental component does not persist over time.

Summary

The one dinucleotide and three tetranucleotide microsatellite loci used here were very effective in resolving parentage in both the initial 12 X 1 and the 36 X 1 simulated full-sib crosses, with over 99.5 % of individuals being unambiguously traced to a single set of parents. Analyses based on uncorrected data, or corrected data at the 3 tetranucleotide loci were similarly efficient, with approximately 97 % of offspring matching a single cross. In a final 12 X 12 simulated analysis, involving considerably more full-sib, but also multiple half-sib crosses, 82 % of offspring were matched to a set of parents, and 95% to three or fewer of 144 possible crosses.

Individuals incompatible with any of the original set of 12 parents were analyzed to determine the source of the incongruity. Both mutations and scoring errors contributed to incompatibilities . Mutations were rare, occurring in only 2 of 794 individuals assayed at 4 loci, or at an overall rate of 3.4 $X10^{-4}$ per gamete. Observed scoring errors were more frequent, and were observed in 97 of 674 individuals assayed, or at a rate of 2-3 % per allele scored. Analysis of the

impact of errors suggests that, on average, greater than 89 % of scoring errors and mutations will be detected through incompatibility analysis, and of those that are not, only 0.1-0.3 % (depending on the locus) will lead to inaccurate parentage assignments. Several lines of evidence, including observed scoring error rates and impact analyses, suggests that nearly all (>99.75 %) offspring were correctly matched to their true set of parents. Based on types and rates of errors, study-specific and general experimental design recommendations are given.

CONCLUDING REMARKS

The primarily tetranucleotide microsatellite based-genetic marker system developed here is highly informative, efficient, accurate and reliable. Although the system has only been available for the last two years, it has been used exclusively in multiple aquaculture and fisheries management and conservation studies (McConnell *et al.* 1997, in press; McConnell *et al.*, manucscript in prep.; O'Connell *et al.*, manuscript in prep; Galvin *et al.*, manuscript in prep.; Jones *et al.*, manuscript in prep.; Herbinger *et al.* 1997, manuscript submitted; O'Reilly *et al.* 1997, manuscript in prep.). However, there are two serious limitations with this technology: the complexity of the methodology, and the nearly exclusive availability to researchers with access to either automated sequencing equipment, or facilities for the handling and disposing of radioisotopes. The former shortcoming results in increased opportunity for procedural mistakes leading to either inconsistent product or scoring errors. Developing simpler non-isotopic methods, as has been done for several human tetranucleotide microsatellites, can considerably reduce these drawbacks. In such analyses, PCR-amplified alleles are size-fractionated in fine sieving metaphore agarose gels, and visualized using standard ethidium bromide (or SYBR green) staining. The entire procedure could be carried out on a single bench top, well within an average working day. Furthermore, a minimum of specialized equipment would be required (PCR machine, chilling gel electrophoresis unit, portable transilluminator, and poloroid camera plus hood), thereby permitting access to researchers from smaller universities, colleges, and other non-specialized

research institutions. Interest in simple, non-radioactive means of surveying variation at microsatellite loci from researchers from such institutions has been considerable. I recommend that further effort be expended to resolve and visualize length differences at *Ssa* 171, *Ssa* 197, and *Ssa* 202 using such an approach.

Although 6,400 parent-offspring transfers of alleles were assayed, too few mutations were observed to address either modes or mechanisms of mutations at microsatellite loci, and rates could only be roughly approximated. Nevertheless, empirical estimates made here are the first for any fish species that I am aware of, and values are comparable to those typically reported for most human di- and tetranucleotide microsatellite loci. A more fruitful approach to studying mutations, however, may be to assay variation in one of several fish species for which multiple hypervariable microsatellite loci (greater than 95% heterozygosity) have been developed.

The microsatellite system developed here has proven to be very useful in resolving parentage of nearly 800 communally reared Atlantic salmon. Using corrected genotype information from all four loci, parentage of all but a single offspring could be unambiguously resolved to one of twelve possible full-sib crosses. Even under more rigorous conditions (more and half-sib crosses), these four loci were demonstrably very effective in resolving parentage. It would also seem that although errors may be somewhat common, most are detected as parent-offspring incompatibilities, and of those that are not, surprisingly few will result in false parentage assignments.

REFERENCES

Alford, R. L., Hammond, H. A., and Caskey, C. T. 1994. Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am. J. Hum. Genet. **55**, 190-195.

Armour, J. A. L., Neumann, R., Gobert, S. & Jeffreys, A. J. 1994. Isolation of human simple repeat loci by hybridization selection. Hum. Mol. Genet. **3**, 599-605.

Baker, C. S., MacCarthy, M., Smith, P.J., Perry, A.P. & Chambers, G. K. 1992. DNA fingerprints of orange roughy, *Hoplostethus atlanticus*: A population comparison. Marine Biology **113**, 561-567.

Banchs, I., Bosch, A., Guimera, J., Lazaro, C., Puig, A., and Estivill, X. 1994. New Alleles at Microsatellite Loci in CEPH Families Mainly Arise From Somatic Mutations in the Lymphoblastoid Cell Lines. Hum. Mut. **3**, 365-373.

Baron, B., Poirier, C., Simon-Chazottes, D., Barnier, C. & Guenet, J. 1992. A new strategy useful for rapid identification of microsatellites from DNA libraries with large size inserts. Nuc. Acids Res. **20**, 3665-3669.

Beckmann, J. S. & Weber, J. L. 1992. Survey of human and rat microsatellites. Genomics **12**, 627-631.

Beckmann, J. S., Tomfohrde, J., Barnes, R.I., Williams, M., Broux, O., Richard, I., Weissenbach, J., and Bowcock, A.M. 1993. A linkage map of human chromosome 15 with an average resolution of 2 cM and containing 55 polymorphic microsatellites. Hum. Mol. Genet. **2**, 2019-2030.

Behn-Krappa, A. & Doerfler, W. 1994. Enzymatic amplification of synthetic oligodeoxyribonucleotides: Implications for triplet repeat expansions in the human genome. Hum. Mut. **3**, 19-24.

Bell, G. E., Selby, M. J. & Rutter, W. J. 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeated sequences. Nature **295**, 31-35.

Bentzen, P. , Harris, A. S., and Wright, J. M. 1991. Cloning of hypervariable minisatellite and simple sequence microsatellites for DNA fingerprinting of important aquaculture species of salmonids and tilapia. In: DNA fingerprinting: Approaches and applications, edited by T. Burke, G. Dolf, A. J. Jeffreys and R. Wolff. Basel, Switzerland: Birkhauser Verlag, pp. 243-262.

Bentzen, P. & Wright, J. M. 1993. Nucleotide sequence and evolutionary conservation of a minisatellite variable number tandem repeat cloned from Atlantic salmon, *Salmo salar*. Genome **36**, 271-277.

Bentzen, P., Taggart, C. T., Ruzzante, D. E., and Cook, D. 1996. Microsatellite polymorphism and the population structure of Atlantic *cod (Gadus morhua)* in the Northwest Atlantic. Can. J. Fish. Aquatic. Sciences. **53**, 2706-2721.

Blouin, M.S., Parsons, M., Lacaille, V., and Lotz, S. 1996. Use of microsatellite loci to classify individuals by relatedness. Mol. Ecol. **3**, 393-401.

Bowcock, A., Osborne-Lawrence, S., Barnes, R., Chakravarti, A., Washington, S., and Dunn, C. 1993. Microsatellite Polymorphism Linkage Map of Human Chromosome 13q. Genomics **15**, 376-386.

Breen, M., Downs, P., Irvin, Z., and Bell, K. 1994. Intrageneric amplification of horse microsatellite markers with emphasis on the Przewalski's horse *(Equus. przewalskii)* Anim. Genet. **25**, 401-405.

Brooker, A. L., Cook, D., Bentzen, P., Wright, J. M. & Doyle, R. W. 1994. Organization of microsatellites differs between mammals and cold-water teleost fishes. Can. J. Fish Aqu. Sciences **51**, 1959-1966.

Bruno, B., Poirier, C., Simon-Chazottes, D., Barnier, C. & Guenet, J. 1992. A new strategy useful for rapid identification of microsatellites from DNA libraries with large size inserts. Nuc. Acids Res. **20**, 3665-3669.

Buchanan, F. C., Littlejohn, R.P., Galloway, S.M. & Crawford, A. M. 1993. Microsatellites and associated repetitive elements in the sheep genome. Mamm. Genome **4**, 258-264.

Budowle, B., Giusti, A. M., Waye, J. S., Baechtel, F. S., Fourney, R. M., Adams, D. E., Presley, L. A., Deadman, H. A. and Monson, K. L. 1991. Fixed BIN analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci for use in forensic comparisons. Am. J. Hum. Genet. **48**, 841-855.

Burke, T. & Bruford, M. W. 1987. DNA fingerprinting in birds. Nature **327**, 149-152.

Callen, D. F., Thompson, A. D., Shen, Y., Phillips, H. A., Richards, R. I., Mulley, J. C. & Sutherland, G. R. 1993. Incidence and origin of "null" alleles in the $(AC)_n$ microsatellite markers. Am. J. Hum. Genet. **52**, 922-927.

Carter, R. E., Mair, G. C., Skibinski, D. O. F., Parkin, D. T. and Beardmore, J. A. 1991. The application of DNA fingerprinting in the analysis of gynogenesis in tilapia. Aquaculture **95**, 41-52.

Carvalho, G. R. & Hauser, L. 1994. Molecular genetics and the stock concept in fisheries. Rev. Fish Biol. Fish. **4**, 326-350.

Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. G. and Nelson, D. L. Triplet Repeat Mutations in Human Disease. Science **256**, 784-789.

Chuang, S. E., and Blattner, F. R. 1994. Ultrafast DNA recovery from agarose by centrifugation through a paper slurry. Biotechniques **17**, 634.

Chakraborty, R., Kimmel, M., Stivers, D. N, Davison, L. J., and Deka, R. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Nat. Acad. Sci., USA. **94**, 1041-1047.

Colbourne, J. K., Neff, B. D., Wright, J. M. & Gross, M. R. 1996. DNA fingerprinting of bluegill sunfish *(Lepomis macrochirus)* using $(GT)_n$ microsatellites and its application to assessment of the relative fitness of male mating strategies. Can. J. Fish Aqu. Sciences **53**, 342-349.

Coltman, D. & Wright, J. M. 1994. Can SINEs: A family of tRNA-derived retroposons specific to the superfamily Canoidea. Nuc. Acids Res. **22**, 2726-2730.

Crouau-Roy, B., Bouzekri, N., Carcassi, C., Clayton, J., Contu, L., and Cambon-Thomsen, A. 1996. Strong association between microsatellites and an HLA-B, DR haplotype (B18-DR3): Implication for microsatellite evolution. Immunogenetics **43**, 255-260.

Dallas, J. F. 1988. Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc. Nat. Acad. Sci. USA **85**, 6831-6835.

Davidson, W. S., Birt, T. P., and Green, J. M. 1989. A review of genetic variation in Atlantic salmon, *Salmo salar*, and its importance for stock identification, enhancement programmes and aquaculture. J. Fish Biol. **34,** 547-560.

Devlin, B., Risch, N. & Roeder, K. 1991. Estimation of allele frequencies for VNTR loci. Am. J. Hum. Genet. **48**, 662-676.

Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. & Freimer, N. B. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proc. Nat. Acad. Sci. USA **91**, 3166-3170.

Dow, B. D., and Ashley, M. V. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol. Ecol. **5**, 615-629.

Doyle, R. W. & Herbinger, C. 1995. The use of DNA fingerprinting for high-intensity, within-family selection in fish breeding. Proceedings of the Fifth World Congress on Genetic Applications to Live Stock Production **19**, 364-370.

Doyle, R. W., Herbinger, C., Taggart, C. T. & Lochmann, S. 1995. Use of DNA microsatellite fingerprinting to analyse genetic correlations between hatchery and natural fitness. In Transactions of the American Fisheries Society Symposium 15 (Schramm, H. L. & Piper, R. G. eds). pp 205-211. American Fisheries Society, Maryland

Edwards, A., Civitello, A., Hammond, H. A. & Caskey, C. T. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. **49**, 746-756.

Edwards, A., Hammond, H. A., Jin, L., Caskey, C.T. & Chakraborty, R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics **12**, 241-253.

Ellegren, H. 1992. Polymerase chain reaction (PCR) analysis of microsatellites-a new approach to studies of genetic relationships in birds. Auk **109**, 886-895.

Ellegren, H. 1995. Mutation rates at porcine microsatellite loci. Mamm. Genome **6**, 376-377.

Ellegren, H. 1996. Directional evolution in germline microsatellite mutations. Nature Genet. **13**, 391-393.

Estoup, A., Presa, P., Krieg, F., Vaiman, D., and Guyomard, R. 1993. $(CT)_n$ and $(GT)_n$ microsatellites: A new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity **71**, 488-496.

Farr, C. J. & Goodfellow, P. N. 1991. DNA typing: new variations on the theme. Nature **354**, 184.

Ferguson, M. 1994. The role of molecular genetic markers in the management of cultured fishes. Rev. Fish Biol. Fish. **4**, 351-373.

Galvin, P., McGregor, D. & Cross, T. 1995a. A single locus minisatellite as a hypervariable genetic marker in gadoid species. Aquaculture **137**, 31-40.

Galvin, P., Sadusky, T., McGregor, D. & Cross, T. 1995b. Population genetics of Atlantic cod using amplified single locus minisatellite VNTR analysis. J. Fish Biol. **47A**, 200-208.

Galvin, P., McKinnell, S., Taggart, J. B., Ferguson, A., O'Farrell, M., and Cross, T. F. (1995c). Genetic stock identification of Atlantic salmon using single locus minisatellite DNA profiles. J. Fish Biol. **47A**, 186-199.

George, M., Lequarre, A. S., Castelli, M., Hanset, R. & Vassart, G. 1988. DNA fingerprinting in domestic animals using four different minisatellite probes. Cytogenetetics and Cell Genet. **47**, 127-131.

Gill, P., Kimpton, C., D'Aloja, E., Anderson, J. F., Bar, W., Brinkmann, B., Holgerson, S., Johnsson, V., Kloosterman, A. D., Lareu, M. V., Nellemann, L., Pfitzinger, H., Phillips, C. P., Schmitter, H., Schneider, P. M., and Stenersen, M. 1994. Report on the European DNA profiling group (EDNAP)-towards standardisation of short tandem repeat (STR) loci. Forensic Science Int. **65**, 51-59.

Gill, P., Woodroffe, S., Lygo, J. E. & Millican, E. S. 1991. Population genetics of four hypervariable loci. International J. of Legal Medicine **104**, 221-227.

Goff, D. J., Galvin, K., Katz, H., Westerfield, M., Lander, E. S. & Tabin, C. J. 1992. Identification of polymorphic simple sequence repeats in the genome of the zebrafish. Genomics **14**, 200-202.

Goldstein,D.B, and Clark, A.G. 1995. Microsatellite variation in North American populations of *Drosophila melanogaster.* Nuc. Acids Res. **23**, 3882-3886.

Goodier, J. L., and Davidson, W. S. 1994. Identification of minisatellite loci in the genome of Atlantic salmon. J. Mol. Biol. **241**, 26-34.

Guo, S. W. and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics **48**, 361-372.

Hammond, H. A., Jin, L., Zhong, Y., Caskey, C. T., and Chakraborty, R. 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am. J. Hum. Genet. **55**, 175-189.

Harris, A. S. 1995. Organization and evolution of minisatellite VNTR's in the tilapiine genome. Ph.d. thesis. Dalhousie University, Canada.

Harris, A. S. & Wright, J. M. 1995. Nucleotide sequence and genomic organization of cichlid fish minisatellites. Genome **38**, 177-184.

Harris, A. S., Bieger, S., Doyle, R. W. & Wright, J. M. 1991. DNA fingerprinting of tilapia, *Oreochromis niloticus*, and its application to aquaculture genetics. Aquaculture **92**, 157-163.

Hastbacka, J., de-la-Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. & Lander, E. 1992. Linkage disequlibrium mapping in isolated founder populations: Disatrophic dysplasia in Finland. Nature Genet. **2**, 204-211.

Hauge, X. Y. & Litt, M. 1993. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by PCR. Hum. Mol. Genet. **2**, 411-415.

Henderson, S. T. & Petes, T. D. 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol. Cell Biol. **12**, 2749-2757.

Herbinger, C. M., O'Reilly, P. T. , Doyle, R. W., Wright, J. M., and O'Flynn, F, 1997. Early growth performance of Atlantic salmon full-sib families reared in single family tanks versus in mixed family tanks. Aquaculture (submitted).

Herbinger, C. M., Doyle, R. W., Taggart, C. T., Lochmann, S., Wright, J. M. & Brooker, A. L. 1997. Family relationships and effective population size in a natural cohort of Atlantic cod *(Gadus morhua)* larvae. Can. J. Fish Aqu. Sciences **54**,11-18.

Herbinger, C. M., R. W. Doyle, E. R. Pitman, D. Paquet, K. A. Mesa, D. B. Morris, J. M. Wright and D. Cook. 1995. DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. Aquaculture **137**, 245-256.

Hinegardner, R., and Rosen, D. E. 1972. Cellular DNA content and the evolution of teleostean fises. Am. Nat. **106**, 621-644.

Hovey, S. J., King, D. P. F., Thompson, D., and Scott, A. 1989. Mitochondrial DNA and allozyme analysis of Atlantic salmon, *Salmo salar* L., in England and Wales. J. Fish Biol. **35(A)**, 253-260.

Huang, T. H. M., Cottingham, R. W., Ledbetter, D. H. & Zoghbi, H. Y. 1992. Genetic mapping of four dinucleotide repeat loci, DXS453, DXS458, DXS454, and DXS424, on the X chromosome using multiplex polymerase chain reaction. Genomics **13**, 375-380.

Hubert, R., Weber, J. L., Schmitt, K., Zhang, L. & Arnheim, N. 1992. A new source of polymorphic DNA markers for sperm typing: Analysis of microsatellite repeats in single cells. Am. J. Hum. Genet. **51**, 985-991.

Imbert, G. *et al.* 1993. Origin of the expansion mutation in myotonic dystrophy. Nature Genet. **4**, 72-76.

Jarne, P. and Lagoda, P. J. L. 1996. Microsatellites, from molecules to populations and back. Trend. Evol. Ecol. **11**, 424-429.

Jeffreys, A. J., Wilson, V. & Thein, S. L. 1985. Hypervariable "minisatellite" regions in human DNA. Nature **314**, 67-74.

Jeffreys, A. J., Micola, J. R., Wilson, V. & Wong, Z. 1988a. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature **332**, 278-281.

Jeffreys, A. J., Wilson, V. & Keyte, J. 1988b. Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. Nuc. Acids Res. **16**, 10953-10971.

Jeffreys, A. J., Neumann, R. & Wilson, V. 1990. Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell **60**, 473-485.

Jeffreys, A. J., MacLeod, A., Tamaki, K., Neil, D. L., & Monckton, D. G. 1991. Minisatellite repeat coding as a digital approach to DNA typing. Nature **354**, 204-209.

Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L. & Armour, J. A. L. 1994. Complex gene conversion events in germline mutation at human minisatellites. Nature Genet. **6**, 136-145.

Jones, A. G., and Avise, J. C. 1997. Microsatellite analysis of maternity and the mating system in the Gulf pipefish, *Syngnathus scovelli*, a species with male pregnancy and sex-role reversal. Mol. Ecol. **6**, 203-213.

Jordan, W.C., and Youngson, A. F. 1991. Genetic protein variation and natural selection in Atlantic salmon *(Salmo salar, L.)* parr. J. Fish Biol. **39(A)**, 185-192.

Kaukinen, J. & Varvio, S. 1992. Artiodactyl retroposons: Association with microsatellites and use in SINEmorph detection by PCR. Nuc. Acids Res. **20**, 2955-2958.

Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M., and Okada, N. 1991. Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. Proc. Natl. Acad. Science. U.S.A. **88**, 2326-2330 (1991)

Kimpton, C. P., Fisher, S., Walton, A., Urquhart, Millican, E. S. & Adams, M. 1993. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods and Applications **3**, 13-22.

Koorey, D.J., Bishop, G. A., and McCaughan, G.W. 1993. Allele non-amplification: a source of confusion in linkage studies employing microsatellite polymorphisms. Hum. Mol. Genet. **2**, 289-291.

Kristensen, T., Lopez, R., and Prydz, H. 1992. An estimate of the sequencing error frequency in the DNA sequence databases. DNA Seq. **2**, 343-346.

Kwiatkowski, D.J., Henske, E. P., Weimer, K., Ozelius, L., Gusella, J. F. & Haines, J. 1992. Construction of a GT polymorphism map of human 9q. Genomics **12**, 229-240.

La Spada, A. R., Roling, D. B., Harding, A. E., Warner, C.L., Spiegel, R., Hausmanowa-Petrusewicz, I., Yee, W. C. & Fischbeck, K. H. 1992. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat X-linked spinal and bulbar muscular atrophy. Nature Genet. **2**, 301-304.

Laughlin, T. F. & Turner, B. J. 1994. Multilocus DNA fingerprinting detects population differentiation in the outbred and abundant fish species *Poecilia latipinna*. Mol. Ecol. **3**, 263-266.

Levinson, G. & Gutman, G. A. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol. Biol. and Evol. **4**, 203-221.

Li, S., McInnis, M. G., Margolis, R. L., Antonarakis, S. E. & Ross, C. A. 1993. Novel triplet repeat containing genes in human brain: Cloning, expression and length polymorphisms. Genomics **16**, 572-579.

Litt, M. & Luty, J. A. 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. **4**, 397-401.

Litt, M., Hauge, X., and Sharma, V. 1993. Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. Biotechniques. **15**, 280-284.

Luty, J. A., Guo, Z., Willard, D. H. F., Ledbetter, D. H., Ledbetter, S. & Litt, M. 1990. Five polymorphic microsatellite VNTRs on the human X chromosome. Am. J. Hum. Genet. **46**, 776-783.

MacHugh, D. E., R. T. Loftus, D. G. Bradley, P. M. Sharp, and P. Cunningham. 1994. Microsatellite DNA variation within and among European cattle breeds. Proc. Nat. Acad. Sci. U.S.A. **75**, 2868-2872.

Maha, G. C., Mason, J. M., Stuhlmiller, G. M., and Heine, U. 1995. Reply to Pena. Amer. Jour. of Hum. Genet. **56**, 1506-1507.

Mahtani, M. M. & Willard, H. F. 1993. A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: Implications for mechanisms of mutation at short tandem repeat loci. Hum. Mol. Genet. **2**, 431-437.

Matise, T. C., Perlin, M., and Chakravarti, A. 1994. Automated construction of genetic linkage maps using an expert system (MultiMap): A human genome linkage map. Nature Genet. **6**, 384-390

McConnell, S. K., O'Reilly, P., Hamilton, L., Wright, J.M. & Bentzen, P. 1995a. Polymorphic microsatellite loci from Atlantic salmon *(Salmo salar)*: Genetic differentiation of North American and European populations. Can. J. Fish Aqu. Sciences **52**, 1863-1872.

McConnell, S. K, Hamilton, L., Morris, D., Cook, D., Paquet, D., Bentzen, P. & Wright, J. M. 1995b. Isolation of microsatellite loci in Atlantic salmon and their application to the population genetics of Canadian East coast stocks. Aquaculture **137**, 19-30.

McConnell, S. K., Ruzzante, D. E., O'Reilly, P. T., Hamilton, L. and Wright, J. M. 1997. Microsatellite loci reveal highly significant genetic differentiation between Atlantic salmon *(Salmo salar)* populations from the East Coast of Canada. Mol. Ecol. **6**, 282-296.

McElligott, E. A., and Cross, T. F. 1991. Protein variation in wild Atlantic salmon, with particular reference to southern Ireland. J. Fish. Biol. **39(A)**, 35-42.

Morin, P. A., Wallis, J., Moore, J. J., and Woodruff, D. S. 1994. Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. Mol. Ecol. **3**, 469-478.

Morral, N., Nunes, V., Casals, T. & Estevil, X. 1991. CA/GT microsatellite alleles within the cystic fibrosis transmembrane conductance regulator (CFTR) gene are not generated by unequal crossingover. Genomics **10**, 692-698.

Morris, D. B. 1993. The isolation and characterization of microsatellites from rainbow trout *(Oncorhynchus mykiss)*. M.Sc. Thesis, Dalhousie University.

Morris, D. B., K. R. Richard, J. M. Wright. 1996. Polymorphic microsatellites from rainbow trout *(Oncorhynchus mykiss)* and their use for genetic study of salmonids. Can. J. Fish. Aquat. Sci. **53**, 120-126.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harbor Symp. Quant. Biol. **51**, 263.

Murata,S., Takasaki,N., Saitoh,M. and Okada,N. 1993. Determination of the phylogenetic relationships among Pacific salmonids using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc.Natl.Acad.Science. U.S.A. **90**, 6995-6999.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M. & Martin, C. 1987. Variable numbers of tandem repeat (VNTR) markers for human gene mapping. Science **235**, 1616-1622.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89,** 583-590.

Nelson, D. L. & Warren, S. T. 1993. Trinucleotide repeat instability: When and where? Nature Genet. **4**,107-108.

Nielsen, J. L., Gan, C. A., Wright, J. M., Morris, D. B. & Thomas, K. 1995. Biogeographic distributions of mitochondrial and nuclear markers for southern steelhead. Mol. Mar. Biol. and Biotech. **3**, 281-293.

Nielsen, J. L., Gan, C., Wright, J.M. & Thomas, W.K. 1994. Phylogeographic patterns in California steelhead as determinied by mtDNA and microsatellite analyses. CalCOFI REP **35**, 90-92.

O'Connell, M., Skibinski, D. O. F., and Beardmore, J. A. 1995. Mitochondrial DNA and allozyme variation in Atlantic salmon *(Salmo salar)* populations in Wales. Can. J. Fish. Aquat. Sci. **52**, 171-178.

O'Connell, M., Danzmann, R. G., Cornuet, J. M., Wright, J. M., and Ferguson, M. M. 1996a. Differentiation of rainbow trout populations in Lake Ontario and the evaluation of the stepwise mutation and infinite allele mutation models using microsatellite variability. Can. J. Fish. Aquat. Science. **54** (in press).

O'Connell, M., Dillon, M. C., and Wright, J. M. 1996b. Development of microsatellite markers for genetic discrimination of Prince William Sound herring populations. Report to the Alaska Department of Fish and Game. 31 pp.

O'Connell, M. and Wright, J. M. 1997. Microsatellite DNA in fishes. Rev. Fish Biol. and Fisheries. **7**, 1-33.

O'Reilly, P., and Wright, J. M. 1995. The evolving technology of DNA fingerprinting and its application to fisheries and aquacluture. J. Fish Biol. **47A**, 29-55.

O'Reilly, P. T., Hamilton, L. C., McConnell, S. K., and Wright, J. M. 1996. Rapid analysis of genetic variation in Atlantic salmon *(Salmo salar)* by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Can. J. Fish. Aquat. Science **53**, 2292-2298.

Oudet, C., Mornet, E., Serre, J. L., Thomas, F., Lentes-Zengerling, S., Kretz, C., Deluchat, C., Tejada, I., Boue, J., Boue, A. & Mandel, J. L. 1993. Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that fragile X chromosomes are derived from a small number of founder chromosomes. Am. J. Hum. Genet. **52**, 297-304.

Paetkau, D. & Strobeck, C. 1994. Microsatellite analysis of genetic variation in black bear populations. Mol. Ecol. **3**, 489-495.

Park, L. K. & Moran, P. 1994. Developments in molecular genetic techniques in fisheries. Rev. Fish Biol. Fish. **4**, 272-299.

Paxton, R. J., Thoren, P. A., Tengo, J., Estoup, A., and Pamilo, P. 1996. Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using microsatellites. Mol. Ecol. **5**, 511-521.

Pemberton, J. M., Slate, J., Bancroft, D. R., and Barrett, J. A. 1995. Non-amplifying alleles at microsatellite loci: a caution for parentage and population studies. Mol. Ecol. **4**, 249-252.

Petrukhin, K. E., Speer, M. C., Cayanis, E., De Fatima, Bonaldo, M., Tantravahi, U., Soares, M. B., Fischer, S. G., Warburton, D., Gilliam, T. C., and Ott, J. A Microsatellite Genetic Linkage Map of Human Chromosome 13. Genomics **15**, 76-85.

Phillips, H. A., Hyland, V. J., Holman, K., Callen, D. F., Richards, R. I. & Mulley, J. C. 1991. Dinucleotide repeat polymorphism at D16S287. Nuc. Acids Res. **19**, 6664.

Prodohl, P. A., Taggart, J. B. & Ferguson, A. 1992. Genetic variability within and among sympatric brown trout *(Salmo trutta)* populations: Multi-locus DNA fingerprint analysis. Hereditas **117**, 45-50.

Puers, C., Hammond, H. A., Caskey, C. T., Lins, A. M., Sprecher, C. J., Brinkmann, B. & Schumm, J. W. 1994. Allelic ladder characterization of the short tandem repeat polymorphism located in the 5' flanking region to the human coagulation factor XIII A subunit gene. Genomics **23**, 260-264.

Raymond, M., and F. Rousett. 1995. Genepop (Version-1.2). Population genetics software for exact tests and ecumenicism. J. Hered. **86**, 248-249.

Richards, R. I., and Sutherland, G. R. 199 4. Simple repeat DNA is not replicated simply. Nature Genet. **6**, 115-116.

Richards, R. I., Holman, K., Friend, K., Kremer, E., Hillen, D., Staples, A., Brown, W.T., Goonewardena, P., and Tarleton, J. 1992. Evidence of founder chromosomes in fragile X syndrome Nature Genet. **1**, 257-260.

Rico, C., Kuhnlein, U. & Fitzgerald, G. J. 1991. Male reproductive tactics in threespine stickleback: an evaluation by DNA fingerprinting. Symposium on Biochemical Genetics and Taxonomy of Fish. Queen's University, Belfast.

Rico, C., Rico, I. & Hewitt, G. 1994. An optimized method for isolating and sequencing large $(CA/GT)_n$(n>40) microsatellites from genomic DNA. Mol. Ecol. **3**, 181-182.

Rico,C., Zadworny,D., Kuhnlein, U., and Fitzgerald, G.J. 1993. Characterization of hypervariable microsatellite loci in the threespine stickleback *Gasterosteus aculeatus* Mol. Ecol. **2,** 271-272.

Ritter, J. A. 1989. Marine migration and natural mortality of North American Atlantic salmon *(Salmo salar L.).* Can. MS Rep. Fish. Aquat. Science. No 2041. 136p.

Rohrer, G. A., L. J. Alexander, J. W. Keele, T. P. Smith and C. W. Beattie. 1994. A microsatellite linkage map of the porcine genome. Genetics, **136**, 231-245.

Ruzzante, D. E. 1997. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. Can. J. Fish Aqu. Sciences (in press).

Sakagami, M., Tozaki, T., Mashima, S., Hirota, K., and Mukoyama, H. 1995. Equine parentage testing by microsatellite locus at chromosome 1q2.1. Anim. Genet. **26**, 123.

Sakamoto, T., Okamoto, N., Ikeda, Y., Nakamura, Y. & Sato, T. 1994. Dinucleotide-repeat polymorphism in DNA of rainbow trout and its application in fisheries science. J. Fish Biol. **44**, 1093-1096.

Sambrook, J., Fritsh, E. F., and Mauniatis, T. 1989. "Molecular Cloning: A laboratory manual." 2nd Ed., Cold Spring Harbour Laboratory Press, New York

Sanger, F. S., Nicklen, S. and Coulsen, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA . **74**, 5463-5467.

Santos, F. R., Pena, S. D. J. & Epplen, J. T. 1993. Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum. Genet. **90**, 655-656.

Scribner, K. T., Gust, J. R., and Fields, R. L. 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetics applications. Can. J. Fish. Aquat. Sciences. **53**, 833-841.

Schlotterer, C. & Tautz, D. 1992. Slippage synthesis of simple sequence DNA. Nuc. Acids Res. **20**, 211-215.

Shelbourne, P., Wingvist, R., Kunert, E., Davies, J., Leisti, J., Thiele, H., Bachmann, H., Buxton, J., Williamson, B., Johnson, K. 1992. Unstable DNA may be responsible for the incomplete penetrance of the myotic dystrophy phenotype. Hum. Mol. Genet. **1**, 467-473.

Sirugo, G., Keats, B., Fujita, R., Duclos, F., Purohit, K., Koenig, M., Mandel, J. L. 1992. Friedreich ataxia in Louisiana Acadians: Demonstration of a founder effect by analysis of microsatellite-generated extended haplotypes. Am. J. Hum. Genet. **50**, 559-566.

Slettan, A. I, Olsaker, I. & Lic. O. 1993. Isolation and characterization of $(GT)_n$ repetitive sequences from Atlantic Salmon, *Salmo salar* L. Anim. Genet. **24**, 195-197.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**, 503.

Sprecher, J. C., Puers, C., Lins, A. M., and Schumm, J. W. General approach to analysis of short tandem repeat loci. 1996. Biotechniques **20**, 266-276.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E. & Moyzis, R. K. 1991. Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics **10**, 807-815.

Stevens, T. A., Withler, R. E., Goh, S.H. & Beacham, T. D. 1993. A new multilocus probe for DNA fingerprinting in chinook salmon *(Oncorhynchus tshawytscha)*, and comparisons with a single-locus probe. Can. J. Fish Aqu. Sciences **50**, 1559-1567.

Strassmann, J. E., Solis, C. R., Barefield, K., and Queller, D. C. 1996. Trinucleotide microsatellite loci in a swarm-founding neotropical wasp, *Parachartergus colobopterus* and their usefulness in other social wasps. Mol. Ecol. **5**, 459-461.

Strassmann, J. E., Solis, C. R., Barefield, K., Hughes, C. R. and Queller, D. C. 1997. Trinucleotide microsatellite loci for a social wasp, Polistes. Mol. Ecol. **6**, 97.

Straub, R. E., Speer, M. C., Luo, Y., Rojas, K., Overhauser, J., Ott, J., and Gilliam, T. C. 1993. Identification and localization of microsatellite markers covering human chromosome 18. Genomics **15**, 48-56.

Taggart, J. B. & Ferguson, A. 1990. Hypervariable minisatellite DNA single locus probes for the Atlantic salmon, *Salmo salar* L. J. Fish Biol. **37**, 991-993.

Talbot, C. C. Jr., Avramopoulos, D., Gerken, S., Chakravarti, A., Armour,J. A., Matsunami, N., White, R., and Antonarakis, S. E. 1995. The tetranucleotide repeat polymorphism D21S1245 demonstrates hypermutability in germline and somatic cells. Hum. Mol. Genet. **4**, 1193-1199.

Tautz, D. & Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nuc. Acids Res. **12**, 4127-4138

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nuc. Acids Res. **17**, 6463-6471.

Taylor, E. B., Beacham, T. D. & Kaeriyama, M. 1994. Population structure and identification of North Pacific Ocean chum salmon (*Oncorhynchus keta*) revealed by an analysis of minisatellite DNA variation. Can. J. Fish Aqu. Sciences **51**, 1430-1442.

Thomas, M. R., Cain, P., and Scott, N. S. 1994. DNA typing of grapevines: A universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol. Biol. **25**, 939-949.

Urquhart, A., Nicola, J., Oldroyd, Kimpton, C. P. & Gill, P. 1995. Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. BioTechniques **18**, 116-121.

Usha, A. P., Simpson, S. P., and Williams, J. L. 1995. Probability of random sire exclusion using microsatellite markers for parentage verification. Anim. Genet. **26**, 155-161.

Valdes, A. M., Slatkin, M. & Freimer, N. B. 1993. Allele frequencies at microsatellite loci: The stepwise mutation model revisited. Genetics **133**, 737-749.

Waples, R. S. 1897. A multi-species approach to the analysis of gene flow in marine shore fishes. Evolution, **41**: 385-400.

Ward, R. D. & Grewe, P. M. 1994. Appraisal of molecular genetic techniques in fisheries. Rev. Fish Biol. Fish. **4**, 300-325.

Weber, J. L. & C. Wong. 1993. Mutation of human short tandem repeats. Human Molecular Genetetics **2**, 1123-1128.

Weber, J. L. & May, P. E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44**, 388-396.

Weber, J. L. 1990. Informativeness of human (dC-dA)n·(dG-dT)n polymorphisms. Genomics **7**, 524-530.

Weber, J. L. 1989. Length polymorphism in (dC-dA)n- (dG-dT)n sequences detected using the polymerase chain reaction. *In* "Polymerase chain reaction" (H. A. Erlich, R. Gibbs, and H. H. Kazanian, Jr., Eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, New York.

Weir, B. S. and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. Evolution, **38**: 1358-1370.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. & Lathrop, M. 1992. A second-generation linkage map of the human genome. Nature **359**, 794-801.

Westneat, D. F., W. A. Noon, H. K. Reeve and C. F. Aquadro. 1988. Improved hybridization conditions for DNA "fingerprints" probed with M13. Nuc. Acids Res. **14**: 4161.

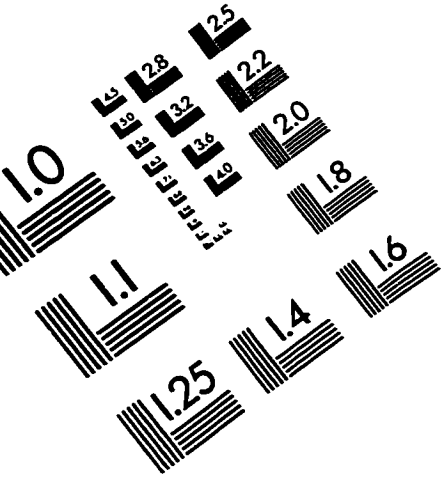
Whitmore, D. H., Elliott, N. G., Grewe, P. & Craft, C. M. 1992. Gene amplification permits minimally invasive analysis of fish mitochondrial DNA. Trans. Amer. Fish. Soc. **121**, 170-177.

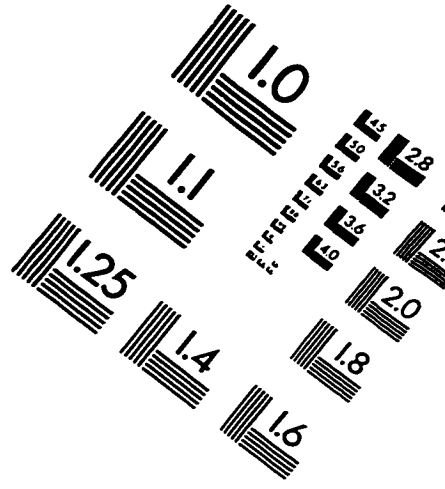
Wilke, K., Jung, M., Chen, Y. & Geldermann, H. 1994. Porcine (GT)n sequences: Structure and association with dispersed and tandem repeats. Genomics **21**, 63-70.

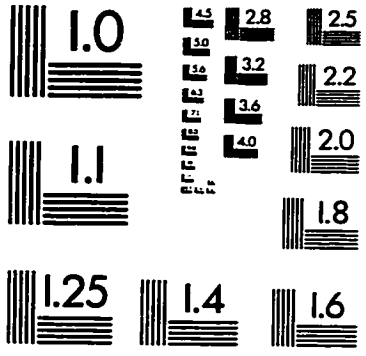
Wright, J. M. 1993. DNA fingerprinting of fishes. In *Biochemistry and molecular biology of fishes*, *volume 2*, edited by Hochachka and Mommsen. Elsevier Science Publishers B.V., pp. 57-91.

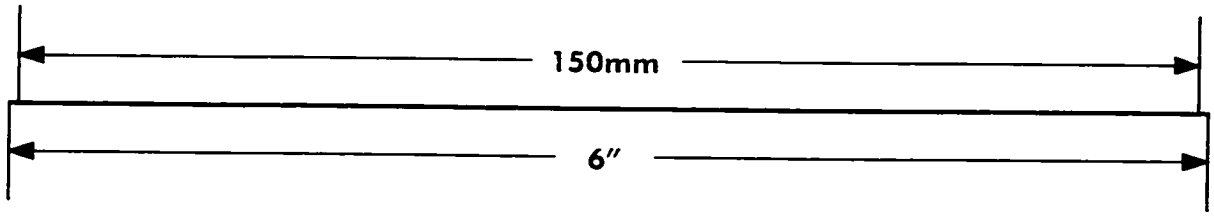
Wright, J. M., and P. Bentzen. 1994. Microsatellites: genetic markers for the future. Rev. Fish Biol. Fish. **4**, 384-388.

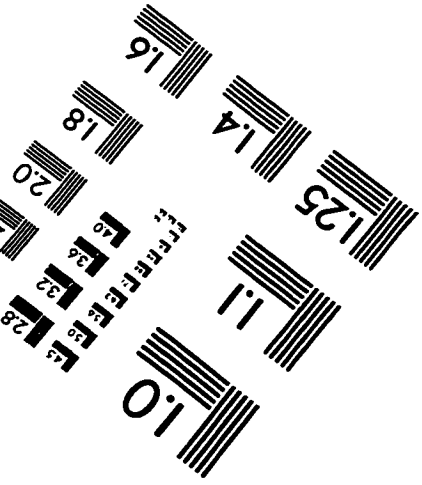
IMAGE EVALUATION TEST TARGET (QA-3)

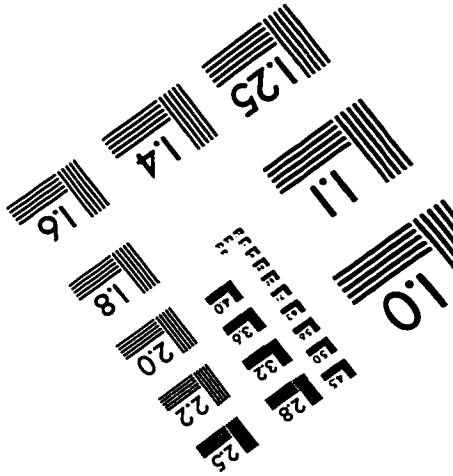












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989