National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

Organization and Evolution of Minisatellite VNTR's in the Tilapiine Genome

by

Andrew S. Harris

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

at

Dalhousie University
Halifax, Nova Scotia

March, 1995

ISBN 0-612-05272-9

Canada

Name _____

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

SUBJECT TERM

SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture	0729
Art History	0377
Cinema	0900
Dance	0378
Fine Arts	0357
Information Science	0723
Journalism	0391
Library Science	0399
Mass Communications	0708
Music	0413
Speech Communication	0459
Theater	0465

EDUCATION

General	0515
Administration	0514
Adult and Continuing	0516
Agricultural	0517
Art	0273
Bilingual and Multicultural	0282
Business	0688
Community College	0275
Curriculum and Instruction	0727
Early Childhood	0518
Elementary	0524
Finance	0277
Guidance and Counseling	0519
Health	0680
Higher	0745
History of	0520
Home Economics	0278
Industrial	0521
Language and Literature	0279
Mathematics	0280
Music	0522
Philosophy of	0998
Physical	0523
Psychology	0525
Reading	0535
Religious	0527
Sciences	0714
Secondary	0533
Social Sciences	0534
Sociology of	0340
Special	0529
Teacher Training	0530
Technology	0710
Tests and Measurements	0288
Vocational	0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language	
General	0679
Ancient	0289
Linguistics	0290
Modern	0291
Literature	
General	0401
Classical	0294
Comparative	0295
Medieval	0297
Modern	0298
African	0316
American	0591
Asian	0305
Canadian (English)	0352
Canadian (French)	0355
English	0593
Germanic	0311
Latin American	0312
Middle Eastern	0315
Romance	0313
Slavic and East European	0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy	0422
Religion	
General	0318
Biblical Studies	0321
Clergy	0319
History of	0320
Philosophy of	0322
Theology	0469

SOCIAL SCIENCES

American Studies	0323
Anthropology	
Archaeology	0324
Cultural	0326
Physical	0327
Business Administration	
General	0310
Accounting	0272
Banking	0770
Management	0454
Marketing	0338
Canadian Studies	0385
Economics	
General	0501
Agricultural	0503
Commerce Business	0505
Finance	0508
History	0509
Labor	0510
Theory	0511
Folklore	0358
Geography	0366
Gerontology	0351
History	
General	0578
Ancient	0579
Medieval	0581
Modern	0582
Black	0328
African	0331
Asia, Australia and Oceania	0332
Canadian	0334
European	0335
Latin American	0336
Middle Eastern	0333
United States	0337
History of Science	0585
Law	0398
Political Science	
General	0615
International Law and Relations	0616
Public Administration	0617
Recreation	0814
Social Work	0452
Sociology	
General	0626
Criminology and Penology	0627
Demography	0938
Ethnic and Racial Studies	0631
Individual and Family Studies	0628
Industrial and Labor Relations	0629
Public and Social Welfare	0630
Social Structure and Development	0700
Theory and Methods	0344
Transportation	0709
Urban and Regional Planning	0999
Women's Studies	0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture	
General	0473
Agronomy	0285
Animal Culture and Nutrition	0475
Animal Pathology	0476
Food Science and Technology	0359
Forestry and Wildlife	0478
Plant Culture	0479
Plant Pathology	0480
Plant Physiology	0817
Range Management	0777
Wood Technology	0746
Biology	
General	0306
Anatomy	0287
Biostatistics	0308
Botany	0309
Cell	0379
Ecology	0329
Entomology	0353
Genetics	0369
Limnology	0793
Microbiology	0410
Molecular	0307
Neuroscience	0317
Oceanography	0416
Physiology	0433
Radiation	0821
Veterinary Science	0778
Zoology	0472
Biophysics	
General	0786
Medical	0760

EARTH SCIENCES

Biogeochemistry	0425
Geochemistry	0996
Geodesy	0370
Geology	0372
Geophysics	0373
Hydrology	0388
Mineralogy	0411
Paleobotany	0345
Paleoecology	0426
Paleontology	0418
Paleozoology	0985
Palynology	0427
Physical Geography	0368
Physical Oceanography	0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences	0768
Health Sciences	
General	0566
Audiology	0300
Chemotherapy	0992
Dentistry	0567
Education	0350
Hospital Management	0769
Human Development	0758
Immunology	0982
Medicine and Surgery	0564
Mental Health	0347
Nursing	0569
Nutrition	0570
Obstetrics and Gynecology	0380
Occupational Health and Therapy	0354
Ophthalmology	0381
Pathology	0571
Pharmacology	0419
Pharmacy	0572
Physical Therapy	0382
Public Health	0573
Radiology	0574
Recreation	0575
Speech Pathology	0460
Toxicology	0383
Home Economics	0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry	
General	0485
Agricultural	0749
Analytical	0486
Biochemistry	0487
Inorganic	0488
Nuclear	0738
Organic	0490
Pharmaceutical	0491
Physical	0494
Polymer	0495
Radiation	0754
Mathematics	0405
Physics	
General	0605
Acoustics	0986
Astronomy and Astrophysics	0606
Atmospheric Science	0608
Atomic	0748
Electronics and Electricity	0607
Elementary Particles and High Energy	0798
Fluid and Plasma	0759
Molecular	0609
Nuclear	0610
Optics	0752
Radiation	0756
Solid State	0611
Statistics	0463

Applied Sciences

Applied Mechanics	0346
Computer Science	0984
Engineering	
General	0537
Aerospace	0538
Agricultural	0539
Automotive	0540
Biomedical	0541
Chemical	0542
Civil	0543
Electronics and Electrical	0544
Heat and Thermodynamics	0348
Hydraulic	0545
Industrial	0546
Marine	0547
Materials Science	0794
Mechanical	0548
Metallurgy	0743
Mining	0551
Nuclear	0552
Packaging	0549
Petroleum	0765
Sanitary and Municipal	0554
System Science	0790
Geotechnology	0428
Operations Research	0796
Plastics Technology	0795
Textile Technology	0994

PSYCHOLOGY

General	0621
Behavioral	0384
Clinical	0622
Developmental	0620
Experimental	0623
Industrial	0624
Personality	0625
Physiological	0989
Psychobiology	0349
Psychometrics	0632
Social	0451

DEDICATION

I would like to dedicate this thesis to my father who sadly did not live to see its conclusion. I know he would have been proud of me.

TABLE OF CONTENTS

RESULTS

LIST OF FIGURES

PAGE

ABSTRACT

This thesis describes the isolation and characterization of three minisatellites or variable number of tandem repeats (VNTR's) from the cichlid fish, *Oreochromis niloticus*. The human-derived polycore probes of Jeffreys, 33.6 and 33.15, generate multilocus DNA fingerprint patterns of hybridization in tilapia, that are consistent with Mendelian inheritance of minisatellite loci. The utitlity of these human polycore probes as tools for describing inheritance in tilapiines was tested by examining the genomic heterozygocity of gynogenetic individuals. Analysis of hybridization patterns showed reductions in the number of fragments produced in mitotic gynogens over meiogynes and controls, consistent with expected reductions in genomic heterozygocity.

A second objective of this study was the isolation and description of minisatellite loci from tilapia. A genomic DNA library of *O. niloticus* sequences was screened with the human-derived 33.6 and 33.15 probes. This revealed that minisatellite VNTRs related by sequence to these probes are present every 1000 kbp in the tilapia genome. Hybridization studies of cloned VNTRs from tilapia revealed many of the properties described for human minisatellites such as the grouping of minisatellite loci into families based on sequence identity and the clustering of minisatellite arrays.

A third objective of this study was to describe the sequence of VNTR's in tilapia. Variability at these loci is low compared to minsatellites previously described from humans. Repeat unit length and variability of the repeated sequence differed considerably between the loci.

Finally, this study examines the evolution of VNTR's from tilapiines. Each minisatellite was hybridized to genomic DNA from a number of related taxa. Detection of similar sequences in other species varied for the three cloned minisatellites. Clones OniMS17 and OniMS32 showed hybridization only to genomic DNA of species in the tilapiine tribe and the closely related African cichlid *Haplochromis moori*, whereas clone OniMS34 showed hybridisation to the genomic DNA of the South American cichlid *Cichlasoma otofasciatus*. The sequence and varient repeat structure of two of the loci suggest mechanisms of generation for these minisatellites.

LIST OF ABBREVIATIONS

bp	nucleotide base pairs
BSA	bovine serum albumin
Ci	Curie
cm	centimeter
dATP	deoxyadenosine-5'-triphosphate
dGTP	deoxyguanosine-5'-triphosphate
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
IPTG	isopropylthio-β-D-galactoside
kbp	kilo base pairs
LINES	Long Interspersed Nucleotide Sequences
MVR	minisatellite variant repeat
ng	nanogram
PCR	polymerase chain reaction
PFU	plaque forming unit
RNA	ribonucleic acid
RFLPs	restriction fragment length polymorphisms
SDS	sodium dodecylsulphate
SINES	Short Interspersed Nucleotide Sequences
VNTRs	Variable Number of Tandem Repeats
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactoside

ACNOWLEDGEMENTS

I wish to extend my sincere thanks to my supervisor, Dr. Jonathan Wright, for his patience, his guidance, his concern and his inspiration. I would like to thank the members of my committee; Drs. Bili Pohajdak, Roger Doyle and Rick Singer for all their advice and comments made in the lab and in the preparation of this thesis. I also owe a debt of gratitude to my friends, Jens with whom I shared many of the ups and downs of graduate student life, Eileen who made us both feel intellectually challenged and the many others who I have had the good fortune to meet. I also thank the staff and researchers associated with the Marine Gene Probe Laboratory for their invaluable technical experience and camaraderie particularly Drs. Paul Bentzen, Bruce Ramsey, Jason Hoffman and Christophe Herbiger. I am grateful to all of the members of the Biology Department's office, stores and technical staff for their help over the years.

I owe a huge debt to my wife, Dawn, for her support and encouragement and to Benjamin and Evan who made many unknowing sacrifices for me.

Finally I would like to thank Professor David Fensom who lit the spark of inquiry in me.

PUBLICATIONS

Much of the work presented in this thesis has been published or is presently "in press"

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

Harris, A.S. and Wright, J.M. 1994. Organization and evolution of minisatellites from tilapia. Gemome. In Press.

Bentzen, P.B., Harris, A.S. and Wright, J.M. 1991. Cloning of hypervariable and simple sequence microsatellite repeats for DNA fingerprinting of important aquacultural species of salmonids and tilapias. In: DNA fingerprinting: Approaches and Applications. (Eds. T. Burke, G. Dolf, A.J. Jeffreys and R.Wolff). Birkhauser Verlag, Basel. pp. 243-262.

Franck, J.P.C., Harris, A.S., Bentzen, P., Denovan-Wright, E.M. and Wright, J.M. 1991. Organization and evolution of satellite, minisatellite and microsatellite DNAs in teleost fishes. In: Oxford Surveys on Eukaryotic Genes. (Ed. N. Maclean). Oxford University Press, Oxford, U.K. pp. 51-85.

CHAPTER I: INTRODUCTION

A Review of DNA Fingerprinting

Repetitive DNA's

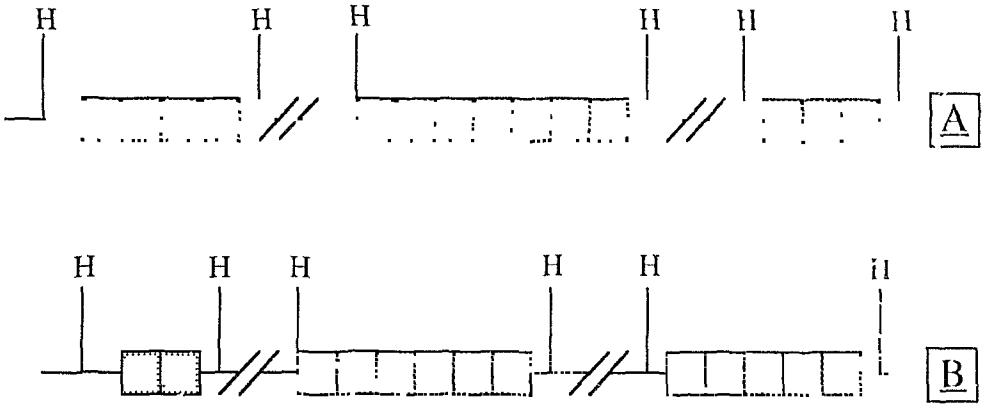
A significant proportion of the eucaryotic genome is composed of repetitive DNA sequences, accounting for 20 to 80% of the total DNA (Britten and Davidson 1971). Originally identified as a major component of eucaryotic genomes by density gradient centrifugation, these classes of sequence were visualized as "satellite" peaks orbiting the bulk of the genomic DNA (Kit 1961). Studies of denatured mouse DNA using reassociation kinetics detailed the extent of repetitive DNA's in the murine genome (Waring and Britten 1966). Further reassociation experiments broadly divided genomic DNA into three distinct classes: highly repetitive, moderately repetitive and unique sequence (Britten and Kohne 1968). More recent studies of mamalian genomes have led to a discrimination between interspersed and clustered repetitive DNAs (Singer 1982). Interspersed repeats, subdivided into Short Interspersed Nucleotide Sequences (SINES) and Long Interspersed Nucleotide Sequences (LINES) are proposed to be of mixed origins. Those of viral origin have dispersed by transposition of RNA polymerase transcript intermediates. Others of non viral origin have resulted from the retrotransposition of RNA transcripts from coding regions as retropseudogenes. Both share a common structural motif in being associated with single copy sequence. Clustered repeats comprise both coding regions such as ribosomal and histone genes and non coding, tandemly repeating sequence termed satellites, minisatellites

or microsatellites. These latter two classes of tandemly repeated DNA sequence have become the center of much recent interest with the advent of DNA fingerprinting techniques (Burke et al. 1991).

Multilocus DNA Fingerprinting

Minisatellites or Variable Number of Tandem Repeats (VNTR's) form some of the most variable sequences in the human genome (Jeffreys et al. 1985a, b, Nakamura et al. 1987). The first demonstration of DNA fingerprinting by Jeffreys used cloned human sequences isolated with a concatenated 33 bp repeat from the human myoglobin gene to detect restriction fragment length polymorphisms (RFLP's) in human genomic DNA (Jeffreys et al. 1985a). When used in Southern blot and hybridization studies, these probes produced individual specific polymorphic patterns. These patterns were presumably generated by mutation at variable number of tandem repeat (VNTR) loci. The novelty and utility of the technique derives from the simultaneous detection of large numbers of loci by virtue of a shared core sequence within the repeating units (Figure 1.1). Sequence analysis of the probes isolated with the 33 bp myoglobin repeat, revealed a common 12 to 15 base pair (bp) 'core' sequence nested within the monomer units. The repeat periodicity of these sequences showed a simple length relationship being equal to, one half of or twice that of the hybridizing 33 bp repeat. Jeffreys has suggested that heteroduplex formation relies not only on sequence similarity but also on periodicity between probe and target DNA repeat structure. The discovery of distinct fingerprint patterns using

Figure 1.1 Schematic showing the generation of multilocus DNA fingerprints by hybridisation of minisatellite 'core' probes to genomic DNA. The top panel shows three minisatellite loci from individuals A and B. A restriction endonuclease (H) that does not cleave at recognition sites within the repeat array is used to digest the genomic DNA. The resulting fragments are fractionated by gel-electrophoresis, Southern transfered to nylon membrane and hybridized under conditions of low stringency to a 'multilocus' probe containing 'core' sequence similar to that of the repeat units. The lower panel shows the pattern of bands generated for the two individuals shown and a putative offspring. All bands in the offspring fingerprint can be assigned to either the male or female parent. For the sake of simplicity only the haploid complement of three VNTR's has been illustrated.


H
H
H
H
H
H
A
H
H
H
H
H
H
B

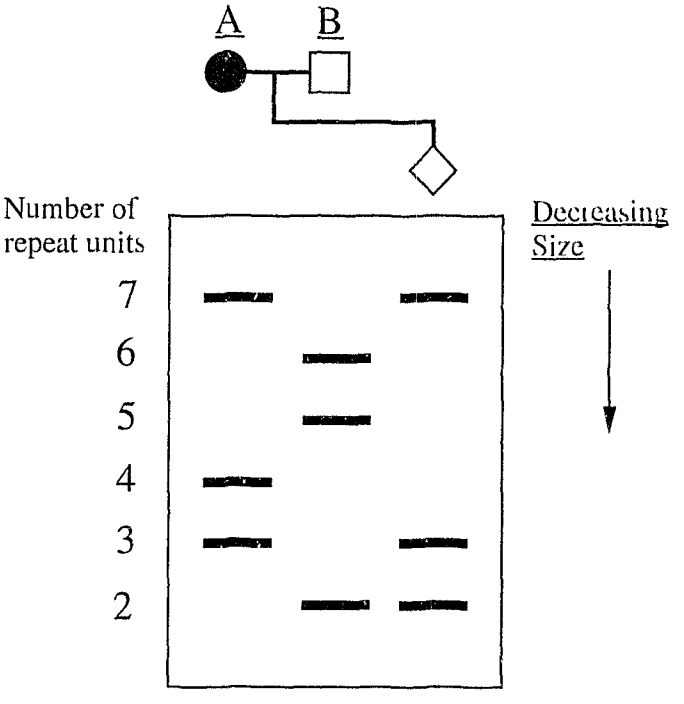

A
B
Number of repeat units
7
6
5
4
3
2
Decreasing Size

Figure 1.1

different core probes was the first indication that VNTR loci fall into groups or families related by the sequence of their repeating units. Extensive pedigree analysis in humans using two of the cloned probes, 33.6 and 33.15 has revealed that the degree of co-detection of loci is as low as 1% (Jeffreys et al. 1986).

The apparent abundance of minisatellite loci in the human genome coupled with the discovery of cross-hybridization of human minisatellite probes to other species (Jeffreys and Morton 1987, Burke 1987) has sparked the development of a number of DNA fingerprinting probes. Vassart and coworkers (1987) demonstrated that a tandemly repeated region from the protein III region of the bacteriophage M13 produced a polymorphic pattern when hybridized to Southern blots of human and animal DNA. Subsequently, this probe was used to generate fingerprints across taxonomic groups including vertebrates, invertebrates, plants, insects and microorganisms (Ryskov et al. 1988, Huey and Hall 1989, Blanchetot 1989, Gatei et al. 1991, Kvarnheden and Engstrom 1992). In a similar manner to the myoglobin derived sequences, the 3' region of the human alpha-globin gene has been shown to produce individual specific DNA fingerprints (Fowler et al. 1988). DNA fingerprints have been generated with various simple sequence repeats such as the Bkm sequence, a W chromosome specific tandem repeat isolated from the banded krait snake (Lloyd et al. 1989) and a synthetic oligonucleotide of $(CAC)_5$ repeats (Schafer et al. 1988). More recently it has been demonstrated that almost any randomly generated tandemly arrayed sequence can detect multiple variable

DNA fragments (Vergnaud 1991). This latter finding attests to the ubiquity of VNTR's in the eucaryotic genome.

Single locus VNTR probes

Multilocus DNA fingerprinting presents a number of inherent technical difficulties. The power of the technique as a tool for providing information on genetic variation relies on the large numbers of loci revealed. This complexity severely limits its utility in contexts where the allelic states of particular loci are required. Because of the number of bands generated by multilocus probes, allelic pairs specific to individual loci cannot be ascertained with any degree of confidence. Information is also likely to be lost through intensity variations between lanes and comigration of bands representing unrelated loci, particularly in the lower molecular weight range.

The solution to these problems has been to examine individual VNTR loci separately. This has been achieved by cloning single minisatellite loci which are then used as hybridization probes in Southern blots of genomic DNA. The resulting pattern of one or two bands, corresponding to the homo- or heterozygosity of the locus, is thus readily interpreted. Several strategies for isolation of minisatellite loci have been used. Some of the most variable human minisatellites have been discovered by selective cloning of gel-isolated DNA fragments into bacteriophage λ vectors (Wong et al. 1986, 1987, Gyllensten et al. 1989). Armour and coworkers (1990) have refined this approach by generating a size-selected charomid library. This library was screened in an ordered manner using a

number of multilocus probes. Others have screened genomic libraries with previously described multilocus probes or oligonucleotides based on known VNTR sequences (Nakamura et al. 1987, 1988). One novel approach utilizes the commonality of VNTR sequence. In a method termed 'probe walking' it has been proposed that isolated VNTR's can be used in sequential rounds of rescreening of the same library to identify VNTR families of similar sequence (Washio et al. 1989).

To date, several hundred cloned minisatellites have been isolated from species including human (Wong et al. 1986, 1987, Nakamura et al. 1987, 1988), mouse (Kelly et al. 1989), bovine (Georges et al. 1991), bird (Gyllensten et al. 1989), fish (Taggart and Ferguson 1990, Bentzen and Wright 1993), insect (Blanchetot 1989) and plant (Daly et al. 1991).

Digital DNA fingerprinting

The variant nature of the repeating units within the minisatellite array was first noted with the sequencing of specific VNTR loci (Wong et al. 1986, 1987). Recently, Jeffreys and co-workers have explored this extremely high degree of internal allelic complexity (Jeffreys et al. 1990, Gray and Jeffreys 1991). In its original form, minisatellite variant repeat (MVR) mapping required minisatellite loci which consisted of two repeat types differing by the presence or absence of a restriction endonuclease recognition site. By a process of amplification by polymerase chain reaction (PCR), isolation and partial restriction a ladder of fragments is generated, increasing in size by the unit length of the repeats. Missing 'rungs' in

Figure 1.2 Schematic illustrating the generation of digital DNA fingerprints by partial restriction endonuclease digestion. (A) The illustration shows an allele consisting of two types of repeating unit. Each repeat unit contains a site for restriction enzyme 1. Interspersed along the array is a second repeat type containing a restriction site to enzyme 2 in addition to enzyme 1. The allele is amplified from genomic DNA using primers flanking the minisatellite array. Amplification products are seperated by gel electroporesis. Each sample is end-labled with α^{32}P-dATP using reverse transcriptase, partially digested with enzyme 1 and enzyme 2 independently. (B) The resulting range of fragments generated by enzyme 1, which cuts at all repeat units, produces a ladder of fragments of incremental size. Restriction digestion with enzyme 2 produces a reduced number of fragments with sizes dependent upon the location of the type 2 repeat within the array. (C) Electrophoretic separation of the fragments followed by Southern transfer and autoradiography reveals a ladder fragments corresponding in size to the repeat unit length generated with enzyme 1 and an interspersed pattern of bands generated by enzyme 2 corresponding to the location of the type 2 repeats within the array. The result can be digitally encoded with a simple binary system where 0 indicates the repeat contains only the type 1 restriction site and 1 indicates that the repeat unit in that position contains an enzyme 2 site in addition to the enzyme 1 site.

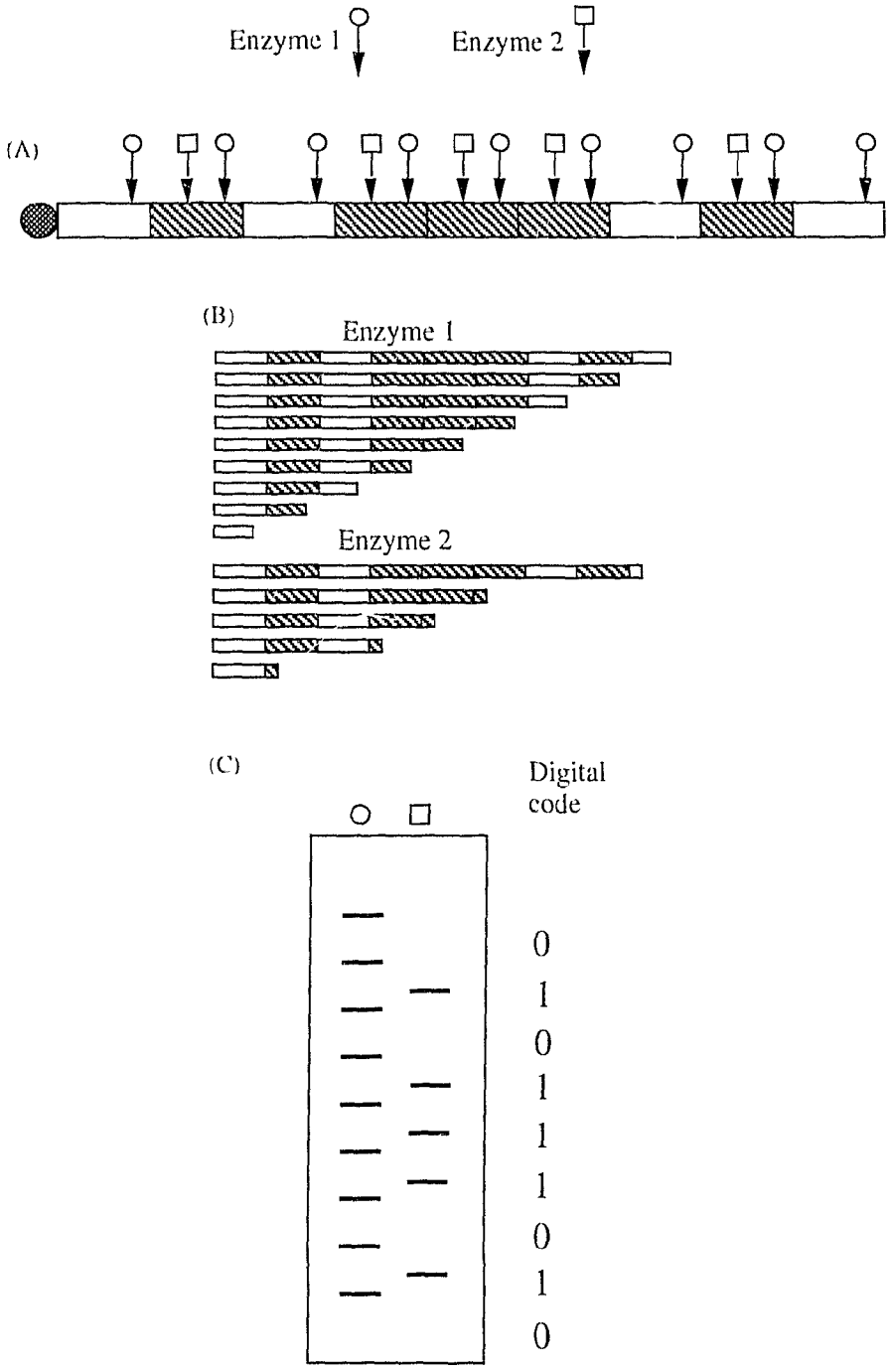

Enzyme 1
Enzyme 2
(A)
(B)
Enzyme 1
Enzyme 2
(C)
Digital code
0
1
0
1
1
1
0
1
0

Figure 1.2

the ladder, correspond to repeat units that did not contain the restriction site (Figure 1.2). By this method the order and position of the two repeat types across the array can be established. The method is restricted however, by the need for minisatellites with two repeat types, each differing by a single restriction endonuclease recognition site. A variation of the technique, developed in Jeffreys' laboratory, was the use of the polymerase chain reaction to amplify different variant types directly from genomic DNA samples (Jeffreys et al. 1991). To acheive this, primers were constructed to a region of unique DNA flanking the array and to the two unit types within the array. Amplification of the locus produced a series of fragments with incremental lengths corresponding to the position of the variant repeat types within the array. The problems associated with this strategy is the tendency of the arrays to collapse to the shortest possible length due to the nature of the annealing and amplification cycles. This problem was circumvented by uncoupling the initial annealing and amplification step from the subsequent amplification steps. This is acheived by generating PCR primers consisting of a 5' pairing region with sequence complementary to the repeat variant and a unique 3' region unrelated to the repeat sequence (Figure 1.3). This primer is added to the PCR reaction in limiting quantities equimolar with the number of annealing sites. Further amplification is performed with a second primer complementary to the unique 3' sequence of the first primer. The subsequent amplification occurs between the 5' 'anchor' primer and the unique sequence at the 3' end of each strand. This prevents unwanted annealing of primers repeats within the array which would result in collapse to the shortest unit

Figure 1.3 Schematic illustrating digital DNA fingerprinting by PCR primer annealing. A primer with sequence complementary to one of the repeat types within the array is annealed to genomic DNA. This primer is present in limiting quantities calculated to anneal on average once per array. The 3' end of this primer is a unique 'tag' sequence which is not present within or outside the array. Following the annealing reaction a mixture of two primer types, an 'anchor' primer annealing to sequence in the region flanking the 5' end of the array and a second primer which complements the 'tag' primer are added in excess and the mixture is amplified. The effect of the 'tag' primer is to uncouple the annealing of the repeat unit primer from the amplification of the array. Similar reactions are run using primers designed to anneal to other repeat types within the array. The resulting products are separated electrophoretically, transfered by Southern blot and visualized by hybridization to labeled array DNA. The resulting pattern reveals an interspersed ladder of fragments corresponding to the positions of the repeat types within the array. The repeat pattern can be digitally encoded to give a tertiary code describing the state of both alleles at the locus simultaneously.

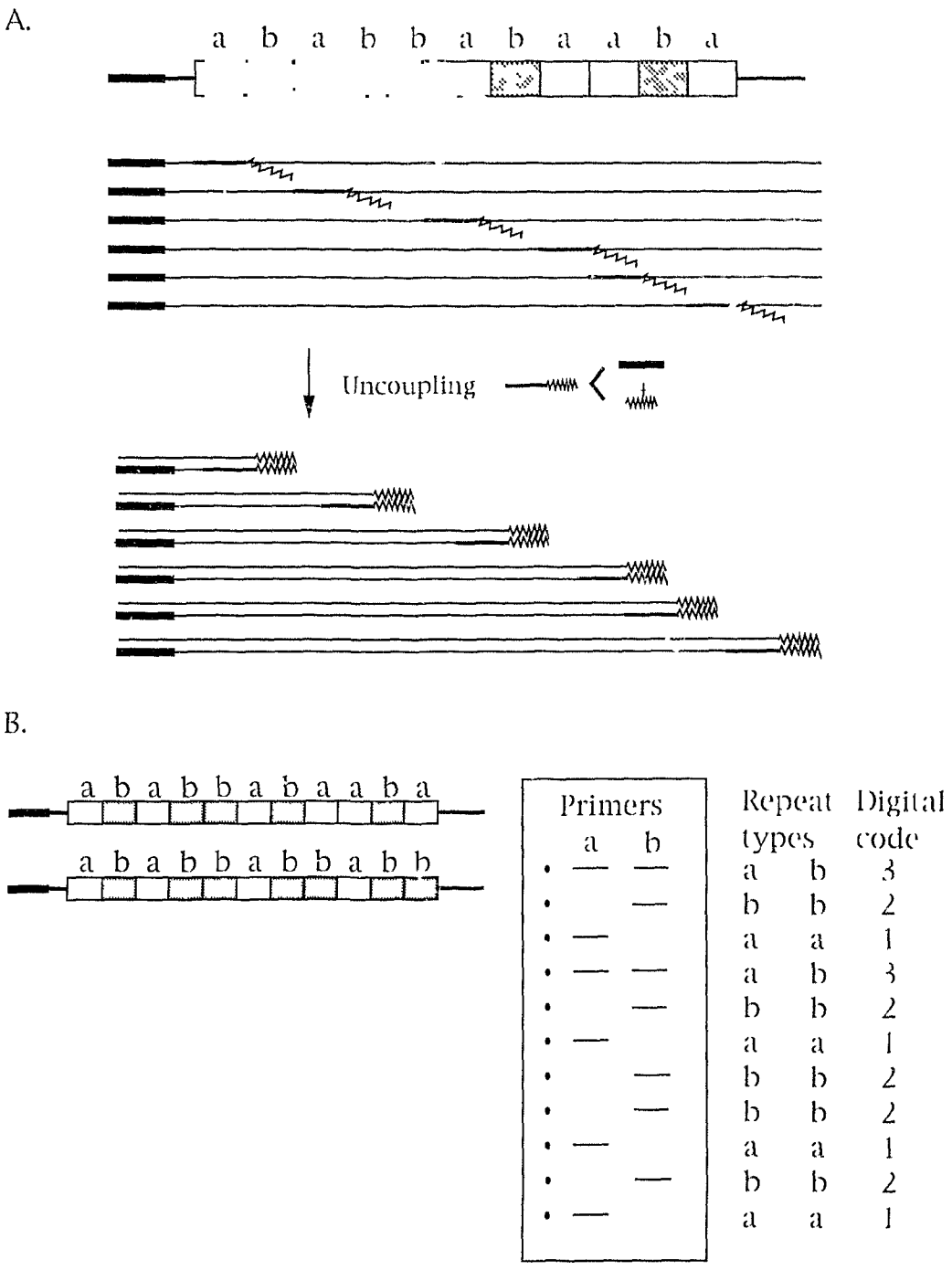

A.
a b a b b a b a a b a
Uncoupling
B.
a b a b b a b a a b a
a b a b b a b b a b b
Primers
a b
Repeat types
Digital code
a b 3
b b 2
a a 1
a b 3
b b 2
a a 1
b b 2
b b 2
a a 1
b b 2
a a 1

Figure 1.3

length. A ladder of size variants is resolved by electrophoretic separation which can be converted easily into a binary code.

Mutation of VNTR's

VNTRs, as their name implies, exhibit length polymorphism through variation in repeat unit copy number. To date these sequences comprise some of the most variable sequences known with heterozygosities at some loci approaching 100% (Wong et al. 1987, Armour et al. 1990). Extreme variation in the size of arrays has been reported with some loci differing in size by up to 25 kilobase pairs (kbp) (Wong et al. 1987). The high allelic length variability at these loci reflects a high rate of mutation producing new length alleles. Measurement of mutation rates at VNTR's from extensive human pedigrees of some loci are as high as 0.05 per gamete (Wong et al. 1987). A direct measurement of germline mutation by PCR amplification of mutant alleles from human sperm has yielded comparable rates to those determined by pedigree analysis (Jeffreys et al. 1990, Jeffreys et al. 1994). Mutation at VNTR's occurs with equal frequency in both males and females and was thought to have an equal likelihood of loss or gain of repeats. Recent studies by Jeffreys et al. (1994), however, have shown that for at least two human minisatellites a bias exists towards gain of repeats. Similar mutation frequencies have been reported for a mouse minisatellite, Ms6-hm, with rates of 2.5% per gamete (Kelly et al. 1989).

The processes underlying the changes in allele length have been the subject of much speculation. The group of human

minisatellites isolated by Jeffreys and coworkers, using the 33 bp myoglobin repeat showed similarities between the 'core' repeat and the *chi* sequence from bacteriophage λ (Jeffreys et al. 1985a). This sequence serves as a signal for homologous recombination probably by binding *Rec* BCD protein to cause local nicking and unwinding (Smith 1987). The observation led Jeffreys' group to propose that the minisatellite core was a eucaryotic recombination signal. Several other observations supported this hypothesis. Assuming that processes of mispairing and unequal exchange are responsible for length changes at minisatellite loci, the rates of new allele generation are too high to be accounted for by normal levels of recombination. The minisatellites must therefore form local hotspots for recombination. This is supported by mapping studies of human VNTR's, which show clustering of minisatellites in subtelomeric regions of chromosomes (Royle et al. 1987, 1988). Additionally, hybridization studies of minisatellite sequences to meiotic chromosome spreads have shown an association of probe with chiasmata, generally taken as physical indicators of crossing over (Chandley and Mitchell 1988).

More recent studies however, have shown that unequal crossing over between homologous chromosomes during meiosis is not the major mechanism responsible for generating new length alleles (Wolff et al. 1989). Examination of markers flanking the human minisatellite MS1 showed no evidence of recombination in 12 new mutant alleles. Similarly, examination of the ultrastructure of MS32 alleles by variant repeat mapping has shown only two cases of interallelic unequal exchange between two paternal alleles in 572

meioses (Jeffreys et al. 1991). In both events, flanking markers were too uninformative to determine whether distal regions had been exchanged. Other mutation events examined by this method appeared to have arisen by unequal sister chromatid exchange or replication slippage. These findings agreed with previous studies of MS32 which showed no unequal crossover between progenitor alleles in 19 germline and 30 somatic mutants analyzed (Jeffreys et al. 1990).

Other evidence has been presented that suggests that the mutational events at VNTR loci are not restricted to a single stage of gametogenesis or meoisis. Armour and coworkers demonstrated the presence of somatic mutations in clonal tumor cell populations (Armour et al. 1989). This study also provided evidence against unequal mitotic recombination as a major mechanism of new length generation. Analysis of 15 mutant tumors showed no reciprocal exchange products where at least 30% were expected. Somatic mutation has also been demonstrated in a hypervariable mouse minisatellite locus. Tissue mosaicism, with allele dosages ranging from 0.1 to 0.6 was seen in 2.8% of mice. This compared to germline mutation rates at the same locus of 2.5% per gamete indicating approximately equal rates of length change in both tissue types. Interestingly, the mutation events at this locus appear to be largely paternal in origin. It has been suggested that this results from the the large number of cell divisions involved in spermatogenesis and thus reflects a mitotic generation event (Kelly et al. 1989).

More recently, examination of the human minisatellite loci MS 32 and MS 31a by MVR-PCR, has shown that a complex conversion-

like event may be responsible for the generation of mutant alleles (Jeffreys et al. 1994). At least 50% of the alleles studied showed evidence of interallelic transfer. This transfer appears to be non-reciprocal suggesting a conversion event rather than reciprocal exchange. A number of alleles, however, showed complex pattern of repeat unit exchange with incorporation of anomalous repeats of unknown origin. All of the minisatellite loci investigated showed mutational polarity; the gain or loss of repeats tends to occur at one end of the array preferentially. Similar observations have been made for other loci (Jeffreys et al. 1990, 1991, Armour et al. 1993). Jeffreys et al. (1994) have speculated that this indicates the presence of a modulating element outside the repeat array which is responsible for controlling the activation of transfer by introduction of double stranded gap breaks.

VNTR location, generation and evolution

In humans, minisatellite arrays appear to localize at the proterminal regions of chromosomes (Royle et al. 1988). This discovery has been used in support of the recombinator role of minisatellites because of the reported association with chiasmata during meiosis (Chandley and Mitchell 1988). In mouse, however, the distribution of VNTR loci is dispersed over the autosomes and does not show localization for centromeric or telomeric regions (Jeffreys et al. 1987, Julier et al. 1990). Similar findings, by examination of somatic cell hybrids have been made within the bovidae (Georges et al. 1991).

Analysis of human minisatellites has shown frequent association with other tandemly repetitive and dispersed sequences. To date four sets of paired minsatellites have been described in human DNA (Armour et al. 1989). These pairs of minisatellites do not appear to show any pattern of similarity in repeat unit length, sequence or orientation, possibly indicating different origins. Clustering of VNTR loci has also been reported in the bovidae (Georges et al. 1991) and in fish (Bentzen et al. 1991, Wright 1993) . However, these reports are based on linkage analysis and not direct sequence information. Examination of the sequences flanking human minisatellites has revealed that they are often associated with dispersed repeat elements such as *Alu* repeats and retroviral LTRs (Armour et al. 1989). Other tandem repeat arrays have been found within dispersed and interspersed repeat regions of humans and mouse (Mermer et al. 1987, Kelly et al. 1989). In Atlantic salmon, minisatellites have been found adjacent to microsatellites (Bentzen and Wright 1993 and unpublished data). It has been suggested that the repetitive nature of these residual flanking sequences originally played a role in the amplification and dispersion of some minisatellite arrays (Jeffreys et al. 1985a, Wright 1993, 1994).

The processes responsible for the *de novo* generation of minisatellites remain unclear. Armour and coworkers (1989) have reported a minisatellite that appears to have arisen from within a retroviral LTR. Similar examples of expansion of tandem repeat sequences has been shown in other human dispersed repeats (Mermer et al. 1987) and in interspersed repeats in mouse (Kelly et al. 1989). Sequence analysis showed that this latter minisatellite

repeat sequence was derived from that of a diverged MT (mouse transcript) repeat. Two iterations of the repeat unit sequence GCAGG are present in the MT element. It has been suggested that the amplification of these repeats gave rise to the minisatellite. Whether these examples of VNTR's within other repeats have arisen through similar mechanisms is unknown. It has been hypothesized that minisatellite amplification and retroposon insertion may mark particularly unstable regions of the genome (Gray and Jeffreys 1991). If this is correct, they may be merely the byproduct of some other process of genome turnover.

Few studies have examined the evolutionary dynamics of minisatellites. Comparison of homologous loci in humans and primates for the locus detected by human minisatellites MS32 has shown that a relatively stable precursor state exists in both Old world monkeys and great apes consisting of the extreme 5' and 3' regions of arrays (Gray and Jeffreys 1991, Armour et al. 1992). However, while a second locus, detected by MS1 is monomorphic in great apes, the homologous locus in Old world monkeys is highly variable. This suggests that the variability of homologous loci over short periods of evolutionary time is largely unpredictable. Studies of a number of VNTR loci in the Indian peafowl showed conservation of some loci across species boundaries while others showed hybridization only within the peafowl (Hanotte et al. 1991). This may indicate the presence of a short allele in the distant species which is not detectable by Southern blot analysis (Armour et al. 1992). Alternatively, the lack of hybridization may be due to sequence divergence at these loci or indicate an origin more recent than the

evolutionary split between the taxa. The persistence of a large number of short relatively invariant minisatellite alleles is evident in multilocus DNA fingerprints (Jeffreys et al. 1985, Harris et al. 1991). These fragments, it is suggested, are minisatellite arrays with few repeat copies which have remained stable in the genome over extended periods of evolutionary time (Armour et al. 1992). The factors which initiate the expansion of these arrays to high copy number, and presumably their collapse, remain unknown.

Function of VNTR's

Whatever processes lead to the generation of VNTR's the persistence of minisatellites containing similar core sequence across the eucaryotes remains puzzling. It has been proposed that the shared repeats are causally related to the polymorphic nature of these loci (Jeffreys et al. 1985a, b). A number of lines of evidence have emerged to support this theory. The first of these is the discovery of proteins that bind to minisatellite sequences (Collick and Jeffreys 1990, Wahls et al. 1991, Collick et al. 1991). At least one of these has been shown to be a single stranded binding protein. This finding is consistent with the involvement of this class of protein in stabilizing single stranded DNA during recombination processes. A second line of evidence is the correlation between the the presence of a GGCAGG motif in multilocus fingerprint probes and the extent of polymorphism that they detect in Southern analyses (Mitani et al. 1990). Other studies have shown the presence of a minisatellite array near the murine major histocompatability complex, a meiotic recombination hotspot (Steinmetz et al. 1986). In humans,

minisatellites have been shown to cluster preferentially at proterminal regions of chromosomes (Royle et al. 1988) and *in situ* studies in man have shown the association of minisatellite sequences with chiasmata (Chandley and Mitchell 1988). Whether these observations indicate some functional role in genome turnover or are the result of turnover events is presently unknown.

Applications of DNA fingerprinting to Aquaculture Genetics

Aquaculture and fisheries management has in recent years undergone dramatic change. Many areas of capture fisheries are in crisis from lack of effective management and harvest practices. Aquaculture and culture based fisheries have come under scrutiny for the possible long term effects of introduction of exotic species and hatchery produced fish into wild populations (Lannan et al. 1989). In the past many of the tools available to fish genetics research lacked sufficient resolving power to allow rapid and efficient analysis. Protein electrophoresis and restriction fragment length polymorphisms (RFLPs) are hampered by the paucity of polymorphism revealed by these markers. The advent of VNTR analysis has provided a remarkable new source of genetic markers.

One of the primary uses of VNTR's is in the assessment of the genetic resource within populations of fish. Intensive pressures on wild populations through over fishing as well as poor management practices in the aquaculture industry have created bottlenecks with potentially severe reduction of genetic variability. Artificial constraints on variability have been studied in inbred domestic chickens using multilocus DNA fingerprinting. Kuhnlein and

coworkers (1990) have determined a coefficient of inbreeding for a number of populations through the degree of band sharing in DNA fingerprints. Other studies have used a measure termed similarity index to determine the degree of relatedness between genetically isolated populations of foxes and within and between groups of wild lions (Gilbert et al. 1990, Packer et al. 1991, Gilbert et al. 1991). Hillel et al. (1990) have proposed the application of multilocus fingerprinting to increase the efficiency of classical selective breeding processes for agricultural species. In many breeding regimens the number of backcrosses necessary to restore the recipient genome represent a substantial investment in time and money (Beckmann and Soller 1986). The use of DNA fingerprinting allows the selection of individuals for backcrossing to be based on a "genome similarity" assessed by the degree of band sharing. This provides a more precise evaluation of genetic relatedness than the previously used level of similarity in one or more quantitative traits.

Numerous other applications exist in aquaculture genetics for both research and commercial uses. These include such measures as registering specific brood stocks with genetically desirable traits to protect proprietary rights. An extension of this stock registry will be the ability to monitor the impact of escaped or released stocks on wild populations. Several interesting possibilities have been proposed in the areas of genotype/environment interaction. In classical quantitative genetic experiments the progeny of individual crosses must be maintained in separate enclosures due to the difficulty of tagging newly emerged fry. This segregation has the obvious effect of introducing confounding factors because of

potentially significant environmental variations between tanks (Uraiwan and Doyle 1986). Additionally, such studies require greater resources in the form of holding facilities and equipment. The use of DNA fingerprinting allows progeny to be hatched and reared in common enclosures. Individuals showing traits of interest can be traced back to particular breeding pairs by comparison of fingerprint patterns. Familial relationships can be established without the need for broodstock to be preselected for particular genetic or phenotypic markers (Wright 1993).

Objectives of the study

At the outset of this study, few investigations of VNTR's had been performed outside of humans and mice. There were therefore, several objectives to this study: the first of these was to determine if VNTR's were present in the genome of the tilapiines and whether their abundance approximated that described in other species. The second objective was the isolation of a number of VNTR's from these fish. These cloned minisatellites provided sequence information for comparison to other known VNTR's. Hybridization of these sequences to genomic DNA demonstrated the variability and evolutionary transience in related species within the cichlid family.

CHAPTER 2: MATERIALS AND METHODS

Sources of Fish

Oreochromis niloticus were obtained from breeding stocks at Dalhousie University and facilities in Thailand. Some samples of *O. niloticus* , *Oreochromis aureus*, *Oreochromis mossambicus*, *Tilapia zillii* and *Sarotherodon galilaeus* were provided by Dr. Brendan MacAndrew from the Institute of Aquaculture at the University of Stirling, Scotland. Samples of African (*Haplochromis moori*), Asian (*Etroplus maculatus*) and South American (*Cichlasoma otofasciatus*) cichlids were purchased from a local aquarium store. Atlantic salmon samples were provided by Dr. Paul Bentzen.

Genomic DNA extractions

Genomic DNA was isolated from tissue from the caudal peduncle region or from blood samples taken by caudal vein puncture of anesthetized fish. For blood extractions, needles and syringes were flushed with 0.5 M EDTA to prevent clotting. 0.5 ml of whole blood was subjected to centrifugation, washed once with 0.85% saline and resuspended in 5 ml of high T.E. buffer (10 mM Tris-Cl, pH 7.6, 40 mM EDTA). Five ml of lysis buffer (10 mM Tris-Cl, pH 7.6, 40 mM EDTA, 0.5 % SDS) was injected into the suspension through an 18 guage needle. The solution was then extracted with phenol and chloroform and DNA precipitated with ethanol. The DNA was dried briefly and resuspended in T.E. buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA). Tissue specimens (approximately 0.1g) were frozen in liquid nitrogen and crushed in 1.5 ml microfuge tubes using a glass pestle.

Disrupted tissue was resuspended in 1.0 ml of lysis buffer (10 mM Tris-Cl, pH 7.6, 40 mM EDTA,0.5 % SDS) and subjected to centrifugation to remove debris. The homogenate was immediately extracted with one volume of Tris-Cl saturated phenol (pH 7.6), one volume of chloroform. DNA was precipitated with 0.1 volumes of ^{5} M ammonium acetate and 2 volumes of ethanol, dried briefly under vacuum and resuspended in T.E. buffer.

Restriction Endonuclease Digestion

Genomic DNA samples (approximately 30 µg) were digested in accordance with manufacturer's recommendations using 4µl of a 25 units/µL stock of the restriction endonuclease *Hae* III or *Pal* I in the appropriate restriction buffer for 1 hour at 37°C. Samples were thoroughly mixed, a further 2 µl (50 units) of restriction enzyme added and the reaction incubated for a further 2 hours at 37°C. Following digestion, DNA was extracted once with one volume of phenol/chloroform, precipitated by ethanol, dried and resuspended in T.E. buffer. These samples formed a stock digest from which aliquots were drawn for electrophoresis.

Agarose Gel Electrophoresis

Digested genomic DNA samples were fractionated by gel electrophoresis on 25 cm, 1% agarose gels using a 1x Tris-Borate-EDTA buffer. Electrophoresis was carried out at 1.5 volts/cm for 36 to 48 hours. Gels were stained with ethidium bromide (0.5 µg/ml) prior to photographing.

Southern Transfer

DNA fractionated by gel electrophoresis was transferred to nylon Membrane (Hybond-N, Amersham) by vacuum blotting (Vacu-Blot, Pharmacia). DNA was depurinated for 15 minutes with 0.2 M HCl, denatured for 20 minutes with 1.5 M NaCl, 0.5 M NaOH, and neutralized for 20 minutes with 1.5 M NaCl, 0.5 M Tris-Cl, pH 8.0. DNA was then transferred for 1 hour with 20 x SSC (3.0 M NaCl, 0.3 M sodium citrate). Nylon membranes were washed briefly in 2 x SSC, air dried and baked under vacuum.

Radiolabeling of Probes

Jeffreys' probes, 33.6 and 33.15 (Cellmark Diagnostics) were labelled as specified by the supplier using α^{32}P-dGTP (3000 Ci/mmol). Cloned DNA fragments were labelled by random oligonucleotide priming (Feinberg and Vogelstein, 1983, 1984) using α^{32}P-dGTP or α^{32}P-dATP (3000 Ci/mmol). All probes were purified using a Sephadex G50 column prior to denaturation.

Hybridization conditions

DNA fingerprint probes 33.6 and 33.15 were denatured by heating to 100°C immediately prior to addition to the hybridization reaction. Nylon membranes were prehybridized for at least one hour in 0.5 M sodium phosphate, pH 7.4, 7% SDS, 1% BSA, 1 mM EDTA. Hybridization was carried out in the same solution for 16 to 24 hours at 58°C, either in sealed polyethylene bags or in a Hybaid rotating hybridization oven. Membranes were washed in 2 x SSC, 0.1% SDS for

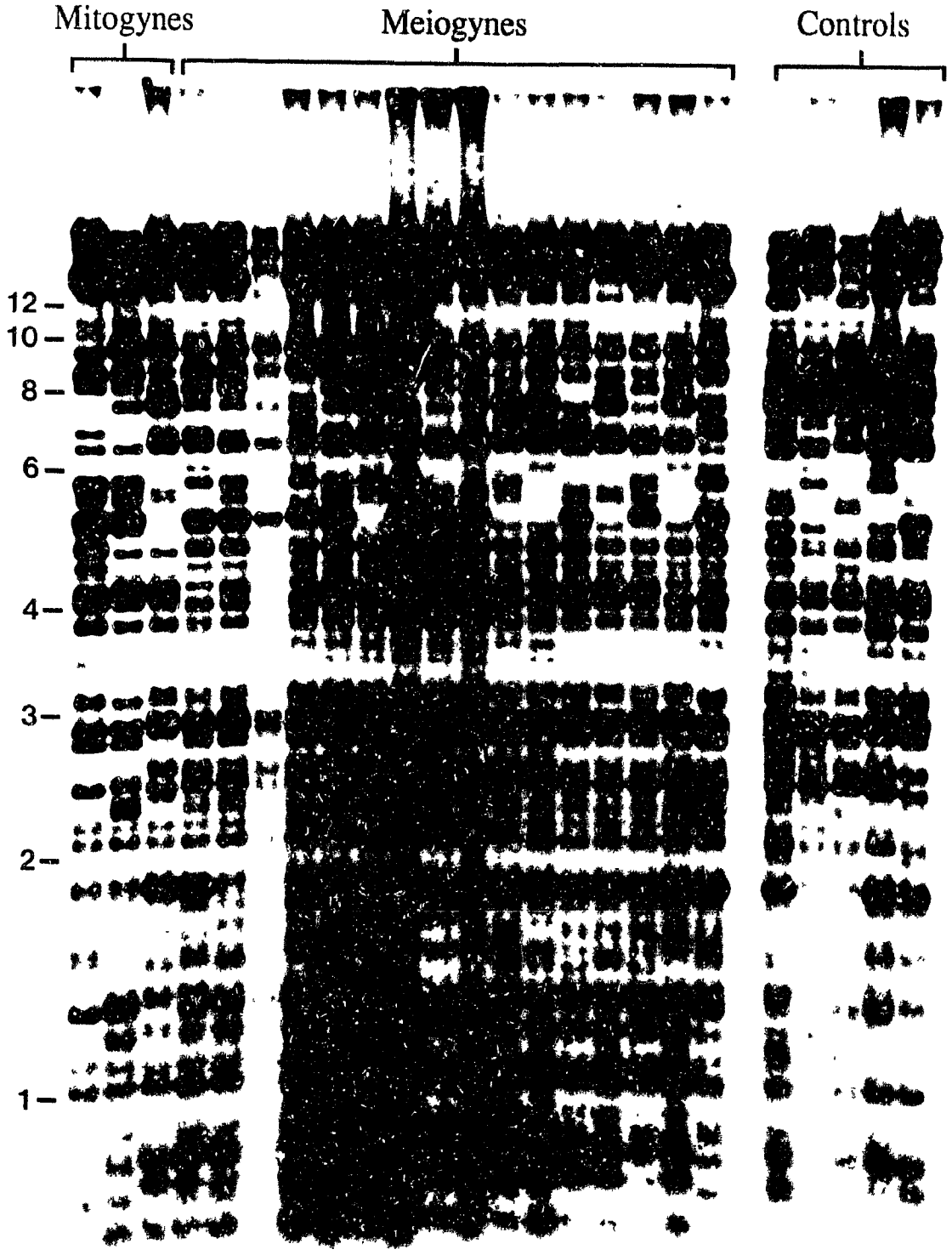


Figure 3.2

20 minutes at room temperature followed by 0.5 x SSC, 0.1% SDS at 58°C for 20 minutes.

Prior to addition to the hybridization reaction, cloned minisatellite probes were denatured and incubated at 58°C with 100 mg/ml of sheared, denatured tilapia genomic DNA in 0.5 M sodium phosphate buffer (pH 7.4). The addition of the sheared DNA served to block repetitive regions of the probe and leave only unique flanking regions available for hybridization with the genomic DNA. Membranes were prehybridized and hybridized in 0.5 M sodium phosphate, pH 7.4, 7% SDS, 1% BSA, 1 mM EDTA. Following washes filters were blotted dry and subjected to autoradiography overnight using Kodak AR film, in the presence of an intensifying screen.

Library Screening

A commercially constructed genomic DNA library in the λ replacement vector EMBL 3 was purchased (Clonetech). In its construction, DNA from a single *O. niloticus* individual was partially digested with the restrictrion endonuclease *Mbo* I and size-selected fragments of 8 to 21 kbp were ligated to vector arms. Aliquots of this library were plated on *E.coli* NM538. Plaques were transferred to Hybond-N filters by standard techniques (Maniatis et al. 1982). DNA was denatured by soaking filters in 0.5 M NaOH, 1.5 M NaCl followed by neutralization in 0.5 M Tris-Cl, pH 8.0, 1.5 M NaCl. Filters were washed for 5 minutes in 2 x SSC and bacterial debris was removed by gently scrubbing the filters with a Kimwipe. The filters were probed with a 1:1 mixture of the Jeffreys' probes 33.6 and 33.15 as described above. Fifty positively hybridizing plaques were randomly

selected for further analysis. DNA was isolated by large scale liquid lysate (Maniatis et al. 1982) and purified by cesium chloride centrifugation.

Mapping of EMBL 3 clones

Five hundred ng of DNA from each clone was digested with the restriction endonuclease *Sma* I, *Hind* III, *Bam*H I and *Kpn* I according to manufacturerss specifications (Pharmacia). Digested samples were subjected to electrophoresis through 1% agarose gels at 1.5 volts/cm using a 1 x TBE buffer. Gels were stained, photographed and DNA was transferred to filters as previously described. Filters were probed with both human core probes 33.6 and 33.15 individually. Hybridization and subsequent exposure of autoradiographs was as described above.

Probe Nomenclature

Naming of cloned sequences follows the convention described in Hanotte et al. (1991). Initial letters indicate genus and species. The letters MS are incorporated to designate that the probe detects a minisatellite locus. The final numerals are arbitrary, denoting the clone number from the original library screening. EMBL clones 17, 32 and 34 refer to the original library isolates containing the minisatellite loci OniMS17, OniMS32 and OniMS34, respectively.

Cloning Methodology

Fragments of EMBL-3 clones, identified as containing minisatellite sequences by hybridization with the Jeffreys' 33.6 or

33.15 probes, were subcloned into either pUCBM20/21 (Boehringer-Mannheim) or M13mp18/19 vectors (Yanis-Perron et al. 1985). Five hundred ng of DNA from the EMBL clones was digested with the appropriate restriction endonuclease and fractionated in a 1% agarose gel using a Tris-Acetate-EDTA buffer. Fragments were visualized by ethidium bromide staining and excised from the gel. DNA was eluted from the agarose using a glass matrix method (Geneclean). The cloned locus OniMS32 contained within a 1.9 kbp *Sma* I/ *Bam*H I fragment was ligated between the *Sma* I and *Bam*H I sites of M13mp18 RF. DNA fragments containing the locus OniMS17 were ligated into a pUCBM21 vector digested with the restriction endonuceases *Sma* I and *Hind* III. Fragments containing OniMS34 were blunt end ligated into the *Sma* I site of pUCBM21. The ligation products of M13 vectors were used to transform competent *E.coli* (TG1) and plated on 2 x YT media containing X-gal and IPTG. Recombinant plaques, indicated by lack of colour, were selected and prepared for DNA sequencing by standard protocols (Maniatis et al. 1989). pUCBM21 ligation reactions were used to transform competent *E.coli* DH5α (recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1). Competent cells were either purchased from a commercial supplier (Stratagene) or made by the calcium chloride method (Maniatis et al. 1982).

Sequencing protocols

Both single-stranded recombinant M13 DNA and double-stranded plasmids were sequenced by the chain termination method (Sanger et al. 1977) using either α^{35}S-dATP or α^{32}P-dATP with

either Klenow (Promega) or T7 polymerase (Pharmacia) sequencing kits with supplied protocols. Products were fractionated by electrophoresis through 6% polyacrylamide ionic gradient or non-gradient gels. Clones were sequenced with the M13 universal primer or using synthetic oligonucleotides to regions within the clone.

CHAPTER 3 : MULTILOCUS DNA FINGERPRINTING OF TILAPIA

DNA fingerprinting of tilapia

Genomic DNA from several tilapia was digested with the restriction endonuclease *Hinf* I, fractionated by agarose gel-electrophoresis, transferred to nylon membrane and hybridized to the human polycore probes 33.6 and 33.15 (Figure 3.1). Lanes 1 and 2 contain DNA from the male and female parents and lanes 3 to 6 contain DNA from four full-sib offspring randomly selected from the brood. DNA fragments that hybridized to the 33.6 probe were detected over the whole effective fractionation range of the gel (0.5 to 30 kbp). The intensity of the autoradiographic signal differed amongst fragments within lanes and did not correlate to fragment length. The intensity of the autoradiographic signal indicates the degree of hybridization between the probe and the fragment based on sequence similarity or to the number of repeat units within a fragment. In the two parental fish, a total of 40 distinctly hybridizing bands were detected, the majority of which were less than 2.0 kbp in length. At shorter exposures, poor resolution of fragments below 2.0 kbp in size was evident which made analysis of this region difficult. Most fragments in this range appear to be common between individuals. The fragment distribution does not appear to be uniform throughout the size range. A distinct cluster of fragments appears between 6.0 kbp and 30 kbp. Similar clustering has been reported in cetaceans (Amos and Dover 1990).

Figure 3.1 DNA fingerprint pattern generated in tilapia. Southern blot of genomic DNA from male and female *O.niloticus* individuals and four full sib offspring, digested with the restriction endonuclease *Hinf* I. Left panel: hybridized to the human polycore probe 33.6. Right panel: hybridized to the human polycore probe 33.15. Restriction fragments smaller than 0.5 kbp have migrated off the gel. Molecular size markers are 1 kb ladder (BRL).

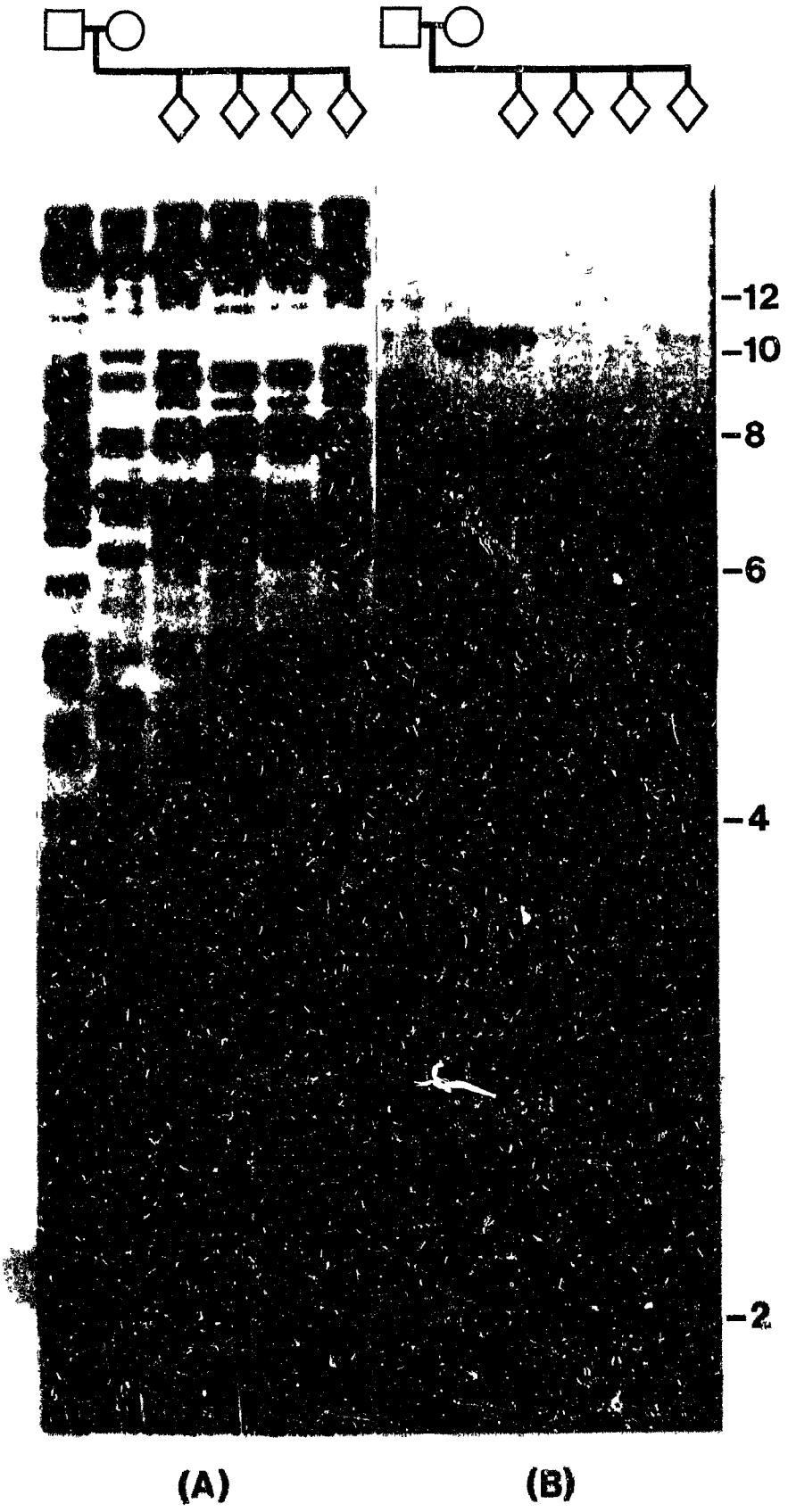


Figure 3.1

Comparisons of parental and offspring band patterns show that the hypervariable fragments were inherited in a Mendelian fashion with all but one offspring fragment being present in one or other of the parents. This anomalous fragment (Figure 3.1 left panel, lane 5) is unique to this individual and probably is the result of mutational changes in the number of repeats in the array.

The pattern of *Hinf* I fragments which hybridized to the 33.15 polycore probe was distributed between 0.5 and 8.0 kbp (Figure 3.1). This differed from the pattern produced by the 33.6 probe both in fragment size and in the distribution of fragments. Extensive hypervariability is apparent with approximately 24 different variable fragments in the two parental fish. Again, a number of bands are shared by all individuals with the majority of these occuring below 2.0 kbp in length. Comparison of offspring and parental lanes showed Mendelian inheritance with all offspring fragments being attributable to one parent or the other.

We can conclude from these results that multilocus DNA fingerprints can be generated in tilapia. These fish appear to contain minisatellite-like sequences that are similar to VNTR's found in other eucaryotes. As has been demonstrated in a number of other species at least two classes of VNTR's exist in the tilapiine genome. These classes show differences in sequence and size distribution in tilapia and may indicate clustering of particular size ranges within the class.

DNA fingerprinting of gynogenetic lines.

The use of artificial gynogenesis to generate highly inbred lines has important applications in immunological, endocrinological and genetic studies. In the field of aquaculture, gynogenetic lines have potential uses in the selection of commercially important traits and in the control of unwanted breeding. In many artesian operations with tilapia, problems of reduced individual size occur due to crowding effects resulting from precocious breeding. The production of monosex cultures by artificial gynogenesis is one method of reducing unwanted breeding and thus limiting the number of fish in a facility (Penman et al. 1987, Mair et al. 1987). The production of monozygotic lines is accomplished through the *in vitro* activation of oocytes by irradiated sperm. Diploidization of the egg can be induced by retention of the second polar body, by heat, cold or pressure shock, during the second stage of meiosis. The resulting gynogens are termed meiogynes. Similarly, suppression of first mitosis with the resultant lack of chromosomal segregation can produce diploid zygotes or mitogynes (Thorgaard 1986, reviewed in Stanley and Sneed 1974, Purdom 1983). The two processes differ in the degree of heterozygosity expected in the progeny. In meiogynes this will depend upon the extent of recombination at first meiotic prophase. In this case polymorphic loci located distal to the centromere may exhibit heterozygocities as high as 100%. Conversely, mitotic gynogenesis should produce clonal progeny that are individually fully homozygous. In practical terms the success of gynogenesis is often uncertain because of the variability of timing in oocyte development and also, potentially, from homologous

recombination of male DNA fragments (Carver et al. 1991). The use of DNA fingerprinting has proven to be more informative than previously used allozyme studies to estimate heterozygosity (Mair et al. 1987) because of the ability to simultaneously detect a number of loci dispersed throughout the genome. It was felt that this would provide a useful test of the utility of DNA fingerprinting in tilapiines.

Gynogenetic lines were established from a single female *O. niloticus* by Drs. Penman and McAndrew at the Institute of Aquaculture, University of Stirling, Scotland. DNA samples from 3 mitogynes, 15 meiogynes and 10 controls were stored in ethanol and shipped to Dalhousie University for analysis. Five microgram aliquots of each sample were digested with the restriction endonuclease *Pal* I and fractionated through a 1% agarose gel. The DNA was transferred to nylon membrane and hybridized to the human polycore probe 33.15 (Figure 3.2). The pattern of hybridization was complex and individual specific. Fragments were detected that spanned the whole fractionation range of the gel.

Assessment of gynogenetic origin

One predictor of gynogenetic individuals is an overall decrease in genomic heterozygosity. This should be evident as a reduction in the number of bands in the DNA fingerprint of an individual. The total number of loci in meiogyne individuals should be less than that of a normally interbred fish. The remaining heterozygous loci will be a result of crossover during prophase of meiosis I. Correspondingly, mitogynes will show even fewer bands as they should be homozygous at all loci excluding those mutant bands generated by

Figure 3.2 DNA fingerprint produced from gynogenetic lines of tilapia *O. niloticus*. Southern transfer of *Pal* I-digested DNA from 24 individuals hybridized to the human polycore probe 33.15. Lanes 1-3; putative mitotic gynogens. Lanes 4-19; putative meiotic gynogens. Lanes 20-24; Controls, generated by a normal cross between the maternal fish used to produce gynogenetic lines and a single male. Molecular size markers shown to the left of the panel are 1 kb Ladder (BRL).

Figure 3.3 Frequency of fingerprint bands in meiotic and mitotic gynogens. A histogram plotting the number of hybridizing bands generated in each individual lane by both 33.6 and 33.15 probes against frequency. A legend identifying corresponding shading patterns is shown in the figure.

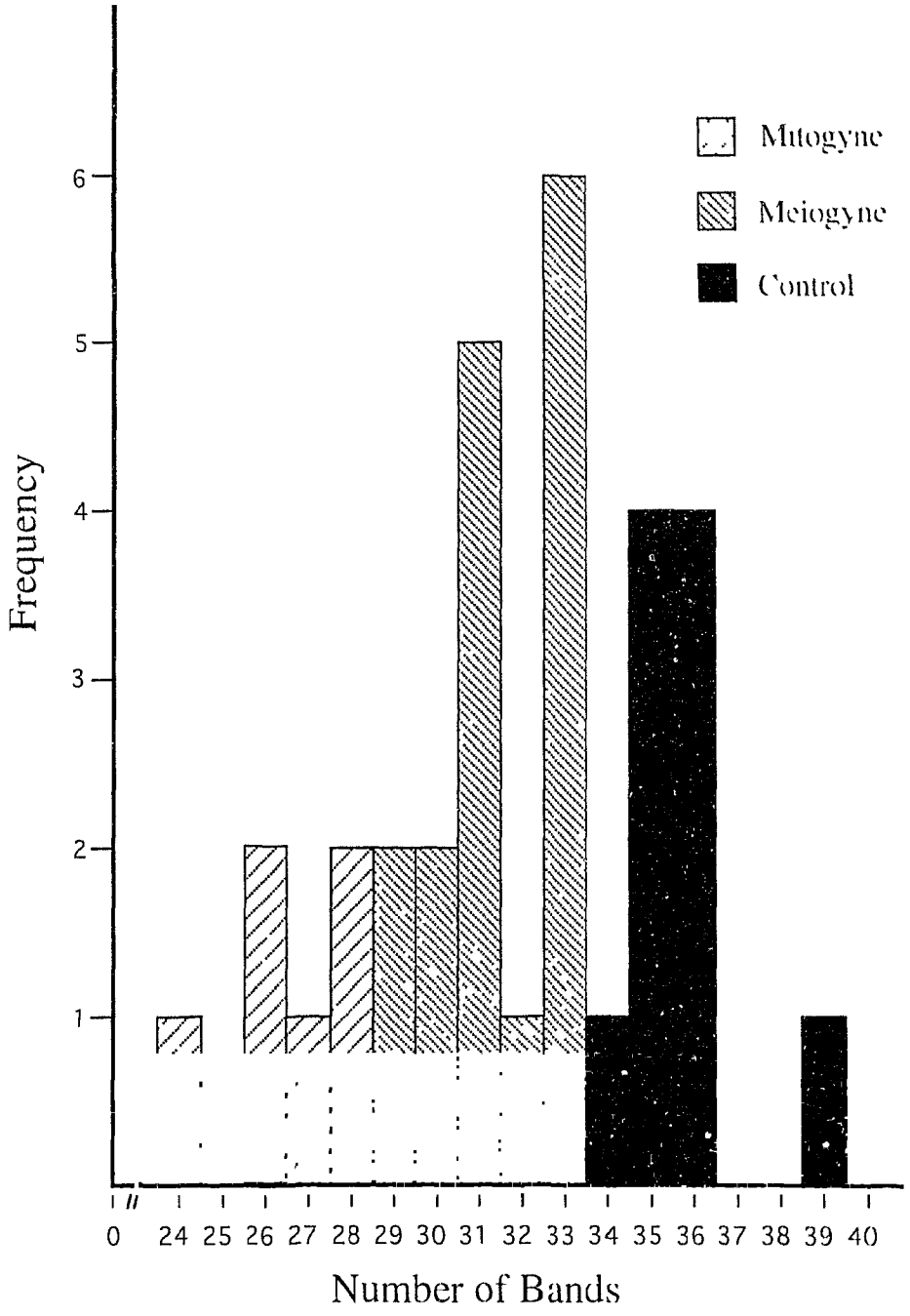

Mitogyne
Meiogyne
Control
Frequency
1
2
3
4
5
6
0
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Number of Bands

Figure 3.3

slipped-strand mispairing or unequal sister chromatid exchange either during early embryogenesis or oogenesis. Figure 3.3 shows the distribution of bands for the three mitogynes, fifteen meiogynes and five full sib controls using the 33.15 polycore probe. The three individual types fall into three distinct classes. Calculation of the numbers of bands show a 14% reduction for meiogynous individuals over controls and a 27% reduction for mitogynes compared to controls. In a study published subsequent to this analysis, Carter and coworkers (1991) used DNA fingerprinting to evaluate gynogenetic lines of *Oreochromis aureus* and *O. niloticus*. Although their analysis allowed them to identify individuals inheriting male DNA, they were unable to unambiguously distinguish mitogynes from meiogynes. This may have resulted from the high level of inbreeding in their captive population causing low levels of variation.

Difficulties in the estimation of individual relatedness

The analysis of gynogenetic origin illustrated some of the technical difficulties inherent in the technique of multilocus DNA fingerprinting. A substantial number of problems arise during the generation of the data. These include ambiguities in the amount of DNA loaded on the gel or in the salt concentration of the loading buffer which lead to aberrant migration of fragments. The low stringency hybridization conditions are difficult to reproduce from one experiment to another. The analysis of DNA fingerprints poses another series of problems. First, the upward bias of band sharing when compared to actual relatedness is due to the finite number of

alleles in a population. This skewing cannot be effectively corrected because it is not consistent from individual to individual. Second, a number of potential biases exist due to comigration of non-allelic bands. This is a greater problem at lower molecular sizes where large numbers of fragments tend to be ...nt. Third, linkage between marker loci which is readily apparent in systems with one or a few alleles, is frequently masked in complex fingerprint patterns. Bias may also be introduced by the linkage of loci with other loci under selection. Finally, one confounding factor in the interpretation of DNA fingerprints is the effects of mutation. Jeffreys and coworkers have estimated that for the 33.6 and 33.15 probes, 27% of human offspring will exhibit one mutant band, 1.2% two and <0.3% three bands (Jeffreys et al. 1991).

CHAPTER 4: CLONING AND ASSAY OF VNTRs FROM TILAPIA

Library screening and mapping of clones

Genomic DNA was extracted from a single tilapia *Oreochromis niloticus*. A genomic library was constructed in the λ replacement vector EMBL 3. This library was screened by hybridization to a 1:1 mixture of the two human core probes, 33.6 and 33.15. Approximately 3 x 10^5 pfu's were screened, representing about 2 genome equivalents assuming a mean insert size of 15 kbp. Autoradiography revealed several thousand positively hybridizing plaques representing approximately 1% of the total genome. Given an even distribution throughout the genome, this suggests that minisatellite arrays related in sequence to the 33.6 and 33.15 probes are present every 1000 kbp in the tilapia genome. Fifty positively hybridizing plaques were randomly selected. Those plaques which showed the most intense hybridization signals were selected on the assumption that they contained the longest minisatellite arrays. Previous evidence has shown that the longest arrays are more prone to length changes and thus are more likely to be highly polymorphic (Jeffreys et al. 1985a, Jarman et al. 1986). These plaques were purified through two further rounds of screening. Of the original 50 isolates, a total of nineteen clones were selected for further analysis. Figure 4.1 shows the restriction maps for these 19 clones. Three pairs of clones, 50 and 51, 45 and 49 and 1 and 17 produced very similar restriction maps. Comparison of clones 1 and 17 shows that they

Figure 4.1 Restriction maps of EMBL 3 clones. A schematic representation of nineteen EMBL 3 clones. Restriction endonucleases used were *Bam*H I (B), *Hind* III (H), *Kpn* I (K) and *Sma* I (S). Hatched regions represent fragments hybridizing to the 33.6 probe alone. Shaded regions represent fragments hybridizing to the 33.15 probe alone. Stippled regions represent fragments hybridizing to both polycore probes. Broken bars common to all clones represent λ arms

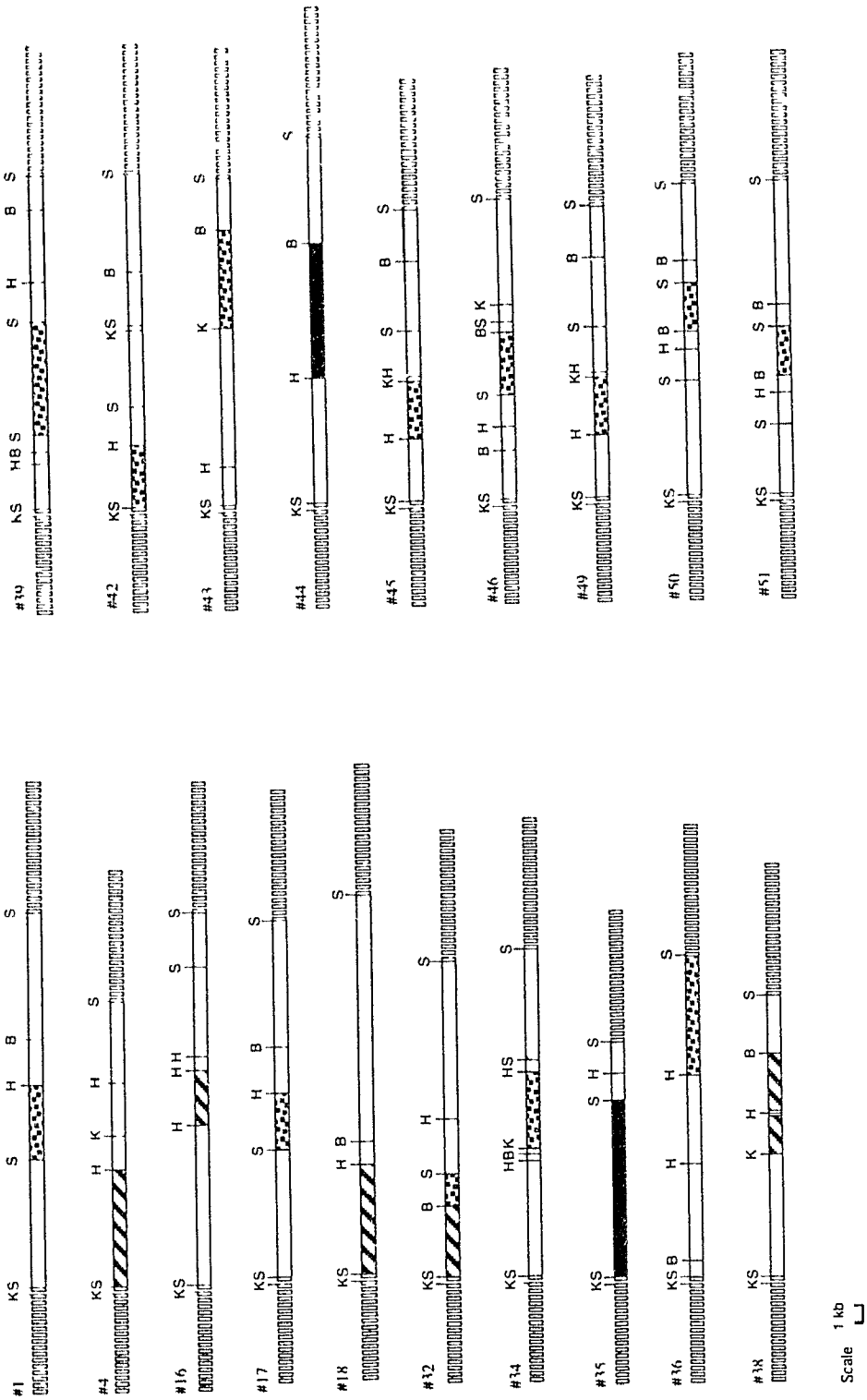


Figure 4.1

differ only by 1.0 kbp within the minisatellite bearing *Sma* I/ *Hind* III fragment. It is tempting to speculate that this indicates the surreptitious cloning of two alleles of this minisatellite locus. However, Kelly et al. (1989) have reported the instability of minisatellite arrays in λ cloning vectors. It appears possible therefore that the size difference may be a result of loss or gain of repeating units during the cloning process. Restriction maps of clones 50 and 51 differ by the size of the *Sma* I/*Sma* I and *Bam*H I/*Sma* I fragmants adjacent to the vector arms. This may be a result of the ligation of different *Mbo* I fragments into the original cloning vector or may represent a restriction fragment length polymorphism (RFLP) due to an additional *Mbo* I site. Clones 45 and 49 probably represent two identical clones selected coincidentally. Similar results have been obtained when cloning minisatellite sequences from other species. The large scale isolation of human minisatellites from a charomid library produced repeated isolates from already cloned loci (Armour et al. 1990). The authors suggest that this is due to only a limited number of loci, from those present in the genome, being readily clonable in λ vectors. They propose that the shortening of clones during propagation due to loss of repeats, prevents them from being repackaged into phage particles.

Following visualization of restriction fragments by ethidium bromide staining, DNA was transferred to nylon membranes and hybridized to the 33.6 and 33.15 polycore probes individually. From the ninteen EMBL 3 clones that were mapped, 5 showed fragments which hybridization exclusively to the 33.6 probe (clones 4, 16, 18 and 38), two to the 33.15 probe (clones 35 and 44) and 12 to both

probes (clones 1, 17, 32, 34, 36, 39 42 ,43, ,45, 46, 49, 50 and 51)(Figure 4.1).

Possible clustering of minisatellite-like sequences in the tilapia genome.

Two of the ninteen clones mapped showed a pair of minisatellite-like sequences in close association (Figure 4.1). Clone 32 contained a 4.0 kbp *Sma* I/*Bam*H I fragment that hybridized to the 33.6 probe immediately adjacent to a 1.9 kbp *Bam*H I/*Sma* I fragment that hybridized to the 33.15 polycore probe. This is consistent with other reports of clustered minisatellites from human, bovine and salmonid genomes (Armour et al. 1989, Georges et al. 1991, Bentzen and Wright 1993).

Restriction endonuclease mapping of clone 38 revealed two fragments that hybridized to the 33.6 polycore probe. This may represent a similar clustering of two minisatellites arrays as those of clone 32. However, the fact that both fragments show sequence similarity with the 33.6 probe does not exclude the possibility that this represents one minisatellite array containing a single repeat unit containing a *Hind* III recognition site.

Identification of polymorphic probes

Restriction fragments from each clone, which hybridized to the minisatellite core sequences, were isolated from agarose gels by the glass milk method. The DNA was quantified, labeled with [^{32}P]-dATP and hybridized to nylon filters containing digested genomic DNA from a panel of tilapia. Three types of hybridization patterns were

observed. A number of clones showed no hybridization to the genomic filters or produced only a general background smear. This may be the result of endonuclease recognition sites within the repeat units of the minisatellite array. Several of the clones produced a multilocus pattern, similar to, but less complex than that generated in tilapia with the polycore probes (Figure 3.2). There was no evidence of a hypervariable locus overlaying a background multilocus pattern. However, the locus may have been contained in the portion of fragments less than 1.0 kbp which migrated off the gel. Interestingly, several different clones showed a very similar pattern of hybridization (Figure 4.2). This suggests that, under the conditions used, these probes detect a group or family of minisatellites with highly conserved core sequence. The hybridizations were carried out at high stringency (42°C, 55% formamide) indicating an extremely high degree of sequence similarity between these arrays. Similar results have been reported for minisatellites isolated from Indian peafowl (Hanotte et al. 1991). In this case, however, the pattern consisted of either a single locus overlying a multilocus fingerprint-like pattern which exhibited a single locus pattern in the presence of genomic competitor DNA, or a simple multiband pattern containing more than two bands. This is consistent with a continuum of minisatellite families each containing a conserved core sequence and having repeat units with varying degrees of sequence similarity to one another. The finding of identical patterns of hybridization for different cloned loci in tilapia suggests that the minisatellite families may form more discrete or unrelated groups in this species compared to peafowl.

Figure 4.2 Autoradiograph showing hybridization of ^{32}P-labeled fragments from two EMBL 3 clones. *Pal* I digested DNA from a pedigree of tilapia individuals consisting of a male and female and full sib offspring. Lanes marked with male and female symbols contain *Hind* III digested DNA from the same male and female as in the pedigree. Left panel: Hybridized to a *Sma* I/ *Hind* III fragment from clone 4. Right panel: hybridized to an 8.0 kbp fragment from clone 16. Markers shown between the panels are a 1 kb Ladder (BRL).

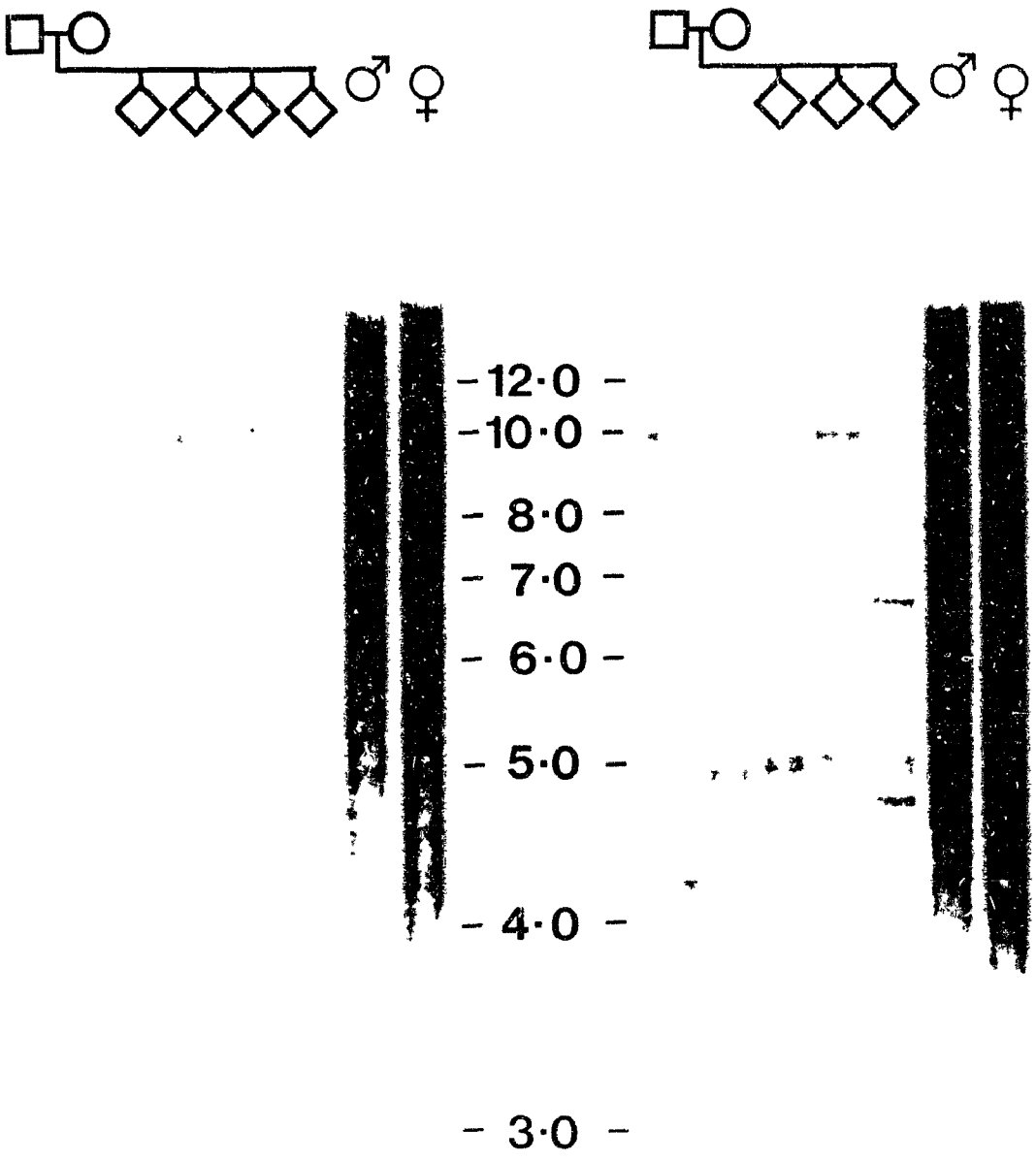


Figure 4.2

Three of the clones produced simplified patterns of hybridization in tilapia (Figures 4.3, 4.4 and 4.5). Clone 34 generated one or two bands in each individual corresponding to the homozygous or heterozygous state at a particular locus. Clone 17 produced two distinct groups of bands whereas clone 32 gave a simplified pattern of hybridization with between two and ten bands per individual. Four clones (1, 17, 32 and 34) were selected for more detailed analysis.

Polymorphism and inheritance patterns of cloned loci

To establish the inheritance pattern and the degree of heterozygosity at these loci, each clone was hybridized to DNA from 22 fish, including a pedigree of male, female and three full sib offspring. Clones 1 and 17 produced identical patterns of hybridization with this tilapiine pedigree. Restriction maps of the EMBL-3 clones of these two loci showed that they differed only by the size of a *Sma* I/*Hind* III fragment containing the minisatellite array (Figure 4.1). Hybridization of the *Sma* I/*Hind* III fragment of each clone to Southern blots of *Pal* I digested DNA detected two distinct loci (Figure 4.3). A strongly hybridizing locus with alleles ranging in size between 4.0 and 5.0 kbp and a second more weakly hybridizing locus with alleles of between 1.8 and 2.5 kbp. Both loci showed Mendelian inheritance of alleles where each offspring band corresponds to one or other of the parental bands. The larger locus shows at least 4 size-variants in the individuals tested with the smaller locus showing 5 allele sizes. The low level of allelic variability may be misrepresentative, due to the likelihood of

Figure 4.3 Inheritance pattern and polymorphism of minisatellite loci detected by OniMS17. Hybridization of a restriction fragment from clone 17 to *Pal* I restricted genomic DNA from 22 fishes. Lanes a - e contain DNA from male and female tilapia and three full sib offspring randomly selected from a brood. Lanes g - n contain DNA from individuals collected from artesinal facilities in Thailand. Lanes o - q contain DNA from individuals kept at Dalhousie University originating from Thailand (Chilandra strain). Lanes r - v contain individuals from stocks at the Institute of Aquaculture, University of Stirling, Scotland. Molecular size markers shown at the left of the panel are a 1 kb ladder (BRL).

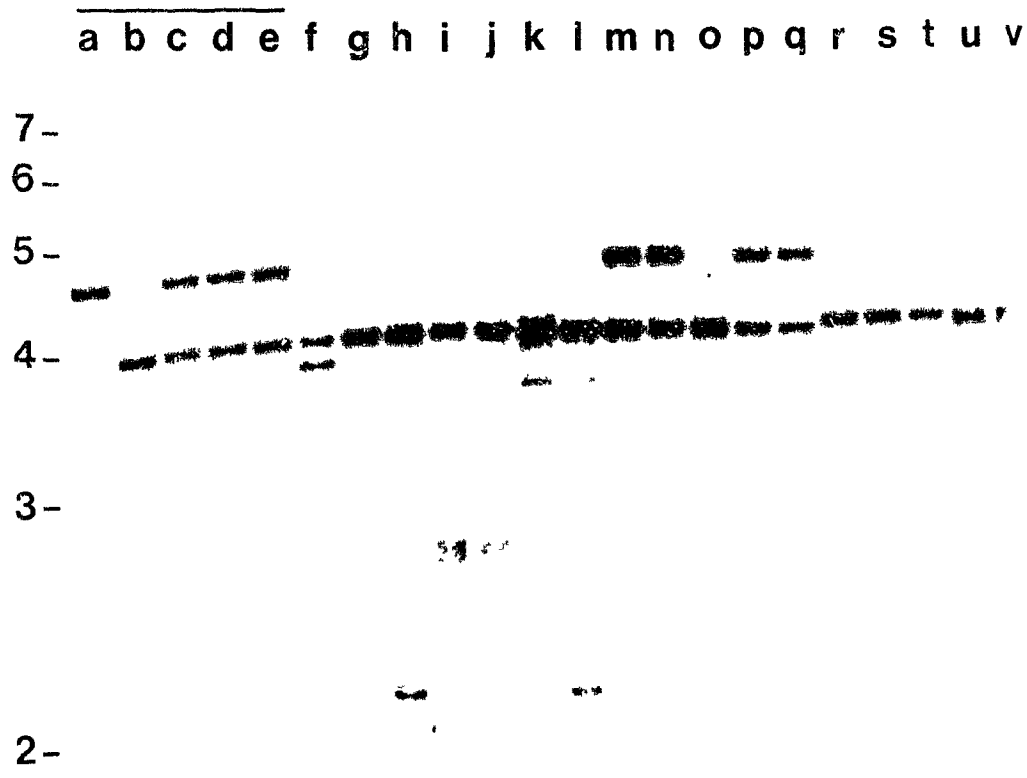


Figure 4.3

Figure 4.4 Inheritance pattern and polymorphism of minisatellite loci detected by OniMS32. Hybridisation of a 1.9 kbp BamH I/ Sma I restriction fragment from clone 32 to *Pal* I digested genomic DNA from 22 fishes. See Figure 4.3 for explanation of samples. Molecular size markers shown to the left of the panel are a 1 kb ladder (BRL).

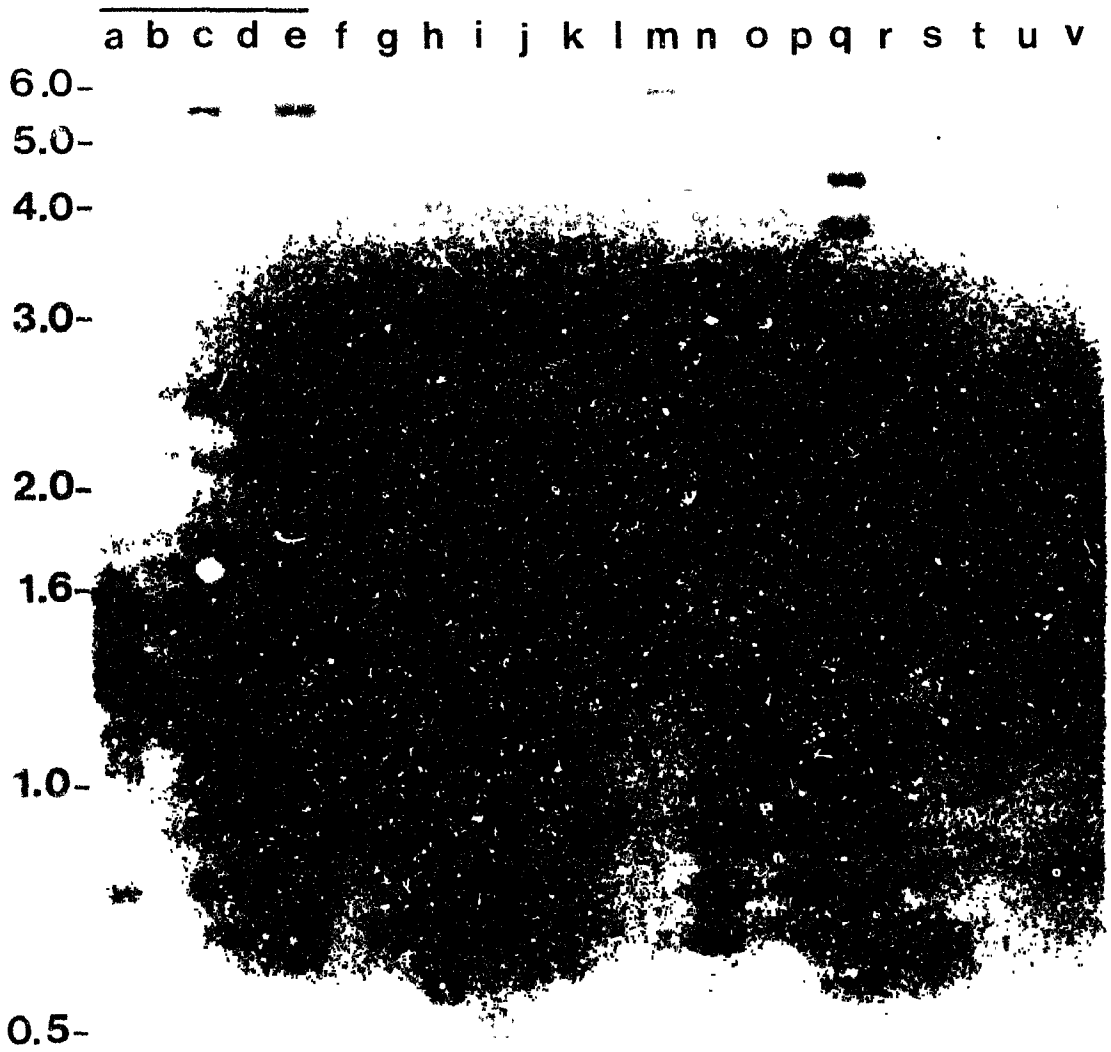


Figure 4.4

Figure 4.5 Inheritance pattern and polymorphism of minisatellite loci detected by OniMS34. Hybridization of a 4.3 kbp *Kpn* I/ *Hind* III restriction fragment from clone 34 to *Pal* I restricted genomic DNA from 22 fishes. See Figure 4.3 for explanation of samples. Molecular size markers shown to the left of the panel are a 1 kb ladder (BRL).

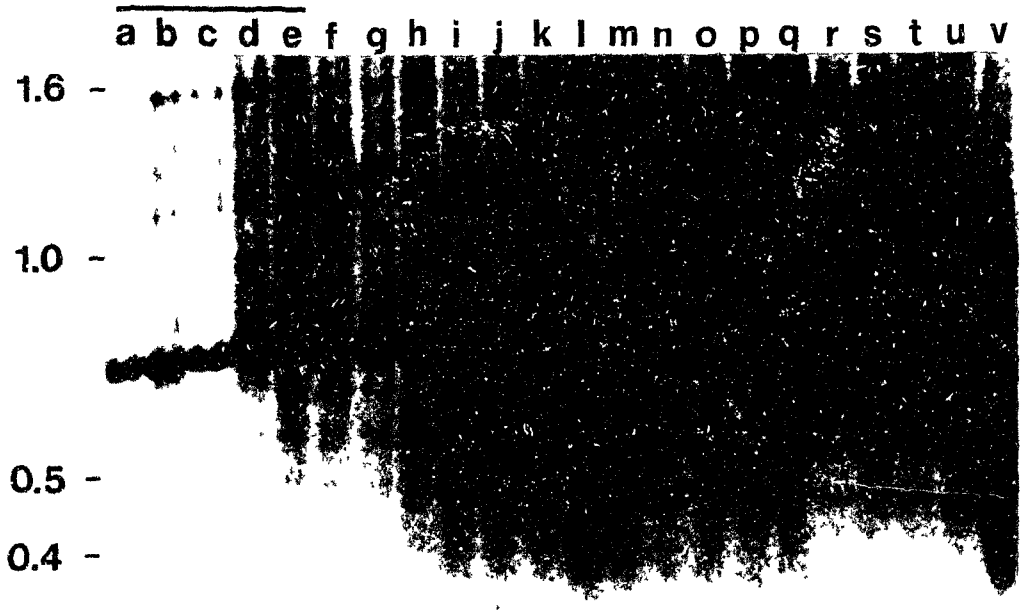


Figure 4.5

inbreeding in the original stock from which these individuals were taken.

The restriction map of clone 32 showed that the insert of the EMBL 3 vector comprised two minisatellite arrays in close association. One of the arrays is contained within a 4.0 kbp *Sma* I/*Bam*H I fragment. This minisatellite hybridized to the 33.6 human core probe. When hybridized to tilapiine genomic DNA, it showed a generalized smear of hybridization in all lanes (data not shown). A second locus, contained within a 1.9 kbp *Bam*H I/*Sma* I restriction fragment hybridized to the 33.15 core probe. This locus generated a multi-band pattern in tilapia individuals with between one and 10 bands per lane (Figure 4.4). This multi-band pattern was reproduced under very high stringencies of hybridization and with the use of competitor genomic DNA in the hybridization reaction. No clear pattern of inheritance could be established for the locus detected by this cloned fragment.

The minisatellite contained within clone 34 was located within a 4.3 kbp *Kpn* I/ *Hind*III fragment of the insert of the EMBL-3 vector. Hybridization of this fragment to tilapiine genomic DNA produced a simple pattern of one (homozygous) or two (heterozygous) bands per individual (Figure 4.5). At least six size variants were detectable, varying in size from approximately 650 bp up to 1.3 kbp. The exact number of variants is uncertain because of electrophoretic anomalies within the gel in the low molecular weight range. Because of this, the inheritance pattern cannot be determined with certainty, however, it is consistent with Mendelian patterns. The locus appears to exhibit a low level of polymorphism but as stated

above, this may be a reflection of the potentially inbred state of the captive fish surveyed.

CHAPTER 5: NUCLEOTIDE SEQUENCE OF VNTR's FROM TILAPIA

Instability of cloned minisatellite arrays

To facilitate sequence analysis of the cloned minisatellite loci, fragments of those EMBL 3 clones containing minisatellite arrays were subcloned into plasmid vectors. EMBL clone 32 contained two *Bam* HI/ *Sma* I fragments bearing minisatellite arrays, one of 4.0 kbp that hybridized to the 33.6 probe and the other of 1.9 kbp that hybridized to the 33.15 probe. Each of these was subcloned into the replicative form of the M13mp18 sequencing vector and propagated in *E.coli*, TG1. Repeated attempts at sequencing the 4.0 kbp fragment proved largely unsuccessful due to the following:. In the few subclones which produced positive plaques under X-gal selection, there was no evidence of tandemly arrayed sequence. In several clones which did contain insert DNA between the cloning sites, this was found to be fragments of the M13 vector originating distally from the cloning site. These sequences had presumably been inserted at the cloning site by some type of rearrangement event. Whether this was the result of an enhanced rate of recombination in this construct due to the presence of a minisatellite array is unknown. Wahls and coworkers (1990) have shown that the Jeffreys' derived 'core' sequence stimulated recombination by 13.5 fold when inserted into circular plasmids. This enhanced recombination has been shown to occur in both directions and at a distance from the core array.

Problems of instability of cloned minisatellite arrays have been reported in other species. The murine minisatellite λMm3 showed

considerable instability during cloning in a λ replacement vector, generating an heterogeneous set of fragment sizes from an original 7.0 kbp array down to 400 bp (Kelly et al. 1989). Similarly, the Atlantic salmon minisatellite, Ssa1, produced M13 recombinants which varied in size from < 300 bp to 3.0 kbp (Bentzen and Wright 1993). Sequence analysis of the subclones revealed that the length changes were due to variation in the number of units within the repetitive array. The regions flanking the array appeared to show no change in sequence (data not shown).

In light of the difficulties encountered with subcloning minisatellites into single stranded M13 vectors it was decided to ligate fragments from clones 17 and 34 into the plasmid pUCBM21. Sequencing was then performed on the double stranded form using the universal sequencing primers and a T7 polymerase protocol. The host used for the propogation of these sub-cloned minisatellites was *E.coli* DH 5α. Using this system the problems of unwanted instability were resolved. Whether this was the result of the double stranded nature of the vector or some property of the bacterial host is unknown.

Nucleotide sequence of VNTR's

The 1.9 kbp fragment from clone 32 was subcloned in both orientations in the M13mp18 and mp19 vectors. Each clone was sequenced to the limits of resolution from the universal priming site of M13. Oligonucleotides were then designed to prime at positions within the insert DNA flanking the minisatellite and the sequence extended into the repeat region. The locus, designated OniMS32,

consists of a 24 bp iteration repeated 20 to 22 times in the cloned sequence (Figure 5.1). The array appears to be highly homogenized, the sequenced portion consisting of only four different repeat types (A,B,C and D) amongst the thirteen present. Each repeat type differs by a single base deletion or substitution. The 24 bp repeat unit does not appear to show a high degree of sequence similarity with the 33.6 polycore probe with which it was isolated.

The minisatellite arrays from clones 17 and 34 were subcloned into the plasmid vector pUCBM20/21. Subclones were sequenced using a similar approach to the sequencing of OniMS32, but because of the double-stranded nature of the plasmid vector, sequencing reactions could be primed from either the forward or reverse directions. Again, clones were sequenced to the limits of resolution. Synthetic oligonucleotide primers were then constructed to prime at points within the insert DNA to allow extension into the repeat region.

Because of the small size of the cloned minisatellite from OniMS34, it was possible to determine the sequence of the entire array (Figure 5.2). The array consists of 31 repeats of a 7 to 9 bp unit. There was a high degree of homogenization of the repeat with a total of five different repeat types evident, differing by only one or two deletions or substitutions. The variant units seem to cluster towards the 3' end of the array, a finding similar to that observed in some human minisatellites (Jeffreys et al. 1991). In these instances MVR mapping of human minisatellite locus, MS32, showed strong similarity in the 5' region between different allele sizes. The 3' end of

Figure 5.1 Nucleotide sequence of minisatellite OniMS32. The figure shows the repeat units of the array aligned for sequence comparison. Sequence above and below the repeats corresponds to 5' and 3' flanking regions. The repeat number and repeat type are indicated to the right of the column. The sequenced portion of the array shows four repeat types designated A,B,C and D. Asterix have been added to maintain sequence alignment.

```
GCCTGAATGGCTGAGTGTAGTAATGGTTGAGGTGGAAGAGAAACAGGTGG
TAGTGTGGATGATGGAGTCAGCGGGCACACAGGAGAC

                                    Repeat Type

     AGAGCTTCAGGCTGGGCAGCTAAG       1.        A
     AGAGCTTCAGGCTGGGCAGCTAAG       2.        A
     AGAGCTTCAGGCTGGGCAGCTAAG       3.        A
     AGAGCTTCAGGCTGGGCAG*TAAG       4.        B
     AGAGCTTCAGGCTGGGCAGCTAAG       5.        A
     AGAGCTTCAGGCT...
                              (7-9 24 bp repeats)
        ...TCAGGCTGGGCAGCTAAG
     AGAGCTTCAGGCTGGGCAGCTAAG       6.        A
     AGAGCTTCAGGCTGGGCAGCTAAG       7.        A
     AGAGCTTCAGGCTGGGCAGCTAAG       8.        A
     AGAGCTTCAGGCTGGGCA*CTAAG       9.        C
     AGAGCTTCAGGCTGGGCAGCTAAG       10.       A
     AGAGCTTCAGGCTGGGCAGCTAAG       11.       A
     CGAGCTTCAGGCTGGGCAGCTAAG       12.       D
     CGAGCTTCAGGCTGGGCAGCTAAG       13.       D

AGTAGTCTCAGGAAAACAGTCTTTGAAGTGTTTATTACCTGCTCTGAGTAAGCCAAGAAGAA
CGTCGATAGTCACAGTCGGATGATCCATCGTTCAGGTCAGACATCGAAT
```

Figure 5.1

Figure 5.2 Nucleotide sequence of minisatellite OniMS34. The figure shows the entire array with repeat units aligned for sequence comparison. Sequence above and below the repeats correspond to the 5' and 3' flanking regions. Repeat number and repeat type are shown to the right of the column. The array consists of five repeat types A, B, C, D and E. Asterix have been added to maintain sequence alignment.

GCCCGTACCGCGGCCGCCTTAAGACGCGTGGATCCCCCCAAGAAAAACATTTGTTTTTATGCAGC
AGCTTTTGTCCTACTGAAACAGCTCGTCCTGGTACGGACGTCACACAAACAGTTGGACTCACTGT
TTCTGCTGCTTTATCTCCAAAACTTTCTTTAGTTTTGCTGAAATTCAGGATTTCACTGTGACGAG
ACTGTTTTATCTGTATCAGACAAACAGCATGAATAACTCATTCATTTGCTGACACAGCAACACAC
TTTGTCAATCCAGCA

	Repeat	Type
GGGGGCTGT	1.	A
GGGG*CTGT	2.	B
GGGGGCTGT	3.	A
GGGG*CTGT	4.	B
GGGGGCTGT	5.	A
GGGG*CTGA	6.	C
GGGG*CTGT	7.	B
GGGG*CTGT	8.	B
GGGGGCTGT	9.	A
GGGG*CTGA	10.	C
GGGG*CTGT	11.	B
GGGGGCTGT	12.	A
GGGGGCTGT	13.	A
GGGG*CTGT	14.	B
GGGGGCTGT	15.	A
GGGGTTTTA	16.	D
GGG**CTGT	17.	E
GGGG*CTGT	18.	B
GGGGGCTGT	19.	A
GGGG*CTGA	20.	C
GGGG*CTGT	21.	B
GGGGGCTGT	22.	A
GGGG*CTGT	23.	B
GGGGGCTGT	24.	A
GGGGTTTTA	25.	D
GGG**CTGT	26.	E
GGGG*CTGT	27.	B
GGGGGCTGT	28.	A
GGGGTTTTA	29.	D
GGG**CTGT	30.	E
GGGG*CTGT	31.	B

CAGAATTTCATTGCTACACTGTGTGTGATGATATTTTATTAATCAATATACTTTTTGCTTTATCT
ATATGACATTTCTTACAGTGCATTTTGCCGCCCATTCACACCAGGGCTTGTTAATAATGACAGAA
ATGTAATCAGCGATATTTGACTCGATACTGTAACCTGTGGGAAAGAA

Figure 5.2

the array however, showed far greater interallelic variability. Jeffreys et al. (1991, 1994) have suggested that this polarity in mutation events reflects the preferential location of length changes in the 3' region. Jeffreys et al. (1991) speculate that mutation at minisatellites is modulated by a *cis* acting sequence 5' to the array which controls the generation of a double-strand break and results in a gene conversion event. Whether similar processes occur in tilapia has not been investigated.

The sequence of OniMS17 displays a much more complex repeat pattern than the other cloned tilapiine minisatellite arrays (Figure 5.3). The cloned array of OniMS17 consists of an underlying 15 base pair repeat which contains 150 to 155 iterations within the array. At least 27 different repeat types were present within the array. Individual pairwise comparison shows widely varying degrees of sequence identity, ranging from 93% (D and J) to 40% (D an F). Alignment of the 15 bp array shows blocks of conservation within all repeat units. These consist of a GA pair at positions 4 and 5, a pyrimidine residue at position 8 and GAG/T at positions 13, 14 and 15.

Nested repeat structure of OniMS 17

The novelty of the OniMS17 array arises from the presence of a second overlying repeat pattern. Examination of the aligned 15 bp repeats reveals a 90 to 120 bp periodicity of certain repeat types (Figure 5.4). At the 5' end of the array comparison of the repeat units at similar positions within the larger array shows them to either be of identical sequence or have a high degree of sequence similarity.

Figure 5.3 Nucleotide sequence of minisatellite OniMS17. Repeating sequence is aligned for sequence comparison. The sequenced portion of the array consists of 26 repeat types of 15 bp labeled A through Z. Repeat number and repeat type are shown to the right of the column. Sequence above and below the repeats corresponds to the 5' and 3' flanking regions.

GCGCAAATTAATGCAAGAACTAGTAACATAACACCTGAGTTTATTCCATTCCACCTGTTGATGCT
ACAAAATGTGCCAGGACTGAATGGAAATCTGACCTTGGTTTTGGAGACAATTTAGTCTGTGGCTC
TGGTTCCATTTTGGGTTCAGGTTTTGTGTTCAGTC

	Repeat	Type
CAGATTCTTTCACAG	1	A
CAGGCACTGGGACAG	2.	B
C'GACTTGGTGACAG	3.	C
TGGAGGTTGGGACAG	4.	D
GAGACACTTTGACAT	5.	E
CTGATTTTTCAGGAG	6.	F
TTGAAGTCACATCAG	7.	G
GGGACACAGGTGCAG	8.	H
GTGACTTGGTGGCAG	9.	I
TGGAGGTCGGGACAG	10.	J
TGGAGGTCGGGACAG	11.	J
TGGAGGTCGGGACAG	12.	J
GAGACACTTTGACAT	13.	E
CTGATTTTGCAGGAG	14.	K
TTGAAGTCACGTCAG	15.	L
*GGACACAGGTGCAG	16.	M
GTGACTTGGTGACAG	17.	N
TGGAGGTTGGGACAG	18.	D
TGGAGGTCGGGACAG	19.	J
GAGACACTTTGACAT	20.	E
CTGATTTTGCAGGAG	21.	K
TTGAAGTCACATCAA	22.	O
GGGACACAGGTGCAG	23.	H
GTGACTTGGTGACAG	24.	N
TGGAGGTTGGGGCAG	25.	P
TGGAGGTCGGGACAG	26.	Q
TGGAGGTTGGGACAG	27.	D
T----------		
	(110-115 15 bp repeats)	
------TCCGGACTG		
GACACTTTGACCATC	28.	R
TGGATTTTGCAGGAG	29.	S
TTGAAGTCACCTCAG	30.	T
GACACAGGTGGCCAG	31.	U
ATGGTGGTGGACCAG	32.	V
TGGAGGTTGGGGCAG	33.	P
TGGAGGTCGGGGCAG	34.	W
TGGAGGTCGGGACTG	35.	X
GAGACACTTTGACAT	36.	E
CTGATTTTGCAGGAG	37.	K
TTGAAGTCACATCAG	38.	G
GGGACACAGGTGTAG	39.	Y
GTGATTTGGTTACAG	40.	Z
AGGTTGGGACAGGAG	41.	AA

ACACTGCAGTGGTCGACTCAGGGGACTTTTGATTAGATGATATTGGGGGTTTTGATGGTGGCGAA
AGATCAACCAGAGGCGTTTTAGTCTGCCCAGACTCTTTTTGTTGTGGCTCTTTAGAGGCTACTTG
AAGACTATCTTTGATTAGTTCAGGGAGGGGAGCAGCCACCTTGGTGTCAAAAGTT

Figure 5.3

Figure 5.4 Schematic representation of the nested repeat structure of OniMS17. The 5' and 3' regions of the array are depicted. Each box represents a single 15 bp repeat unit. Repeat unit labels correspond to those shown in Figure 5.3. The 90 - 120 bp periodicity of the nested repeating units are shown beside each itteration. The 15 bp units which appear responsible for the differences in nested repeat size are hilighted. Hatched boxes correspond to regions of unique sequence flanking the array.

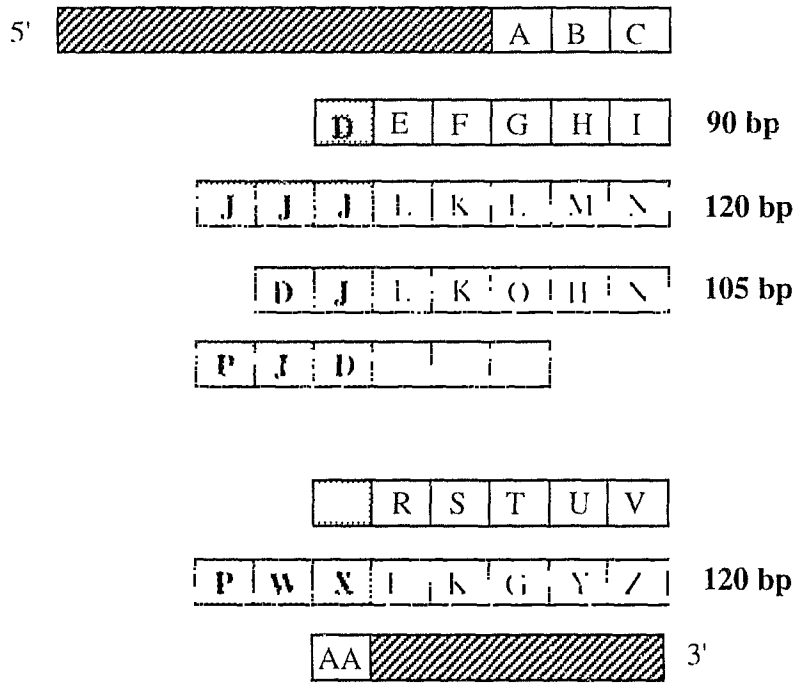


Figure 5.4

The variation in the larger repeat iterations appears to be generated at a particular single 15 bp repeat type (designated type D, J or P). The repeat units at these positions differ by only a single base substitution. These r ,eats are present in one, two or three copies, thus generating the repeat unit size variations in the overlying larger iteration. Interestingly, the human polycore probe (33.6) used to isolate OniMS17 consists of a diverged trimer of a basic 12 bp repeat within its 33 bp repeating unit (Figure 5.5). To date these are the only reported examples of this type of nested structure. The complex nature of the OniMS17 array may provide some insight into the mechanism involved in the generation of this minisatellite. It appears that the ancestral sequence originated as a 15 bp block of sequence which underwent duplication and amplification. This array may have undergone a period of amplification and decay prior to reaching a stable copy number of six to eight repeats or it may have remained inactive. The considerable sequence divergence between the 15 bp repeats within each 90/120 iteration suggests that the minisatellite was maintained at low copy number for a period of time. It is possible that the 90/120 bp array, at some later time, underwent a second duplication and amplification event to generate the present minisatellite locus.

Conservation of a common core sequence

The fact that certain 15 bp repeat types in OniMS17 are involved in the process of array length change may have some significance. Upon isolation of the first minisatellite arrays with the myoglobin-derived 33 bp probe, Jeffreys et al. (1985a) described an

Figure 5.5 Comparison of nucleotide sequences of minisatellite arrays. The Nakamura core sequence G***GTGGG or a degenerate version is underlined in each sequence (Nakamura et al. 1987). Numbers to the right of the sequences indicate where more than one repeat has been concatenated to illustrate the core sequence. The sequences shown are: CHI, *E.coli chi* sequence, 33.6 and 33.15 polycore probes (Jeffreys et al. 1985a, b), MTHIRTEEN, bacteriophage M13 gene III region (Vassart et al. 1987); MYOGLOBIN (Jeffreys et al. 1985a); ZETAG, zeta-globin pseudogene (Proudfoot et al. 1982); INSULIN, tandem repeat region adjacent to the insulin gene (Bell et al. 1982); PλG3, clone pλg3 (Wong et al. 1986); HBV-1 and HBV-2, 16 mer and 20 mer oligonucleotides corresponding to a portion of the hepatitis B virus X gene region (Nakamura et al. 1987); λMS1, λMS31, λMS32, λMS8(1)(2) and λMS43, cloned human minisatellite loci (Wong et al. 1987); YNZ2, YNZ22, YNZ132, YNI10, TTH59 and YNH24, cloned human minisatellite loci (Nakamura et al. 1987); DXYS14, human minisatellite adjacent to the pseudoautosomal telomere (Inglehearn and Cooke 1990); MOUSEKO, mouse minisatellite (Kominani et al. 1988); LTHIRTEEN and LEIGHTEEN, bird minisatellites (Gyllensten et al. 1989); ATLSAL, minsatellite Ssa1 from Atlantic salmon (Bentzen and Wright 1993); OniMS17, OniMS32 and OniMS34. In the figure R represents a purine residue, Y represents a pyrimidine residue and N represents any nucleotide.

Minisatellite	Sequence
CHI	GCTGGTGGGCTGGTGG[2]
33.6	GGAGGAGGGCTGGAGGAAGGGCTGGAGGAGGGCTCC
33.15	GGGGCAGGTGGAGAGGT
MTHIRTEEN	GAGGGTGGNGGNTCT
MYOGLOBIN	GGAGGTGGGCAGGAACGACCGAGGTCTAAAGCT
ZETAG	GGTTGTGAGTGGGGCACAGGTTGTGAGTGGGGCACA[2]
INSULIN	ACAGGGGTGTGGGGACAGGGGTGTGGGG[2]
PλG3	AGAAAGGCGGGYGGTGTGGGCAGGGAGRGGCAGGAAT
HBV-1	GGAGTTGGGGGAGGAG
HBV-2	GGGGACTGGGAGGAGTTGGG
λMS1	GTGGAYAGGGTGGAYAGG
λMS31	TGGGAGGTGGRYAGTGTCTGTGCGAGGTGGRYAGTGTCTG
λMS32	GGYGRCCAGGGGTGAATGGAGCAGGYGRCCAGGGGT
λMS8 (1)	GGAGGTGTTGGGGGCTGGGGACATGGTGGAGGAGGTGTTGG
(2)	AGGCTGGGGAGATGGTGGAGGAAGAGTAC
λMS43	TGTGTGTAATGGGTATAGGGAGGGCCCCGGGAAGGGGGTGTGGYG
YNZ2	GAGGCTCATGGGGCACAGAGGCTCATGGGGCACA[2]
YNZ22	TGGAGTCTCTCGGTGTCGTGCGTCAGAGT
YNZ132	TGCAGGCTGTGGGTGTGATGGGTGA
YNI10	CCAGGCTGTCAGATGCTCACCTGGGGTGTCGGTGCTGCTCCAGGCTGTCAGATGCTCAC
THH59	CTGGGGAGCCTGGGGACTTTCCACACC
YNH24	CAGGAGCAGTGGGAAGTACAGTGGGGTTGTT
DXYS14	GCGGGGGTCTGGGGTGGTCCCGAGGGGAGGA
MOUSEKO	CTGGGCAGGGAGGACTGGGCAGGGAGGA[2]
LTHIRTEEN	GGGGACAGGGGACACCC
LEIGHTEEN	CAGGGAAGGGCTCAGGGAAGGGCT[2]
ATLSAL	CTTCAGGCTGGGCAGCTAGAGA
ONIMS17	CTTTCACAGCAGGCACTGGGACAGGTGACTTGGTGACAGTGGAGG[4]
ONIMS32	AGAGCTTCAGGCTGGGCAGCTAAGAGAGCTTCAGGCTGGGCAGCTAAG[2]
ONIMS34	GGGGGCTGTGGGGCTGTGGGGGCTGT[3]

Figure 5.5

almost invariant core sequence shared by all of the clones. This sequence GGGCAGGAXG was suggested to bear resemblance in both length and G content to the *chi* sequence, a generalized recombination signal in *E.coli*. Nakamura and coworkers (1987) have postulated a somewhat different core sequence, G---GTGGG. This core sequence bears some similarity to that found within the myoglobin family of repeats and an examination of some of the G-rich family of minisatellites indicates that the Nakamura core, or a degenerate form of it, is present in most miniatellite VNTR's, including human, mouse and fish (Figure 5.5). This same core sequence is also found in the M13 gene III repeat, another commonly used multilocus fingerprinting probe. It is perhaps more noteworthy that the G---GTGGG motif is also present in the X-gene region of the hepatitis B virus (HBV). This region has been purported to be involved in the integration of HBV into the host genome (Koshy et al. 1983). This finding provides further support for Jeffreys' initial hypothesis that minisatellite sequences act as recombinational hotspots. This agrees with evidence provided by Wahls and coworkers (1990) of enhancement of homologous recombination by cloned minisatellite core sequence. More recently, using the technique of MVR mapping, Jeffreys et al. (1991) have demonstrated interallelic unequal exchange involved in length changes within arrays of the human minisatellite MS32 (Jeffreys et al. 1991). The frequency of mutational events measured in this minisatellite is achieved by an apparent 700-fold enhancement in recombination frequency over the mean frequency for the human genome. Whether or not similar mechanisms are prevalent in other species remains unknown.

However, within the OniMS17 sequence the Nakamura core is contained within the particular 15 bp repeating subunits that determine the 90 or 120 bp length of the larger array iteration. If this particular 15 bp repeat type does form a focal point at which the mutational processes act within this array, then this provides further evidence of the significance of the Nakamura core in array length changes.

Possible mechanism of generation of OniMS34

The question of how minisatellites arise in the genome still remains unanswered. Wright (1994) has proposed one mechanism responsible for generation of some minisatellites from simple sequence (microsatellite) repeats. Mutational processes such as substitution and transition can generate stretches of cryptic repeats which then become duplicated by the process of slipped-strand mispairing (Levinson and Gutman 1987). An association between minisatellites and microsatellite repeats has been observed in the Ssa1 locus from Atlantic salmon where a cryptic microsatellite composed of polypurine-polypyrimidine repeats is present near the 3' end of the array (Bentzen and Wright 1993). Similar associations between minisatellites and simple sequence repeats have been reported in the bovine genome (Kashi et al. 1990). The sequence of OniMS34 shows a degenerate run of CA repeats in the region immediately flanking the array, suggesting that this array may have arisen from a microsatellite. A number of other possible simple sequence-derived minisatellites have been reported in the literature including the mouse minisatellite locus Ms6-hm which shows a

repeating GGGCA motif (Kelly et al. 1989) and the willow warbler minisatellite L 13 which has a GGGGACAGGGGACACCC consensus repeat sequence (Gyllensten et al. 1989).

At least three other possible mechanisms for minisatellite generation have been proposed. Jeffreys and coworkers (1985) from observations of a 9 bp direct repeat flanking the myoglobin locus suggested that some minisatellites may be related by retroviral mediated transposition. A second hypothesis is that the core sequence of the repeating units is inherently able to expand (Wolff et al. 1991). This expansion from a single copy of the core, which presumably arises by random mutation, occurs at multiple dispersed loci independently. A third possibility is that minisatellites may expand from other repetitive sequences. A number of examples of this have been presented in the literature (Armour et al. 1989, Mermer et al. 1987). No strong evidence presently exists to either refute or support a particular theory of minisatellite propogation and maintainance. It appears possible then that several diverse mechanisms exist by which minisatellite loci are generated and that these act in concert within the genome.

CHAPTER 6: EVOLUTIONARY CONSERVATION IN RELATED CICHLIDS

Homologous loci in other Cichlids

Several studies have demonstrated the conservation of minisatellite loci in related taxa. Georges and coworkers (1991) showed that bovine VNTR's hybridized to genomic DNA of sheep. They speculate that these sequences represent evolutionary homologues in both bovine and sheep species. In the primate genome, two human-derived minisatellite loci have been characterized in Old world and New world monkeys and great apes (Gray and Jeffreys 1991). To determine whether the cloned *O. niloticus* minisatellite loci were conserved in related taxa, DNA from a number of species within and outside the cichlid family was hybridized to each of the three minisatellite probes. Genomic DNA from individuals of three species of the genus *Oreochromis* (*O. niloticus, O. aureus and O. mossambicus*), one species from each of the closely related genera *Tilapia* (*T. zillii*) and *Sarotherodon* (*S. galilaeus*) was included. One other species of Old World cichlid (*Haplochromis moori*), one representative of Asian cichlid (*Etroplus maculatus*) and one representative of South American cichlasomine species (*Cichlasoma otofasciatus*) as well as an Atlantic salmon (*Salmo salar*) which represented a species outside the tilapiine tribe.

Hybridization of the cloned minisatellite locus OniMS17 to these genomic DNA's generated a simple pattern of bands (Figure 6.1). For all of the tilapiine species and the haplochrome *H. moori*, the Southern blot revealed that the individuals exhibit one or two

Figure 6.1 Evolutionary conservation of a minisatellite locus detected by OniMS17 in Old and New world cichlids. Cloned minisatellite OniMS17 was hybridized to *Pal* I-digested genomic DNA from one to three individuals of *Oreochromis niloticus* (O.nil), *Oreochromis aureus* (O.aur), *Oreochromis mossambicus* (O.mos), *Tilapia zilli* (T.zil), *Sarotherodon galilaeus* (S.ga), *Haplochromis moori* (H.mor), *Etroplus maculatus* (E.m), *Cilchlasoma otofasciatus* (C.oto) and a salmonid, *Salmo salar* (S.sal). Molecular size markers shown at the left of the panel are a 1 kb ladder (BRL).

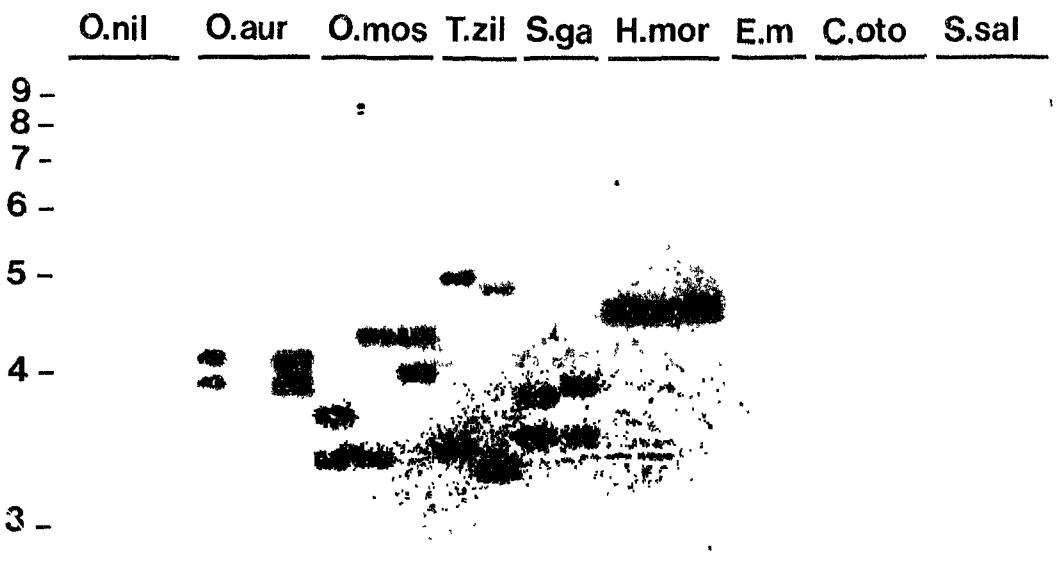


Figure 6.1

hybridizing bands. Band sizes range from approximately 3.2 kbp to 5.0 kbp, falling within the range of sizes found in the *O. niloticus* individuals assayed (Figure 4.3). Hybridization of OniMS17 to genomic DNA was clearly evident in all of the tilapiine tribe and in the closely related African cichlid *H. moori*. However, no hybridization was observed in the Asian or South American cichlids or the Atlantic salmon, under the conditions used.

In contrast, the hybridization pattern generated by OniMS34 is distinctly different from that of OniMS17 (Figure 6.2). Band patterns were variable between species. The *Oreochromis* species showed one or two strongly hybridizing bands on a diffuse background of weaker hybridizing bands. The more distantly related *Tilapia* and *Sarotherodon* species showed four or five hybridizing bands with no specifically hybridizing locus. However, the *H. moori* individuals produce one or two bands as did the South American cichlid (*C. otofasciatus*). Interestingly, a large number of distinct hybridizing bands was generated in the Atlantic salmon. The reason for this is not clear, although it may indicate a family of minisatellite loci with sequence identity similar to the OniMS17 core sequence in this salmonid genome.

A mixed pattern of hybridization was produced by the OniMS32 probe (Figure 6.3). Individuals showed several distinct bands on a diffuse weakly hybridizing background. These strongly hybridizing bands appear to be limited to the *Oreochromis* and *Tilapia* genera unlike the pattern observed for OniMS34. The greater intensity of signal in the larger molecular size

Figure 6.2 Minisatellite loci detected by OniMS34 in Old world and New world cichlids and a salmonid. Cloned minisatellite probe OniMS34 was hybridized to *Pal* I-digested genomic DNA from one to three individuals of *Oreochromis niloticus* (O.nil), *Oreochromis aureus* (O.aur), *Oreochromis mossambicus* (O.mos), *Tilapia zilli* (T.zil), *Sarotherodon galilaeus* (S.ga), *Haplochromis moori* (H.mor), *Etroplus maculatus* (E), *Cilchlasoma otofasciatus* (C.o) and a salmonid, *Salmo salar* (S.sal). Molecular size markers are shown to the left of the panel.

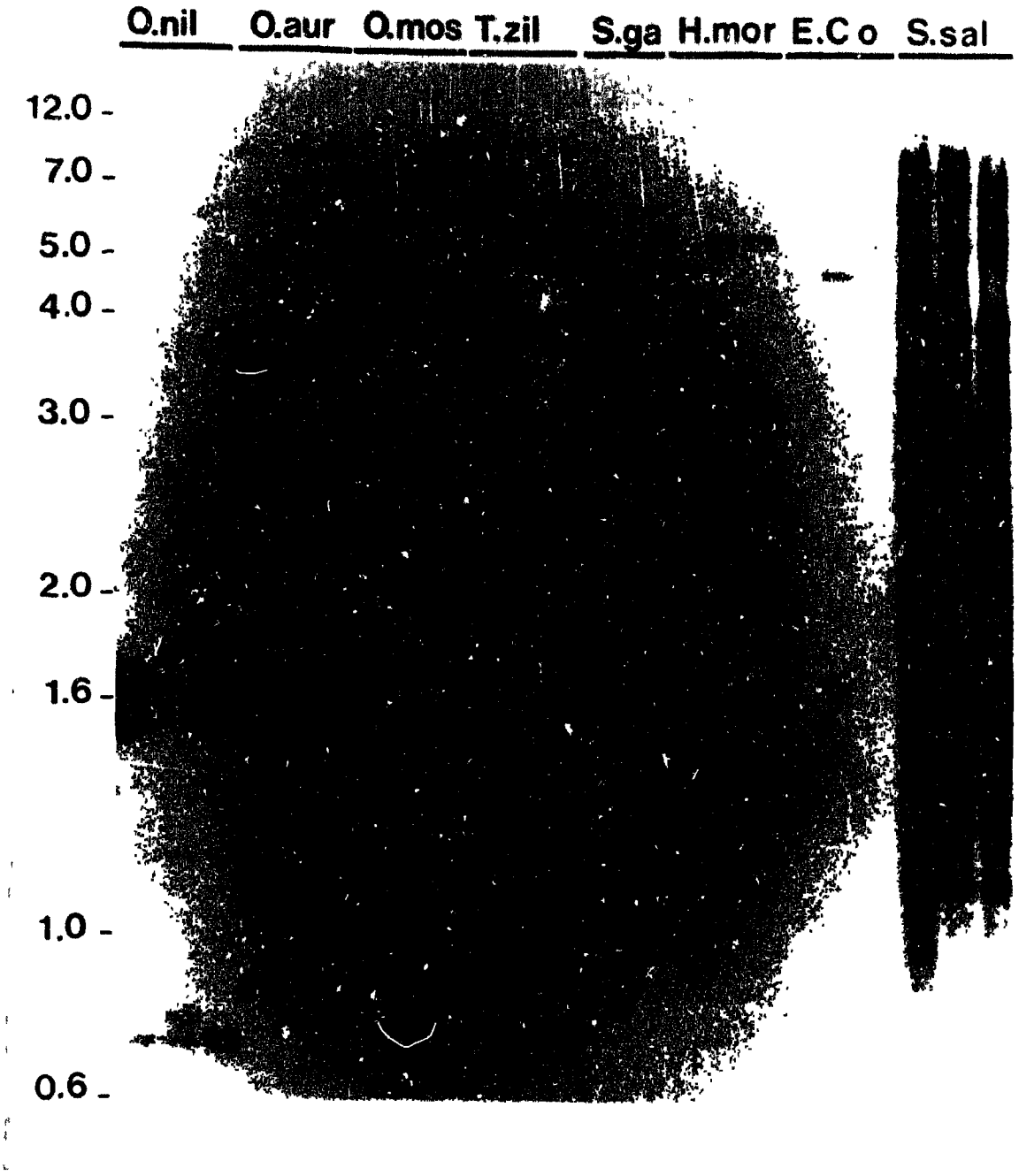


Figure 6.2

Figure 6.3 Detection of minisatellite loci with related core sequence to OniMS32 in Old world and New world cichlids. Shown are one to three individuals of *Oreochromis niloticus* (O.nil), *Oreochromis aureus* (O.aur), *Oreochromis mossambicus* (O.mos), *Tilapia zilli* (T.zil), *Sarotherodon galilaeus* (S.ga), *Haplochromis moori* (H.mor), *Etroplus maculatus* (E), *Cilchlasoma otofasciatus* (C.ot) and a salmonid, *Salmo salar* (S.sal). Molecular size markers shown to the left of the panel are a 1 kbp ladder (BRL).

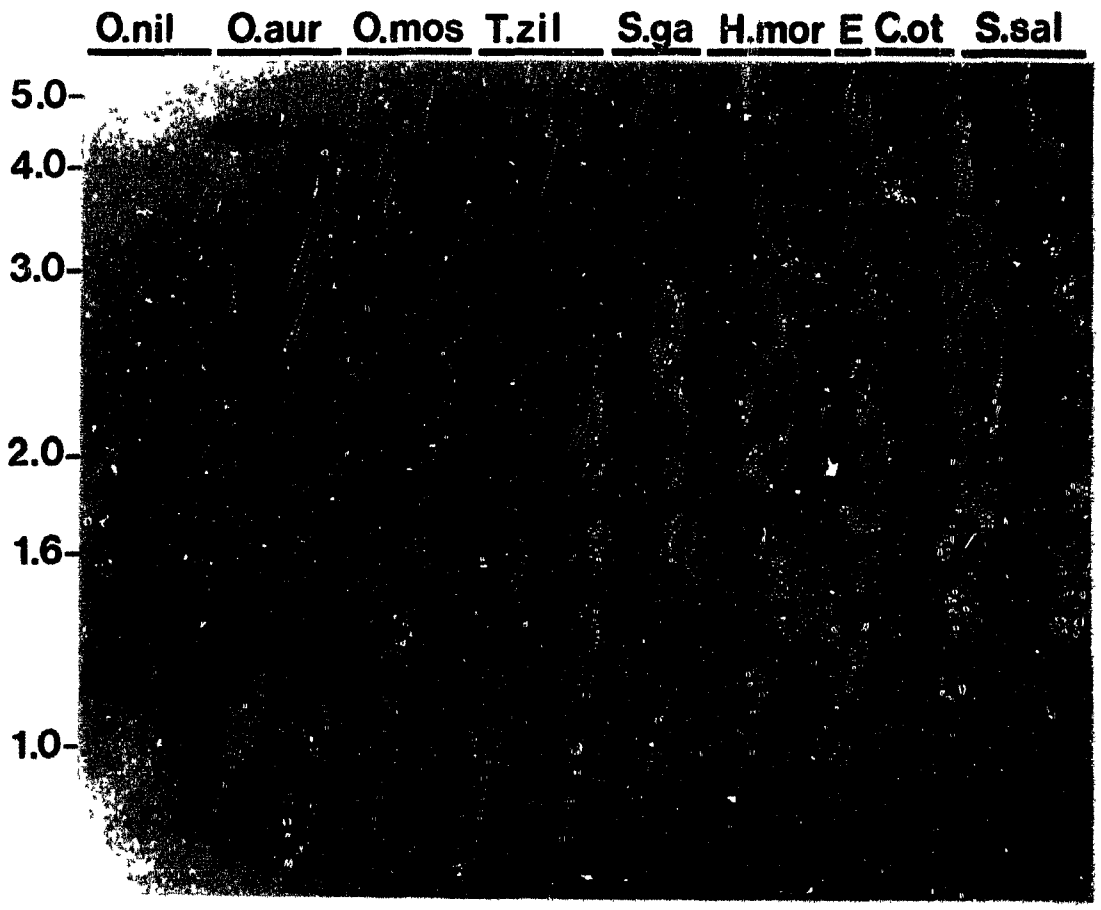


Figure 6.3

range is consistent with longer repeat arrays. However, whether these bands represent hybridization between arrays of similar repeat sequence or with sequence flanking the probe array is uncertain.

Similar patterns of hybridization have been reported for minisatellites isolated from the Indian peafowl (Hanotte et al. 1991). Two probes, cPcr5 and cPcr10 detected loci only in closely related peafowl species. In comparison, probe cPcr1 detects loci in more distantly related Phasianidae species. Bentzen and Wright (1993) have shown that a minisatellite probe, Ssa1, isolated from Atlantic salmon (*S.salar*) hybridizes to loci in species within the same subfamily (*Salmonidae*) and in more distantly related salmonids in the subfamily *Coregoninae*. From this observation they propose that the locus recognized by the Ssa1 probe has been conserved for at least 25 million years since the split of the two subfamilies. This is similar to estimates made by Gray and Jeffreys, of 25 million years for the conservation of minisatellite locus MS1 in human and Old world monkeys. Loci, with sufficient sequence similarity to show hybridization to the OniMS17 probe, have only been conserved within the tilapiine tribe and in the closely related haplochromine lineage. The presence of bands that hybridized to the OniMS34 probe in Southern blots of genomic DNA, may indicate homologous loci in the Asian and South American cichlids. This would suggest conservation of these repeats at least since the separation of North and South America and possibly predating the separation of the Gondwanan land mass (Starr and Taggart 1987). A second possibility is that the OniMS34 sequence acts as a multilocus fingerprint probe.

The nonrepetitive clone L17 isolated from the willow warbler (*Phylloscopus trochilus*) has also been shown to generate multilocus patterns of hybridization in birds, mammals and fish (Gyllensten et al. 1989). This may be due to the sequence showing cryptic similarities with minisatellite repeats at other loci because of its guanine rich nature. Similar findings have been described by Bentzen and Wright (1993) for a non-repetitive 5' flanking region of the Ssa1 probe which generates polymorphic bands in a number of salmonids. They have suggested that this may reflect a common relationship between these loci by decent, either through transposition and/or duplication.

Evolutionary transience of polymorphism

Gray and Jeffreys (1991), using studies of human and primate minisatellite loci, have suggested that even though minisatellite loci persist over evolutionary time, extreme polymorphism as seen in some of the human minisatellites may be a transient phenomenon. Wolff and coworkers (1991) investigated a number of loci comparing human and primate sequence. They concluded that there were no factors inherent to the loci that accelerated variability, suggesting that generation of the array was contingent upon the rate limiting step of the initial duplication. It seems then, that the degree of polymorphism at a given locus is dependent upon unknown factors. In three of the salmonine genera, homologues of the Ssa1 locus have apparently remained polymorphic for more than 25 million years (Bentzen and Wright 1993). A similar situation may exist in the cichlids. The OniMS17 probe detects polymorphic loci across the

tilapiines and in the haplochromine *H. moori* (Figure 6.1). Individuals of the *Oreochromis*, *Tilapia* and *Sarotherodon* genera tested, show between two and four allele sizes. Most individuals appear to be heterozygous at this locus. The haplochromine species shows less variation than the Tilapiines with only two alleles appearing to be homozygous in the three individuals. However, the degree of inbreeding in this stock is unknown and could account for the low degree of polymorphism. The heterozygosity in the tilapiine genera suggests that the locus has remained polymorphic at least since the divergence of the tilapiine and haplochromine lineages and possibly as long ago as the divergence of the African and South/Central American lineages.

By contrast the OniMS34 probe shows hybridization to genomic DNA of species across a much broader phylogenetic range (Figure 6.2). Individuals of the Haplochromine species show differences between alleles of up to 3 kbp. This suggests that the array detected by OniMS34 is significantly larger in *H. moori* than in *O. niloticus*. This increase in the size of the locus may reflect an increased number of repeats or a loss of restriction sites for *Pal* I flanking the minisatellite. In the Asian cichlid, *E. maculatus*, two distinct hybridizing bands appear at around 2 kbp. Although it would be tempting to speculate that these represent a polymorphic locus in this species, this cannot be determined without surveying more individuals. In the Soutth American cichlid *C. otofasciatus* fish assayed a single faint band of 5.5 kbp was detected, possibly indicating a locus of similar sequence. Again in these individuals uncertainty about inbreeding in the sample stock precludes any

speculation about variability in this species. It appears then that the OniMS34 locus has remained polymorphic for at least as long as the OniMS17 locus and possibly since the split of the African and South American cichlids.

CHAPTER 7: SUMMARY

The purpose of this thesis was to investigate the structure, genomic organization and evolution of minisatellite VNTRs in tilapia. The presence in the tilapiine genome of minisatellites related by core sequence to those found in humans has been demonstrated here by the generation of DNA fingerprint patterns using the human polycore probes 33.6 and 33.15. These probes generate complex patterns of hybridization similar to those observed in a number of other species. In the tilapiine genome, repetitive minisatellite sequences related to the core probes account for approximately 1% of the total genome size. This compares to figures of 1.6 % for Atlantic salmon (Bentzen and Wright 1993). The utility of DNA fingerprinting has been tested in a study of the production of gynogenetic lines of tilapia. Gynogenetic breeding systems occur naturally in a number of species (Dawley and Bogart 1989). The induction of artificial gynogenesis has been proposed as a method of producing all female stocks (Thorgaard et al. 1983). This prevents problems of precocious sexual maturity of male offspring which often results in uncontrolled reproduction and hence, leads to overcrowding with resulting loss of productivity in aquaculture. Minisatellite probes offer a potentially useful tool for the evaluation of gynogenetic lines.

The isolation, sequencing, genomic structure and evolution of three minisatellite loci from tilapia constitutes one of the first extensive studies of these sequences in fishes. As has been reported in human and mouse studies, the subcloning of these sequences is made difficult by their propensity for rearrangement, frequently by

deletion, during propogation and cloning. Restriction maps of 19 randomly selected minisatellite clones revealed three pairs of isolates with very similar structure. One pair differed by the size of the minisatellite containing fragment possibly due to sequence rearrangement during cloning. A second pair showed identical patterns possibly due to coincidental selection of the same clone or contamination during isolation. A third pair of clones contained different *Mbo* I fragments of the same locus. Similar results of repeated isolates have been reported in other species (Armour et al. 1990). Two of the nineteen clones contained minisatellites in close association. Hybridization studies of clone 32 revealed two arrays with sequence similarity to different polycore probes. This is consistent with clustering of minisatellites reported in humans (Armour et al. 1989).

Sequence information determined from three clones, OniMS17, OniMS32 and OniMS34, demonstrated a wide diversity in array size and organization. OniMS34 contained the shortest repeating unit of seven to nine nucleotides. The complete sequence of the cloned array was determined and indicated a high degree of homogenization of the repeat units with a total of five repeat types. The variations in repeat type appear to cluster towards the 3' end of the array in a similar fashion to that described for human minisatellite loci MS32 (Jeffreys et al. 1991). The sequence of these arrays have also provided some insight into the *de novo* generation of minisatellites in tilapia. OniMS34 is flanked by a degenerate run of CA repeats suggesting that this array may have expanded from a microsatellite

repeat. Similar mechanisms have been proposed for minisatellites from other species (Wright 1994).

The DNA sequence of OniMS32 revealed a 24 bp repeating unit which exhibits a high degree of homogenization within the array. Only four different repeat types were detected in the sequenced portion. Polarity in the repeat unit sequence could not be readily determined because of the limited data from the 3' end of this array.

OniMS17 showed some marked differences in structure from either OniMS32 or OniMS34. The repeat pattern consisted of a 15 bp unit nested within a larger 90 to 120 bp iteration. The 15 bp repeat is extremely variable within the array with sequence identity between repeats ranging from 40% to 93%. The structure of this array possibly reflects the process of amplification involved in the generation of this minisatellite. A putative model might implicate an ancestral 15 bp repeat generated by duplication and amplification. At some later date a 6 to 8 unit array of these 15 bp repeats, underwent a second amplification which generated the present array. The maintenance of a high degree of sequence conservation in the first few repeat units of the 90-120 bp iteration indicates the importance of these sequences in the process of amplification.

The apparent ubiquity of minisatellites, related by core sequence, across the eucaryotes has raised questions about the evolutionary persistence of this family of repetitive sequences. Hybridization of the three cloned tilapiine minisatellites to genomic DNA from closely and distantly-related cichlids reveals some variety in the degree of conservation in these species. Sequences that hybridized to OniMS17 appear to have a high degree of sequence

similarity with loci in closely related tilapiine genera and the more distant African cichlid, *Haplochromis moori*. The minisatellite loci detected by Southern blot hybridization display similar degrees of variability in all of the tilapiine fishes. Each individual tested displayed heterozygosity at this locus suggesting that the minisatellite has remained polymorphic in these species, at least since the divergence of the tilapiine and haplochrome lineages. The multilocus pattern of hybridization generated by OniMS32 suggests that, although these loci show weak hybridization in the *Oreochromis* and *Tilapia* genera, they are not well conserved in other related species. In contrast, OniMS34 when hybridized to genomic DNA from other cichlids generated distinct bands in all of the tilapiines as well as the African and South American cichlids. The appearance of hybridizing bands in the salmon outgroup is more likely due to the presence a family of minisatellite loci with sequence similar to that of the OniMS34 core, than some conserved array.

LITERATURE CITED

Amos, W. and Dover, G. A. 1990. DNA fingerprinting and the uniqueness of whales. Mamm. Rev. **20**: 23-30.

Armour, J. A. L., Patel, I., Thein, S. L., Fey, M. F. and Jeffreys, A. J. 1989. Analysis of somatic mutations at human minisatellite loci in tumors and cell lines. Genomics. **4**: 328-334.

Armour, J. A. L., Wong, Z., Wilson, V., Royle, N. J. and Jeffreys, A. J. 1989. Sequences flanking the repeat arrays of human minisatellites: association with tandem and dispersed repeat elements. Nuc. Acids Res. **17**: 4925-4936.

Armour, J. A. L., Povey, S., Jeremiah, S. and Jeffreys, A. J. 1990. Systematic cloning of human minisatellites from ordered array charomid libraries. Genomics. **8**: 501-512.

Armour, J. A. L., Crosier, M. and Jeffreys, A. J. 1992. Human minisatellite alleles detectable only after PCR amplification. Genomics. **12**: 116-124.

Armour, J. A. L., Harris, P. C. and Jeffreys, A. J. 1993. Allelic variation at minisatellite MS205 (D16S309): evidence for polarized variability. Hum. Mol. Genet. **2**: 1137-1145.

Beckmann, J. S. and Soller, M. 1986. Restriction fragment length polymorphisms in plant genetic improvement. Oxf. Surv. Plant Mol. Cell Biol. vol III. pp 225-227.

Bell, G. I., Selby, M. J. and Rutter, W. J. 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature (London). **295**: 31-35.

Bentzen, P., Harris, A. S. and Wright, J. M. 1991. Cloning of hypervariable and simple sequence microsatellite repeats for DNA fingerprinting of important aquacultural species of salmonids and tilapias. In: DNA Fingerprinting: Approaches and Applications. Birkhauser Verlag. Basel. pp. 243-262.

Bentzen, P. and 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.

Blanchetot, A. 198^. Detection of highly polymorphic regions in insect genomes. Nucl. Acids Res. **17**: 3303.

Britten, R. J. and Davidson, E. H. 1971. Repetitive and nonrepetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. **46**: 111-133.

Britten, R. J. and Kohne, D. E. 1968. Repeated sequences in DNA. Science. **161**: 529-540.

Burke, T. and Bruford, M. W. 1987. DNA fingerprinting in birds. Nature (London) **327**: 149-152.

Burke, T., Hanotte, O., Bruford, M. W. and Cairns, E. 1991. Multilocus and single locus minisatellite analysis in population biological studies. In: DNA Fingerprinting: Approaches and Applications. Birkhauser Verlag. Basel. pp. 154-168.

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.

Chandley, A. C. and Mitchell, A. R. 1988. Hypervariable minisatellite regions are sites for crossing-over at meiosis in man. Cytog. Cell Genet. **48**: 152-155.

Collick, A. and Jeffreys, A. J. 1990. Detection of a novel minisatellite-specific DNA-binding protein. Nuc. Acids Res. **18**: 625-629.

Collick, A., Dunn, M. G. and Jeffreys, A. J. 1991. Minisatellite binding protein MSBP-1 is a sequence-specific single-stranded DNA-binding protein. Nucl. Acids Res. **19**: 6399-6404.

Daly, A., Kellam, P., Berry, S. T., Chojecki, A. J. S. and Barnes, S. R. 1991. The isolation and characterization of plant sequences homologous to human hypervariable minisatellites. In: DNA Fingerprinting: Approaches and Applications. Birkhauser Verlag. Basel. pp. 330-341.

Dawley, R.M. and Bogart, J.P. 1989. Evolution and ecology of unisexual vertebrates. New York State Museum, Albany, Bull.

Feinberg, A. P. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**: 6-13.

Feinberg, A. P. 1984. Addendum. Anal. Biochem. **137**: 266-267.

Fowler, S. J., Gill, P., Werrett, D. J. and Higgs, D. R. 1988. Individual specific DNA fingerprints from a hypervariable region probe: alpha-globin 3'HVR. Hum. Genet. **79**: 142-146.

Gatei, M. H., Chen, P. M., Daniel, R. C. W. and Lavin, M. F. 1991. DNA fingerprints of sheep using and M13 probes. Anim. Genet. **22**: 285-290.

Georges, M., Gunawardana, A., Threadgill, D. W., Lathrop, M., Olsaker, I., Mishra, A., Sargeant, L. L., Schoeberlein, A., Steele, M. R., Terry, C., Threadgill, D. S., Zhao, X., Holm, T., Fries, R. and Womack, J. E. 1991. Characterization of a set of Variable Number of Tandem Repeat markers conserved in Bovidae. Genomics. **11**: 24-32.

Gilbert, D. A., Lehman, N., O'Brien, S. J. and Wayne, R. K. 1990. Genetic fingerprinting reflects population differentiation in the California Channel Island fox. Nature (London) **344**: 764-767.

Gilbert, D. A., Packer, C., Pusey, A. E., Stephens, J. C. and O'Brien, S. J. 1991. Analytical DNA fingerprinting in lions: parentage, genetic diversity and kinship. J. Hered. **82**: 378-386.

Gray, I. C. and Jeffreys, A. J. 1991. Evolutionary transience of hypervariable minisatellites in man and the primates. Proc. Roy. Soc. Lond. B. **243**: 241-253.

Gyllensten, U. B., Jakobsson, S., Temrin, H. and Wilson, A. C. 1989. Nucleotide sequence and genomic organization of bird minisatellites. Nuc. Acids Res. **17**: 2203-2214.

Hanotte, O., Burke, T., Armour, J. A. L. and Jeffreys, A. J. 1991. Cloning, characterization and evolution of Indian peafowl Pavo christatus minisatellite loci. In: DNA Fingerprinting: Approaches and Applications. Birkhauser Verlag. Basel. pp. 193-216.

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

Hillel, J., Schaap, T., Haberfield, A., Jeffreys, A. J., Plotzky, Y., Cahaner, A. and Lavi, U. 1990. DNA fingerprints applied to gene introgression in breeding programs. Genetics. **124**: 783-789.

Huey, B. and Hall, J. 1989. Hypervariable DNA fingerprinting in *Echerischia coli*: minisatellite probe from bacteriophage M13. J. Bact. **171**: 2528-2532.

Inglehearn, C. F. and Cooke, H. J. 1990. A VNTR immediately adjacent to the human pseudoautosomal telomere. Nucl. Acids Res. **18**: 471-476.

Jarman, A. P., Nicholls, R. D., Weatherall, D. J., Clegg, J. B. and Higgs, D. R. 1986. Molecular characterization of a hypervariable region downstream of the human a-globin gene cluster. EMBO. J. **5**: 1857-1863.

Jeffreys, A. J., Wilson, V. and Thein, S. L. 1985a. Hypervariable 'minisatellite' regions in human DNA. Nature (London). **314**: 67-73.

Jeffreys, A. J., Wilson, V. and Thein, S. L. 1985b. Individual specific 'fingerprints' of human DNA. Nature. **316**: 76-79.

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. and Ponder, B. A. J. 1986. DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. Am. J. Hum. Genet. **39**: 11-24.

Jeffreys, A. J. and Morton, D. B. 1987. DNA fingerprints of dogs and cats. Anim. Genet. **18**: 1-15.

Jeffreys, A. J., Wilson, V., Kelly, R., Taylor, B. A. and Bulfield, G. 1987. Mouse DNA "fingerprints": analysis of chromosome localization and germ-line stability of hypervariable loci in recombinant inbred strains. Nuc. Acids Res. **15**: 2832-2836.

Jeffreys, A. J., Neumann, R. and 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. and Monckton, D. G. 1991. Minisatellite repeat coding as a digital approach to DNA typing. Nature (London). **354**: 204-209.

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

Julier, C., de Gouyon, B., Georges, M., Guenet, J.-L., Nakamura, Y., Arner, P. and Lathrop, G. M. 1990. Minisatellite linkage maps in the mouse by cross-hybridization with human probes containing tandem repeats. Proc. Natl. Acad. Sci. USA. **87**: 4585-4589.

Kashi, Y., Iraqi, F., Tikochinski, Y., Ruziysky, B., Nave, A., Beckmann, J. S., Friedmann, A., Soller, M. and Gruenbaum, Y. 1990. (TG)n Uncovers a sex-specific hybridization pattern in cattle. Genomics. **7**: 31-36.

Kelly, R., Bulfield, G., Collick, A., Gibbs, M. and Jeffreys, A. J. 1989. Characterization of a highly unstable mouse minisatellite locus: evidence for somatic mutation during early development. Genomics. **5**: 844-856.

Kit, S. 1961. Equilibrium centrifugation in density gradients of DNA preparations from animal tissues. J. Mol. Biol. **3**: 711-716.

Kominami, R , Mitani, K. and Muramatsu, M. 1988. Nucleotide sequence of a mouse minisatellite DNA. Nuc. Acids Res. **16**: 1197.

Kornfield, I. L., Ritte, U., Richler, C. and Wahrman, J. 1979. Biochemical and cytological differentiation among cichlid fishes of the sea of galilee. Evolution. **33**: 1-14.

Koshy, R., Koch, S., Freytag von Loringhoven, A., Kahmann, R., Murray, K. and Hofshneider, P. 1983. Integration of hepatitis B virus DNA: Evidence for integration in the single stranded gap. Cell. **34**: 215-223.

Kuhnlein, U., Zagworny, D., Dawe, Y., Fairfull, R. W. and Gavora, J. S. 1990. Assessment of inbreeding by DNA fingerprinting: development of a calibration curve using defined strains of chicken. Genetics. **125**: 161-166.

Kvarnheden, A. and Engstrom, P. 1992. Genetically stable, individual-specific differences in hypervariable DNA in Norway spruce, detected by hybridization to a phage M13 probe. Canadian J. Forest Res. **22**: 117-123.

Lannan, J. E., Gall, G. A. E., Thorpe, J. E., Nash, C. E. and Ballachey, B. E. 1989. Genetic resource management of fish. Genome. **31**: 798-804.

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

Lloyd, M. A., Fields, M. J. and Thorgaard, G. H. 1989. *Bkm* minisatellite sequences are not sex associated but reveal DNA fingerprint polymorphisms in rainbow trout. Amer. Trans. Fisheries. **32**: 865-868.

Lynch, M. 1990. The similarity index and DNA fingerprinting. Mol. Biol. Evol. **7**: 478-484.

Mair, G. C., Scott, A. G., Beardmore, J. A. and Skibinski, D. O. F. 1987. A technique for the induction of diploid gynogenesis in Oreochromis niloticus by supression of the first mitotic division. (Ed. K. Tiews).World Symposium on Selection, Hybridization and Genetic Engineering in Aquaculture. vol II. pp. 289-301.

Mermer, B., Colb, M. and Krontiris, T. G. 1987. A family of short interspersed repeats is associated with tandemly repetitive DNA in the human genome. Proc. Natl. Acad. Sci. USA. **84**: 3320-3324.

Mitani, K., Takahashi, Y. and Kominami, R. 1990. A GGCAGG motif in minisatellites affecting their germline instability. J. Biol. Chem. **265**: 15203-15210.

Nakamura, Y., Lathrop, M., O'Connell, P., Leppert, M., Barker, D., Wright, E., Skolnick, M., Kondoleon, S., Litt, M., Lalouel, J.-M. and White, R. 1988. A mapped set of DNA markers for human chromosome 17. Genomics. **2**: 302-309.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. 1987. Variable number tandem repeat (VNTR) markers for human gene mapping. Science. **235**: 1616-1622.

Packer, C., Gilbert, D. A., Pusey, A. E. and O'Brien, S. J. 1991. A molecular genetic analysis of kinship and cooperation in African lions. Nature (London). **351**: 562-565.

Penman, D. J., Shah, M. S., Beardmore, J. A. B. and Skibinski, D. O. F. 1987. Sex ratio of gynogenetic and triploid tilapia. (Ed. K. Tiews).World Symposium on Selection, Hybridization and Genetic Engineering in Aquaculture. vol II. pp. 267-276.

Proudfoot, N. J. and Gil, A. 1982. The structure of the human Zeta-globin gene and a closely linked, nearly identical pseudogene. Cell. **31**: 553-563.

Purdom, C. E. 1983. Genetic engineering by manipulation of chromosomes. Aquaculture. **33**: 287-300.

Royle, N. J., Clarkson, R. E., Wong, Z. and Jeffreys, A. J. 1987. Preferential localisation of hypervariable minisatellites near human telomeres. Human Gene Mapping 9: Ninth International Workshop on Human Gene Mapping. Cyto. Cell. Genet. **46**: 685.

Royle, N. J., Clarkson, R. E., Wong, Z. and Jeffreys, A. J. 1988. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics. **3**: 352-360.

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

Schafer, R., Zischler, H. and Epplen, J. T. 1988. $(CAC)_5$, a very informative oligonucleotide probe for DNA fingerprinting. Nuc. Acids. Res. **16**: 5196.

Singer, M. F. 1982. Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. **76**: 67-112.

Smith, G. R. 1987. Mechanism and control of homologous recombination in *Esherichia coli*. Ann. Rev. Genet. **21**: 179-201.

Stanley, J. G. and Sneed, K. E. 1974. Artificial gynogenesis and its applications in genetics and selective breeding of fishes. Springer-Verlag. Berlin.

Starr, C. and Taggart, R. 1987. Biology: the unity and diversity of life. Wadsworth Publishing Company, Belmont, California. p. 565.
Steinmetz, M., Stephan, D. and Lindahl, K. F. 1986. Gene organization and recombinational hotspots in the murine major histocompatability complex. Cell. **44**: 895-904.

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

Thorgaard, G. H. 1986. Ploidy manipulation and performance. Aquaculture. **57**: 57-64.

Uraiwan, S. and Doyle, R. W. 1986. Replicate varience and the choice of selection procedure for tilapia (*Oreochromis niloticus*) Stock improvement in Thailand. Aquaculture. **57**: 27-35.

Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A. S. and Christophe, D. 1987. A sequence of M13 phage detects hypervariable minisatellites in human and animal DNA. Science. **235**: 683-684.

Vergnaud, G., Mariat, D., Zoroastro, M. and Lauthier, V. 1991. Detection of single and multiple polymorphic loci by synthetic tandem repeats of short oligonucleotides. Electrophoresis. **12**: 134-140.

Wahls, W. P., Swenson, G. and Moore, P. D. 1991. Two hypervariable minisatellite DNA binding proteins. Nucl. Acids Res. **19**: 3269-3274.

Wahls, W. P., Wallace, L. J. and Moore, P. D. 1990. Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. Cell. **60**: 95-103.

Waring, M. and Britten, R. J. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. Science. **154**: 791-794.

Washio, K., Misawa, S. and Ueda, S. 1989. Probe walking: development of novel probes for DNA fingerprinting. Hum. Genet. **83**: 223-226.

Wolff, R. K., Plaetke, R., Jeffreys, A. J. and White, R. 1989. Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics. **5**: 382-384.

Wolff, R.K., Nakamura, Y., Odelberg, S., Shiang, R. and White, R. 1991. Generation of variability at VNTR loci in human DNA. In: DNA Fingerprinting: Approaches and Applications. Birkhauser Verlag. Basel. pp. 20-38.

Wong, Z., Wilson, V., Jeffreys, A. J. and Thein, S. L. 1986. Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. Nuc. Acids. Res. **14**: 4605-4616.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. 1987. Characterisation of a panel of highly variable minisatellites cloned from human DNA. Ann. Hum. Genet. **51**: 269-288.

Wright, J.M. 1993. DNA fingerprinting of fises. In: Biochemistry and molecular biology of fishes, vol. II. (T. Mommsen and P.W. Hochachka Eds). Elsevier Press. Amsterdam. pp.57-91.

Wright, J.M. 1994. Mutation at VNTRs: Are minisatellites the evolutionary progeny of microsatellites? Genome **37:** 345-347.

Yanis-Peron, C., Vieira, J. and Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene. **33**: 103-119.