National Library of Canada

Bibliotheque nationale du Canada

Acquisitions and Bibliographic Services Branch

Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa Ontario
K1A 0N4

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.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

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.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Organization and Evolution of Two Satellite DNA Families, SATA and SATB, from the Tilapiine and Haplochromine Genome (Pisces: Cichlidae)

by

Jens P.C. Franck

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

at

Dalhousie University

Halifax, Nova Scotia

January, 1993

© Copyright by Jens P.C. Franck, 1993

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

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-87454-6

Canadä

Name Jens P.C. Franck

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.

Biology (Molecular) — SUBJECT TERM

0307 — SUBJECT CODE

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

TABLE OF CONTENTS

PAGE

TABLE OF CONTENTS iv

LIST OF FIGURES viii

LIST OF TABLES xi

ABSTRACT xii

LIST OF ABBREVIATIONS xiii

ACKNOWLEDGEMENTS xiv

PUBLICATIONS xv

CHAPTER 1: INTRODUCTION 1

Repetitive DNA Sequences 1

Terminology and definitions 1

Function of satellite DNAs 3

Maintenance and propagation of long tandem arrays 5

Hierarchical organization of satellite arrays . 6

Phylogenetic utility of satellite DNA sequences 7

Satellite DNA families in teleost genomes . . . 8

Cichlid Biology and Taxonomy 11

Phylogeny of cichlids determined from morphological characters 12

Brooding behaviour of African tilapiine fishes 15

Molecular studies of cichlid and tilapiine evolution 18

Objective of Study 20

PAGE

CHAPTER 2: MATERIALS AND METHODS 21

Source of Fish Specimens 21

Genomic DNA Extractions 21

Restriction Endonuclease Digestions 22

Polyacrylamide and Agarose Gel Electrophoresis . . . 22

Southern and Slot Blot Techniques 23

Cloning Methodology 23

Radiolabelling Techniques 25

Hybridization Conditions 25

Sequencing Protocols 26

Computer Sequence Analyses 27

RESULTS 29

CHAPTER 3: SATA SATELLITE DNA FAMILY 29

SATA Satellite DNA Family Includes Three
Size-Variants 29

Copy Number and Species Distribution of
the SATA Repeat 40

SATA Sequence Cloned from *H. similis* 43

Polymorphic Hybridization Patterns 50

Interspecific Comparison of SATA Sequences 53

Concerted Evolution and Homogenization
of the SATA Repeats 56

Molecular Phylogenetic Analyses 62

CHAPTER 4: SATB SATELLITE DNA FAMILY 73

Sequence Analysis of Oho750#1 Clone 73

PAGE

Genomic Organization of Oho750#1 Repeat 76

Cloning of 1900 bp Repetitive Sequence from *Oreochromis niloticus* 81

Copy Number Estimation and Species Distribution 94

CHAPTER 5: DISCUSSION 99

The SATA Satellite DNA Family 99

Species specificity of the SATA repeat . . . 103

Higher order organization 104

Concerted evolution and phylogenetic implications of the SATA repeat 105

SATA sequences: molecular cause or effect of speciation? 109

The SATB Satellite DNA Family 112

Evidence for monophyly of neotropical/African cichlid assemblage 112

SATB direct repeat sequences may be conserved by selective pressures 113

Summary and Future Prospects 115

APPENDIX 1: Consensus Alignments of SATA satellite DNA sequences for *O. placidus* I, *O. placidus* II, *O.niloticus*, *S. galilaeus*, *T. zillii*, *T. rendalli*, and *H. similis*. 117

APPENDIX 2: Multiple alignment of six tilapiine and *H. similis* SATA consensus sequences 125

PAGE

LITERATURE CITED 127

LIST OF FIGURES

PAGE

Figure 1.1 Summary cladogram of cichlid taxonomy based on morphological data 13

Figure 1.2 Cladograms representing the two models for tilapiine evolution 16

Figure 3.1 6.0% polyacrylamide gel of *Eco*RI-digested tilapiine DNA 30

Figure 3.2 Southern blot and hybridization of a cloned satellite repeat from *O. niloticus* to *Eco*RI digests of various tilapiine genomic DNAs 33

Figure 3.3 Dot blot hybridization of nick-translated *O. niloticus* genomic DNA to recombinant M13 clones 36

Figure 3.4 Multiple alignment of six tilapiine SATA consensus sequences 38

Figure 3.5 Slot blot hybridization of radiolabelled OplI-5 (SATA) insert to graded amounts of denatured cichlid samples 41

Figure 3.6 Southern hybridization of radiolabelled OniIII-14 (SATA) clone to *Eco*RI-digested cichlid genomic DNA samples 46

Figure 3.7 Southern hybridization of radiolabelled OniIII-14 (SATA) clone to genomic DNA of *O.mossambicus/hornorum* and *H. similis* 48

PAGE

Figure 3.8 Slot blot hybridization of radiolabelled OplI-5 (SATA) insert to recombinant M13 clones for *H.similis* SATA cloning experiment 51

Figure 3.9 Polymorphic DNA fingerprint-like patterns detected by a satellite probe (SATA) from tilapia 54

Figure 3.10 Schematic representation of transition stages during the spread of mutations 59

Figure 3.11 Bootstrapped DNA parsimony phylogenetic tree generated by the DNABOOT program of PHYLIP 65

Figure 3.12 Cladograms of alternative topologies for the parsimony tree as tested by the MacClade phylogenetic program 67

Figure 3.13 Neighbour-Joining phylogenetic tree generated for the tilapiine SATA sequences 71

Figure 4.1 Southern hybridization of radiolabelled Oho750#1 clone to *Hin*dIII-digested tilapiine genomic DNA 74

Figure 4.2 Restriction map and sequencing strategy for Oho750#1 clone 77

Figure 4.3 Nucleotide sequence of the Oho750#1 clone 79

Figure 4.4 Southern hybridization of radiolabelled Oho750#1 clone to *Hin*dIII and *Eco*RI partial digests of *O. niloticus* genomic DNA 82

PAGE

Figure 4.5 Restriction map and sequencing strategy for the Oni1900#8 clone 85

Figure 4.6 Nucleotide sequence of clone Oni1900#8 . 87

Figure 4.7 Genomic organization of the 1900 bp tandem repeat 90

Figure 4.8 Alignment of internal direct repeat sequences for SATB repeat 92

Figure 4.9 Slot blot hybridization of radiolabelled Cho750#1 insert to graded amounts of denatured genomic DNA samples 95

Figure 5.1 Schematic models of proposed mechanisms for the generation of high molecular weight polymorphic SATA fragments 106

LIST OF TABLES

PAGE

Table 1.1 Summary of satellite DNAs cloned and characterized from genomes of teleost fishes 10

Table 3.1 Distribution of SATA size-variants in selected tilapiine fish genomes 32

Table 3.2 SATA copy number estimation for selected cichlid species 44

Table 3.3 Cross-homology comparison of tilapiine and *Haplochromis similis* SATA consensus sequences . . . 57

Table 3.4 Partitioning of the distribution of DNA mutations (%) at individual nucleotide positions for all clones from *Oreochromis niloticus* and six other cichlid species' satellite DNAs 61

Table 3.5 Estimation of time elapsed since last major amplification event for SATA sequences from selected cichlid species 63

Table 4.1 SATB copy number estimation for selected cichlid species 97

ABSTRACT

I have cloned and sequenced members of two satellite DNA families, SATA and SATB, from the genomes of the tilapiine and haplochromine fishes. The SATA satellite DNAs were visualized as intensely staining bands after electrophoretic separation of *Eco*RI-digested (tilapiine species) and *Hin*fI-digested (*H. similis*) genomic DNA. Monomer repeats were cloned from tilapiine fishes of the three major genera: *Oreochromis*, *Sarotherodon*, and *Tilapia* as well as the haplochromine species *Haplochromis similis*. Three size-variants of approximately 237 bp (type I), 230 bp (type II), and 209 bp (type III) were identified. Pairwise comparison of derived consensus sequences for the SATA monomers revealed identities ranging from 74 to 97%. The type II and type III size-variants appear to have arisen by deletions of 9 and 29 bp, respectively, within different regions of the type I satellite. Molecular phylogenetic analyses of the tilapiine SATA sequences support the close relationship of the two mouthbrooding genera, *Oreochromis* and *Sarotherodon*, with a clear dichotomy between this clade and the substrate spawning *Tilapia* genus.

The SATB satellite DNA is a 1900 bp *Eco*RI fragment cloned from the genome of the tilapiine fish *Oreochromis niloticus*. A 760 bp *Hin*dIII fragment, an internal component of SATB, was cloned and sequenced from the related tilapiine species, *Oreochromis hornorum*. The SATB repeat is more widely conserved than the SATA repeat and was detected in the genomic DNA of African and neotropical cichlid species. Four imperfect 21 bp direct repeat sequences were present within the cloned 1900 bp *Eco*RI repeat. Alignment of the four direct repeats revealed a core motif of 11 bp that exhibits 100% sequence identity between all of the direct repeats. The conservation of this motif in the SATB repeat suggests that it is under selective constraint.

LIST OF ABBREVIATIONS

bp	nucleotide base pairs
Ci	Curie
dATP	deoxyadenosine-5'-triphosphate
dGTP	deoxyguanosine-5'-triphosphate
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
IPTG	isopropylthio-β-D-galactoside
Kbp	kilo base pairs
RNA	ribonucleic acid
tRNA	transfer RNA
V	Volts
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

ACKNOWLEDGEMENTS

I thank my supervisor, Jonathan Wright, for his support throughout my tenure at Dalhousie, and for his intellectual inspiration and guidance. I would also like to thank the members of my committee; Bill Pohajdak, Tom MacRae, and Paul Bentzen for their invaluable advice and comments on my research as well as the preparation of this thesis. I am indebted to my fellow graduate students for their friendship and support especially Andrew Harris and Eileen Denovan-Wright with whom I have shared many enjoyable times. I am grateful to Marty Ball for both his technical expertise with computer related problems and his helpful informal tutorials on molecular systematics. I thank Dawn Thornham for her photographic skills, and all members of the office, stores and technical staff in the Biology Department.

I am profoundly grateful for the constant moral support of both my parents, and for my father's ability to empathize with the frustrations and problems encountered by graduate students. I would lastly like to acknowledge my sincere gratitude for Nicole's support and friendship and to her family for making me feel at home in Nova Scotia.

The research for this thesis was supported by a grant from the Natural Sciences Engineering and Research Council of Canada to Jonathan Wright.

PUBLICATIONS

Much of the research discussed in this thesis has appeared in the following publications.

Franck, J.P.C., and Wright, J.M. 1993. Conservation of a satellite DNA sequence (SATB) in the tilapiine and haplochromine genome (Pisces: Cichlidae). Genome **36**: 187-194.

Franck, J.P.C., Wright, J.M., and McAndrew, B.J. 1992. Genetic variability in a family of satellite DNAs from tilapia (Pisces: Cichlidae). Genome **35**: 719-725.

Franck, J.P.C., Harris, A.S., Bentzen, P.B., 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 **7**. *Edited by* Norman Maclean. Oxford University Press, Oxford, pp. 51-82.

CHAPTER I

INTRODUCTION

Repetitive DNA Sequences

Repetitive DNA sequences, a common constituent of all eukaryotic genomes, were first identified as peaks flanking the main genomic fraction in buoyant density gradient analyses due to their biased base composition (Kit 1961), hence the term satellite DNA. DNA reassociation experiments showed that a proportion of the mouse genome (10%) reannealed extremely rapidly (Waring and Britten 1966) Later, reassociation experiments revealed that the genomic DNA of higher eukaryotes can be subdivided into three major classes: highly repetitive, moderately repetitive, and unique DNA sequences (Britten and Kohne 1968).

Terminology and definitions

Repetitive DNAs are broadly classified as interspersed or clustered tandemly arrayed sequences. Interspersed DNA sequences are further distinguished as Short Interspersed Nucleotide Sequences (SINES) or Long Interspersed Nucleotide Sequences (LINES) (Singer 1982). Interspersed sequences may be of viral or non-viral origin (reviewed in Weiner *et al* 1986). Repeat sequences of viral origin possess coding regions for their own reverse transcriptase and integrase as well as long terminal repeats. These sequences are believed to have dispersed throughout the genome via transposition of

RNA polymerase transcript intermediates. Non-viral interspersed sequences lack flanking direct repeat sequences and are thought to represent retropseudogenes derived from RNA polymerase II and III transcripts. One of the most extensively characterized non-viral SINEs is the 300 bp Alu sequence of primates (reviewed in Schmid and Shen 1985). The sequence homology of the Alu element to 7S RNA transcripts suggests that the Alu sequences are the result of the retrotransposition of the 7S transcripts into the genome. Clustered repetitive DNAs include tandemly repeated coding sequences such as the histone and ribosomal genes and non-coding regions of DNA comprised of tandemly repeated satellite sequences (reviewed in Brutlag 1980; Singer 1982).

Satellite DNAs are subdivided into three classes based on the size of the monomer unit: satellites, minisatellites (Jeffreys 1985a), and microsatellites (Dover 1989; Tautz 1990). Satellite DNA sequences range from 100 to 2000 bp in length, while minisatellite DNAs normally range from several to 100 bp. Microsatellite DNA sequences, sometimes termed simple sequence repeats, include repeat units from one to ten base pairs (Tautz 1989). Minisatellite and microsatellite monomers occur in clusters of tandem arrays often dispersed throughout the genome. The hypervariability of these arrays or "loci" has provided an abundant source of genetic markers for population and pedigree studies (Jeffreys 1985b; Tautz 1989). The arrays are believed to

increase and decrease in size due to intrastrand replication slippage (Levinson and Gutman 1987). The modes of satellite DNA propagation and maintenance are discussed in greater detail in a later section. Satellite DNA sequences are normally localized to the centromeric and telomeric heterochromatic regions of the chromosomes.

Function of satellite DNA

The ubiquitous presence and maintenance of satellite DNAs within all eukaryotic genomes has engendered two opposing schools of thought, which may be generalized as selectionist and non-selectionist. The non-selectionist school includes the theories of junk (Ohno 1972) or selfish DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). The selfish DNA model interprets the existence of repetitive DNAs as a result of sequence-specific strategies for their propagation and maintenance, independent of organismal phenotype. Although this model circumvents the traditional Darwinian theory of natural selection, the proponents of selfish DNA suggest that the repetitive DNA component contributes towards the plasticity of the eukaryotic genome. The "excess" DNA may serve as the raw material necessary for the production of new genes or allow the reshuffling of coding domains resulting in the ability to accommodate a greater range of morphological and behaviourial adaptations (Doolittle 1982). The selectionist school of thought argues

that satellite DNAs are maintained due to their direct contribution to the functioning of the genome. Because of their localization to the centromeres of chromosomes, satellite DNAs have been implicated in both structural and functional roles for mitotic and meiotic processes (Pardue and Hennig 1990). This proposal is substantiated by several studies that provide evidence for centromeric satellite DNAs serving as specific recognition sites for DNA binding proteins. Masumoto and colleagues (1989) describe a conserved 17 bp core region, termed the CENP-B box within the monomer unit of the human alphoid satellite DNA. This sequence apparently serves as a specific binding domain for the CENP-B binding protein. Additionally, Strauss and Varshavsky (1984) provide evidence that a high-mobility-group-like protein specifically binds to several AT rich sites in human alpha satellite DNA. Recently, more evidence has been presented that supports the role for satellite DNA domains in the three dimensional structure and architecture of chromatin. Subsequent work indicated that the specific regions of mouse satellite DNA that bind to the HMG-I protein possess a stable DNA curvature (Radic *et al.* 1992). It is implied that the combination of the specific DNA curvature and the binding of the chromatin-specific proteins are essential for the condensation of the heterochromatin. In addition to their role in binding specific proteins, human alphoid satellite DNA sequences are implicated as the

DNA component in the chemically resistant non-histone DNA-polypeptide complexes (Pfutz *et al.* 1992).

Maintenance and propagation of long tandem arrays

The apparent rapid amplification and persistence of satellite DNA arrays within eukaryotic genomes is attributed to various genomic processes, including unequal crossing over (Tartof 1975; Smith 1976), gene conversion (Baltimore 1981), replication slippage (Levinson and Gutman 1987), and extrachromosomal replication (Rossi *et al.* 1990). These processes of genomic turnover form the basis for the theory of molecular drive (Dover 1982; 1986), the process whereby mutations are homogenized throughout a satellite array within and between chromosomes. The recombination of misaligned arrays of satellite DNA monomers, termed unequal crossing over, is a major tenet of the molecular drive theory, and was considered sufficient to account for the rapid amplification and homogenization of satellite sequences. Walsh (1987) challenges this model, claiming that unequal crossing over alone cannot cause the growth and homogenization of tandem arrays, and that the net effect would be the reduction and ultimate loss of entire tandem arrays. Since satellite DNA sequences persist in the heterochromatic centromere, and telomeres, regions of low recombinational activity, the role for unequal crossing over in the homogenization of the tandem arrays is further

challenged (Charlesworth *et al.* 1986). To counter the predicted net loss of satellite sequences by deletion biased intrastrand forces, rolling circle replication is proposed as a mechanism for the maintenance of the arrays (Walsh 1987). The extrachromosomal plasmids resulting from intrastrand recombination processes undergo rolling circle replication followed by reintegration of the amplified array by homologous recombination into the chromosome (Rossi *et al.* 1990).

Hierarchical organization of satellite arrays

Satellite DNAs are organized into higher order arrays that are often chromosome-specific (reviewed in Willard 1989). Most studies have focused on the alpha satellite DNA sequences of primate genomes. One of the first observations of a higher order organization for alpha satellite sequences was made by Wu and Manuelidis (1980) who identified a 680 bp tetrameric unit operationally defined by the restriction endonuclease *Eco*RI. The alphoid satellite DNA arrays on human chromosomes 1, 11, 17, and X are believed to be derived from a common pentameric unit (Willard and Waye 1987). Vissel and Choo (1991) described an alpha satellite DNA subfamily for chromosome 14, as well as a common higher order array of alphoid sequences on chromosomes 13, 14, and 21. Superimposed on the higher order arrays of dimeric, tetrameric, and pentameric units is a long range periodicity

of the satellite arrays revealed by pulsed field gel electrophoresis (PFGE) techniques. The long range arrays are polymorphic and are meiotically stable components of an individual's genome such that they may be used as Mendelian markers for pedigree analyses (Wevrick and Willard 1989).

Phylogenetic utility of satellite DNA sequences

Satellite DNAs have been used in systematic studies. The resolving power of these sequences appears to be inherent to the specific sequence; ranging from the identification of conspecific populations, to the analysis of interfamilial relationships. The major criticism of using satellite DNAs for phylogenetic analyses is the extremely high copy number and dynamic nature of the sequences themselves. Concerted evolutionary processes (Dover 1982), however, whereby satellite DNAs maintain a closer sequence identity to each other within species than between species, suggests intriguing possibilities for their utility in phylogenetic analyses. Homogenization of the *Hin*dIII defined satellite DNA of the sheepshead minnow has resulted in perfect sequence identity for every one of the individual monomer sequences cloned from the fish within specific populations (Turner *et al.* 1991). This satellite DNA is therefore fixed at the population level and has been used to distinguish between conspecific populations of sheepshead minnow from the Gulf of Mexico. In contrast, the common

cetacean satellite DNA component appears to have evolved at an extremely slow rate and has been employed for the phylogenetic analysis of interfamilial relationships (Arnason *et al.* 1992; Gretarsdottir and Arnason 1992). A 1750 bp satellite DNA sequence is found in all cetaceans including the odontocetes (toothed whales) and mysticetes (whalebone whales). A phylogenetic comparison was performed to establish the systematic position of the Irrawaddy dolphin (*Orcaella brevirostris*) within the superfamily Delphinoidea using the Baird's beaked whale (*Berardius bairdii*) from the superfamily, Ziphioidea, as an outgroup. This analysis placed the *O. brevirostris* within the Delphinidae family but could not definitively determine its relative position with respect to other species within this family.

Satellite DNA families in teleost genomes

While satellite DNA sequences have been extensively studied in many organisms including invertebrates (Miklos 1982; 1985), amphibians (Hummel *et al.* 1984), mammals (Singer 1982; Arnason *et al.* 1984; 1988), and plants (Flavell *et al.* 1983), until recently, there have been few reports of satellite DNAs from teleost fishes. The first discovery of satellite DNAs in teleosts came from buoyant density gradient analyses for several species of fishes (Helleiner 1974). The satellite DNAs that have been cloned

and characterized are summarized in Table 1.1.

Satellite DNAs in tilapiine fishes were first observed as intensely staining bands of approximately 200 bp in the genomic smears of *Eco*RI and *Hae*III digested DNA of the hybrid strain *Oreochromis mossambicus* x *Oreochromis hornorum* size-fractionated by agarose gel electrophoresis (Wright 1989). The monomer repeat is approximately 237 bp and constitutes 7% of the haploid genome (3 x 10^5 copies). An imperfect 9 bp AT rich motif, regularly spaced every 30 nucleotides within the repeat, may represent the underlying oligonucleotide progenitor sequence for the SATA repeat.

Hybridization of a radiolabelled satellite clone to *Hae*III partial digests of *O. mossambicus/hornorum* genomic DNA reveals a ladder of bands with a multiple of approximately 240 bp. This result is consistent with the tandemly arrayed organization of the repeat. A faintly hybridizing series of bands was also detected in the partial digestion, consisting of multimers of approximately 214 bp. This provided the first evidence that there exist at least two variants of the tilapiine satellite DNA sequence. The sequence of one of the cloned satellite DNAs, Ti-14, revealed a 30 bp deletion. The presence of the two distinct ladders of bands based on multiples of 240 bp and 214 bp, respectively, provides indirect evidence that the two satellite DNA variants are components of individual arrays and that the deletion variant must have arisen in a separate

Table 1.1: Summary of satellite DNAs cloned and characterized from genomes of teleost fishes.

Species	Monomer Length (bp)	Copy Number/ Haploid Genome	Genome Equivalent (%)	Reference
Cyprinus carpio (Carp)	245	6×10^5	8	Datta *et al.* 1988
Notropsis lutrensis	174	2.1×10^4	8	Moyer *et al.* 1988
Oreochromis mossambicus/ hornorum (Tilapia)	237	3×10^5	7	Wright 1989
Pollachius virens (Pollock)	200	5×10^5	13	Denovan and Wright 1990
Salmo salar (Atlantic Salmon)	~450	-	-	Goodier and Davidson 1991
Cyprinodon variegatus (Sheepshead minnow)	~170	-	-	Turner *et al.* 1991
Brachdanio rerio Type I (Zebra fish)	186	8×10^5	8	Ekker *et al.* 1992
Type II	165	2.2×10^4	0.2	

amplification event. Cleavage of genomic DNA with restriction enzymes which lack recognition sites in the consensus sequence for the *O. mossambicus/hornorum* repeat exposed a hierarchical organization for the satellite DNA. Digestion of the genomic DNA with *Hin*dIII revealed fragments that hybridized to the SATA clone with integral multimers of 2,3,4,5,7, and 9 monomeric units, whereas digestion with the restriction endonuclease *Pst*I produced bands corresponding to multimers of 3,5,6,8, and 9 monomeric units. This result provides indirect evidence for the existence of a higher ordered organization for the SATA sequence.

Cichlid Biology and Taxonomy

Cichlids are the most diverse percoid fishes of the freshwater tropical regions of the world. Members of the cichlid family are broadly classified as members of the Old World (Africa) or new world (Neotropical; Central and South America) assemblages (Fryer and Iles 1972). The cichlid fishes of the great lakes of east Africa have undergone an explosive radiation with each species becoming specialized in order to adapt to specific trophic niches (Fryer and Iles 1972). In Lake Malawi alone there are believed to be 500 to 1000 endemic species (Stiassny 1991). The startling explosive radiation of the cichlids has resulted in a wide diversity in size, shape, and colour. The underlying anatomy, however, has remained relatively static. The

individual species' ability to adapt to specialized niches is not based on dramatic reorganization or innovation, but rather on subtle modifications to their anatomy (Stiassny 1991).

Phylogeny of cichlids determined from morphological characters

The majority of research on cichlid evolution comes from comparative morphological studies (Fig. 1.1). All of the African and Neotropical cichlids with the exception of the Zairean genus, *Heterochromis*, are included in a single monophyletic lineage on the basis of two derived characters. These include a distinctive feature of the ventral branchial apparatus and a modification to the frontal bones of the neurocranium specific to members of the Afro-Neotropical lineage (Stiassny 1991). The monophyly of the African cichlids excluding the Zairean genus, *Heterochromis*, is determined by four derived morphological features primarily in relation to their cranial anatomy.

The first attempt at classification of the African cichlid assemblage was by Regan (1920). He used distinctive features of the neurocranial anatomy to define two major lineages namely the *Tilapia* and *Haplochromis* genera. With specific regard to the tilapiine cichlids, Stiassny (1991) identified two morphological features that may be specific for this lineage. One feature describes the presence of two

Figure 1.1 Summary cladogram of cichlid taxonomy based on morphological data. Generated from a data matrix of 28 binary coded morphological characters. Adapted from Stiassny (1991).

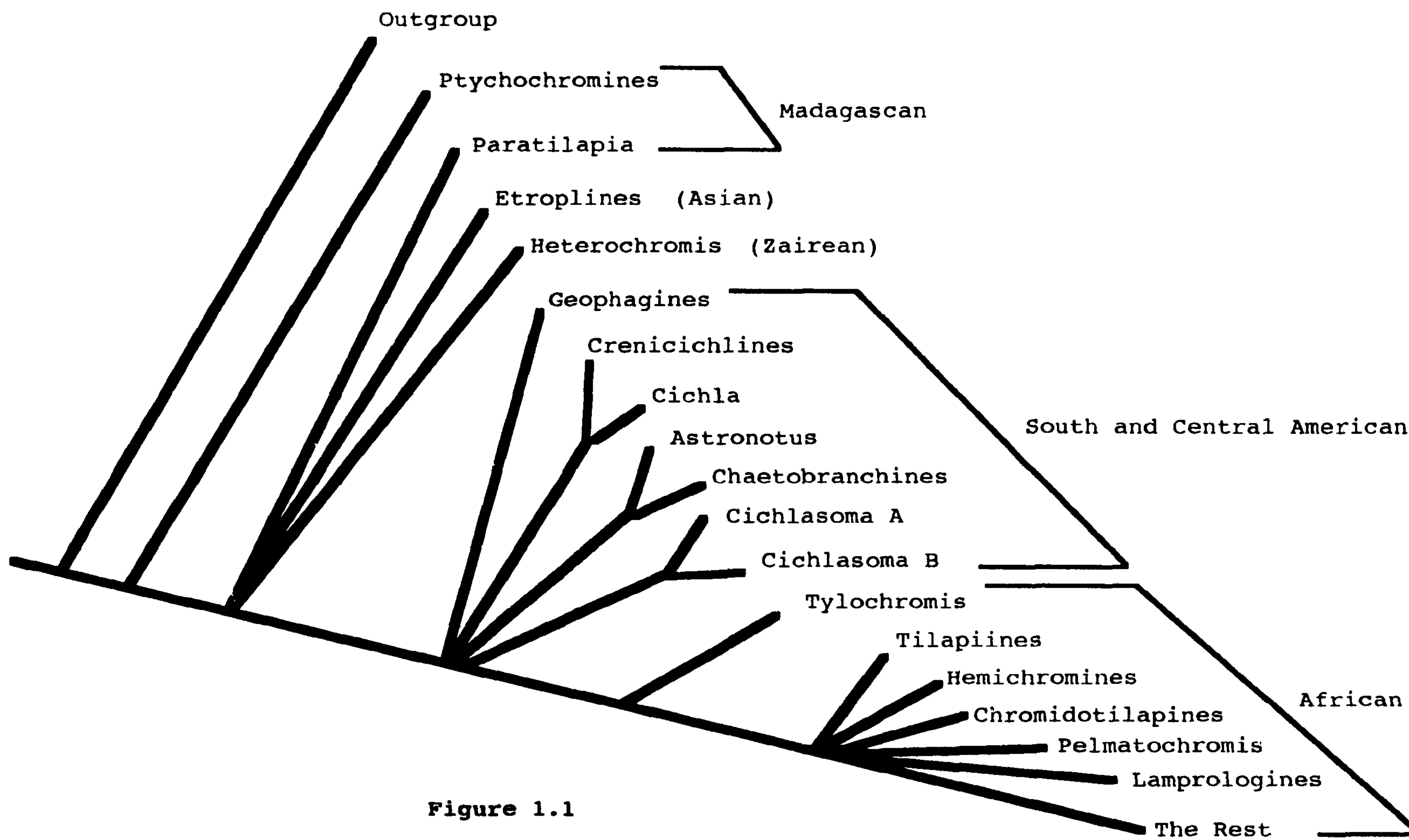


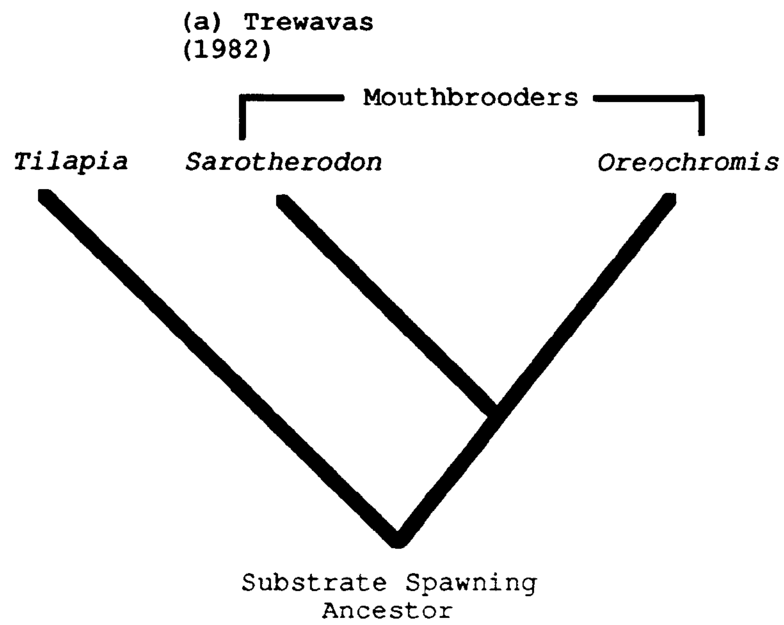
Figure 1.1

foramina on either side of the median suture of the lower pharyngeal jaw (LPJ). The foramina are found in members of the *Oreochromis*, *Sarotherodon*, and *Tilapia* genera, although one of the *Tilapia* species examined, *T. busumana* is polymorphic for the feature, supporting the notion that it is an ancestral tilapiine (Greenwood 1987). The second morphological feature also involves the LPJ in which tilapiines, to the exclusion of all other cichlids, possess a median ridge of bone on the dorsal aspect of the keel (Stiassny 1991). The ridge is most prominent in the *Oreochromis* and *Sarotherodon* genera, and less so in members of the *Tilapia* genus.

Brooding behaviour of African tilapiine fishes

In addition to morphological characteristics, distinctive brooding and breeding behaviours distinguish tilapiine fishes. Trewavas (1982; 1983) identified three major tilapiine genera as exclusively maternal mouthbrooders (*Oreochromis*), biparental and paternal mouthbrooders (*Sarotherodon*), and substrate spawners (*Tilapia*). Two different models have been advanced to explain the evolution of brooding behaviours (Fig 1.2). Peter and Berns (1982) propose that the mouthbrooding strains periodically diverged from the substrate spawning lineage, with the more ancient lineage representing the maternal mouthbrooders and the more recently diverged species representing the younger

Figure 1.2 Cladograms representing the two models for tilapiine evolution. Trewavas (1982) proposes a monophyletic origin for the maternal mouthbrooders (a) while Peters and Berns propose a polyphyletic origin for the mouthbrooders with *Sarotherodon* representing the most recently diverged genus (b).



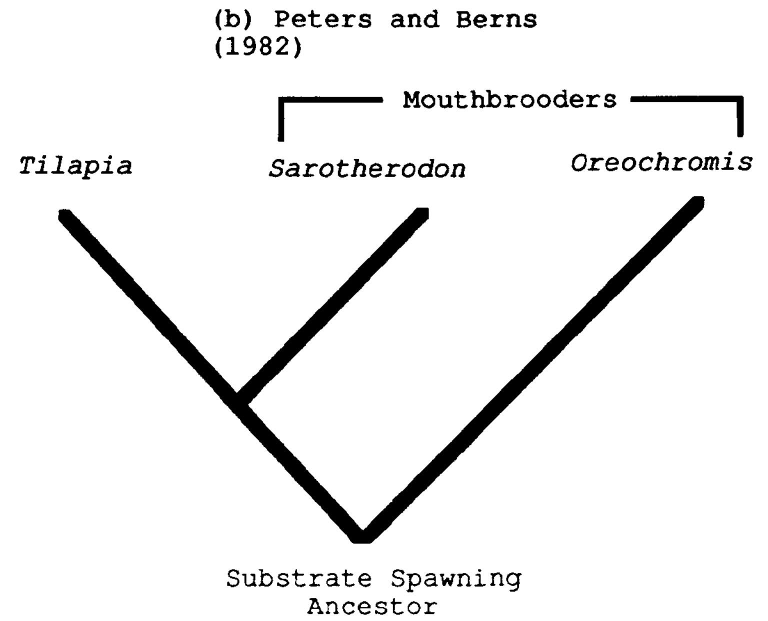


Figure 1.2

biparental mouthbrooders. Alternatively, Trewavas (1982; 1983) proposes that the mouthbrooding lineage originated from a *Tilapia*-like substrate spawner, which subsequently split into the maternal and biparental mouthbrooding lines.

Molecular studies of cichlid and tilapiine evolution

The advent of allozyme and mitochondrial DNA analyses has provided information about the evolution of tilapiine fishes. Kornfield *et al.* (1979) analyzed twenty-one allozyme loci for six species of African cichlids. At the time of this study, the biparental and maternal mouthbrooders were grouped together in the *Sarotherodon* genus. The resulting phylogenetic tree from this analysis grouped the two mouthbrooding species, *Sarotherodon (Oreochromis) aureus* and *Sarotherodon galilaeus*, in the same clade with a clear dichotomy between this clade and the *Tilapia zillii* species. A more comprehensive allozyme analysis using twenty-five enzyme loci for nine tilapiine species was performed by McAndrew and Majumdar (1984). Three different methodologies for producing phylogenetic trees were used in order to resolve the relationship of *Sarotherodon galilaeus* to the other mouthbrooding species. The authors determined that the *S. galilaeus* species was more closely related to the *Oreochromis* species than to *Tilapia zillii*, which refuted the model proposed by Peters and Berns (1982).

Mitochondrial DNA (mtDNA) phylogenetic analyses of the

African cichlids have concentrated primarily on the haplochromine and tilapiine tribes. Seyoum's mtDNA analysis (1989) resolved a trichotomy between the three major tilapiine genera; *Oreochromis*, *Sarotherodon*, and *Tilapia*, concordant with the model of Trewavas (1983). In addition, the restriction endonuclease analysis of mtDNAs has been used to clarify the taxonomy of the *Oreochromis* subspecies complex (Kornfield 1991; Seyoum and Kornfield, 1992). Two recent mtDNA studies have used sequence information to elucidate the relationships of old world cichlids. Meyer *et al.* (1990) sequenced both the control region and the cytochrome b gene of several species belonging to the Lake Victoria haplochromine lineage, which confirmed the monophyletic origin of this assemblage. A similar study was conducted on the *Tropheus* lineage of Lake Tanganyika (Sturmbauer and Meyer 1992). The Lake Tanganyikan cichlid species are morphologically similar, and are distinguished on the basis of colouration differences. The *Tropheus* lineage includes six species distinguished in this manner. This mtDNA analyses revealed a surprisingly high level of genetic divergence accompanied by the relatively low morphological diversity. Geological evidence for fluctuating lake levels implicate this as an instrumental factor in the speciation for this flock of fishes.

Objective of Study

The objectives of this study were two fold. First, the molecular characterization of the satellite DNAs, their organization and evolution within the tilapiine genome were investigated. Second, the implications of the satellite DNA families were examined with respect to the evolution and systematic relationships of the tilapiine tribe within the African cichlid assemblage. This study represents the first attempt at using nuclear DNA markers to examine the phylogenetic relationships of the tilapiine fishes.

CHAPTER 2

MATERIALS AND METHODS

Source of Fish Specimens

All tilapiine specimens with the exception of the *O. niloticus* individuals which are kept in aquaria at Dalhousie University, were obtained from Dr. Brendan McAndrew from the Institute of Aquaculture at the University of Stirling, Scotland. All other cichlid specimens were purchased from a local aquarium store.

Genomic DNA Extractions

Genomic DNA samples were extracted from either tissue of the caudal peduncle region or from blood samples collected by caudal puncture of anaesthetized fish. For extraction from blood, needles and syringes were flushed with 0.5 M EDTA to prevent clotting and 500 μL to 1 ml of blood was collected and brought to a final volume of 5 ml with high TE 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 gauge needle. The solution was then extracted with phenol and chloroform and DNA was precipitated with 2.5 volumes of ethanol. The DNA was dried briefly and resuspended in TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA). Extraction of genomic DNA from tissue was done according to a protocol

developed in the Marine Gene Probe Laboratory at Dalhousie University. Approximately 1 mg of tissue from the caudal peduncle region was homogenized in 500 μL of extraction buffer (0.1 M Tris-OH, pH 8.0, 0.1 M EDTA, 0.25 M NaCl, 1.0% SDS). The homogenate was extracted once with one volume of TE-saturated phenol, followed by a one extraction with a single volume of phenol/chloroform. The DNA was precipitated with 1/10 volume of 5 M ammonium acetate and one volume of isopropanol. The precipitated and vacuum dried DNA was resuspended in TE buffer.

Restriction Endonuclease Digestion

Genomic DNA was digested by high concentrations of restriction endonucleases. Following digestion, the DNA was precipitated by ethanol, dried, redissolved in buffer and further treated with the indicated restriction endonuclease to ensure complete digestion.

Polyacrylamide and Agarose Gel Electrophoresis

Digested genomic DNA samples were fractionated by gel electrophoresis on either 6% polyacrylamide gels at 3.0 V/cm for 18 hrs or on 1.0% agarose gels at 2.5 V/cm for 18-36 hrs. Gels were then stained with 0.5 μg/ml ethidium bromide solution.

Southern and Slot Blot Techniques

DNA was transferred from agarose gels to Hybond N nylon membranes (Amersham) using a Vacu-Blot apparatus (Pharmacia). The DNA was depurinated with 0.25 M HCl for 20 min, denatured with 1.5 M NaCl, 0.5 M NaOH for 20 min, and neutralized for 20 min with 1.5 M NaCl, 1.0 M Tris-HCl, pH 8.0, followed by transfer for one hour with 20 X SSC (1 X SSC = 0.15 M NaCl, 0.015 M sodium citrate). Slot blots were prepared using a Minifold II apparatus (Schleicher and Schuell). The nylon membrane and two sheets of 3 MM chromatography paper were wetted in 10 X SSC before use. The wells were washed once with 100 μl of 20 X SSC before applying the heat-denatured DNA samples. Genomic DNA samples were denatured by heating at 95°C for 5 min in 16 X SSC. Following application of the DNA samples, the wells were washed once with 100 μl of 20 X SSC. The membrane was then washed in 6 X SSC solution.

Cloning Methodology

For the cloning of satellite DNA sequences, genomic DNA samples were digested with an appropriate restriction endonuclease that revealed an intensely staining band on an ethidium stained agarose or polyacrylamide gel. The region of the gel containing the band was excised and the DNA was eluted from polyacrylamide gels by the crush and soak method

(Maniatis *et al*. 1982) or from agarose gels using low melt agarose or by using a silica glass matrix (Geneclean II). For the cloning of the SATA sequences genomic DNA was digested with either *Hae*III (*O. niloticus*), *Eco*RI, or *Hin*fI (*H. similis*) restriction endonucleases. DNA digested with *Hae*III was blunt end ligated into the *Sma*I site of M13mp8. *Eco*RI digested DNA was ligated into the *Eco*RI sites of either M13mp18 or M13mp19. *Hin*fI digested DNA was incubated with Mung Bean nuclease for 20 min at room temperature to produce blunt ended fragments before being ligated into the *Sma*I site of M13mp18. Competent *E. coli* (TG1) were transformed with the ligation products and plated on 2X YT medium, with X-gal and IPTG. Recombinant plaques (clear) were picked and prepared for DNA sequencing by standard protocols (Maniatis *et al*., 1982). M13 phage lysates were dot blotted onto nylon membranes using a vacuum blotting manifold and hybridized to nick-translated DNA from the same species as the cloned DNA. Clones that produced strong hybridization signals were selected for further study based on the assumption that they contained sequences highly reiterated in the genome. Large scale isolations of the replicative form (RF) of the M13 phage clones were prepared from 250 mL of 12 to 18 hour old cultures of TG1 bacterial cells infected with the appropriate clone. The RF was purified on a cesium chloride buoyant density gradient

according to Maniatis *et al.* (1982).

Radiolabelling Techniques

Genomic DNA was labelled by nick translation using ^{32}P-dATP (3000 Ci/mmol) (Rigby *et al.* 1977). M13 clones and the PCR amplified OniIII-14 insert DNA were labelled by primer extension using the universal sequencing primer and ^{32}P-dATP and/or ^{32}P-dGTP (3000 Ci/mmol) (Messing 1983). Insert DNAs used for hybridization to the DNA slot blots were either labelled by random priming with α^{32}P-dATP (Feinberg and Vogelstein 1987) or by Klenow filling of the recessed 5′ ends. DNA was routinely labelled to a specific activity of 10^8 cpm/μg.

Hybridization Conditions

Nylon membranes were incubated for at least two hours in a pre-hybridization mixture (50% formamide; 5 X SSPE; 1 X Denhardt's solution; 100 μg/ml yeast tRNA; 0.1% SDS) (1 X SSPE = 0.015 M NaCl, 10 mM sodium dihydrogenorthophosphate, 1 mM EDTA) at 42°C. Radiolabelled probe was added to a final concentration of 10^6 cpm/ml and hybridization was allowed to proceed for 24 to 48 hours. Membranes were washed at varying stringencies. Genomic DNA blots were typically washed at high stringency conditions (0.1 X SSC, 0.1% SDS) at 65°C, while dot blots and slot blots were washed under low

stringency conditions (0.2 X SSC, 0.1% SDS) at room temperature. Membranes were exposed to Kodak XAR5 X-ray film with or without intensifying screens from two hours to six days. To estimate the copy number of the repetitive DNAs the slot blot autoradiographs were digitally converted to TIFF file formats for the Apple Macintosh computer. Quantification of signal from the digitized image was conducted with the program: ScanAnalysis Densitometry for the Macintosh (Release 2.20; Burcham, 1987).

Sequencing Protocols

Single-stranded recombinant M13 templates were sequenced by the chain terminating method (Sanger *et al.* 1977) using α^{35}S-dATP (1000 Ci/mmol) with either Klenow (Promega) or T7 DNA polymerase (Pharmacia) sequencing kits. The reaction products were fractionated by electrophoresis on a 6.0% polyacrylamide ionic gradient gel (Bankier and Barrel 1983).

Nested deletion series of the replicative forms (RF) were generated according to the procedure of Henikoff (1984). Regions of clones not spanned by deletion derivatives were sequenced using synthetic oligonucleotide primers.

Computer Sequence Analyses

Overlapping deletion clones were merged using Microgenie sequence software (Queen and Korn 1984). DNA sequences were aligned using the multiple sequence-alignment program CLUSTAL (Higgins and Sharp 1988) for both the pairwise and multiple alignments. Percent sequence similarity was calculated on a matches/length basis with large deletions being counted as a single mutational event. Phylogenetic trees were generated from the alignments using the PHYLIP software package for the IBM PC (Felsenstein 1991). Programs used were the following: Bootstrapped DNA parsimony algorithm (DNABOOT) version 3.4, the distance matrix program DNADIST version 3.4, and the Neighbour-Joining method (NEIGHBOUR) version 3.41. The parsimony tree generated by PHYLIP was analyzed using the program MacClade version 2.1 (Maddison and Maddison 1987). The spread and fixation of the SATA sequences were analyzed using the method of Strachan *et al.* (1985) for classifying mutations. The variation is classified into six different categories.

Class 1: Completely homogeneous for the position in all clones of both species. The ancestor base N_1 is therefore shared by both species.

Class 2: A minority of clones possesses a new mutation (N_2) at a position. The remainder of the clones remain homogeneous for the ancestor base (species A: N_1 only;

species B: $N_1>N_2$).

Class 3: Positions where the ancestor bases and mutations appear in equal frequency in one species whereas the other species remains homogeneous for one base (species A: N_1 only; species B: $N_1=N_2$).

Class 4: Positions where one species is homogeneous for a base that occurs in the minority of clones in other species (species A: N_1 only; species B: $N_2>N_1$).

Class 5: Each species is homogeneous for bases that are diagnostic for that species (species A: N_1 only; species B: N_2 only).

Class 6: All subsequent mutations beyond class 5 (species A: N_1 only; species B: $N_2>N_3$).

CHAPTER 3: RESULTS

SATA SATELLITE DNA FAMILY

SATA Satellite DNA Family Includes Three Size-Variants

Digestion of tilapiine genomic DNA with the restriction endonucleases *Eco*RI and *Hae*III revealed an intensely staining band of approximately 200 bp after fractionation by agarose gel electrophoresis. Size-fractionation by polyacrylamide gel electrophoresis of *Eco*RI-digested genomic DNA from eight tilapiine species representative of the three major genera, resolved three size-variants of the satellite DNA (Fig. 3.1). These size-variants, designated type I (237 bp), type II (230 bp) and type III (209 bp), were differentially distributed in the species examined (Table 3.1). To determine the species distribution of the tilapiine satellite DNA one of the cloned satellite monomers, OniIII-14, was radiolabelled and hybridized to a Southern blot of *Eco*RI digested DNA from eight tilapiine species (Fig. 3.2). The OniIII-14 clone hybridized to fragments in each of the species thereby establishing the presence of the satellite DNA in the genomes of the three major tilapiine genera.

To determine the extent of sequence relatedness among the satellite DNAs and the molecular basis for the monomer size variation, four or five independent clones from selected tilapiine species were sequenced. The initial cloning experiment was the cloning of the type III size-

Figure 3.1 6.0% polyacrylamide gel of *Eco*RI-digested tilapiine DNA. *Eco*RI-digests of genomic DNA from: *O. aureus* (O. aur.); *O. hornorum* (O. horn); *O. niloticus* (O. nil.); *O. placidus* (O. placi.); *T. zillii* (T. zillii); *T. rendalli* (T. rend.); *S. galilaeus* (S.gal.) and the hybrid strain, *O. mossambicus* x *O. hornorum* (O. moss./horn.) were fractionated by electrophoresis and stained with ethidium bromide. Molecular size markers in bp are shown at the left of the panel.

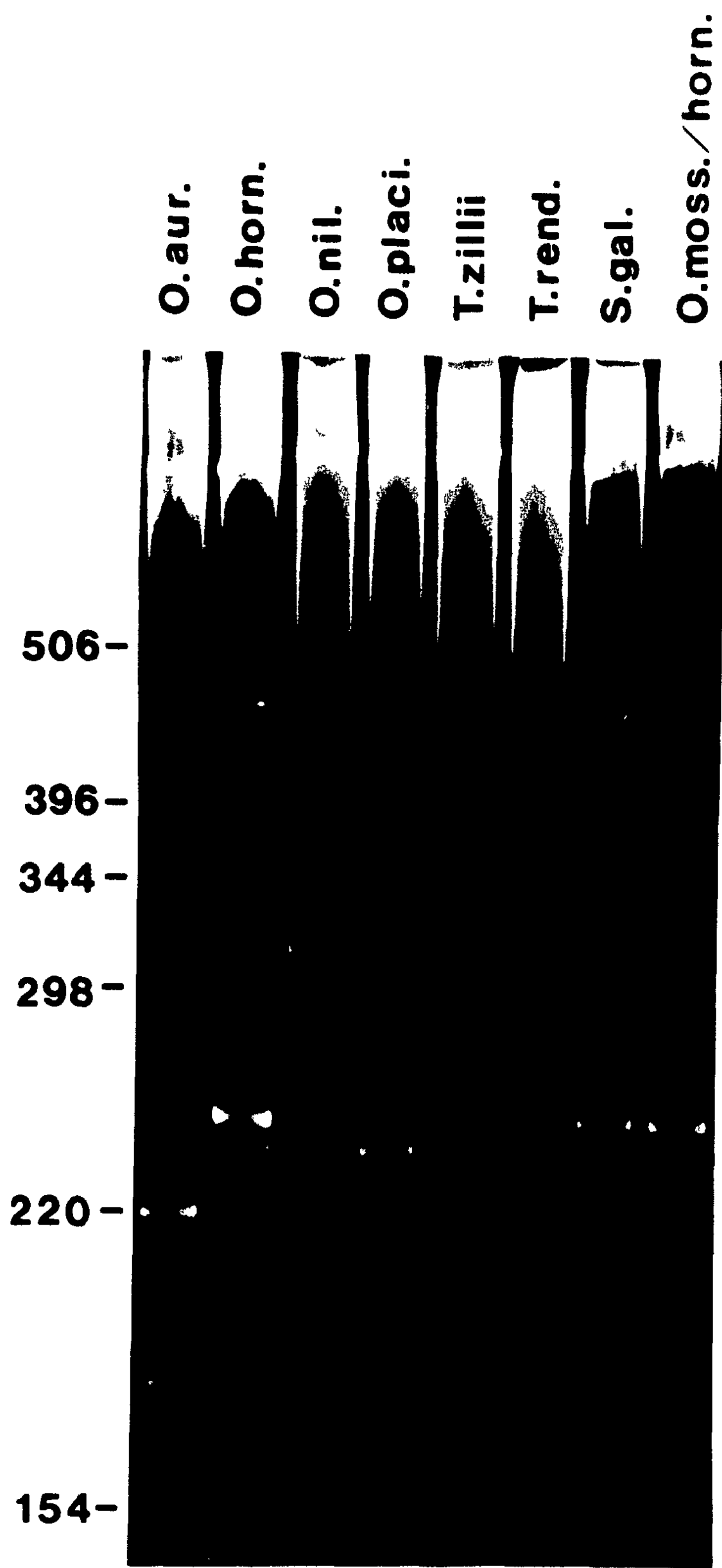


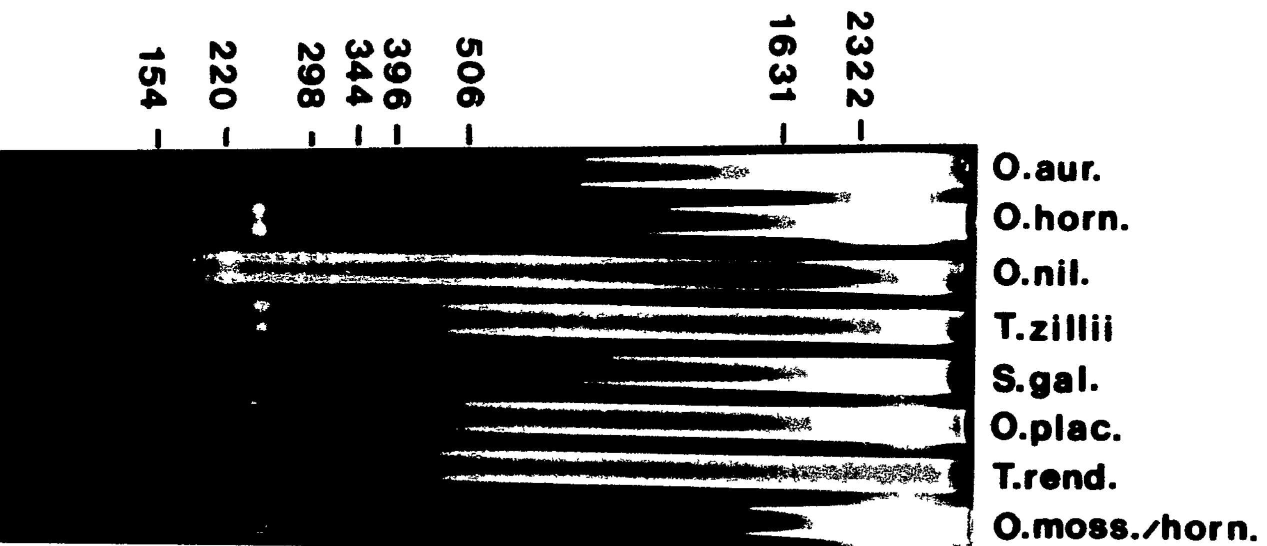
Figure 3.1

Table 3.1: Distribution of SATA size-variants in selected tilapiine fish genomes.

Species	Size-Variant		
	I	II	III
Oreochromis aureus	+	-	++
Oreochromis hornorum	++	++	-
Oreochromis niloticus	-	-	++
Oreochromis placidus	++	++	-
Tilapia zillii	++	-	-
Tilapia rendalli	++	-	-
Sarotherodon galilaeus	++	-	-
Oreochromis mossambicus/hornorum	++	-	+

Distribution of tilapiine SATA size-variants as visualized after digestion of genomic DNA with *Eco*RI and fractionation on a 6.0% polyacrylamide gel. Size-variants are I (~237 bp), II (~230 bp), and III (~209 bp). Relative quantities of the repeats are given as very abundant (++), abundant (+), and not detectable (-).

Figure 3.2 Southern blot and hybridization of a cloned satellite repeat from *O. niloticus* to *Eco*RI-digests of various tilapiine genomic DNAs. Left panel: 2% agarose gel electrophoresis of *Eco*RI-digests of genomic DNA from *O. aureus* (O. aur.); *O. hornorum* (O. horn.); *O. niloticus* (O. nil.); *O. placidus* (O. placi); *T. zillii* (T. zillii); *S. galilaeus* (S. gal.) *T. rendalli* (T. rend); and the hybrid strain, *O. mossambicus* x *O. hornorum* (O. moss./horn.). The resolution of the agarose gel is lower than that of the 6.0% polyacrylamide gel in Fig. 1 such that the type I and type II size-variants cannot be resolved. Right panel: Southern-blot hybridized to a cloned satellite repeat from *O. niloticus* (type III). Molecular weight markers in base pairs are shown at the left of the figure.



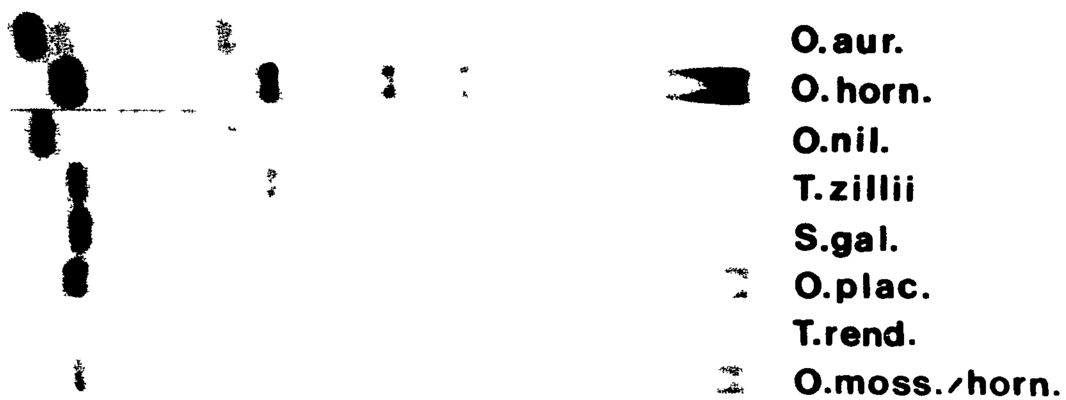
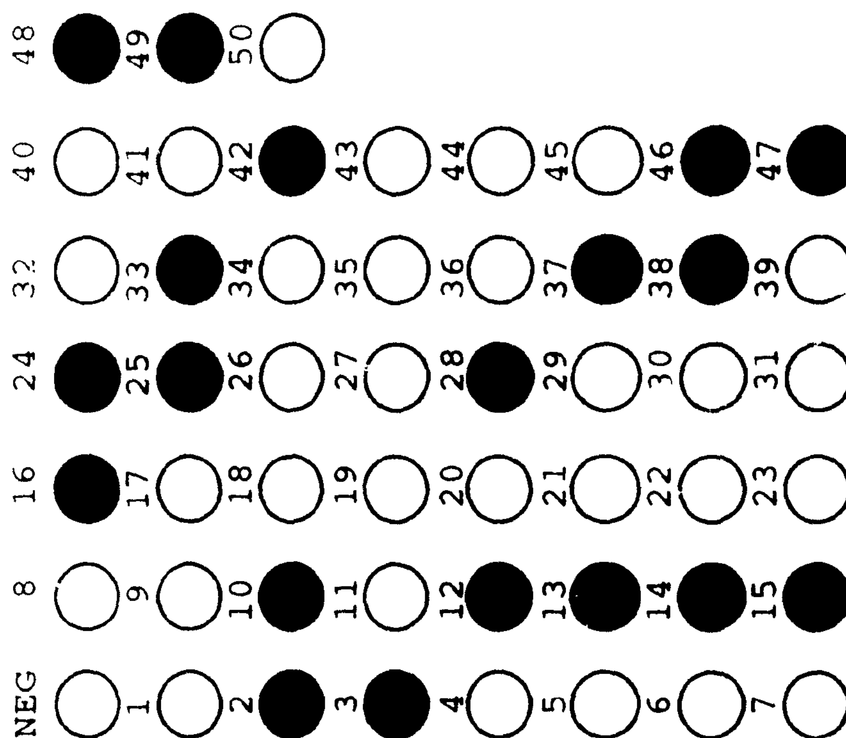


Figure 3.2

variant from the species *Oreochromis niloticus*. *O. niloticus* genomic DNA was digested with *Hae*III and fractionated on a 2.0% low melt agarose gel. The DNA was eluted from an intensely staining band of approximately 210 bp visualized in the ethidium stained gel. The eluted DNA was ligated into the *Sma*I site of the cloning vector M13mp9. Fifty of the recombinant M13 clones were screened with radiolabelled *O. niloticus* genomic DNA (Fig. 3.3). Nineteen of the clones hybridized strongly to the probe and were assumed to represent cloned repetitive DNA sequences. Four of the clones were sequenced to generate a consensus sequence (Appendix 1). Satellite DNA size-variants were subsequently cloned from *Oreochromis placidus* (size-variants I and II), *Sarotherodon galilaeus* (size-variant I), *Tilapia rendalli* (size-variant I) and *Tilapia zillii* (size-variant I). *Eco*RI-digested DNA for each species was size-fractionated by polyacrylamide gel electrophoresis and the eluted DNA was ligated into the *Eco*RI sites of M13 vectors. A consensus sequence was derived for each of the satellite DNAs based on four or five independently cloned monomer repeats (Appendix 1) and the six consensus sequences were arranged in a multiple alignment (Fig. 3.4). The type II and type III size-variants of the SATA sequence apparently have arisen from the type I variant by a deletion of 9 bp between residues 72 and 80 inclusive (in *O. placidus*) and a deletion

Figure 3.3 Dot blot hybridization of nick-translated *O. niloticus* genomic DNA to recombinant M13 clones. Left panel: Dot blot of M13 phage lysates hybridized to nick-translated genomic DNA from *O. niloticus*. Right panel: Schematic representation of dot blot with clone numbers. The lysate from a non-recombinant M13 was included as a negative control (NEG). The nineteen strongly hybridizing lysates are assumed to contain inserts of highly repetitive DNA.



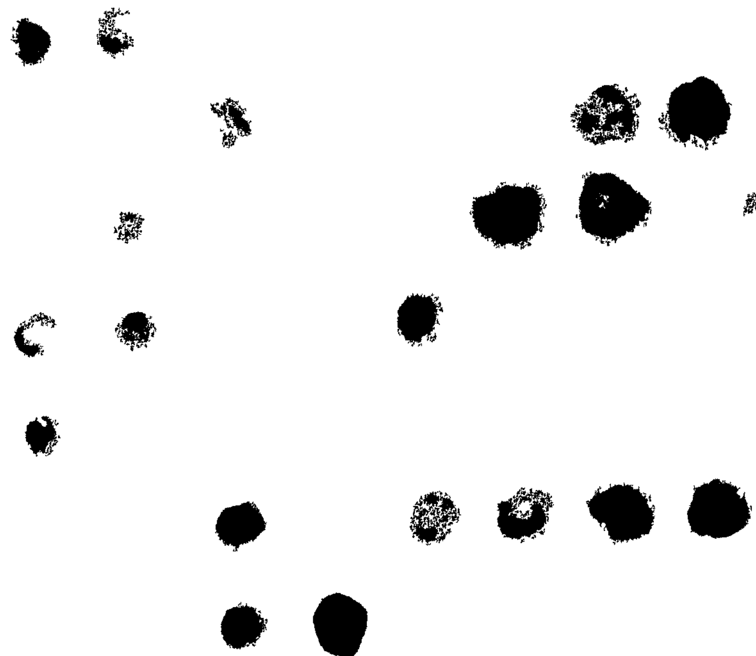


Figure 3.3

Figure 3.4 Multiple alignment of six tilapiine SATA consensus sequences for: *O. placidus* (type I and type II), *O. niloticus* (type III), *S. galilaeus* (type I), *T. zillii* (type I) and *T. rendalli* (type I). Consensus sequences were derived from four or five (*T. rendalli*) individual cloned sequences. N denotes no consensus, a dash indicates identity between consensus sequences, while deletions are represented by asterisks. The position of *Eco*RI and *Pal*I (isochizomer *Hae*III) recognition sites are shown above the consensus sequences. Consensus bases are assigned if 50% or greater of the clones possess the same nucleotide.

```
                           10        20        30        40        50
                  EcoRI    . PalI    .         .         HinfI     .
O. placi I CON:   AATTCTAT*AAGGCCAAG*CCTGAAATATGTGTGTCCGAGTCTTCTATCA
O. placi II CON:  --------G---------G-------------------------------
O. nil III CON:   --------*---------*------------------------C------
S. gal I CON:     --------*---------*-------------------------------
T. zillii I CON:  --------*-----*--AG---T------C-----G--------------
T. rend I CON:    -------NT---**---AG---T------C-----N--------------

                           60        70        80        90       100
                           .         .         .         .         .
OplI CON          AAAGTTACAGCTGTCTTTATGGACCTGGTG*AAAATCGCCTTATTTCGGG
OplII CON         -C-----------C-------*********G--------T---------*
OniIII CON        ------------------------T-----G------------------*
SgaI CON          --*------------------A--T-A---*--------T---------*
TziI CON          --*------------------C-GT-AA--*--------T--C-----CC
TreI CON          ---------------------C-GT-AA--*--------T--------CC

                          110       120       130       140       150
                           .         .         .         .         .
OplI CON          CGAGACAGTGCGTTTCTCGCTATTACATGCATTTGAATG*AGTTCTCGCC
OplII CON         -----------------G---------------------G----------
OniIII CON        --------------------*****************************-
SgaI CON          ---------------------------------------G----------
TziI CON          *A-------------------------------------G-C--------
TreI CON          -A-------------------------------------G----------

                          160       170       180       190       200
                           .         .         .         .         .
OplI CON          TGAAACACATTATGGG*TTTTCATTTTGTGAATAACTTGAAAATCTTAGC
OplII CON         ----------------*---------------------------------
OniIII CON        ----------------*---------------------------------
SgaI CON          ----------------*---------------------------------
TziI CON          -------A-G------A---------------------------------
TreI CON          -------A-G------*------------- -------------------

                          210       220       230       240
                           .         .         .         .  EcoRI
OplI CON          TCAAACAGCTGCAAAACCTATTTCCCCAGCAT*GGAAATGGTG
OplII CON         --------------------------------*-------C--
OniIII CON        -----------------------N--------*-------C--
SgaI CON          --------------------------------*-------C--
TziI CON          --------------------------------AA-G----C--
TreI CON          --------------------------*-----A-------C--
```

Figure 3.4

of 29 bp between residues 121-149 inclusive (in *O. niloticus*), respectively. The size-variants in *O. aureus* and *O. hornorum* are presumed to be derived from the same ancestral molecular events.

Copy Number and Species Distribution of the SATA Repeat

To determine the copy number of the SATA repeat and establish its species distribution, a radiolabelled satellite DNA insert (OplI-5) was hybridized to a slot blot with graded quantities of denatured genomic DNA from various fish species, as well as OplI-5 RF DNA. Genomic DNA was used from three tilapiine genera; *Oreochromis*, *Sarotherodon*, and *Tilapia*, as well as representative from both Old world and New world cichlids (Fig. 3.5). Cichlid species other than the tilapiines represented on the blot, include three haplochromine fishes, *Haplochromis (Protomelas) similis*, *Haplochromis moori*, and *Melanochromis auratus*, two west African cichlids, the hemichromine *Hemichromis bimaculatus*, and a member of the chromidotilapiine tribe, *Pelvicachromis pulcher*, as well as the Asian cichlid, *Etroplus maculatus*, and the South American cichlid, *Cichlasoma meeki*. Genomic DNA samples were also included from rainbow trout (*Oncorhynchus mykiss*) and haddock (*Melanogrammus aegelfinus*). The hybridization of the probe to the denatured DNA samples was quantified by densotimetric scanning of the

Figure 3.5 Slot blot hybridization of radiolabelled OplI-5 (SATA) insert to graded amounts of denatured cichlid genomic DNA samples. 500, 50, and 5 ng aliquots of the genomic DNA and 50, 5, and 0.5 ng aliquots of the OplI-5 (Opl5) denatured replicative form (RF) were immobilized on a nylon membrane. Tilapiine species included are: *O. hornorum* (O. horn); *O. niloticus* (O. nil); *O. placidus* (O. placi); *O. mossambicus* (O. moss); *O. andersonii* (O. ander); *O. mortimeri* (O. mort); *O. aureus* (O. aur); *S. galilaeus* (S. gal); *T. rendalli* (T. rend); *T. zillii* (T. zillii); *T. mariae* (T. mar) and *T. tholloni* (T. tholl). The three haplochromine cichlid samples included are: *Haplochromis similis* (H. sim); *Haplochromis moori* (H. moor) and *Melanochromis auratus* (M. aur). *Hemichromis bimaculatus* (H. bimac) is a west African hemichromine species; *Etroplus maculatus* (E. mac) an Asian cichlid; *Pelvicachromis pulcher* (P. pulch) is a west African chromidotilapiine species, and *Cichlasoma meeki* (C. meeki) is a South American cichlasomine species. Genomic DNA samples from rainbow trout, *Oncorynchus mykiss* (O. mykiss) and haddock, *Melanogrammus aegelfinus* (M. aegel) were included on the blot.

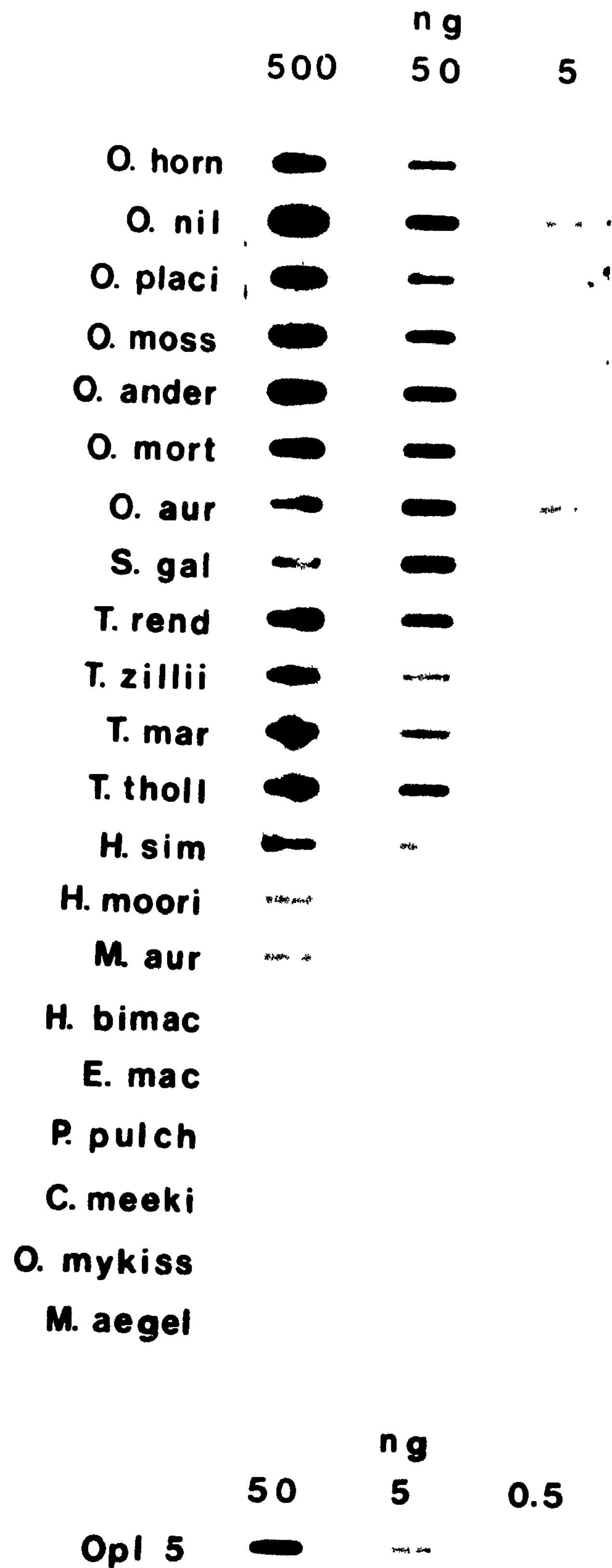


Figure 3.5

autoradiograph. The copy number for each species was estimated based on the haploid DNA content for cichlids as 1.0 pg (Majumdar and McAndrew 1986) (Table 3.2). The SATA sequence could be detected in all of the tilapiine samples as well as the three haplochromine species. The copy number of the SATA repeat is significantly lower in the three haplochromine fishes than it is in the tilapiine fish genomes. This may be indicative of a major amplification event of the SATA array after the divergence of the tilapiine fishes from the haplochromine lineage.

SATA Sequence Cloned from *H. similis*

The estimates for the lower copy number of the SATA repeat in the haplochromine genomes could be attributed to a low level of sequence identity to the tilapiine SATA probe. The hybridization and wash conditions for the slot blot should prevent the hybridization of the tilapiine SATA probe to sequences with less than ~75-80% sequence identity. The calculation of the copy number for the SATA sequences in the tilapiine and haplochromine genomes are therefore conservative estimates. To establish the basis for the reduced signal in the haplochromine species the SATA sequence from one of the haplochromine species, *Haplochromis similis* was determined. The monomer repeat could not be detected in digests of genomic DNA with the restriction

Table 3.2: SATA copy number estimation for selected cichlid species.

Species	Copies/Haploid Genome	Genome Equivalent (%)
Oreochromis hornorum	3.6×10^4	1.0
Oreochromis niloticus	5.4×10^4	1.5
Oreochromis placidus	2.6×10^4	0.7
Oreochromis mossambicus	3.9×10^4	1.1
Oreochromis andersonii	4.9×10^4	1.4
Oreochromis mortimeri	4.4×10^4	1.2
Oreochromis aureus	5.6×10^4	1.6
Sarotherodon galilaeus	5.9×10^4	1.6
Tilapia rendalli	3.9×10^4	1.1
Tilapia zillii	7.4×10^3	0.2
Tilapia mariae	1.6×10^4	0.4
Tilapia tholloni	2.8×10^4	0.8
Haplochromis similis	5.1×10^3	0.1
Melanochromis moori	1.1×10^3	0.03
Haplochromis auratus	1.2×10^3	0.03

Copy number estimation of SATA repeat based on densitometry of slot blot hybridization experiment (Fig. 3.5). Copy number is estimated based on a haploid DNA content of 1.0 pg (Majumdar and McAndrew 1986).

endonucleases *Eco*RI (Fig. 3.6) nor *Pal*I (*Hae*III isochizomer, data not shown). The monomer SATA sequences hybridized to the OniIII-14 probe in the tilapiine species, whereas hybridization of the tilapiine SATA probe could only be detected in the high molecular weight DNA for the *Eco*RI-digests of the haplochromine species' genomic DNA. A computer analysis (Microgenie sequence software) of the consensus sequences from the six tilapiine satellite DNAs identified potential unique restriction endonuclease sites for the isolation of the *H. similis* SATA sequence. The restriction endonucleases *Hin*fI, *Mbo*I, and *Hinc*II were used to digest the genomic DNA of *H. similis*. The Southern blot of the fractionated DNA was hybridized to a radiolabelled SATA clone (Fig. 3.7). The *Mbo*I digest showed a ladder of hybridizing bands corresponding to multiples of the monomer SATA sequence, but no monomer size fragment could be visualized, whereas, the *Hin*dII digest revealed only high molecular weight DNA that hybridized to the probe. A monomer fragment, however, could be visualized in the *Hin*fI digest of the *H. similis* DNA that hybridized to the tilapiine SATA probe. The multimer ladders observed in this Southern hybridization may be explained by the loss of recognition sites by mutation or the sensitivity of both *Mbo*I and *Hin*fI to methylation of the genomic DNA. The DNA from the intensely staining band of monomer size in the *Hin*fI-

Figure 3.6 Southern hybridization of radiolabelled OniIII-14 (SATA) clone to *Eco*RI-digested cichlid genomic DNA samples. There are four tilapiine species: *O. mossambicus* (O. moss); *O. aureus* (O. aur); *T. zillii* (T. zillii) and *S. galilaeus* (S. gal); two haplochromine species: *H. moori* (H. moori) and *H. similis* (H. sim); the Asian species: *Etroplus maculatus* (E. mac); and three South American cichlids: *Cichlasoma otofasciatum* (C. oto); *Aequidens pulcher* (A. pulch) and *Cichlasoma meeki* (C. meeki). Molecular size markers are given at the left of the figure in Kbp.

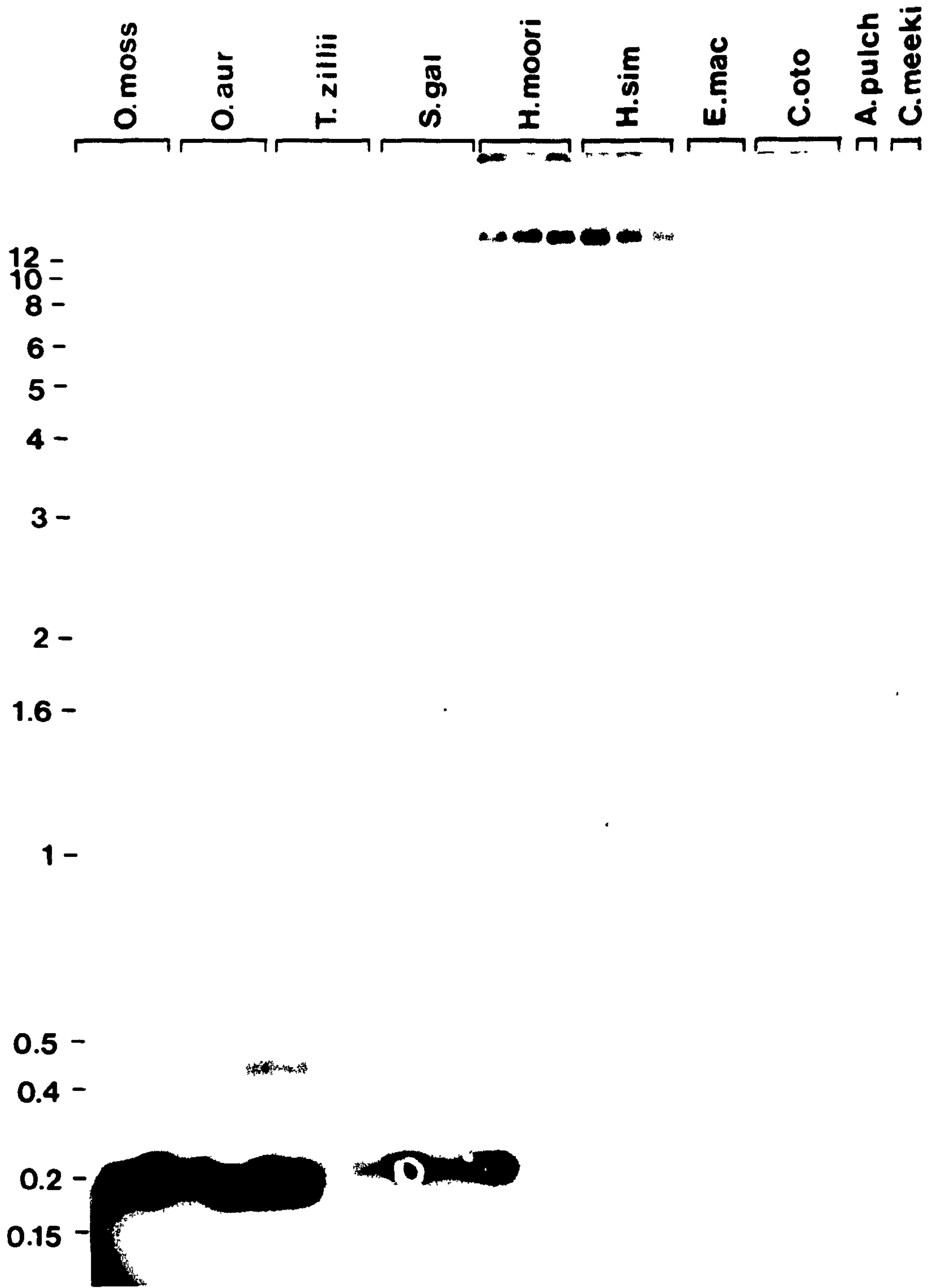


Figure 3.6

Figure 3.7 Southern hybridization of radiolabelled OniIII-14 (SATA) clone to genomic DNA of *O. mossambicus/hornorum* and *H. similis. H. similis* genomic DNA was digested with *Mbo*I-, *Hin*fI-, and *Hin*dII restriction endonucleases. The first lane contains *Eco*RI-digested *O. mossambicus/hornorum* genomic DNA to serve as a positive control. The two size-variants of the hybrid strain both hybridized to the cloned insert as do their two respective dimers. A fragment of monomer size hybridized to the OplI-5 insert in the *Hin*fI digest of the *H. similis* genomic DNA. Molecular size markers are given at the left of the figure in Kbp.

Figure 3.7

digested *H. similis* DNA was eluted from an agarose gel. The eluted DNA was treated with Mung Bean nuclease and blunt end ligated into the *Sma*I site of M13mp18. Forty-one recombinant clones were isolated and prepared by the procedures described in Chapter 2. Phage lysates from the clones were dotted onto a nylon membrane and hybridized to the radiolabelled insert of the *O. placidus* SATA clone, OplI-5 (Fig. 3.8). Eight of the forty-one clones hybridized to the SATA probe of which five were sequenced. The five clones were aligned to produce a consensus alignment (Appendix 1). The *H. similis* SATA consensus sequence is 230 bp in length. The consensus length is shorter than the tilapiine SATA type I size-variant due to the cloning methodology which used Mung Bean Nuclease to blunt end the insert DNA before cloning. This resulted in the loss of the overhanging single stranded DNA and the subsequent truncation of the SATA repeat for each clone.

Polymorphic Hybridization Patterns

The cloned OniIII-14 satellite DNA sequence was hybridized to *Pal*I digests of genomic DNA from several species in the genera, *Oreochromis*, *Sarotherodon*, and *Tilapia*, in which the *Pal*I released monomer had migrated off the end of the agarose gel. Long exposures of the membrane revealed a series of polymorphic, high molecular weight

Figure 3.8 Slot blot hybridization of radiolabelled OpIf-5 (SATA) insert to recombinant M13 clones for the *H. similis* SATA cloning experiment. Right panel: Schematic representation of the slot blot with the corresponding clone number above each slot. The lysate from a non-recombinant M13 phage (NEG) was included as a negative control. Eight of the forty-one clones hybridized to the radiolabelled SATA insert.

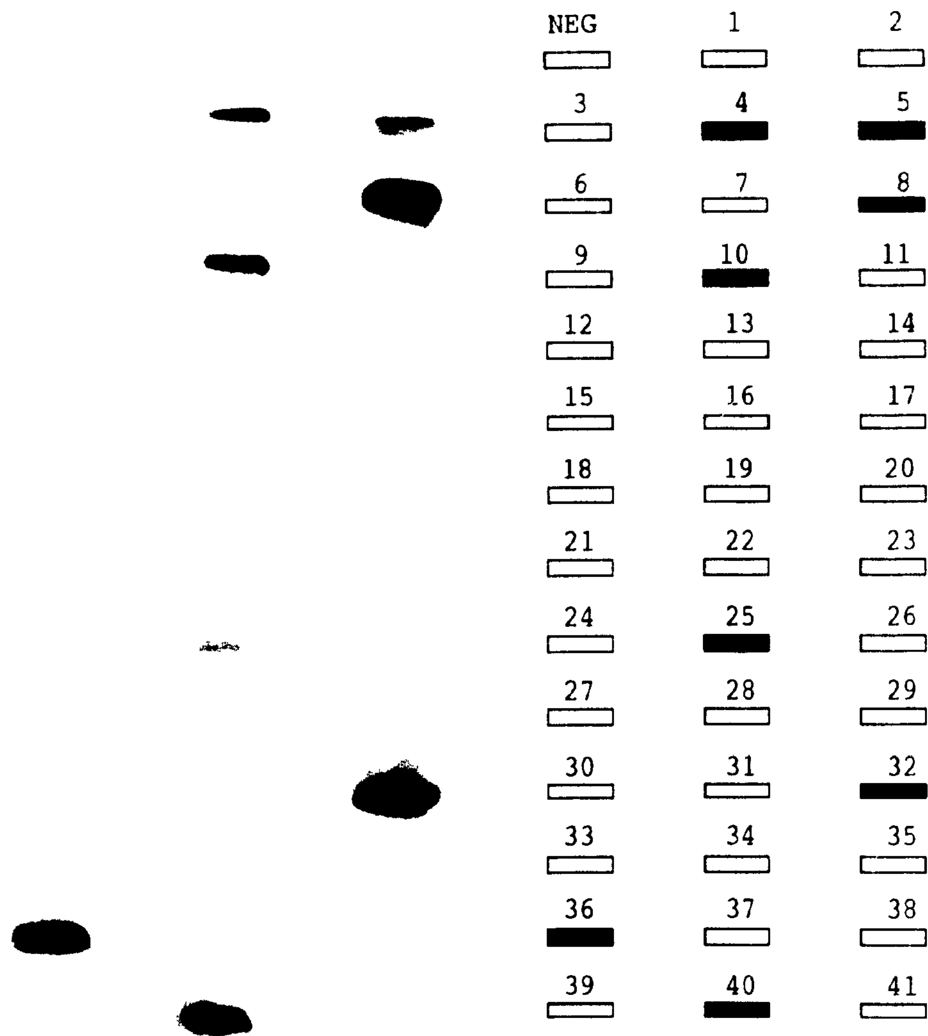


Figure 3.8

fragments or fingerprint-like patterns (Jeffreys *et al.* 1985a, 1985b) (Fig. 3.9). The polymorphism detected in these high molecular weight restriction fragments varied from species to species. To calculate the extent of polymorphism, a pairwise comparison of band-sharing between individuals was determined by the percentage of co-migrating bands. The calculated values ranged from 27% for *O. mossambicus* to approximately 90% for *S. galilaeus*. The lack of the *Pal*I recognition site in the sequence of the cloned satellite monomers from the *Tilapia* species (Fig. 3.4) explains the absence of fragments <4 kbp that hybridized to the *O. niloticus* probe in the genomic DNA of these species. The *Tilapia* DNA fingerprints that were observed are most readily explained as polymorphism in the higher-order structure of the satellite tandem array, *i.e.* the appearance of a *Pal*I recognition sequence by mutation in subsets of the tandem array. As such, this *Pal*I polymorphism reveals a generic- and possibly species-specific profile.

Interspecific Comparison of SATA Sequences

The seven consensus sequences for the tilapiine species and *Haplochromis similis* were aligned using the CLUSTAL multiple alignment program (Higgins and Sharp 1988) and compared in a pairwise fashion to estimate sequence identity between the different sequences. Sequence identity was

Figure 3.9 Polymorphic DNA fingerprint-like patterns detected by a satellite probe (SATA) from tilapia. A cloned *O. niloticus* (OniIII-14) satellite monomer was hybridized to *Pal*I-digested genomic DNA from either two or three individuals of *O. aureus* (Oa), *O. mossambicus* (Om), *O. mortimeri* (Or), *O. andersonii* (Os). *O. niloticus* (On) and *S. galilaeus* (Sg). Right panel: Polymorphic DNA fingerprint-like patterns detected by hybridization of the OniIII-14 insert to *Pal*I-digested DNA from three individuals each of *T. zillii* (Tz), *T. rendalli* (Tr), *T. tholloni* (Tt), and *T. mariae* (Tm). Molecular size markers are given at the left of each panel in Kbp.

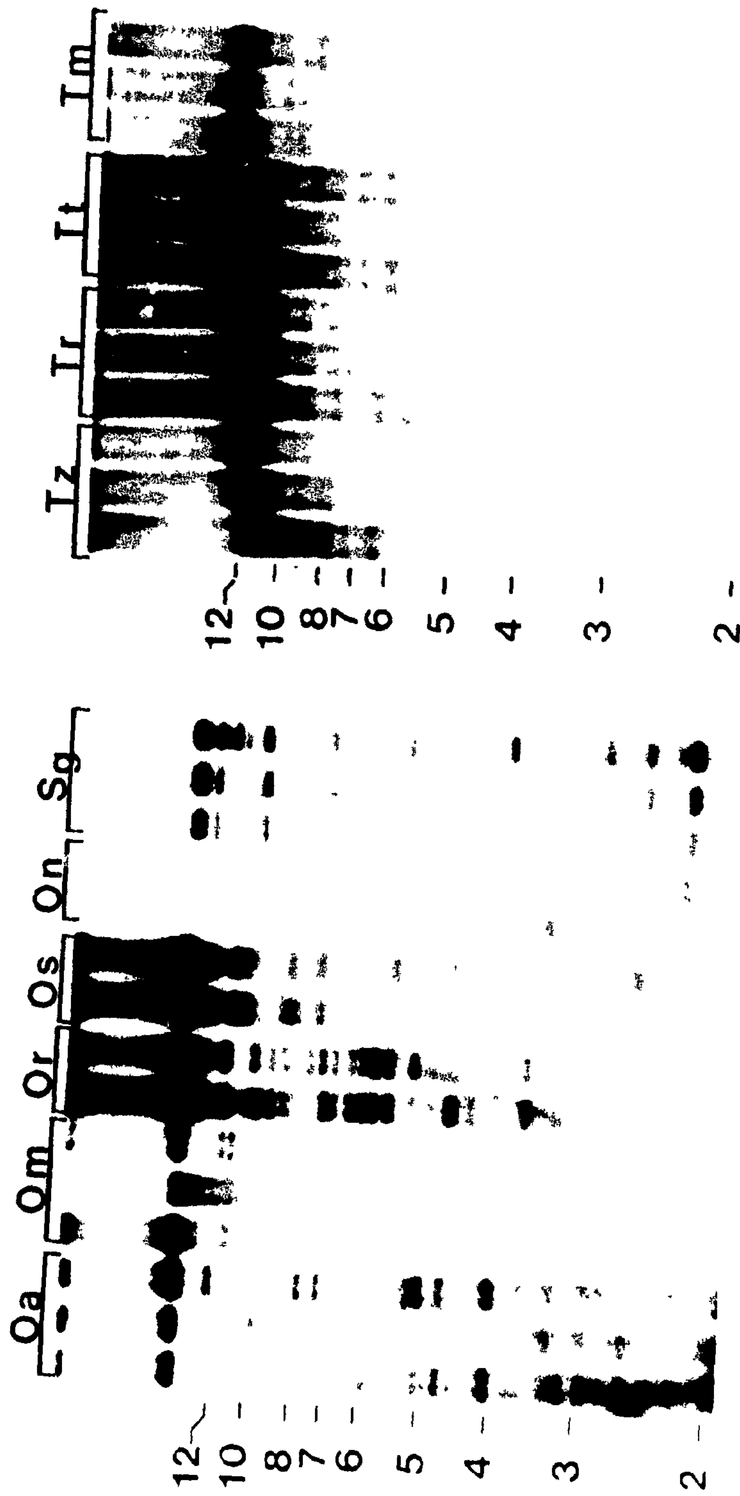


Figure 3.9

calculated on a matches/length basis and the two major deletion events were counted as single mutational events (Table 3.3). The *Oreochromis* and *Sarotherodon* sequences have the highest sequence identity with matches/length values of approximately 96%. The *Oreochromis* and *Sarotherodon* sequences show similar identity to both of the *Tilapia* genera SATA sequences with matches/length values of approximately 90%. The *H. similis* consensus sequence shows significantly lower sequence identity values in comparison with all of the tilapiine sequences ranging from 73.5% for *O. placidus* II to 80.1% for the *T. zillii* SATA sequence.

Concerted Evolution and Homogenization of the SATA Repeats

The homogenization and fixation of species-diagnostic mutations in satellite DNAs is defined as concerted evolution. Molecular drive is the collective term invoked to describe the genomic forces that result in this outcome (Dover 1982; Dover 1986). The loss or gain of restriction endonuclease recognition sites is an indirect result of the fixation of satellite DNA monomers for specific nucleotides. The absence of recognition sites for the restriction endonucleases *Pal*I for the *Tilapia* genera's individual SATA monomers, and both *Eco*RI and *Pal*I restriction endonucleases for the *H. similis* SATA monomer repeats is evidence of concerted evolution for the SATA family. Strachan *et al.*

Table 3.3: Cross-homology comparison of tilapiine and *Haplochromis similis* SATA consensus sequences.

	Percentage Sequence Identity						
	O. placidus I	*O. placidus II*	*O. niloticus III*	*S. galilaeus I*	*T. rendallii I*	*T. zillii I*	*H. similis I*
O. placidus I	-	97.0	97.1	95.8	89.5	89.2	78.3
O. placidus II	-	-	96.0	96.1	90.9	89.5	73.5
O. niloticus III	-	-	-	96.2	88.1	88.2	78.6
S. galilaeus I	-	-	-	-	90.8	92.0	77.4
T. rendalli I	-	-	-	-	-	94.2	80.0
T. zillii I	-	-	-	-	-	-	80.1
%GC	40.5	40.7	41.1	39.0	36.3	38.2	34.9
Size (bp)	237	231	209	236	237	238	230

Percentage sequence identity was calculated on a match/length basis following pairwise alignment.

(1985) have devised a methodology for analyzing the transition stages of molecular drive at the nucleotide level in highly repetitive DNAs. The six classes of mutational distribution as described in Chapter 2 reveal the transition stages of monomer repeats from homogeneity at a given nucleotide position to partial and full fixation for a mutation in one species. Six classes of variation are used to describe the stage of transition for the species' satellite DNA sequence, as described in Chapter 2. The transition stages are diagrammed schematically in Figure 3.10 to illustrate the spread of mutations, and the methodology used for classifying invidual positions. The majority of positions are homogeneous for all clones (class 1) ranging from 59.7% for the *O. niloticus* vs. *H. similis* comparison to 79.1% for the *O. niloticus* vs. *O. placidus* (I) comparison (Table 3.4). Conversely, class 5 mutations which indicate complete fixation for different bases in each species are highest in frequency for the *O. niloticus* vs. *H. similis* comparison (14.2%) and lowest for the intrageneric *O. niloticus* vs. *O. placidus* comparison (0.5%). Class 5 represents the classic case of concerted evolution for a satellite DNA sequence within a species (Dover 1982).

To estimate the time elapsed since the last major amplification event for the SATA arrays the individual monomer sequences were compared to their respective

Figure 3.10 Schematic representation of transition stages during the spread of mutations. The six classes of patterns of distribution of mutations are represented for four clones a-d each for Species A and Species B. The circles represent nucleotides that may be A, T, G, or C. Following the multiple alignment of the individual cloned sequences each position is classified according to this scheme. For the class 1 pattern the aligned position is homogeneous for a base. In classes 2-5 only two bases are found at a given position across all clones of a pair of species. Class 5 represents the total fixation for a nucleotide within a species. All subsequent mutations beyond this point are represented by the pattern of class 6 (Adapted from Strachan *et al*. 1985).

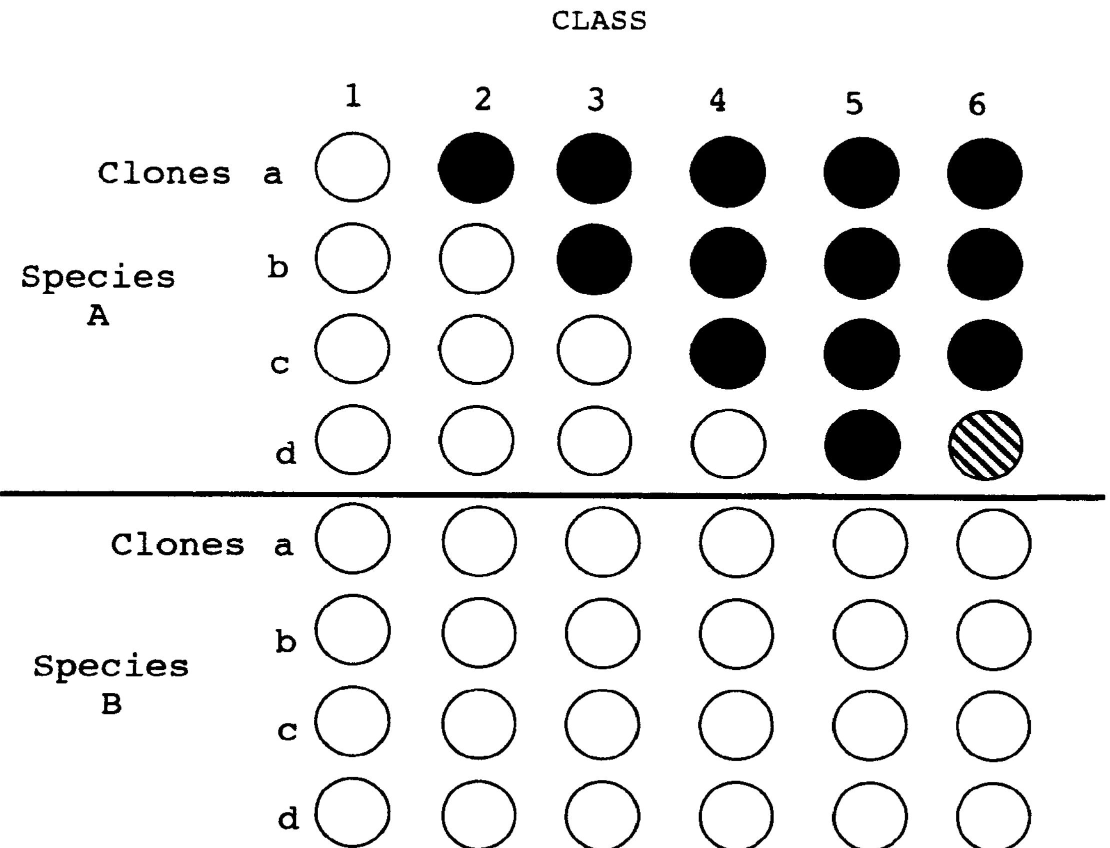


Figure 3.10

Table 3.4: Partitioning of the distribution of DNA mutations (%) at individual nucleotide positions for all clones from *Oreochromis niloticus* and six other cichlid species' satellite DNAs.

	Oreochromis niloticus versus					
Class of mutational distribution	*Oreochromis placidus I*	*Oreochromis placidus II*	*Sarotherodon galilaeus*	*Tilapia rendalli*	*Tilapia zillii*	*Haplochromis similis*
1	79.1	72.7	79.0	61.5	57.2	59.7
2	14.0	19.0	15.4	14.7	20.5	15.6
3	-	1.0	0.9	-	2.8	-
4	-	1.0	0.5	3.9	0.9	3.8
5	0.5	2.4	1.9	4.8	3.7	14.2
6	0.9	1.0	0.5	3.9	2.8	3.8

Individual cloned monomer sequences from *O. niloticus* were aligned in a multiple alignment with individual monomer sequences of the six cichlid species shown. Individual positions were classified according to the scheme of Strachan *et al.* (1985). Deletions were counted as a single class 5 mutation.

consensus sequences. The number of substitutions in relation to the consensus sequence were summed and divided by the total number of bases compared. The calculation for the time elapsed since the last homogenization event was based on a rate of 1.55×10^{-9} directional substitutions (AT-GC) per site per year calculated for cold blooded vertebrates (Bernardi and Bernardi 1990). The results are summarized in Table 3.5. The estimates range from 4.8 million years ago (mya) for the *S. galilaeus* SATA sequence to 11.7 mya for the *T. zillii* and *H. similis* SATA sequences.

Molecular Phylogenetic Analyses

The species-specific patterns produced by hybridization of the SATA sequence to *Pal*I digests of tilapiine genomic DNA indicates the possible utility of the SATA sequences for molecular phylogenetic analyses. The phylogenetic relationships of the six tilapiine satellite DNA consensus sequences were analyzed using the haplochromine *H. similis* sequence as the designated outgroup. The multiple alignment (Appendix 2) generated by the CLUSTAL multiple alignment program was used as the input for the Bootstrapped DNA Parsimony and Neighbour-Joining phylogenetic programs of PHYLIP (PHYLogeny Inference Package; Felsenstein 1991). The bootstrapping feature of the DNA Parsimony program resamples the data set 500 times with replacement and gives confidence

Table 3.5: Estimation of time elapsed since last major amplification event for SATA sequences from selected cichlid species.

SATA Sequence	Number of Directional Base Substitutions	Total Bases in Alignment	Time Elapsed Since Major Amplification (mya)
O. placidus I	13	949	8.8
O. placidus II	15	899	10.8
O. niloticus III	8	830	6.2
S. galilaeus I	7	941	4.8
T. rendalli I	12	1207	6.4
T. zillii I	16	885	11.7
H. similis I	21	1154	11.7

Individual cloned monomer sequence were aligned to their respective consensus sequence to determine the number of directional base substitutions. Positions with more than one cloned monomer sequence having the same base substitution were counted as one since they may be co-amplified in the same array. Time elapsed since the last major amplification event was calculated with a directional substitution rate of 1.55×10^{-9} per site per year (Bernardi and Bernardi 1990) is given as millions of years ago (mya).

limits on the nodes in the parsimony tree by calculating the number out of 500 replicates that identify the particular node under question (Felsenstein 1985). The tree clearly identifies two distinct clades with the two *Tilapia* sequences in one and the *Oreochromis* and *Sarotherodon* sequences united in the other (Fig. 3.11), with a confidence limit on this node of 87.2%. The resolution between the *Sarotherodon* and *Oreochromis* sequences is not as clear with only 62.4% of the 500 replicates identifying this node. The unrooted parsimony analysis was repeated on the data set with the corresponding deleted regions on *O. placidus* II and *O. niloticus* III size-variants being removed with no significant differences from the previous analysis. To test the validity of the parsimony tree generated by the bootstrapped parsimony program the input data were re-analyzed using the program MacClade to test alternative tree topologies (Maddison and Maddison 1987). The parsimony tree generated by PHYLIP was re-created by MacClade which gives a treelength of 72 steps and a consistency index of 0.90 (Fig. 3.12). The treelength is calculated as the sum of the number of steps for each of the characters. Less parsimonious tree topologies will result in increased treelengths. The consistency index of the tree is calculated as the sum of the minimum conceivable number of evolutionary steps for each character in the matrix divided by the number of

Figure 3.11 Bootstrapped DNA parsimony phylogenetic tree generated by the DNABOOT program of PHYLIP. The SATA consensus sequence was used as a designated outgroup to root the tree for the six tilapiine satellite sequences. 500 bootstrapped replicates of the input data were conducted in order to calculate the confidence limits on the major nodes of the tree. The bootstrap values are given as percentage values on the tree.

Bootstrapped Parsimony Tree

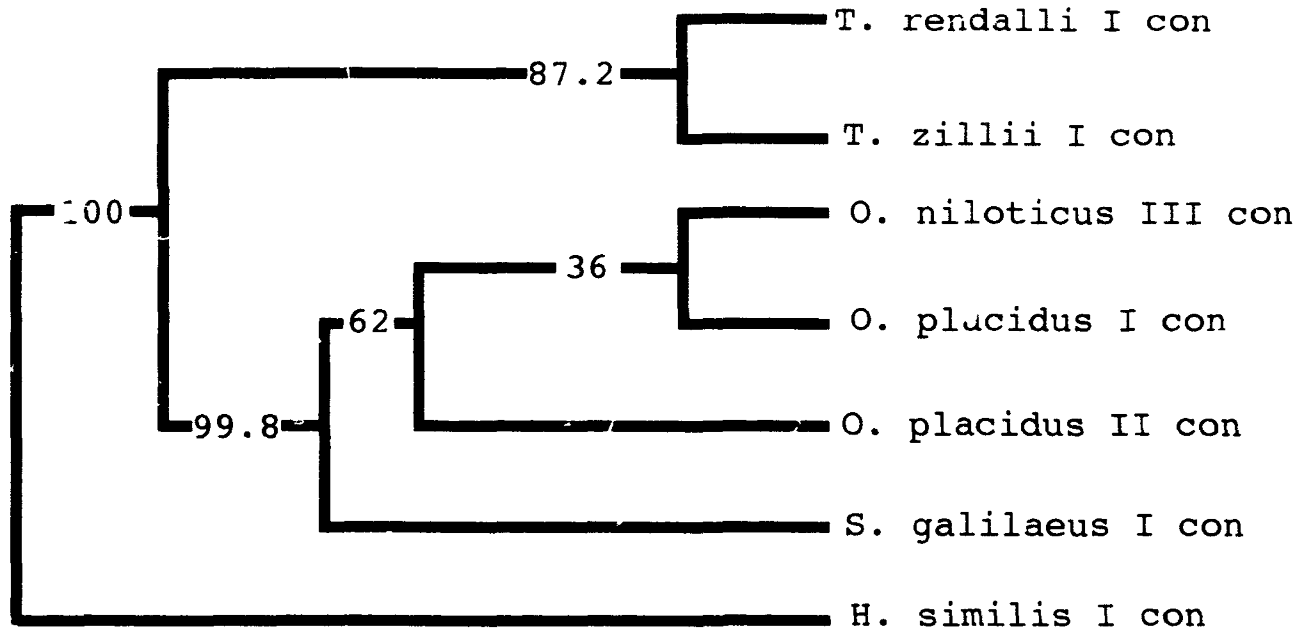
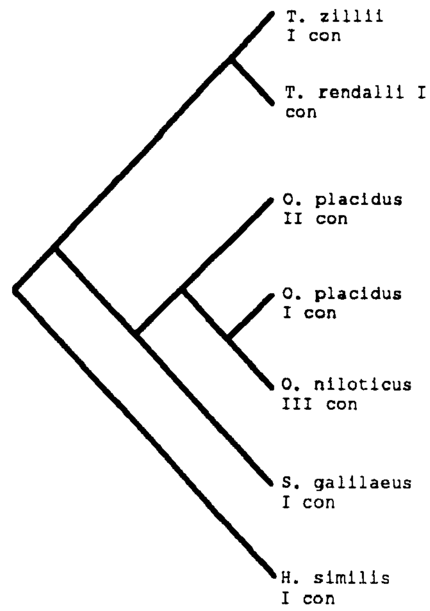
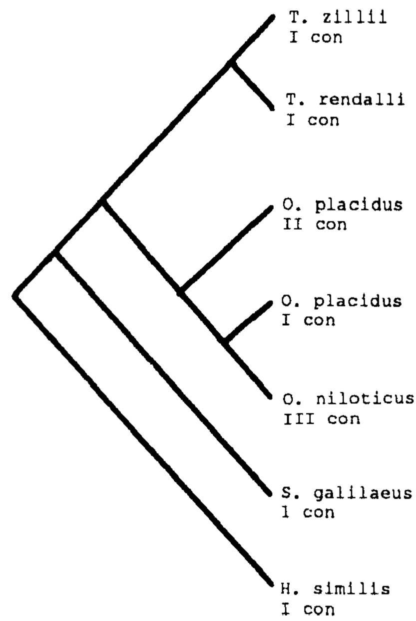


Figure 3.11

Figure 3.12 Cladograms of alternative topologies for the parsimony tree as tested by the MacClade phylogenetic program. The top tree is identical to the tree generated by the DNABOOT program of PHYLIP. This tree has a treelength of 72 steps and a consistency index (CI) of 0.90. The bottom left tree is an alternative topology created by the MacClade phylogenetic program which removes the *Sarotherodon galilaeus* SATA consensus sequence from the monophyletic assemblage of the above tree. This tree has a treelength of 79 steps and a CI of 0 82. The bottom right tree topology unites the *S. galilaeus* SATA consensus sequence with the two *Tilapia* sequences and also results in a treelength of 79 steps and a CI of 0.82.





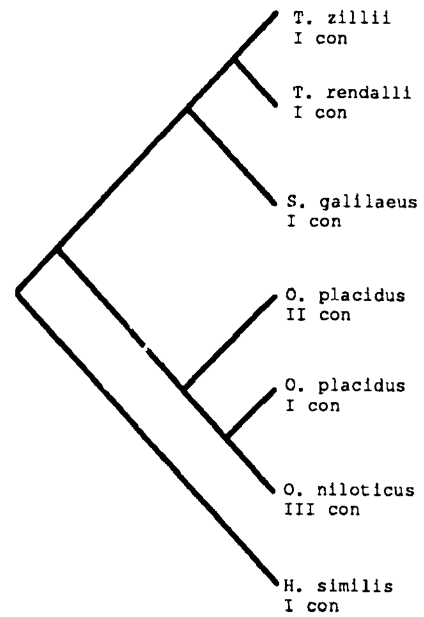


Figure 3.12

observed steps. The consistency index does not count character positions which are monomorphic for all taxa in the sequence alignment. Therefore, as the denominator of observed number of steps increases the consistency index will decrease. The consistency index can vary from 0 to 1. To test the validity of the *Oreochromis/Sarotherodon* monophyletic lineage produced by the parsimony analysis the *Sarotherodon galilaeus* SATA sequence was situated on a separate branch from the *Oreochromis* sequences (Fig. 3.12), resulting in a treelength of 79 steps and a consistency index of 0.82. Another alternative topology united the *Tilapia* and *Sarotherodon* sequences to the exclusion of *Oreochromis*, effectively making the *Tilapia* and *Sarotherodon* sequences members of the same monophyletic lineage. This tree topology, representative of the polyphyletic model of tilapiine evolution proposed by Peter and Berns (1982), resulted in a tree of 79 steps and a consistency index of 0.82. The parsimony tree created by the bootstrapped DNA parsimony algorithm is therefore the best supported topology based on the principle of parsimony since it has the shortest treelength and the highest consistency index.

Neighbour-Joining analysis (Saitou and Nei 1987) of the seven SATA consensus sequences was performed using the NEIGHBOUR program of PHYLIP (Felsenstein 1991). The input data for the Neighbour-Joining method is a distance matrix

of the seven SATA sequences calculated using the DNADIST program of PHYLIP. The Neighbour-Joining tree is constructed by joining the least distant modes on the tree followed by the joining of these two nodes to their common ancestral node. This analysis therefore groups taxon together that have the greatest genetic identity as calculated in the distance matrix. The tree generated by the Neighbor-Joining method (Fig. 3.13) is similar to the one produced by the bootstrapped Parsimony analysis. The three *Oreochromis* sequences are grouped together in a separate clade from the *Sarotherodon* SATA sequence. The *Oreochromis/Sarotherodon* clade as in the parsimony analysis is also clearly grouped away from the *Tilapia* clade.

Figure 3.13 Neighbour Joining phylogenetic tree for the tilapiine SATA sequences. Generated by the NEIGHBOUR program of the phylogenetic software package PHYLIP (Felsenstein 1991). The SATA sequence from the *H. similis* species was designated as an outgroup to root the tree. The tree was calculated from a distance matrix generated by the program DNADIST of PHYLIP.

Neighbour Joining Tree

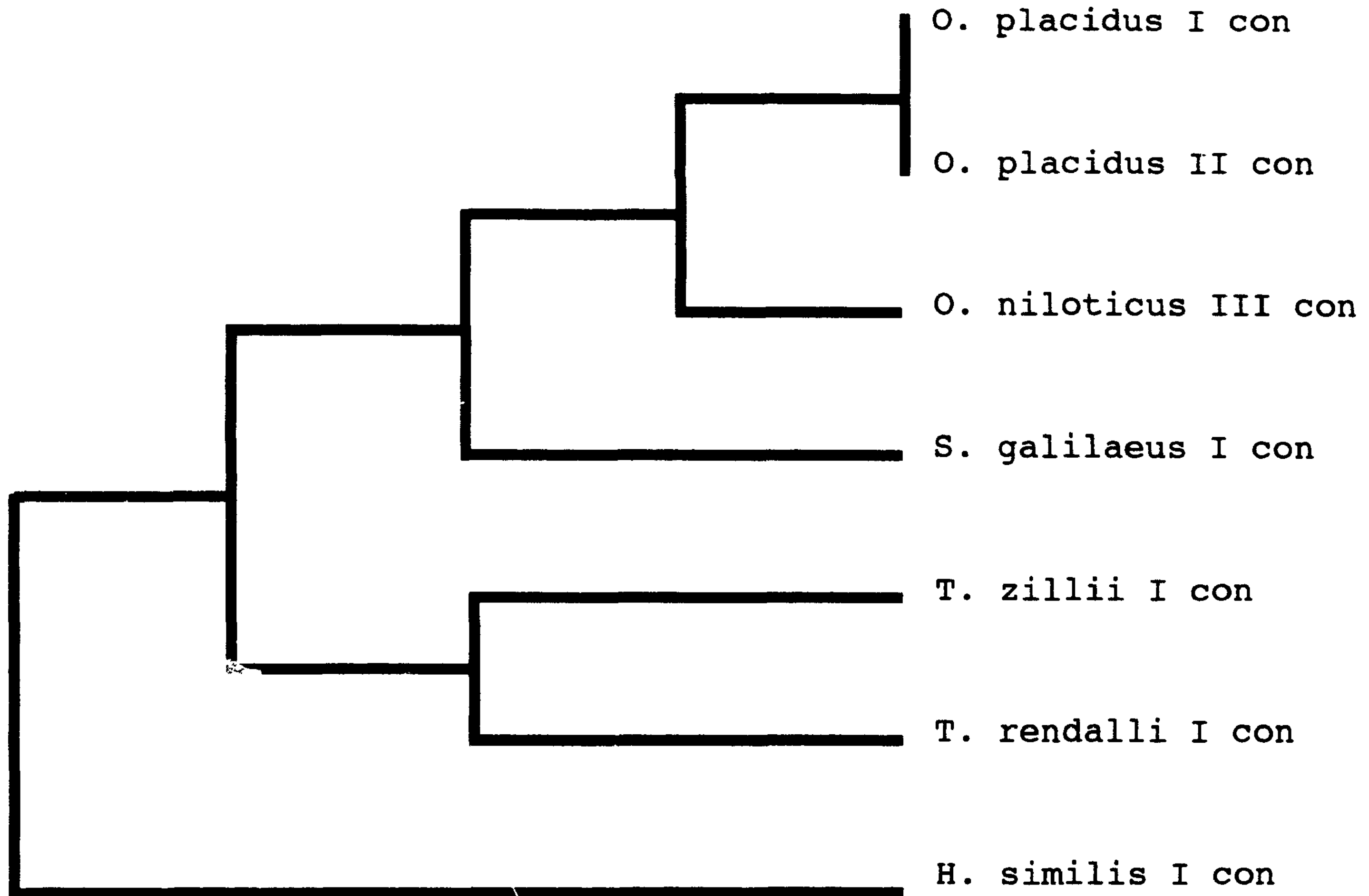


Figure 3.13

CHAPTER 4: RESULTS

SATB SATELLITE DNA FAMILY

Digestion of genomic DNA from several tilapiine species of fishes with the restriction endonuclease *Hin*dIII revealed two intensely staining bands of approximately 750 bp and 1900 bp when fractionated on a 1.0% agarose gel (data not shown). The 750 bp band was most easily detected in the genomic digest of *O. hornorum*. The DNA from this band was eluted from the gel and ligated into the *Hin*dIII site of M13mp19. Recombinant phage were prepared as described in Chapter 2. Fifty μl of the phage lysate from each clone was dot blotted onto a nylon membrane and hybridized to nick translated genomic DNA from *O. hornorum*. Clones that hybridized strongly to the genomic DNA were assumed to contain repetitive DNA elements. One of the clones, Oho750#1, was chosen for further analysis. The Oho750#1 clone was labelled by primer extension and hybridized to a blot of *Hin*dIII digested DNA from various tilapiine species (Fig. 4.1). The autoradiograph revealed three bands that hybridized to Oho750#1 of approximately 750 bp, 800 bp and 900 bp as well as a cluster of bands around 1900 bp.

Sequence Analysis of the Oho750#1 Clone

The RF of the Oho750#1 clone was purified on a CsCl density gradient and used for restriction analysis. Enzymes

Figure 4.1 Southern hybridization of radiolabelled clone Oho750#1 to *Hin*dIII-digested tilapiine genomic DNA from: *O. mossambicus/hornorum* (O. moss./horn.), *S. galilaeus* (S. gal), *O. mossambicus* (O. moss), *O. placidus* (O. placi), *O. niloticus* (O. nil), *O. aureus* (O. aur), and *O. hornorum* (O. horn). Molecular size markers in Kbp are given at the left of the figure.

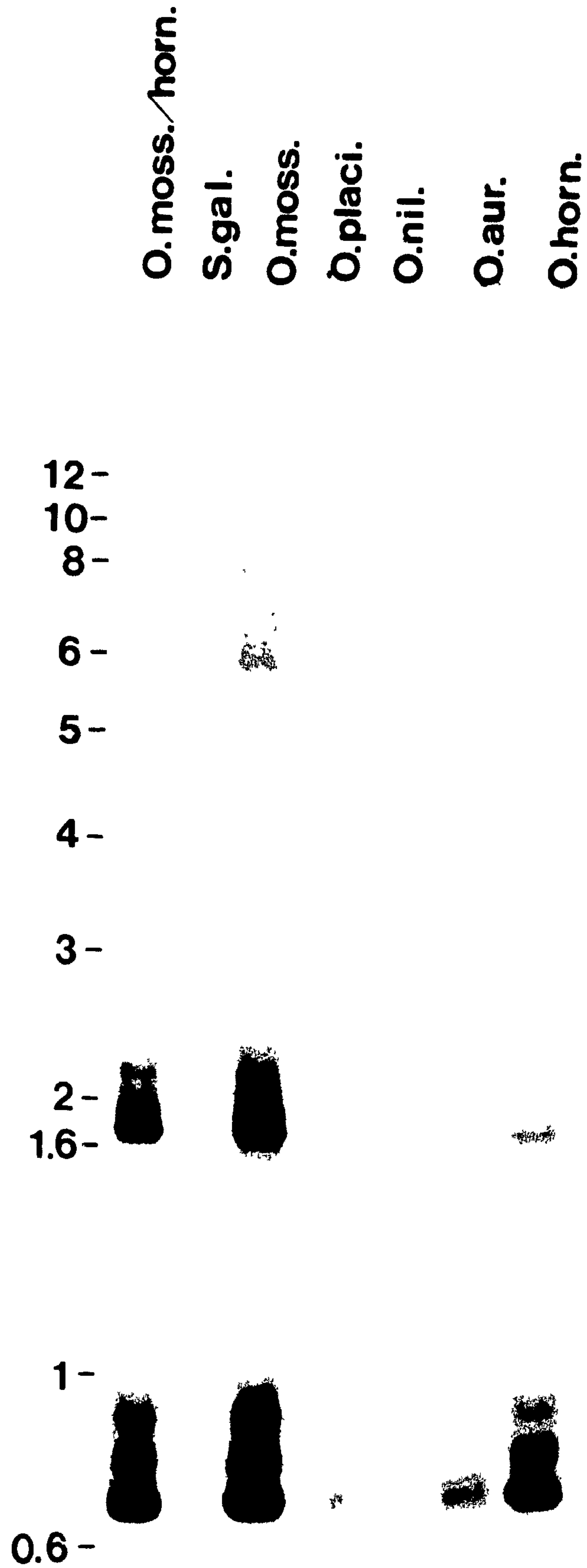


Figure 4.1

were chosen that cleave within the M13mp19 polylinker region to serve as landmarks for the restriction analysis. Two restriction enzymes, *Acc*I and *Eco*RI, each cut once within the 760 bp insert, and were exploited for the sequencing strategy of the Oho750#1 clone. The RF was cut with either *Acc*I or *Eco*RI and the M13 vector was re-ligated on itself to generate nested deletion derivatives of the clone (Fig. 4.2). The sequence information obtained from the Oho750#1 clone and the two deletion derivatives were merged to determine the complete sequence of the Oho750#1 insert DNA (Fig. 4.3).

Distinctive features of the Oho750#1 sequence are the presence of a trinucleotide AAT simple sequence repeat present from nucleotide position 371 to 409; and two imperfect 21 bp repeats at positions 220 to 241 and 564 to 583.

Genomic Organization of Oho750#1 Repeat

To determine the genomic organization of the Oho750#1 repetitive sequence, a partial digestion of genomic DNA from *O. niloticus* was performed with the restriction endonucleases *Hin*dIII and *Eco*RI (recognition site at position 508 in Oho750#1). Five μg aliquots of *O. niloticus* genomic DNA, digested with decreasing quantities of each restriction enzyme were fractionated on a 1.0% agarose gel.

Figure 4.2 Restriction map and sequencing strategy for Oho750#1 clone. The insert is represented by the hatched area and is flanked by the two *Hin*dIII restriction endonuclease recognition sites. The M13 clone was sequenced in three segments by digesting the replicative form (RF) with the restriction endonucleases *Acc*I and *Eco*RI and religating the clone on itself. This created two nested deletions starting from the internal *Acc*I (A) and *Eco*RI (E) restriction sites.

Oho750#1

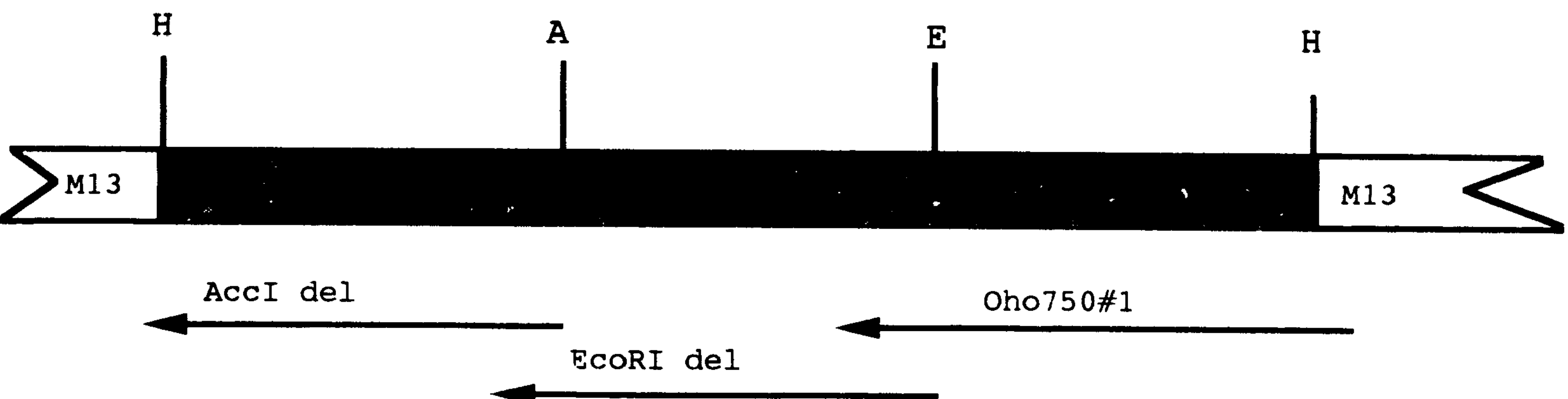


Figure 4.2

Figure 4.3 Nucleotide sequence of the Oho750#1 clone. Restriction endonuclease recognition sites are indicated above the sequence. Arrows indicate the position of the two imperfect direct repeats. A trinucleotide AAT simple sequence repeat is shown in bold face.

HindIII

1 AAGCTTGTAG CAGTAAATTC CAAGTAGGAG TTCGCAAAAG TTCGAGGCAT GGAAATGGCA 60

61 AAATTAGAGC CAAAATGTCA CCTTCACTTC CAAATGGCGG ACTTCCTGTT CGGTTTAGGA 120

121 CATGGCCATA ATAGACTTTT TTGTGCGTCT CCGAACGTTA GATATGTGTT CCCAGTTTTA 180

181 TACATGTAGG TCATACCCAT CTCGGGGGCT CGCGTTTCTC ATTTTTCTAG GTGGCCCTAC 240

241 TGAGCCAATT TTCCCTGGAC ATGTCTACAG ACATTAAAAT ACGTAAATTT TCACCACACC 300

301 TGACGCGTGT GCAAAATTTC ATGAGTTTTT GAGCATGTTT AGGCCCTCAA AAGGCCCATT 360

361 CATTTGCCGA **AATAATAATA ATAATAATAA TAATAATAAT AATAATAATT** ATACGCAAGC 420

421 AGCGATATGA GGGCCCTCGC ACCCCCAGCG GCGTCGAGCC AAGCCCAACA CCTAAAACCA 480

EcoRI

481 GCGCTTCCGA TCTCATGTCA CATTGATGAA TTCATGCATT AACCACCAGC AAAATTTCAT 510

541 CTCATTCAAC CATTATTATA GTCGTCCACT AGGTGGCGCT CTAACCATTA CTGACAAATG 600

601 GCATAACAAA CATTTCCGAG CTCGAGTCTC ATCACGCCTG CGACCTTTGG CAGACATTCG 660

661 ACATTGTGTT TTTGAGCGAC AGCAGGTTAC TGCTTTTTGG CGAGTGATTG AAACTCCACG 720

HindIII

721 TGCCGCCATG ACCACCCCGT TTCCCTGAAC GTAAAAGCTT

Figure 4.3

The blot of the gel was hybridized to the primer extension radiolabelled Oho750#1 clone (Fig. 4.4). High concentrations of *Hin*dIII released fragments of approximately 750 bp and 1600 bp that hybridized to the Oho750#1 repeat (lanes 1-4). Successively lower concentrations of enzyme produced hybridizing bands of approximately 950 bp and 1900 bp and a ladder of bands based on multiples of 950 bp (lanes 5-9). High concentrations of *Eco*RI produced a hybridizing band of approximately 1900 bp (lanes 1-6). Successively lower concentrations of *Eco*RI revealed a ladder of hybridizing fragments based on multiples of 1900 bp (lanes 7-9). The generation of the integral ladders with the two enzymes indicates an underlying tandem organization of the Oho750#1 repeat. The results of this experiment suggest that the Oho750#1 repeat may represent an internal region of a larger repeat of approximately 1900 bp defined by the restriction endonuclease *Eco*RI.

Cloning of 1900 bp Repetitive Sequence from *O. niloticus*

To establish whether the Oho750#1 clone represented an internal region of a larger repetitive element, a 1900 bp *Eco*RI fragment was cloned. Genomic DNA from *O. niloticus* was digested with *Eco*RI and fractionated on a 1.0% agarose gel. Genomic DNA from *O. niloticus* was used as this species was most readily accessible at Dalhousie for analysis of large

Figure 4.4 Southern hybridization of radiolabelled Oho750#1 clone to *Hin*dIII and *Eco*RI partial digests of *O. niloticus* genomic DNA. Aliquots of *O. niloticus* DNA (5 μg) were digested with decreasing quantities of enzyme. The DNA aliquots were digested with (lanes from left to right) 75, 37.5, 18.75, 7.5, 3.75, 1.88, 0.75, 0.38, and 0.19 units for 3 h at 37°C. Molecular size markers in Kbp are given at the left of the figure.

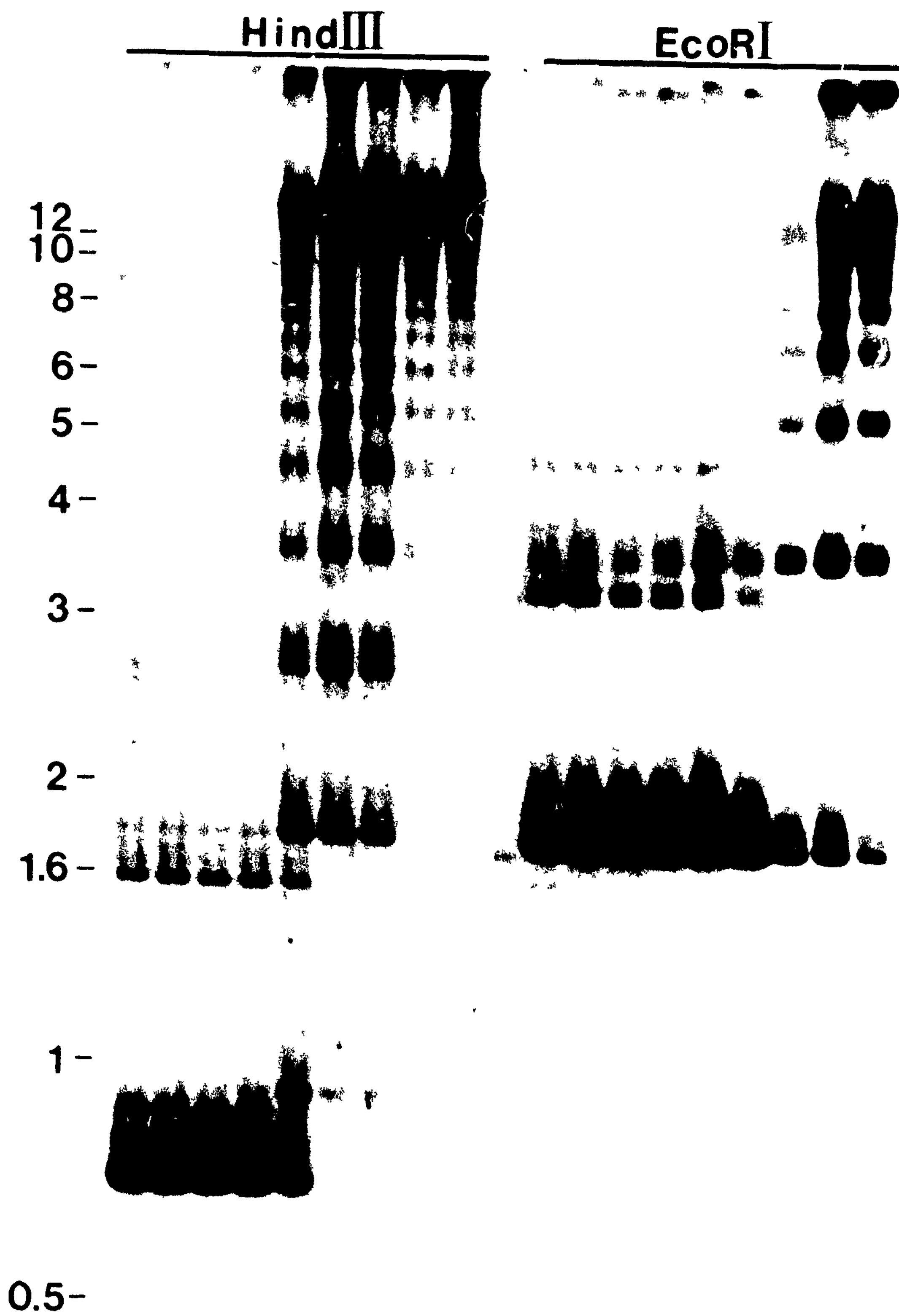


Figure 4.4

sample sizes. An intensely staining band of approximately 1900 bp was eluted from the gel and ligated into the *Eco*RI site of M13mp18. Fifteen recombinant phage were isolated, from which minipreparations of the RF were prepared. Aliquots of the RF minipreparations were digested with *Eco*RI and fractionated on a 1.0% agarose gel. Six of the clones contained 1900 bp inserts. Sequencing of the six clones found that three contained inserts of similar sequence to clone Oho750#1 (>80% identity). A large scale cesium chloride gradient preparation of the RF was performed for clone Oni1900#8. A nested series of deletions were generated from the RF by unidirectional digestion with exonuclease III to facilitate sequencing of the entire insert DNA. To obtain complete sequence from both strands, the orientation of the insert was reversed by subcloning the insert in the opposite direction into a M13mp18 vector (clone Oni1900#8d). Regions of the insert that were not spanned by the nested deletion series were sequenced using synthesized oligonucleotide primers (Fig. 4.5).

The insert of clone Oni1900#8 is 1904 bp in length and contains three *Hin*dIII recognition sites (Fig. 4.6). Two of the *Hin*dIII sites define a 763 bp sequence that possesses 85% sequence identity to the clone Oho750#1 repeat sequence. As suggested by the results of the partial digestion experiment the clone Oho750#1 is an internal component of

Figure 4.5 Restriction map and sequencing strategy for the Oni1900#8 clone. The insert is represented by the hatched area and is flanked by the two *Eco*RI endonuclease recognition sites. The insert DNA was sequenced using nested deletion derivatives generated by exonuclease III, or with restriction endonucleases with internal recognition sites such as *Hin*dIII (H) and *Pst*I (P). Regions not spanned by the deletion derivatives were sequenced using synthesized oligonucleotide primers (primer 381, primer 681).

Oni1900#8

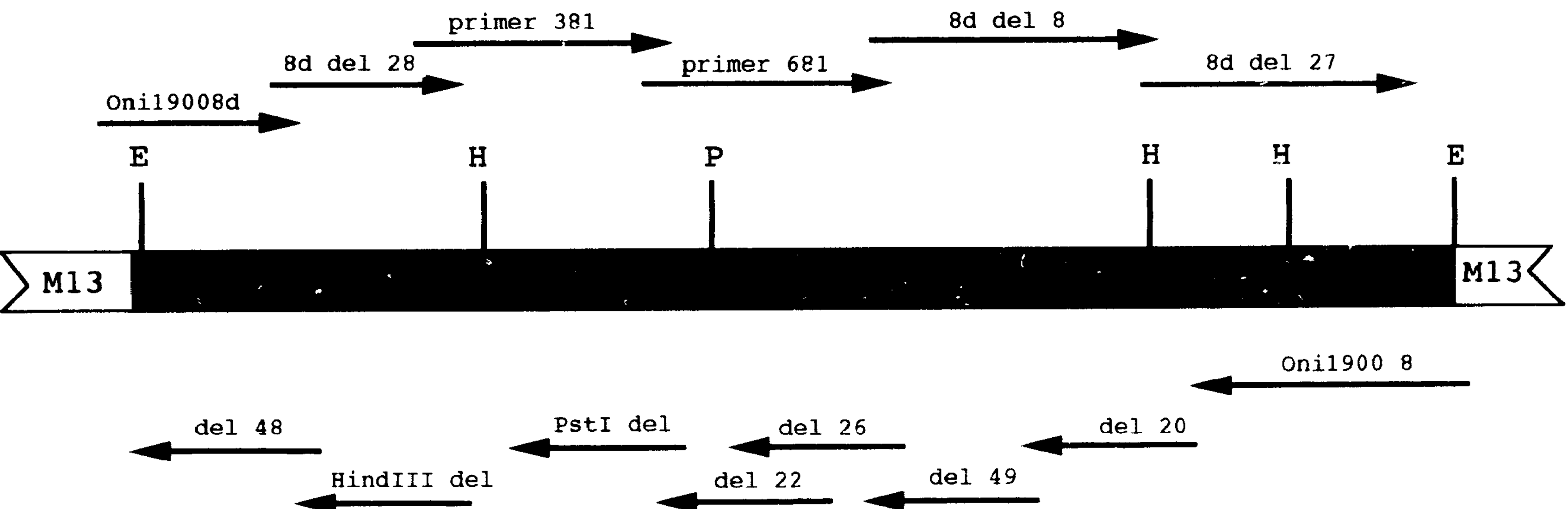


Figure 4.5

Figure 4.6 Nucleotide sequence of clone Oni1900#8. Restriction endonuclease recognition sites are displayed above the sequence. Arrows above the sequence indicate the position of the four imperfect direct repeats in Oni1900#8. The trinucleotide AAT simple repeat sequence is shown in bold face.

```
      EcoRI
1     AATTCACGCA TTTAACCACC AGCTAAAATT TCATCTCATT CAACCATTAT TATAGTCGTC    60
61    CCACTAGGTG GCGCTCTAAC CATTACTGAC AAATGGCATA ACAAACATTT CCGAGCTCGA   120
121   GTCTCATCAC GCCTGCGACC TTTAGTGAGA GATTGGACAT TGTGTTTTTG AGCGACAGCA   180
181   GGTTACTGCT TTTTGGCCAG TGATTGAAAC TCCACGTGCC GCCATGACCA CCCCGTTTCC   240
                 HindIII
241   CTGAACGTAA AAAGCTTCGC AATTTAACAT CACAATGGTC TTTAGATTCC ACACACAGGA   300
301   GATCGCGTAA ATCTAATGAA ATCCCTTGGA GGAGTTCGTC AAAGTATGAG GCCTGAAAAG   360
361   AGGGGAAAAT GTCGCCAAAT TTACACATTA ATTTTAAAAT GGCTGACTTC CTGTTGGATT   420
                                                         HindIII
421   TGGGATATTG CTCCAAGAGA CGTTTTTGTA CATCTTGATA TCTCCAATAA GCTTGCCAGG   480
481   TTTCAAACTC ATACATAAAA CACAGAGCAG GGGCTGTTGT TTTGAAATTT TGTTGGGGGC   540
541   GCTGTGGAGC CATTTTGCGC TGTTCAAGGA AAATGAACAT ATAAAAAGAA ATCCCTCATC   600
601   ACATTAGAGG TGTGCGCCAA ATTTCAAGAC TTTTTAAACT TTTCAAGCCC CTCAAAAGCC   660
661   ACTTCATCTT TCATGGTGAA CTGCGTTGCC ACCAGGGTGC GCCGCCGTTC AGTTTATAGT   720
721   CACAATTTTC TCACTGAAGC ATCAAGAGGG ACTGATGGTG ATATTCACCG CTTTTGAGGT   780
781   GGCCACGATG AACCTGTGAA AATCAGTACA ACAAAATGAA AGACATGACA TTTCCTGGTG   840
841   CCACTAGGTG GCGCTCTACG TATTCCTGAC AATGGGCATA TCAATCTGTT CAGGGCGGGT   900
901   CTGACATCAT CCCTGTAAAA TCTGGTACTG ATCACATTGG ATTTCATTGA GTTACACTAA   960
961   TTTGTTTCTT CATGGCAAGA CCTCAATGTT CGCCATGCTG CTGGGGTCAC ACCCTTCAGC  1020
1021  GAAAACTCAC AGTTTTCACT GTAACCCAAG ACCACTCCAA CTCCTTAAGG CTTTCCTGGA  1080
1081  GAAATTTGAG GCTGCAGATG TCACCCATTA CAATACTGTA CCCCAAAGTG TAAAACATGA  1140
1141  CATTTCCTGT TCCCACTAGG TGGCGCTGTC CCTGATGTCA AATATGGCAG TTGAAATATG  1200
1201  TTCAGGGGTG GAGCCTTATG ATATGTGTGT CCACTTTGGT CAAGGTCGGA CAATGTATGA  1260
1261  CATAACGAGA GGCAATAAGA TTTTCATGGC GAGTCATCGA AATTCGCCGT GGGCCCACGG  1320
1321  CCTCACCGTA TCCCGAAAAC TCAAAAGCTC CGCAATTTAA CATGGCCCAG GTGTGTAGTT  1380
                            HindIII
1381  GACACGTGAC CAAATATGAA GCTTGGTACG ATGTAAATTC CAATGAGGAG TTCGCAAAAG  1440
1441  TTTGAGGCAT GGAAATGGCA AAATTAGAGC CAAAATGTCA CCTTCACTTC CAAATGGCGG  1500
1501  ACTTCCTGTT GGGTCTGCGT CAATGGTCCC ACTGACTTTT TTGTTCGTCT TGGCATGATA  1560
1561  CACATGTGTG CCAAATTTCA TACATGTAGC TCAAACGATG TGTTCGTAGG GCTGCATTTA  1620
1621  ACAAGGCATA GGTGGCGCTA CAGAGCCATT TCCCAGTGCT CATATGTAAA ACCATTAAAA  1680
1681  TACAAAATTT TTCACCAGAC CTGGCATGTG TGCAAAATTT CATGAGTTTT TGAGCATGTT  1740
1741  AAAGCCCTCA AAAAGGCCCT TGTTTTGCCT GAATAATAAT AATAATAATA ATAATAATAA  1800
1801  TAATTAAAGC TGCAAGCAGC GTTATGAGGG CCCTCGCACC CCGGCACTGC GGGCCACGCA  1860
                                                   EcoRI
1861  CAACACCTAC AACCAGCACG TCCGATCTGA TGTCACATTG ATGG
```

Figure 4.6

the 1900 bp repeat. The fact that the homologous 763 bp fragment within the Oni1900#8 clone is contiguous from the 3′ to the 5′ end of the clone confirms that the 1900 bp sequence represents the basic monomer repeat within a tandem array. The simple trinucleotide repeat sequence is also present within this internal *Hin*dIII repeat. There are 11 reiterations of the AAT sequence motif, whereas there are 13 repeats within the Oho750#1 clone. The polymorphism is probably the result of slipped-strand mispairing or unequal crossing over events (Levinson and Gutman 1987; Tartof 1975). As in clone Oho750#1, 21 bp imperfect direct repeats were found in clone Oni1900#8. Four imperfect direct repeats were found at positions 58 to 78, 838 to 858, 1149 to 1170, and 1625 through 1642. These four direct repeat sequences show high sequence identity to the two sequences identified within the Oho750#1 clone. A schematic representation of tandem array of the Oni1900#8 repeat (Fig. 4.7) illustrates the relationship of the internal Oho750#1 clone to the larger repeat. The repeat sequence from both the Oho750#1 clone and the Oni1900#8 clone have been aligned to illustrate the level of sequence identity (Fig. 4.8). A core region within the 21 bp repeat sequences reveals a perfectly conserved 11 nucleotide stretch.

Figure 4.7 Genomic organization of the 1900 bp tandem repeat. The relationship between the *Hin*dIII repeat (Oho750#1) and *Eco*RI repeat (Oni1900#8) sequences is shown. The position of the *Eco*RI (E) and *Hin*dIII (H) restriction endonuclease recognition sites are indicated above each sequence. The tandem organization of the 1900 bp *Eco*RI repeats is shown with an exploded view of the oni1900#8 sequence. The Oho750#1 repeat is aligned under Oni1900#8 to illustrate the contiguous nature of the internal sequence within the larger repeat array. The numbers beneath the Oho750#1 schematic indicate the relative nucleotide position within the sequence. The direct repeat sequences are indicated by arrows overhead blank regions within each sequence.

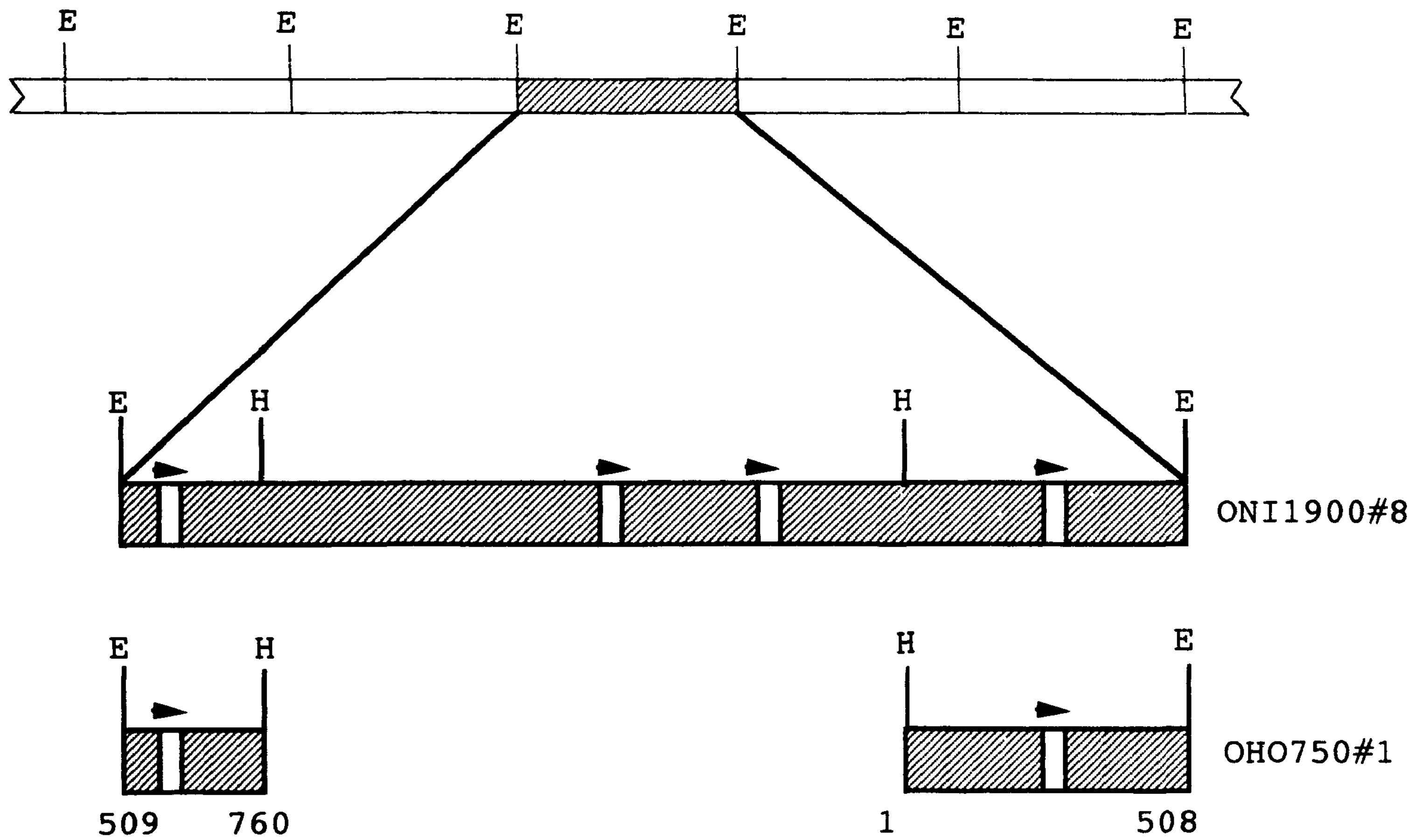


Figure 4.7

Figure 4.8 Alignment of internal direct repeat sequences for SATB repeat. The numbers flanking each sequence indicate the positions of the first and last nucleotide in the Oni1900#8 and Oho750#1 clones. The bold nucleotides highlight an 11 bp core region that shows 100% sequence identity between each repeat.

DIRECT REPEATS

ONI1900#8	58	GTCCC AC**TAGGTGGCGCT**CTA	78
ONI1900#8	838	GTGCC AC**TAGGTGGCGCT**CTA	858
ONI1900#8	1149	GTTCCCAC**TAGGTGGCGCT**GTC	1170
ONI1900#8	1625	G GC A **TAGGTGGCGCT**ACA	1642
OHO750#1	220	CATTTTTC**TAGGTGGCGCT**ACT	241
OHO750#1	564	GT CC AC**TAGGTGGCGCT**CTA	583

Figure 4.8

Copy Number Estimation and Species Distribution

To estimate the copy number of the SATB repeat within the genome and establish the distribution of the repeat sequence within the cichlid family of fishes a slot blot was prepared using genomic DNA samples from several cichlid species as well as two non-cichlid fishes, Rainbow trout (*Oncorhynchus mykiss*), and haddock (*Melanogrammus aegelfinus*). The DNA samples included from outside the tilapiine tribe were from cichlids native to east Africa; *Haplochromis (Protomelas) similis*), *Haplochromis moori*, *Melanochromis auratus*, west Africa; *Hemichromis bimaculatus* and *Pelvicachromis pulcher*, Asia; (*Etroplus maculatus*), and Central and South America (*Cichlasoma meeki*). The radiolabelled insert DNA from clone Oho750#1 hybridized to the genomic DNA of all the African cichlid species as well as the South American *C. meeki* species (Fig. 4.9). The probe did not hybridize to genomic DNA from the Asian representative, *E. maculatus* nor to the rainbow trout or haddock genomic DNAs. The copy number for each species was estimated by densitometric scanning of the autoradiograph using the Oho750#1 clone as a standard and assuming a haploid DNA content for the African cichlid species of 1.0 pg (Majumdar and McAndrew 1986) (Table 4.1). The copy number estimate for the *C. meeki* species was calculated assuming a haploid DNA content of 1.4 pg (Hinegardner and Rosen 1972).

Figure 4.9 Slot blot hybridization of radiolabelled insert of Oho750#1 clone to graded amounts of denatured genomic DNA samples. 500, 50, and 5 ng aliquots of the genomic DNA and 50, 5, and 0.5 ng aliquots of the Oho750#1 denatured M13 replicative form (RF) were applied to the nylon membrane. Tilapiine species included are: *O. hornorum* (O. horn); *O. niloticus* (O. nil); *O. placidus* (O. placi); *O. mossambicus* (O. moss); *O. andersonii* (O. ander); *O. mortimeri* (O. mort); *O. aureus* (O. aur); *S. galilaeus* (S. gal); *T. rendalli* (T. rend); *T. zillii* (T. zillii); *T. mariae* (T. mar) and *T. tholloni* (T. tholl). The three haplochromine cichlid samples included are: *Haplochromis similis* (H. sim); *H. moori* (H. moor) and *Melanochromis auratus* (M. aur). *Hemichromis bimaculatus* (H. bimac) is a west African hemichromine species; *Etroplus maculatus* (E. mac) an Asian cichlid; *Pelvicachromis pulcher* (P. pulch) is a west African chromidotilapiine species, and *Cichlasoma meeki* (C. meeki) is a South American cichlasomine species. Genomic DNA samples from rainbow trout, *Oncorynchus mykiss* (O. mykiss) and haddock, *Melanogrammus aegelfinus* (M. aegel) were also included on the blot.

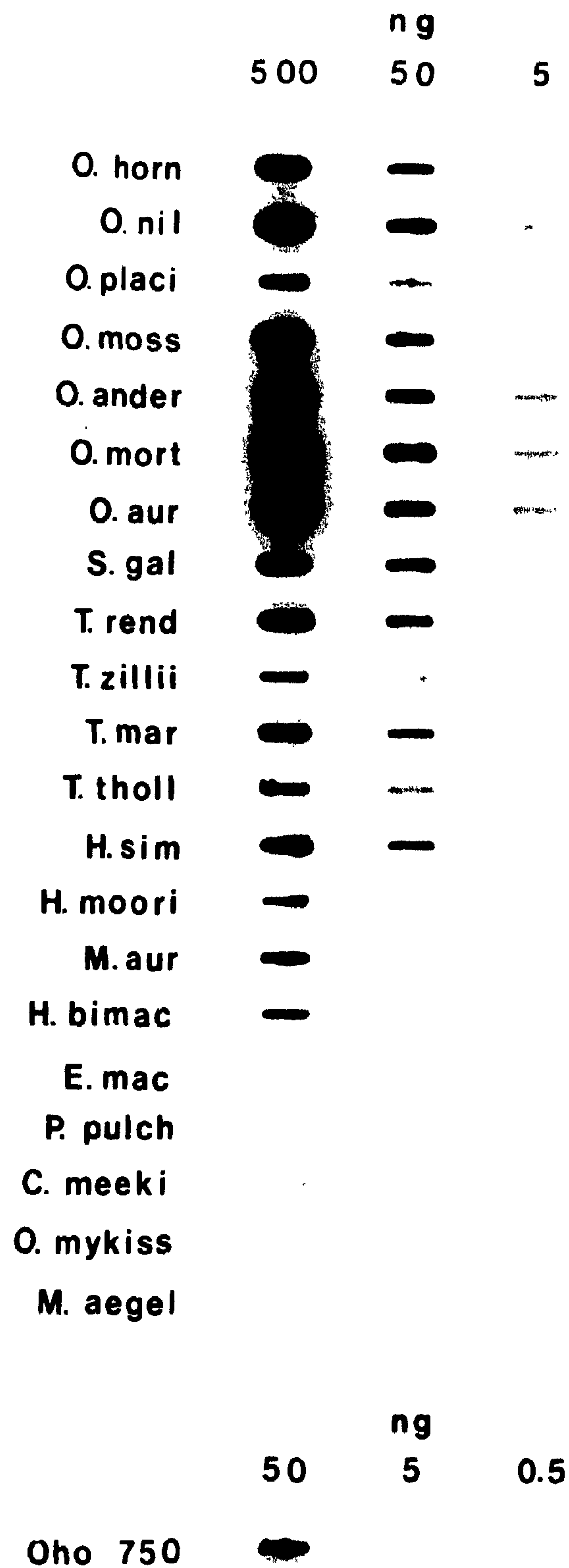


Figure 4.9

Table 4.1: SATB copy number estimation for selected cichlid species.

Species	Copies/Haploid Genome	Genome Equivalent (%)
Oreochromis hornorum	6.2×10^3	1.3
Oreochromis niloticus	8.8×10^3	1.8
Oreochromis placidus	2.9×10^3	0.6
Oreochromis mossamibicus	7.7×10^3	1.6
Oreochromis andersonii	8.4×10^3	1.8
Oreochromis mortimeri	1.1×10^4	2.3
Oreochromis aureus	8.7×10^3	1.8
Sarotherodon galilaeus	6.1×10^3	1.3
Tilapia rendalli	4.6×10^3	1.0
Tilapia zillii	1.1×10^3	0.2
Tilapia mariae	2.8×10^3	0.5
Tilapia tholloni	1.2×10^3	0.2
Haplochromis similis	2.4×10^3	0.4
Melanochromis moori	6.3×10^2	0.1
Haplochromis auratus	5.9×10^2	0.1
Hemichromis bimaculatus	6.1×10^2	0.1
Pelvichachromis pulcher	5.0×10^1	0.01
Cichlasoma meeki	6.0×10^1	0.009

Copy number estimation of SATB repeat based on densitometry of slot blot hybridization experiment (Fig. 4.9). Copy number for all species is estimated based on a haploid DNA content of 1.0 pg (Majumdar and McAndrew 1986), except for *C. meeki* with a haploid DNA content of 1.4 pg (Hinegardner and Rosen 1972)

The copy number varied from only 50 copies for the chromidotilapiine, *P. pulcher*, to 11,000 copies for the tilapiine species, *O. mortimeri*. The stringency conditions for both the hybridization and washes should prevent hybridization of the Oho750#1 SATB probe to sequences with less than 80 to 85% sequence identity. The copy number is therefore a conservative estimation.

CHAPTER 5

DISCUSSION

The SATA and SATB satellite DNAs represent two discrete families of repeats within the genomes of the cichlid fishes. The distribution of the SATA sequence is limited to the tilapiine and haplochromine lineages while the SATB repeat is distributed more widely and was detected in all of the African cichlids and in the genome of the neotropical cichlid *Cichlasoma meeki*. Conservative estimates of the copy number for the SATA repeat range from 100 copies per haploid genome for *Melanochromis moori* to 59,000 copies per haploid genome for *Sarotherodon galilaeus*, equivalent to 0.03% and 1.6% respectively of the total genomic DNA. The SATB repeat ranges from an estimate of 50 copies for *Pelvicachromis pulcher* to 11,000 copies for *Oreochromis mortimeri* per haploid genome, equivalent to 0.01% and 2.3% respectively of the total genomic DNA. The copy number calculations represent conservative estimates since sequences that have less than ~80% sequence identity do not hybridize to the satellite probe.

The SATA Satellite DNA Family

The first representative of the SATA family to be described was a 237 bp sequence cloned from the hybrid tilapiine strain, *Oreochromis mossambicus/hornorum* (Wright

1989). In my research I have endeavoured to more fully characterize the organization and distribution of this satellite DNA family in the tilapiine tribe of fishes. *Eco*RI-digestion of genomic DNA from seven tilapiine species in addition to the *O. mossambicus/hornorum* strain with representatives from the three major genera, revealed three size-variants of the SATA repeat (Fig. 3.1). The size-variants have apparently arisen from two deletion events of 9 bp (II) and 29 bp (III) from the ancestral repeat of ~237 bp (I). The size-variants are differentially distributed with some species possessing more than one of the size-variants *e.g. Oreochromis hornorum*, *Oreochromis aureus*, and *Oreochromis placidus* (Table 3.1). A survey of multiple individuals for each species revealed that the satellite DNA profiles are consistent. The type I and type III size-variants detected in the digest of the genomic DNA of *O. aureus* were disproportionate (Fig. 3.1). The type III size-variant was present in detectably higher quantities than the type I size-variant, which may be attributed to the loss of the *Eco*RI recognition sites from the majority of the type I size-variants. Alternatively, the observation could be explained by selective amplification of the type III size-variant due to stochastic forces. Selective amplification of satellite DNA size-variants was observed previously by Fry and Salser (1977) for which they proposed the "library

hypothesis" model. This model states that every species possesses each of the size-variants (library) which are not detectable as repetitive sequences until they are amplified.

The SATA sequence of the haplochromine cichlid *Haplochromis similis* was isolated by digestion of the genomic DNA with the restriction endonuclease *Hin*fI. The consensus monomer sequence for this species does not possess the restriction endonuclease recognition sites for *Pal*I or *Eco*RI. A survey of genomic DNAs from a variety of old world and new world cichlid species revealed that the SATA sequence is confined to the haplochromine and tilapiine African cichlid tribes. The origin of the SATA repeat therefore occurred after the Gondwanan separation and most likely after the division of India from Africa 65 million years before present (BP) (Starr and Taggart 1987). Interestingly, the copy number for the three haplochromine species was significantly lower than for most of the tilapiine species examined. The hybridization conditions were such that only sequences with greater than 75% to 80% sequence identity hybridized to the tilapiine (OplI-5) SATA sequence. Pairwise comparison of the derived consensus sequences from the tilapiine fishes to the *Haplochromis similis* SATA sequence calculated the sequence identity to the *Oreochromis placidus* (size-variant I) as 78.3% (Table 3.2). If the two other haplochromine fishes included for the

slot blot hybridization, *Melanochromis moori* and *Haplochromis auratus*, are assumed to possess SATA sequences with similar sequence identity then this estimate of copy number is supported. It appears therefore, that the major amplification event for the SATA family occurred following the divergence of the tilapiine and haplochromine tribes which conservative estimates place at ten million years ago (Sage *et al.* 1984). This proposal is supported by the copy number estimation that revealed an inverse relationship of copy number to estimated elapsed time since the last major amplification event (Table 3.5). The two species with the greatest elapsed time estimates of 11.7 mya for the substrate spawning species *T. zillii* and *H. similis*, both had copy number estimates of less than 10,000 copies. A similar observation has been made for the light satellite DNA component of *Mus musculus* with an ancestrally related species possessing a significantly lower copy number than more recently diverged species (Dod *et al.* 1989). The calculations for elapsed time since a major amplification event are based on a directional substitution rate (AT-GC) for cold blooded vertebrates of 1.55×10^{-9} per site per year (Bernardi and Bernardi 1990). The calculation for elapsed time is based on the assumption that the amplification process effectively homogenizes the monomer sequences, which is followed by the process of random

genetic drift. The longer the elapsed time since the amplification event, the greater the level of inter-monomer divergence.

Species-specificity of the SATA repeat

Polymorphic patterns revealed after long exposures of hybridizations of the SATA probe to tilapiine DNA samples digested with the restriction endonuclease *Pal*I imply a higher order organization for the satellite DNA array (Fig. 3.9). Species-specific hybridization patterns are observed in the genomes of several other organisms (Arnason *et al.* 1988; Hoang-Tang *et al.* 1991; Quinn *et al.* 1992; Bachman *et al.* 1992). A satellite DNA repeat, pMAT, cloned from the passerine bird, *Molothrus ater*, revealed species-specific patterns in hybridization to genomic DNA of related avian species (Quinn *et al.* 1992). In this study it was assumed, as may be the case for the SATA sequences, that the periodicity of repeat restriction ladders, the complexity of the ladder, size distribution and relative intensity are markers of phylogenetic relatedness. Although the same degree of complexity in the polymorphic patterns was found in different strains of *O. niloticus* (maintained at Dalhousie University and The University of Stirling), it is not possible to determine whether these values indicate an intrinsic feature of the genomic organization of the

satellite DNA in each species, or whether they simply reflect the level of inbreeding that may have occurred during the maintenance of these strains in captivity.Species-specific restriction patterns have also been made for the felid FA-SAT sequence (Fanning *et al.* 1988), a highly repetitive DNA component of the Baleen whale genera (Arnason *et al.* 1988) and for a satellite repeat cloned from a species of the sorghum plant (Hoang-Tang *et al.* 1991).

Higher order organization

The existence of higher order domains for satellite DNA sequences has previously been documented primarily in relation to the primate alpha satellite DNA family (reviewed in Willard 1989). The first indication of a hierarchical organization for the tilapiine SATA repeats came from the observation of restriction enzyme defined multimers of the satellite DNA sequence of the *O. mossambicus/hornorum* hybrid tilapiine species (Wright 1989). The high molecular weight patterns observed in the *PalI* digested tilapiine genomic DNA samples provides further evidence for the higher order organization of the SATA monomer sequences. It is noteworthy that many of the polymorphic bands observed in the species blot (Fig. 3.9) did not co-migrate with integral multiples of the satellite monomers generated by partial digestion of

genomic DNA (data not shown). Three models are proposed to explain the generation of the non-integer and integer higher order SATA repeats (Fig. 5.1). First, the *Pal*I polymorphism may reside in sequences within DNA adjacent to the satellite tandem array that is either satellite DNA of a sequence different from that characterized here, or unique single copy DNA that represents the boundary of the satellite array. Second, the generation of non-integer polymorphic fragments may be due to the loss or gain of the *Pal*I recognition site near the end of the satellite array. This idea is consistent with the observations that monomer sequences become more heterogeneous towards the ends of a satellite array as they are less prone to the homogenizing effects of unequal crossing-over (Brutlag 1980). Third, the polymorphic satellite DNA fragments may be "orphans", short tandem arrays of the satellite sequence embedded in unique DNA that were created by transposition and saltatory amplification of the satellite monomer (Brutlag 1980).

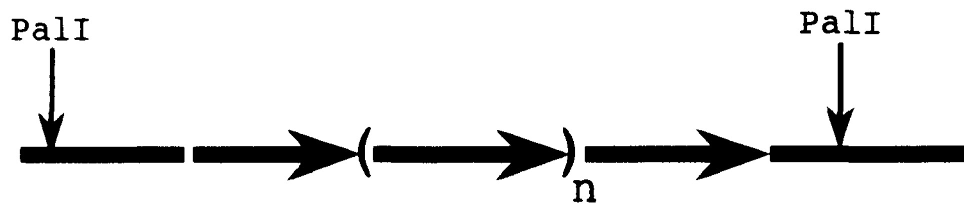
Concerted evolution and phylogenetic implications of the SATA repeat

The species-specific hybridization patterns indicate that the satellite DNAs may provide useful information for phylogenetic analyses of the tilapiine fishes. The tilapiine SATA sequences also exhibits the effects of concerted

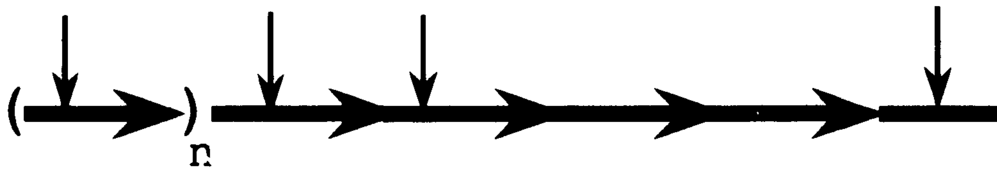
Figure 5.1 Schematic models of proposed mechanisms for the generation of high molecular weight polymorphic SATA fragments.

Non-Integer Fragments

Isolated Blocks of Satellite DNA



Termini of Satellite Arrays



Integer Fragments

Internal Polymorphism

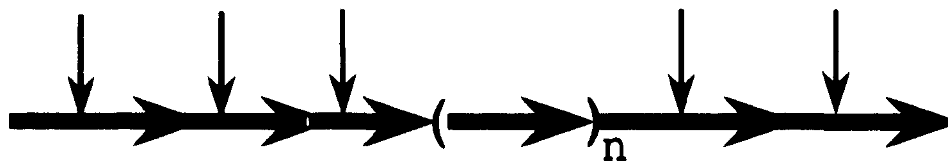


Figure 5.1

evolutionary processes. A pairwise comparison of the individual cloned sequences from different species revealed the species-specific fixation of nucleotides in the alignments. The frequency of the species-specific fixed nucleotide positions ranged from 0.5% for the intrageneric *O. niloticus* III vs. *O. placidus* I comparison to 14.2% for the intergeneric comparison of *O. niloticus* III to the *H. similis* SATA sequence (Table 3.4). The frequency of the concerted nucleotide positions therefore increases with increasing phylogenetic distance.

Two different tree building methodologies were used to test the phylogenetic utility of the SATA sequences. Both the DNA parsimony (Fig. 3.11) and Neighbour-Joining (Fig. 3.13) methods resulted in similar tree outputs. The *Sarotherodon* and *Oreochromis* genera were grouped together in a clade distinct from the *Tilapia* genera's SATA sequences. It was not possible to definitively resolve the *Sarotherodon* genus from the *Oreochromis* genera's SATA sequences with these two methods. The validity of the *Oreochromis/Sarotherodon* clade was tested by designing alternative tree topologies and comparing the relative treelength to the original tree. An alternative tree topology that positions the *Sarotherodon* species diverging recently from a substrate spawning ancestor resulted in a treelength significantly longer than the original tree (Fig.

3.12). This alternative tree topology is representative of the model of tilapiine evolution proposed by Peters and Berns (1982) with polyphyletic origins for the mouthbrooding fishes. The biparental mouthbrooding *Sarotherodon* species represents the most recently diverged lineage and the sole maternal mouthbrooders represent a more anciently diverged lineage. The results of the SATA phylogenetic analyses support the monophyletic origin of the mouthbrooding tilapiines proposed by Trewavas (1982) which was supported by the allozyme studies of McAndrew and Majumdar (1984). The inability to definitively resolve the *Sarotherodon* and *Oreochromis* mouthbrooding tilappiine genera suggests that this divergence occurred too recently to be detected in the evolution of the SATA sequences.

SATA sequences: molecular cause or effect of speciation?

Chromosomal studies of Old World cichlids demonstrate a highly conservative level of karyotypic evolution (Kornfield *et al*. 1979; Majumdar and McAndrew 1984). The most distinctive difference is in the distribution of constitutive heterochromatin detected by C-banding. These studies detected the most distinctive differences in comparisons of the *T. zillii* karyotype to those from species in the *Oreochromis* and *Sarotherodon* genera. Ten to twelve of the chromosomes in the *T. zillii* karyotype did not stain at

all for constitutive heterochromatin, and chromosomes that did stain possessed different banding patterns from chromosomes of the mouthbrooding species. All of the chromosomes from the mouthbrooding species possessed constitutive heterochromatin. While the results of these studies were inconclusive for implicating karyotype evolution as a factor for speciation of the tilapiines, they do imply underlying qualitative differences for the repetitive DNA component of the genomes for the mouthbrooding and substrate spawning assemblages. The molecular characterization of the constitutive heterochromatin (satellite DNAs are a major component) reflects the results of the karyotype studies. The SATA sequences of the mouthbrooding species from the *Oreochromis* and *Sarotherodon* genera showed a high level of sequence identity to each other, but both have significantly diverged from the *Tilapia* genera's SATA sequences. This result coupled with the karyotype observations suggests a potential molecular basis for the speciation of the substrate spawning and mouthbrooding lineages. Models of speciation have been proposed that implicate the divergence of non-coding satellite DNAs as influential forces for hybrid dysgenesis (Dover 1982; Rose and Doolittle 1983). The divergent satellite DNAs result in a reduction of meiotic chromosome recognition, described as a pre-zygotic mechanism of species

isolation. Most likely, the observed level of sequence divergence is an effect and not a cause of the major tilapiine tribe divergence between the mouthbrooders and substrate spawners. To implicate the SATA family in a causal relationship for speciation would require an extremely rapid divergence of the sequences. The divergence of the satellite DNAs may serve as an effective mechanism for maintaining the genetic isolation between these two major lineages.

The SATB Satellite DNA Family

The basic monomer repeat, SATB, is defined by the restriction endonuclease *Eco*RI. Partial digestion of genomic DNA by *Eco*RI revealed a ladder of multimeric bands in a Southern hybridization, which indicates a tandem organization for the repeat sequence (Fig. 4.4). The tandem organization was confirmed by sequence analysis. At least ten multimeric bands can be identified in the Southern hybridization of partial *Eco*RI digests of *O. niloticus* genomic DNA which would represent a block of at least 19 kilobases.

Evidence for monophyly of neotropical/African cichlid assemblage

The cichlid fishes are distinguished primarily on the basis of subtle differences in cranial and dental morphology (Trewavas 1983; Stiassny 1991). The monophyly of the neotropical and African cichlid fishes with the exception of the Zairean and Madagascan species is presently supported by two derived morphological characters of the branchial and neurocranial apparati (Stiassny 1991). The potential occurrence of parallel or convergent evolution makes the designation of monophyletic assemblages based solely on morphological traits error prone. The distribution of the SATB repeat, however, appears to support this monophyletic

assemblage. The SATB repeat could be detected by hybridization in genomic DNA samples from all of the African cichlids examined as well as in the genome of the neotropical fish *Cichlasoma meeki*. The origins of the SATB repeat therefore pre-date the separation of the Gondwanan land mass 135 to 181 million years BP (Starr and Taggart 1987).

Based on the hybridization of clone Oho750#1 to genomic DNA, I have estimated that the copy number of SATB varies from as few as 50 copies in *Pelvicachromis pulcher* to 11,000 copies per haploid genome in *Oreochromis mortimeri*. Unlike the SATA repeat which I propose has undergone a single major amplification event following the divergence of the substrate spawning and mouthbrooding strains, the SATB repeat appears to have undergone two major episodes of amplification. The first amplification event would have occurred after the divergence of the haplochromine tribe followed by a subsequent amplification event in the tilapiine tribe.

SATB direct repeat sequences may be conserved by selective pressures

A distinctive feature of the SATB repetitive DNA sequence is the presence of four imperfect direct repeat sequences. Within the direct repeat sequence is a conserved

stretch of eleven nucleotides with 100% sequence identity for all four repeats. Homologous regions from an internal repeat cloned from the species *Oreochromis hornorum* (clone Oho750#1) also possess the eleven nucleotide block. This may represent a functional region of the repetitive sequence that is conserved by selective pressure. Though the question of satellite DNA function within the genome remains enigmatic, studies have implicated satellite DNA sequences as recognition sites for DNA binding proteins. Masumoto and colleagues (1989) have described a 17 bp core region, termed the CENP-B box within the monomer unit of the human alphoid satellite DNA. This sequence apparently serves as a specific binding domain for the CENP-B binding protein. Strauss and Varshavsky (1984) describe a high-mobility-group-like protein that specifically binds to several AT rich sites in human alpha satellite DNA. Recently, evidence has been presented that supports the role for satellite DNA domains in the three dimensional structure and architecture of the chromatin within the cell nucleus. A study following up the observation of Strauss and Varshavsky (1984) indicates that the specific regions of the mouse satellite DNA sequence that bind to the HMG-I protein are those that possess a stable DNA curvature (Radic *et al.* 1992). It is implied that the combination of the specific DNA curvature and the binding of the chromatin-specific proteins are essential for

the condensation of the heterochromatin. The presence of the trinucleotide AAT simple repeat motif may support this structural role for the SATB repeat. Simple sequence repeats are known to adopt altered DNA confirmations such as Z-DNA when subjected to the torsional stress of negative supercoiling (Ellison *et al.* 1986). A microsatellite sequence upstream of the HMG-T gene in rainbow trout adopts a cruciform structure (Wright and Dixon 1988). In addition to their role in binding specific proteins, human alphoid satellite DNA sequences are implicated as the DNA component in the chemically resistant non-histone DNA-polypeptide complexes (Pfutz *et al.* 1992). I speculate that the conserved direct repeat core sequence of the SATB repeat may serve as a protein or polypeptide-binding domain.

Summary and Future Prospects

The SATA and SATB satellite DNA families provide valuable insights into both the molecular evolution of the tilapiine genome and the phylogenetic relationships of the tilapiine tribe of fishes. The SATB repeat is presumed to be the more ancient of the DNA sequences as it could be detected in all of the African cichlids examined as well as the neotropical cichlid representative. This distribution supports the presumed monophyly of the African/neotropical cichlid assemblage. The SATA repeat is more limited in its

distribution as it is only detected in the genomes of the tilapiine and haplochromine fishes. Molecular phylogenetic analyses of the SATA repeat variants cloned from selected tilapiine and haplochromine species challenge the polyphyletic model for the evolution of the tilapiine mouthbrooders. Two different molecular phylogenetic tree building methodologies resulted in the close association of the *Oreochromis* and *Sarotherodon* genera to the exclusion of the *Tilapia* genera's SATA sequences. This relationship strongly argues for the recent divergence of the two mouthbrooding strains from a common ancestor. It is noteworthy that the allozyme phylogenetic studies of McAndrew and Majumdar (1984) support the close relationship of the *Oreochromis* and *Sarotherodon* genera.

The research that I have conducted on the SATA and SATB satellite DNAs indicates several avenues of future investigation. One aspect would be the sequencing of the SATA repeat from species of the haplochromine tribe of fishes to investigate the phylogenetic relationships of this extreme speciose assemblage. Future studies of the SATB satellite DNA could be focused on its proposed role as a protein binding domain and the identification of these specific regions by using DNA footprinting or gel retardation studies.

APPENDIX 1

Consensus Alignments of SATA satellite DNA sequences for: *Oreochromis placidus* (I), *Oreochromis placidus* (II), *Oreochromis niloticus* (III) *Sarotherodon galilaeus* (I), *Tilapia rendalli* (I), *Tilapia zillii* (I), and *Haplochromis similis* (I). The consensus sequence (CON) is aligned over the individual cloned monomer sequences. Dashes (-) indicate that the presence of the same base in the individual sequence as in the consensus, and asterisks (*) indicate a deletion or gap in the alignment. Consensus bases are designated as the base that occurs in greater than 50% of the individual clones at a given position. The restriction endonuclease recognition sites are presented over the consensus sequence.

O. placidus I Alignment (237 bp)

```
                          10        20        30        40        50
           EcoRI     .PalI     .         .     HinfI           .
CON        AATTCTATAAGGCCAAGCCTGAAATATGTGTGTCCGAGTCTTCTATCAAA
OplI 2     --------------------------------*-----------------
OplI 3     --------------------------------------------------
OplI 5     ---------------------------------------CC---------
OplI 7     --------------------------------------------------

                          60        70        80        90       100
                          .         .         .         .         .
CON        AGTTACAGCTGTCTTTATGG*ACCTGG*TGAAAATCGCCTTATTTCGGGC
OplI 2     --------T-----------*--T-AC*----------T--G--------
OplI 3     --------------------*------*--------------------*-
OplI 5     ---------------C----G--T---G-------A---------T----
OplI 7     --------------------*------*---------------------*

                         110       120       130       140       150
                          .         .         .         .         .
CON        GAG*ACAGTGCGTTTCTCG*CTATTACATGCATTTGAATG*AGTTCTCGC
OplI 2     ---*-------------G-*--------------------G---------
OplI 3     ---*---------------*--G-----------------*---G----G
OplI 5     ---G---------------G--------------------*---------
OplI 7     ---*-------------G-*--------------------*-

                         160       170       180       190       200
                          .         .         .         .         .
CON        CTGAAACACATTATGGGT[illegible]TCATTTTGTGAATAACTTGAAAATCTTAGC
OplI 2     ----C---------------------------------------------
OplI 3     ----------------*---------------------------------
OplI 5     --------------------------------------------------
OplI 7     --------------------------------------------------

                         210       220       230       240
                          .         .         .         .  EcoRI
CON        TCAAACAGCTGCAAAACCTATTTCCCCAGCATGGAAATGGTG
OplI 2     ------------------------------------------
OplI 3     ------------------------------------------
OplI 5     ------------------------------------------
OplI 7     ------------------------------------------
```

O. placidus II Alignment (230 bp)

```
                           10        20        30        40        50
                 EcoRI     .   PalI  .         .         HinfI     .
CON              AATTCTATGAA*GGCCAAGGCCTGAAATATGTGTGTCCGAC CTT*CTAT
OplII 3          --------*--*--------*--------------------  --A----
OplII 5          -----C-----A-----------------G--T--------- -**----
OplII 9          -----------*-------*----CCC---T----G----TG---*----
OplII 14         --------*--*-------*---------------*---------*----

                           60        70        80        90       100
                           .         .         .         .         .
CON              CAACAGTTACAGCTGCCTTTATGGAAAATCGCTTTATTTCGGCGAGACAG
OplII 3          ----------------------*---------------------------
OplII 5          ---A---A-**--------*--A-----------------A---------
OplII 9          -------------------- --*--------------------------
OplII 14         -----------------------*--------------------------

                          110       120       130       140       150
                           .         .         .         .         .
CON              TGCGTTTCTGGCTATTACATGCATTTGAATGGAGTTCTCGCCTGAAACAC
OplII 3          --------------------------------------------------
OplII 5          --------------------------------------------------
OplII 9          --------------------------------------------------
OplII 14         -----------------------------------------T--------

                          160       170       180       190       200
                           .         .         .         .         .
CON              ATTATGGGTTTTCATTTTGTGAATAACTTGAAAATCTTAGCTCAAACAGC
OplII 3          ----------------------*--*-------------------*----
OplII 5          --------------------------------------------------
OplII 9          -----------------------------*--------------------
OplII 14         --------------------------------------------------

                          210       220       230
                           .         .         .    EcoRI
CON              TGCAAAACCTATTTCCCCA*GCATGGAAATGCTG
OplII 3          ------**--------********A------*--
OplII 5          ------*------------C--------------
OplII 9          -------------------*--------------
OplII 14         -------------------*--------------
```

O. niloticus III Alignment (209 bp)

```
                10        20        30        40        50
            EcoRI    .PalI     .         .     HinfI     .
CON         AATTCTATAAGGCCAAGCCTGAAATATGTGTGTCC*GAGTCTCCTATCAA
OniIII 3    -----------------------------------*------T
OniIII 13   -----------------------------------*------ --    -
OniIII 14   ---------*--------*----------------C--- - - -
OniIII 15   -----------------------------------*- --- -  -  -

                60        70        80        90        100
                 .         .         .         .         .
CON         AAGTTACAGCTGTCTTTATGGA*CTTGGTGGAAAATCGCCTTATTTCGGC
OniIII 3    ---------------------G-------*---------  ----
OniIII 13   ----------------------*-------------- ------
OniIII 14   -----------------TA--GCA------*-----G---  TA
OniIII 15   -------------------CC-*----------------  ---

                110       120       130       140       150
                 .         .         .         .         .
CON         GAGACAGTGCGTTTCTCGCC*TGAAACACATTATGGGTTTTCA*TTTTGT
OniIII 3    --------------------*-------------------   *
OniIII 13   --------------------*-----------------     *
OniIII 14   --------------------*---------------- -    A
OniIII 15   --------------------C-------------------   *

                160       170       180       190       200
                 .         .         .         .
CON         GAATAACTTGAAAATCTTAGCTCAAACAGCTGCAAAACCTATTTNCCAG
OniIII 3    ---------------------------------- ---       C
OniIII 13   ---------------------------------------      ATT
OniIII 14   ----------------------AC----------- ** -    **
OniIII 15   ---------------------------------------*     *

                210
                 .  EcoRI
CON         CATGGAAATGCTG
OniIII 3    -------------
OniIII 13   -------------
OniIII 14   -------*-----
OniIII 15   ----A--------
```

***Sarotherodon galilaeus* Alignment (236 bp)**

```
                       10        20        30        40        50
             EcoRI     .PalI     .         .    HinfI          .
CON          AATTCTATAAGGCCAAGCCTGAAATATGTGTGTCCGAGTCTTCTATCAAA
SgaI 1       --------------------------------------------------
SgaI 5       ------------------*---------------G---------------
SgaI 7       -------------------------------T------------------
SgaI 9       ------------------*-------------------------------

                       60        70        80        90        100
                        .         .         .         .         .
CON          GTTACAGCTGTCTTTATGAACTTAGTGAAAATCGCTTTATTTCGGCGAGA
SgaI 1       ------------------------------------------T-------
SgaI 5       --------------------------------------------------
SgaI 7       --------------------------------------------------
SgaI 9       --------------C-----------------------------------

                       110       120       130       140       150
                        .         .         .         .         .
CON          CAGTGCGTTTCTCGCTATTACATGCATTTGAATGGAGTTCTCGCCTGAAA
SgaI 1       --------------------------------------------------
SgaI 5       --------------------------------------------------
SgaI 7       --------------------------------------------------
SgaI 9       --------------------------------------------------

                       160       170       180       190       200
                        .         .         .         .         .
CON          CACAT*TATGGGTTTTCATTTTGTGAATAACTTGAAAATCTTAGCTCAAA
SgaI 1       --A-G*----------------A---------------------------
SgaI 5       -----C--------------------------------------------
SgaI 7       -----*--------------------------------------------
SgaI 9       -----*--------------------------------------------

                       210       220       230
                        .         .         .      EcoRI
CON          CAGCTGCAAAACCTATTTCCCCAGCATGGAAATGCTG
SgaI 1       -------------------------------------
SgaI 5       ----------T--AT------*-----A---------
SgaI 7       -------------------------------------
SgaI 9       ---------------------*---------------
```

Tilapia rendalli Alignment (237 bp)

```
                     10         20         30         40
           EcoRI     .          .          .         HinfI
CON        AATTCTANTAAGCAAAGCCTTAAATATCTGTGT*NCGAGTCTTC*TATC
TreI 1     -------*-------------------------C-*--------*  -
TreI 2     -------TACG------*---------------C-----  -  --    -
TreI 8     -------A-------------------------G----C**---    --
TreI 14    -------TA-G------*---------------GC---      AG-   -
TreI 16    -------*-------------------------T-----  -  -     G

           50         60         70         80         90
           .          .          .          .          .
CON        AAAAG*TTACAGCTGTCTTTATGCAGTTAATGAAAATCGCTTTATTTCGC
TreI 1     -------------------*------------------------- G -
TreI 2     ------------------------------------T------ -  CG
TreI 8     -------*---C---*-------------------------- C
TreI 14    -----A----------------------------- -- ---   * G
TreI 16    -------------------------------------------------

           100        110        120        130        140
           .          .          .          .          .
CON        CCAAGACAGTGCGTTTCTCGCTATTACATGCATTTGAATGGAGT*TCTCG
TreI 1     ------------------------------------- - -     -
TreI 2     ------------------------------------- ----
TreI 8     -*-------------------------------------- -   A
TreI 14    ---------------------------------- ---A C
TreI 16    -**------------------------------------*- - -

           150        160        170        180        190
           .          .          .          .          .
CON        CCTGAAACAAAGT*ATGGG*TTTTCATTTTGTGAA*TAACTTGAAAATCT
TreI 1     ------------------C----------------G----     -
TreI 2     ----------------------------------------    -   **
TreI 8     -------G------T--------------------- -  --
TreI 14    -------------C-----C--------------*---  ---
TreI 16    ------------------*-------------------- -

           200        210        220        230        240
           .          .          .          .          .   EcoRI
CON        T*AGCTCAAACAGCTGCAAAACCTATTTCCC*AGCATAGGAAATGCTG
TreI 1     -------------------------------C--  -   C
TreI 2     ----------*---------**----------- ---*-        *
TreI 8     ------------------------------------CG    -
TreI 14    -G-------------------*----------      **
TreI 16    -------------------------------C--  -  --
```

Tilapia zillii Alignment (238 bp)

```
                    10        20        30        40        50
            EcoRI    .         .         .     HinfI         .
CON         AATTCTATAAGGCAAAGCCTTAAATATCTGTGTGCGAGTCTTCTATCAAA
TziI 8      -----------*---*-------*--*--------------*--------
TziI 9      ------------------*--------------C----------------
TziI 10     ------------------*-----------------------------**
TziI 13     ---------**----*---*-------------*-------*--------

                    60        70        80        90       100
                     .         .         .         .         .
CON         *GTTACAGCTGTCTTTATGCAGTTAATGAAAATCGCTTT*CTTTCGCCAA
TziI 8      *--------------C-----------------------*----------
TziI 9      A------------------T-------------------*----------
TziI 10     *******------------A-------------------AT-----G--C
TziI 13     *--*-------*---*------------------A----*---------G

                   110       120       130       140       150
                     .         .         .         .         .
CON         GACAGTGCGTTTCTCGCTATTACATGCATTT*GAA*TGGACTTCTCGCCTG
TziI 8      A---------C--------------------*---*---G----------*
TziI 9      --G----------------------------*---*--------------*
TziI 10     ----C--*T----------------------T---*----G----------
TziI 13     --*---*------------------------*---A---------------

                   160       170       180       190       200
                     .         .         .         .         .
CON         AAACAAAGTATGGGATTTTCATTTTGTGAATAACTTGAAAATCTTAGCTC
TziI 8      ---G----------*-----------------------------------
TziI 9      ------------------CG--C------------*----*---*-----
TziI 10     --------------C--------------*--*--*----**--------
TziI 13     ------------------*---C---------------------------

                   210       220       230       240
                     .         .         .         .  EcoRI
CON         AAACAGCTGCAAAACCTATTTCCCCAGCATAAGGAA*TGCT*G
TziI 8      ------------------------------G*----*------
TziI 9      --*--*---**--**--------**-------*---*--T---
TziI 10     --*----------**----***-G------**----C--*---
TziI 13     -----C-------------------------*----T------
```

Haplochromis similis I Alignment (230 bp)

```
                   10        20        30        40        50
                    .         .         .         .         .
CON         T**TATCAAAAGTTACAGCTATTTATATGAAGTTGTACAAAAACGCTTTC
HsiI 8      -*****-------------G-----------------------T------
HsiI 10     ---*----------------------------------------------
HsiI 32     *-----------------------C-------------------------
HsiI 36     *-----------------------C-------------------------
HsiI 40     -TC----------------------------------------------A

                   60        70        80        90       100
                    .         .         .         .         .
CON         TTTCGCCAAGACAATGCGTTTCTCACTGTTACATGCATTTGAATGGGAAA
HsiI 8      --------------------------------------------------
HsiI 10     --------------------------------------------------
HsiI 32     -------------G---------------------------------G--
HsiI 36     -------------G---------------------------------G--
HsiI 40     --------------------------------------------------

                  110       120       130       140       150
                    .         .         .         .         .
CON         CTCGNCTGAAACAAAGTATAGGTTTTCATTATGTGAATAATTTTAAAATC
HsiI 8      ----C---------------------------------------------
HsiI 10     ----T---------------------------------------------
HsiI 32     ---CA-C-----------C-A-------------------C--GG-----
HsiI 36     ----A-C-------------A-------------------C--GG-----
HsiI 40     ----C---------------------------------------------

                  160       170       180       190       200
                    .         .         .         .         .
CON         TTAGCTCAAACGGCTGCAAAACCTGTTTGCCCAGCACGGGGATATTGTCT
HsiI 8      ------------------------------------------------.
HsiI 10     --------G-----------------------------------------
HsiI 32     -----------A----------------T----------A--CC------
HsiI 36     -----------A----------------T----------A--CC------
HsiI 40     -----------------------A-------------A------------

                  210       220       230
                    .         .         .
CON         GAGTTCTATAAGATAAAGCCATAAACATGTTT*
HsiI 8      --C-----------------------G-T---*
HsiI 10     ----------------------------T---*
HsiI 32     ----------------------------*----
HsiI 36     ----------------------------*----
HsiI 40     ----------------------------*----
```

APPENDIX 2

Multiple alignment of six tilapiine and *H. similis* SATA consensus sequences. Generated by the CLUSTAL mulltiple alignment program (Higgins and Sharp 1988).

```
                             10        20        30        40        50
                    EcoRI    . PalI     .         .     HinfI        .
O. placi I CON:     AATTCTAT*AAGGCCAAG*CCTGAAATATGTGTGTCCGAGTCTTCTATCA
O. placi II CON:    --------G---------G-------------------------------
O. nil III CON:     --------*---------*------------------------C------
S. gal I CON:       --------*---------*-------------------------------
T. zillii CON:      --------*-----*--AG---T------C-----G--------------
T. rend CON:        -------NT---**---AG---T------C-----N--------------
H. sim CON:         -G------A-GA**T--AG--AT---C----***********--T-----

                             60        70        80        90       100
                             .         .         .         .         .
OplI CON            AAAGTTACAGCTGTCTTTATGGACCTGGTG*AAAATCGCCTTATTTCGGG
OplII CON           -C-----------C-------*********G--------T---------*
OniIII CON          ------------------------T-----G------------------*
SgaI CON            --*------------------A--T-A---*--------T---------*
TziI CON            --*------------------C-GT-AA--*--------T--C-----CC
TreI CON            ---------------------C-GT-AA--*--------T--------CC
HsiI CON            ------------A-T-A----A-GT-*--AC----A---T--C-----*C

                            110       120       130       140       150
                             .         .         .         .         .
OplI CON            CGAGACAGTGCGTTTCTCGCTATTACATGCATTTGAATG*AGTTCTCGCC
OplII CON           -----------------G---------------------G----------
OniIII CON          --------------------*****************************-
SgaI CON            ---------------------------------------G----------
TziI CON            *A-------------------------------------G-C--------
TreI CON            -A-------------------------------------G----------
HsiI CON            -A-----A----------A--G-----------------GGAAA----N-

                            160       170       180       190       200
                             .         .         .         .         .
OplI CON            TGAAACACATTATGGG*TTTTCATTTTGTGAATAACTTGAAAATCTTAGC
OplII CON            ----------------*---------------------------------
OniIII CON          ----------------*---------------------------------
SgaI CON            ----------------*---------------------------------
TziI CON            -------A-G------A---------------------------------
TreI CON            -------A-G------*---------------------------------
HsiI CON            -------A-G---AA-*--------A---------T--T-----------

                            210       220       230       240
                             .         .         .         . EcoRI
OplI CON            TCAAACAGCTGCAAAACCTATTTCCCCAGCAT*GGAA*ATG**GTG
OplII CON           --------------------------------*----*---**C--
OniIII CON          -----------------------N--------*----*---**C--
SgaI CON            --------------------------------*----*---**C--
TziI CON            --------------------------------AA-G-*---**C--
TreI CON            --------------------------*-----A----*---**C--
HsiI CON            ------G------------G---G-------CG--G-T--TGTC--
```

LITERATURE CITED

Arnason, U., Hoglund, M., and Widegren, B. 1984. Conservation of highly repetitive DNA in cetaceans. Chromosoma **89**: 238-242.

Arnason, U., Allderdice, P.W., Lien, J., and Widegren, B. 1988. Highly repetitive DNA in the baleen whale genera *Balenoptera* and *Megaptera*. J. Mol. Evol. **27**: 217-221.

Arnason, U., Gretarsdottir, S., and Widegren, B. 1992. Mysticete (Baleen Whale) relationships based upon the sequence of the common cetacean DNA satellite. Mol. Biol. Evol. **9**: 1018-1028.

Bachman, L., Muller, E., Cariou, M.L., and Sperlich, D. 1992. Cloning and characterization of KM190, a specific satellite DNA family of *Drosophila kitumensis* and *D. microlabis*. Gene **120**: 267-269.

Baltimore, D. 1981. Gene conversion: some implications for immunoglobulin genes. Cell **24**: 592-594.

Bankier, A.T., and Barrel, B.G. 1983. *In* Nucleic Acids Biochemistry, Elseiver Scientific Publishers, Ireland Ltd. pp. 1-34.

Bernardi, G., and Bernardi, G. 1990. Compositional patterns in the nuclear genome of cold-blooded vertebrates. J. Mol. Evol. **31**: 265-281.

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

Brutlag, D.L. 1980. Molecular arrangement and evolution of heterochromatic DNA. Ann. Rev. Genet. **14**: 121-144.

Burcham, T. 1987. Scan Analysis: Densitometry for the Macintosh (Release 2.20). Specom Research, December 1987, published by BIOSOFT.

Charlesworth, B., Langley, C.H., and Stephan, W. 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics **112**: 947-962.

Datta, U., Dutta, P., and Mandal, R.K. 1988. Cloning and characterization of a highly repetitive fish nucleotide sequence. Gene **62**: 331-336.

Denovan, E.M., and Wright, J.M. 1990. A satellite DNA family from pollock (*Pollachius virens*). Gene **87**: 279-283.

Dod, B., Mottez, E., Desmarais, E., Bonhomme, F., and Roizes, G. 1989. Concerted evolution of light satellite DNA in genus *Mus* implies amplification and homogenization of large blocks of repeats. Mol. Biol. Evol. **6**: 478-491.

Doolittle, W.F., and Sapienza, C. 1980. Selfish DNA, the phenotype paradigm and genome evolution. Nature **284**: 601-603.

Doolittle, W.F. 1982. Selfish DNA after fourteen months. *In*: Genome Evolution. *Edited by*: G.A. Dover and R.B. Flavell. Academic Press, London, pp. 3-28.

Dover, G.A. 1982. Molecular drive: a cohesive model of species evolution. Nature **299**: 111-117.

Dover, G.A. 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends in Genet. **2**:159-165.

Dover, G.A. 1989. DNA fingerprints: Victims or perpetrators of DNA turnover. Nature **342**: 347-348.

Ekker, M., Fritz, A., and Westerfield, M. 1992. Identification of two families of satellite-like

repetitive DNA sequences from the Zebrafish (*Brachydanio rero*). Genomics **13**: 1169-1173.

Ellison, M.J., Feigon, J., Kelleher III, R.J., Wang, A.H-J., Habener, J.F., and Rich, A. 1986. An assessment of the Z-DNA forming potential of alternating dA-dT stretches in supercoiled plasmids. Biochemistry **25**: 3648-3655.

Fanning, T.G., Modi, W.S., Wayne, R.K., and O'Brien, S.J. 1988. Evolution of heterochromatin-associated satellite DNA loci in felids and canids (Carnivora). Cytogenet. Cell Genet. **48**: 214-219.

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

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**: 783-791.

Felsenstein, J. 1991. PHYLIP (Phylogeny Inference Package) Version 3.4. From FTP online server at University of Indiana (uindiana.edu)

Flavell, R., Jones, J., Lonsdale, D., O'Dell, M. 1983. Higher plant genome structure and the dynamics of genome evolution. *In*: Advances in Gene Technology: Molecular Genetics of Plants and Animals. *Edited by* K. Downey, R.W. Voellmy, F. Ahmed, and J. Schultz. Academic Press, New York, pp. 47-59.

Fry, K., and Salser, W. 1977. Nucleotide sequences of HS-α satellite DNA from Kangaroo Rat *Dipodomys ordii* and characterization of similar sequences in other rodents. Cell **12**: 1069-1084.

Fryer, G., and Iles, T.D. 1972. The cichlid fishes of the great lakes of Africa: Their biology and evolution. Oliver and Boyd, Edinburgh.

Goodier, J.L., and Davidson, W.S. 1991. A repetitive element in the Atlantic Salmon genome. Abstract for: Symposium on Biochemical Genetics and Taxonomy of Fish, Queens University, Belfast.

Greenwood, P.H. 1987. The genera of pelmatochromine fishes (Teleostei, Cichlidae). A phylogenetic review. Bull. Br. Mus. Nat. Hist. (Zool.) **53**: 139-203.

Gretarsdottir, S. and Arnason, U. 1992. Evolution of the common cetacean highly repetitive DNA component and the systematic position of *Orcaella brevirostris*. J. Mol. Evol. **34**: 201-208.

Helleiner, C.W., Huh, T.Y., and Sangster, S.M. 1974. Satellites and rapidly reassociating fractions in the DNA of *Gadidae* (codfishes). Comp. Biochem. Physiol. **48B**: 11-24.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351-359.

Higgins, D.G., and Sharp, P.M. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**: 237-244.

Hinegardner, R., and Rosen, D.E. 1972. Cellular DNA content and the evolution of teleostan fishes. Amer. Nat. **106**: 621-644.

Hoang-Tang, Dube, S.K., Liang, G.H., and Kung, S.-D. 1991. Possible repetitive DNA markers for *Eusorghum* and *Parasorghum* and their potential use in examining

phylogenetic hypotheses on the origin of *Sorghum* species. Genome **34**: 241-250.

Hummel, S., Meyerhof, W., Korge, E., and Knochel, W. 1984. Characterization of highly and moderately repetitive 500 bp *Eco*RI fragments from *Xenopus laevis* DNA. Nucleic Acids Res. **12**: 4921-4937.

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

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

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

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.

Kornfield, I. 1991. Genetics. *In* Cichlid fishes: Behaviour,

ecology and evolution. *Edited by* M.H.A. Keenleyside. Chapman and Hall, London. pp. 103-128.

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

Maddison, W.P., and Maddison, D. 1987. MacClade version 2.1. A phylogenetics computer program distributed by the authors.

Maniatis, T., Fritsch, E.F., and Sambrook, J. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Majumdar, K.C., and McAndrew, B.J. 1986. Relative DNA content of somatic nuclei and chromosomal studies in three genera, *Tilapia*, *Sarotherodon*, and *Oreochromis* of the tribe Tilapiini (Pisces, *Cichlidae*). Genetica **68**: 175-188.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. **109**: 1963-

1973.

McAndrew, B.J. and Majumdar, K.C. 1984. Evolutionary relationships within three tilapiine genera (Pisces: Cichlidae). Zool. J. Linn. Soc. **80**: 421-435.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. **101**: 20-78.

Meyer, A., Kocher, T.D., Basasibwaki, P., and Wilson, A.C. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature **347**: 550-553.

Miklos, G.L. 1982. Sequencing and manipulating highly repeated DNA. *In* Genome Evolution and Phenotypic Variations. *Edited by* G.A. Dover, and R.B. Flavell. Academic Press, London. pp. 41-68.

Miklos, G.L. 1985. Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. *In* Molecular Evolutionary Genetics. Plenum Press, New York. pp. 241-321.

Moyer, S.P., Ma, D.P. Thomas, T.L., and Gold, J.R. 1988.

Characterization of a highly repeated satellite DNA from the cyprinid fish *Notropis lutrensis*. Comp. Biochem. Physiol. **91B**: 639-646.

Ohno, S. 1972. So much 'junk' DNA in our genomes. *In*: Evolution of genetic systems. *Edited by* H.H. Smith, Gordon and Breach, New York, pp. 366-370.

Orgel, L.C. and Crick, F.H.C. 1980. Selfish DNA: the ultimate parasite. Nature **284**: 604-607.

Pardue, M.L., and Hennig, W. 1990. Heterochromatin: junk or collectors item? Chromosoma **100**: 3-7.

Peters, Von H.M., and Berns, S. 1982. Die Maulbrutpflege der Cichliden: Untersuchungen zur evolution eines verhaltensmusters. Z. Zool. Syst. Evolut.-forsch. **20**: 18-52.

Pfutz, M., Gileadi, O., and Werner, D. 1992. Identification of human satellite DNA sequences associated with chemically resistant nonhistone polypeptide adducts. Chromosoma **101**: 609-617.

Queen, C, and Korn, L.J. 1984. A comprehensive sequence

analysis program for the IBM personal computer. Nucl. Acids Res. **12**: 581-599.

Quinn, J.S., Guglich, E., Seutin, G., Lau, R., Marsolais, J., Parna, L., Boag, P.T., and White, B.N. 1992. Characterization and assessment of an avian repetitive DNA sequence as an iceterid phylogenetic marker. Genome **35**: 155-162.

Radic, M.Z., Saghbini, M., Elton, T.S., Reeves, R., and Hamkalo, B.A. 1992. Hoechst 33258, distamycin A, and high mobility group protein I (HMG-I) compete for binding to mouse satellite DNA. Chromosoma **101**: 602-608.

Regan, C.T. 1920. The classification of the fishes of the family Cichlidae. I. The Tanganyika genera. Ann. and Mag. Nat. Hist. **9**: 33-53.

Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**: 237-251.

Rose, M.R., and Doolittle, W.F. 1983. Molecular biological mechanisms of speciation. Science **220**: 157-162.

Rossi, M.S., Reig, O.A., and Zorzopulos, J. 1990. Evidence for rolling-circle replication in a major satellite DNA from the South American rodents of the genus *Ctenomys*. Mol. Biol. Evol. **7**: 340-350.

Sage, R.D., Loiselle, P.V., Basasibwaki, P., and Wilson, A.C. 1984. Molecular versus morphological change among cichlid fishes of Lake Victoria. *In* Evolution of Fish Species Flocks. *Edited by* A.A. Echelle and I. Kornfield. University of Maine at Orono Press, Orono, Maine. pp. 185-197.

Saitou, N., and Nei, M. 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406-425.

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

Schmid, C.W., and Shen, C.-K.J. 1985. The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. *In* Molecular Evolutionary Genetics. *Edited by* R.J. McIntyre, Plenum Press, New York, pp. 323-358.

Seyoum, S. (1989). Stock identification and the evolutionary relationship of the genera *Oreochromis*, *Sarotherodon*, and *Tilapia* (Pisces: Cichlidae) using allozyme analysis and restriction endonuclease analysis of mitochondrial DNA. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada.

Seyoum, S., and Kornfield, I. 1992. Taxonomic notes on the *Oreochromis niloticus* subspecies-complex (Pisces: Cichlidae), with a description of a new subspecies. Can. J. Zool. **70**: 2161-2165.

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

Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal cross-over. Science **191**: 528-535.

Starr, C. and Taggart, R. 1987. Biology: the unity and diversity of life. Wadsworth Publishing Company, Belmont, California, p. 565.

Stiassny, M.L.J. 1991. Phylogenetic intrarelationships of the family Cichlidae: an overview. *In* Cichlid Fishes: Behaviour, ecology and evolution. *Edited by* M.H.A.

Keenleyside. Chapman and Hall, London. pp. 1-35.

Strachan, T., Webb, D., and Dover, G.A. 1985. Transition stages of molecular drive in multiple-copy DNA families in *Drosophila.* EMBO J. **4**: 1701-1708.

Strauss, F., and Varshavsky, A. 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell **37**: 889-901.

Sturmbauer, C., and Meyer, A. 1992. Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature **358**: 578-581.

Tartof, K.D. 1975. Redundant genes. Ann. Rev. Genet. **9**: 355-385.

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

Tautz, D. 1990. Genomic finger printing goes simple. BioEssays **12**: 44-46.

Trewavas, E. 1982. Generic groupings of tilapiini used in aquaculture. Aquaculture **27**: 79-81.

Trewavas, E. 1983. Tilapiine Fishes of the Genera *Sarotherodon*, *Oreochromis*, and *Danakilla*. British Museum (Natural History), London.

Turner, B.J., Elder, J.F., and Laughlin, T.F. 1991. Repetitive DNA sequences and the divergence of fish populations: some hopeful beginnings. J. Fish Biol. **39** (SA): 131-142.

Vissel, B., and Choo, K.H. 1991. Four distinct alpha satellite subfamilies shared by human chromosomes 13, 14 and 21. Nucl. Acids Res. **19**: 271-277.

Walsh, J.B. 1987. Persistence of tandem arrays: Implications for satellite and simple-sequence DNAs. Genetics **115**: 553-567.

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

Weiner, A.M., Deininger, P.L., and Efstratiadis, A. 1986.

Non-viral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. Ann. Rev. Biochem. **55**: 631-661.

Wevrick, R., and Willard, H.F. 1989. Long-range organization of tandem arrays of α satellite DNA at the centromeres of human chromosomes: High-frequency array-length polymorphism and meiotic stability. Proc. Natl. Acad. Sci. USA **86**: 9394-9398.

Willard, H.F., and Waye J.S. 1987. Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: a survey of alphoid sequences from different human chromosomes. Nucl. Acids Res. **15**: 7549-7569.

Willard, H.F. 1989. The genomics of long tandem arrays of satellite DNA in the human genome. Genome **31**: 737-744.

Wright, J.M., and Dixon, G.H. 1988. Induction by torsional stress of an altered DNA conformation 5′ upstream of the gene for a high mobility group protein from trout and specific binding to flanking sequences by the gene product HMG-T. Biochem. **27**: 576-581.

Wright, J.M. 1989. Nucleotide sequence, genomic organization

and evolution of a major repetitive DNA family in tilapia (*Oreochromis mossamibicus/hornorum*). Nucleic Acids Res. **17**: 5071-5079.

Wu, J.C., and Manuelidis, L. 1980. Sequence definition and organization of a human repeated DNA. J. Mol. Biol. **142**: 363-386.