Title: A Phylogenetic Index of Cichlid Microsatellites

Robert D Kunkle[1], Christian R L Reilly[2], and Suzy C P Renn[1]

1) Biology Dept., Reed College, 3203 SE Woodstock Blvd, Portland OR 97202
2) Santa Catalina School, 1500 Mark Thomas Dr., Monterey, CA 93940

contributing author: Suzy C P Renn
Phone: 503 517 7967
Fax: 503 777 7773
email: renns@reed.edu

SUBMITTED TO COPIA
running title: Cichlid Microsatellite Index

Keywords: microsatellite, cichlid, evolutionary radiation,

ABSTRACT

Microsatellites, or short sequence repeats, abound in most organisms and have proven useful for a range of genetic and genomic studies. Once primers have been created, they can be applied to populations or taxa that have diverged from the source taxon with varying degrees of success. We use PCR amplification, in a 96 well format, to determine the presence and absence of 46 microsatellite loci in 13 cichlid species. At least one primer set amplified product in each species tested, and some loci were present in nearly all species. These results are compared to the known phylogenetic relationships among cichlids. Our results present a phylogenetic index for the presence and absence of microsatellite loci thus providing a collection of primer sets applicable to wide range of species. This information and resource should ease entrance to cichlid genetic studies and increase the potential impact for cross species comparisons.

INTRODUCTION

Microsatellites, or short sequence repeats (SSRs), are short (2-6 bp) DNA motifs repeated at least three and up to hundreds of consecutive times (Tautz *et al*., 1986). SSR's abound in most organisms, and fishes are no exception, with an estimated frequency of one locus per several kb of DNA (Chistiakov, 2006). The repetition of a microsatellite motif makes misalignment of template and newly synthesized strands during DNA replication very likely, resulting in a range of alleles differing by whole numbers of repeats (Ellegren, 2004). Such unstable mutation dynamics hamper sequence-dependent function of a locus, so with few exceptions, such as the human Huntington's locus, observed microsatellite regions are not transcribed. As primarily neutral, polymorphic loci with a signature pattern that facilitates isolation, microsatellites have proven useful for a range of genomic studies.

Despite broad utility, a researcher interested in applying microsatellite-based tools to linkage mapping or phylogenetic analysis faces a significant investment in time and material to isolate repeat regions and create primers that anneal with the microsatellite flanking regions (MFR's) adjacent to the repeats. Their proximity to microsatellites makes MFR's likely to be selectively neutral, to the point that their sequence can be used as a molecular clock for phylogeny studies (Zardoya *et al*., 1996). Since MFR-derived primer pairs do not anneal to repeat regions, this disruption does not interfere with these primers' efficacy. Once MFR primers have been created, therefore, depending on genome-wide mutation dynamics, they can likely be applied

to populations or taxa that have diverged from the source taxon. For example, Rico, Rico and Hewitt (1996) were able to amplify a microsatellite region with the same MFR-derived primer set in two fish species that diverged 470 Mya. However, the pattern of MFR sequence conservation was sufficiently unpredictable to require locus-by-locus confirmation. If, at the outset of work on one taxanomic group, MFR primers are available from previous work on a related group, some expense still must be undertaken to determine which microsatellites are present and informative in the genomes of interest (Zane *et al*., 2002, Glen and Schable, 2005). The current study presents an index of MFR-specific primer sets tested in species representing the major groups within the most speciose family of fish (the Cichlidae). This information should reduce the entrance cost to those interested in applying microsatellite-based analyses to additional cichlid species.

The cichlids of the Great Rift Lakes of Eastern Africa are especially important to research into evolutionary processes because their history has been reconstructed to reveal multiple adaptive radiations, many recent (<2 Mya) and some exceptionally recent (<12,500 ya), resulting in extensive diversification - often exhibiting convergence between lineages - of form, niche and behavior (Barlow, 2000). In the case of such recent speciation, one can assume that drift has done much less than usual to obscure the genomic differences underlying species' unique properties. Furthermore, the presence of many closely related species, often sympatric and quite subtly diverged (Kocher, 2004), enables appropriately subtle analysis of speciation genomics and genetic basis for adaptive traits. For example, jaw morphology-related genes have been studied Malawi cichlids (Albertson *et al*. 2003), and population structure has been addressed in the Tanganyika rock cichlid species (Duftner *et al*., 2006) as well as for the sympatrically speciating Midas cichlids from South America (Bunje, 2007).

The most intensively studied cichlids are the widely-farmed, multi-generic tilapia species, in one of which, *Oreochromis niloticus*, Kocher and colleagues have developed microsatellite isolation methods used to create several hundred sets of MFR-specific primers (Lee and Kocher, 1996), in addition to creating a linkage map of those loci (Lee *et al*., 2005). Kocher and colleagues also isolated microsatellites and created a linkage map for a hybrid of *Labeotropheus fuelleborni* and *Metriaclima zebra*, two closely related species of mbuna or rock cichlid from Lake Malawi. In the creation of the mbuna linkage map, 248 of the primer sets obtained from *O. niloticus* were also tested, and 46 were found to work in the mbuna hybrid (Albertson *et al*., 2003). These 46 loci represent all but three of the *O. niloticus* linkage groups from the available genetic map (Lee et al., 2005), with as many as four loci for groups 3, 10 and 17. As *O. niloticus* and the mbuna species share a common ancestor with most of the Great Lakes cichlids approx. 18-30 Mya (Fig. 1a) (Genner *et al*., 2007), the 46 loci found in both species bore significant chances of being present in other African cichlids. This current study creates a phylogenetic index of the presence and absence of these loci in a wide range of cichlid species to aid cichlid researchers. The 46 primer sets were tested on genomic DNA extracted from 13 cichlid species: *Astatotilapia burtoni, Neoloamprologus brichardi, Perrisodus microlepis, Protomelas similis, Metriaclima esterae, Tylochromis sp., Tropheus duboisi, Xenotilapia flavipinnis, Xenotilapia ochrogenys, Retroculus xinguensis, Cichla temensis, Astronotus sp., and Satanoperca sp*. This sample covers most of the major African clades and some South American clades. No cichlids from Indian or Madagascan were examined. At least one primer set amplified product in each species tested, and some loci were present in nearly all species.

MATERIALS AND METHODS

Fin clips were collected in the field and placed immediately in ethanol. Genomic DNA was extracted from each individual using a standard proteinase K/Phenol protocol. PCR was performed using the standard FastStart Taq protocol in 10μL reactions, with 2.5μM $MgCl_2$, .25μM Forward and Reverse primers, and .5ng template DNA, using the following program: 30 cycles; 30s @ 95C, 30s @ 56C, 1m @ 72C. PCR products were run on 4% agarose gel stained with eithidium bromide. Digital gel images were captured for analysis. Band presence, relative brightness, approximate length and the presence or absence of a doublet were recorded. Failed reactions were repeated for confirmation of negative results.

Clustering of species according to the pattern of successful PCR products was performed using R software v2.0.1 (R Development Core Team 2006). The dissimilarity measures were obtained using the *dist* function in the *stats* package based on Euclidean distance using product presence and absence information only. The consensus tree and bootstrap confidence values for each node were obtained with the *consensus* function in the MAANOVA package (Wu *et al*., 2002). The consensus tree dendogram and confidence values were calculated as the proportion of 1000 trees that agreed with the original tree as obtained by resampling with replacement, again using presence absence data only.

RESULTS

In total 13 species were assayed including 9 from Africa and 4 from South America. In general the number and pattern of successful microsatellite amplification products reflects phylogenetic relationship (Fig.1). Among 9 African cichlid species, 2 are endemic to Lake Malawi and 7 to Lake Tanganyika. The mean number of positive amplifications per species (out of 46) was 33.5 (s.d. $\pm$ 3.5) for Lake Malawi and 35.67 (s.d. $\pm$ 4.03) for Lake

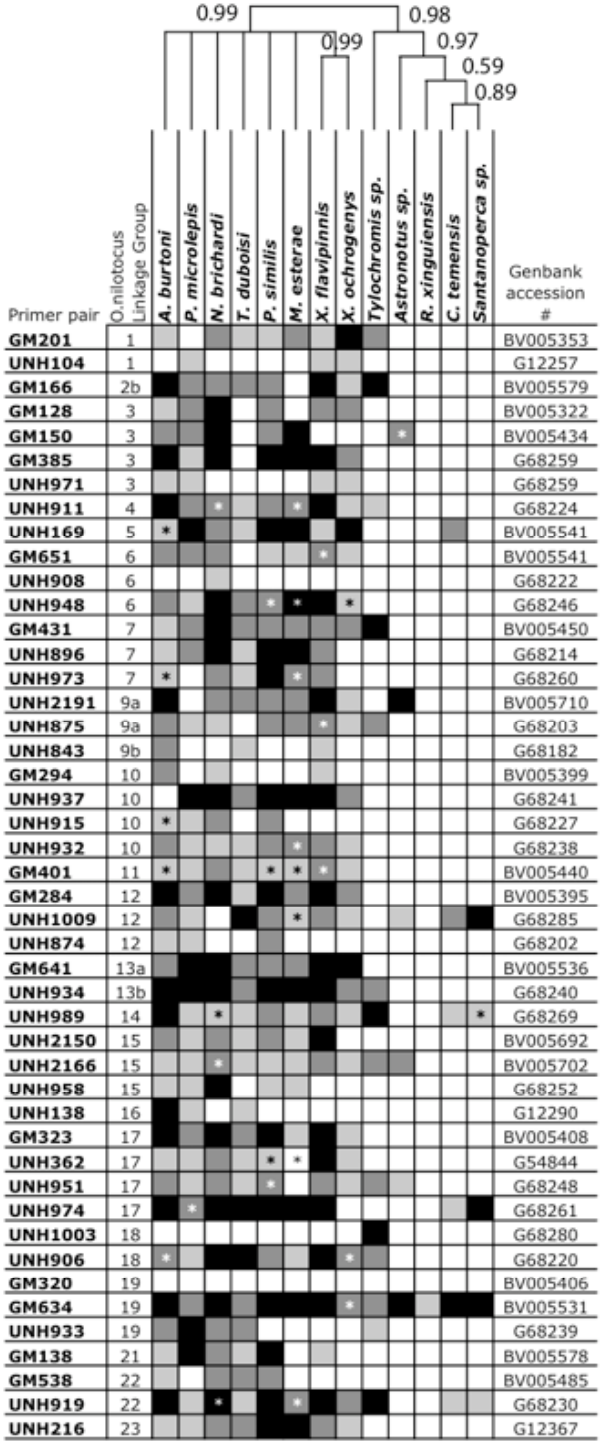

Figure 1: Successful PCR amplification of microsatellite loci in 13 cichlid species using primers designed to Tilapia. Color indicates relative intensity of the PCR band black: brightest, dark grey: visible, light grey: faint, white: absent). Definitive doublets are indicated with an Asterisk. The dendogram represents the consensus cluster confidence values indicated) based upon 1000 trees, with resampling, using presence or absence of product and Euclidean distance measures.

Tanganyika. *T. polylepis* is excluded from this average calculation due to its recent immigrant status and distant relationship to other Tanganyikan cichlids (Koch *et al.*, 2007, Genner *et al.*, 2007). The 14 positive amplifications from *T. polylepis* were the least of any African species, but still more than any of the South American cichlids. The two species from the *Ectodini* tribe showed the most similar pattern of microsatellite amplification products; six primer sets amplified in *X. flavipinnis* and not *X. ochrogenys*, but there were no other differences in their amplification patterns. As expected, the more distantly related South American cichlid species showed significantly fewer successful microsatellite products. On average 4.5 (s.d. $\pm$ 2.06) primers sets amplified in these species, and each species tested had a unique pattern of positive amplifications.

Clustering analysis resulted in a dendogram that separated the African Great Lakes cichlids from their sister genus, *Tylochromis*, and from the South American species. There was insufficient statistical confidence to distinguish relationships within the Great Lakes and accurately capture relationships among the South American clades (Figure 1). These results agree, as far as resolution allows, with the mtDNA phylogeny studies (Farias, 1999, Salzburger, 2005).

DISCUSSION

The data presented here demonstrate that the previously isolated MFR primer sets will be useful for population studies in most cichlid taxa, especially throughout the East African radiation. While the range of species used in this study cannot definitively predict which primer sets will yield informative genetic information for every cichlid species, it does provide a measure of the expected success rate for a given phylogenetic position. Furthermore, the availability of this primer set in a 96 well format will facilitate rapid screening for any species of interest.

The band brightness aspect of the data may estimate sequence divergence in these MFR's. It is possible that highly diverged loci will not amplify as efficiently, and further divergence would prohibit amplification all together; this should be anticipated when a distantly related (e.g. South American) species is studied. As Ellegren (2004) made clear, mutation rates vary between loci, individuals and taxa, due to disabled mismatch repair and proofreading, chromatin structure variation, or other mechanisms. Therefore we cannot infer sequence similarity by a measure of band brightness. Similarly, an allele of a given length may have arisen from either a lengthening or from a shortening mutation, meaning that its exact relationship to other alleles is unclear. In addition, as with absolute mutation rates, the relative frequencies of shortening and lengthening vary within genomes and taxa. Therefore, estimating a given allele's ancestry requires considerable groundwork to describe the variation at that locus for any species of interest. For research over a fairly short scale of divergence, where novel alleles are at a minimum, this groundwork will require amplification from several individuals' genomic DNA to estimate whether enough polymorphism exists to allow for distinction between lineages. Here, (Figure 1) we do report all observed differences in relative brightness (denoted by shading) of the imaged PCR products as well as the presence or absence of a doublet (denoted by the asterisk) in order to provide all possible information regarding polymorphism of each locus. However, it must be noted that only a single individual was assayed in the current study and resolution was ~ 20 bp or greater. Therefore, further work is required to describe polymorphic loci in each species, in order that the application of this primer set to a wide range of studies can facilitate cross species comparisons.

By contacting the corresponding author, the full set of 46 primes used in this study are freely available in a 96 well format, diluted to a working concentration for use in PCR with any species of interest.

ACKNOWLEDGMENTS

We are grateful for the donation of tissue samples by D. Joyce (University of Hull), Malawi species, and S. Willis (University of Nebraska, Lincoln), South American species. We thank H. Machado for comments on the manuscript. This work has been supported by a grant to S.C.P.R. from the M.J. Murdock Charitable Trust (No.:2006253:JVZ:2/22/2007).

LITERATURE CITED

ALBERTSON, R.C., STREELMAN, J.T., KOCHER, T.D. 2003 Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Science USA* 100: 5252–5257.

BARLOW, G.W. 2000 Cichlid Fishes: Nature's Grand Experiment in Evolution, Cambridge, MA: Perseus Books Group.

BUNJE, PME., BARLUENGA, M., MEYER, A. 2007 Sampling genetic diversity in the sympatrically and allopatrically speciating Midas cichlid species complex over a 16 year time series. *BMC Evolutionary Biology* 7: 1-14.

CHISTIAKOV, D.A., HELLEMANS, B., VOLCKAERTM, F.A.M. 2006 Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture* 255: 1-29.

DUFTNER, N., KOBLMULLER, S., STURMBAUER, C. 2005 Evolutionary relationships of the Limnochromini, a tribe of benthic deepwater cichlid fish endemic to lake Tanganyika, East Africa. *Journal of Molecular Evolution*. 60: 277-289.

ELLEGREN, H. 2004 Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 5: 435-445.

FARIAS, I.P., ORTI, G., SAMPAIO, I., SCHNEIDER, H., MEYER, A. 1999 Mitochondrial DNA phylogeny of the family Cichlidae: Monophyly and fast molecular evolution of the neotropical assemblage. *Journal of Molecular Evolution* 48: 703-711.

GENNER, M.J., SEEHAUSEN, O., LUNT, D.H., JOYCE, D.A., SHAW, P.W., CARVALHO G.R., TURNER, G.F. 2007 Age of cichlids: New dates for ancient lake fish radiations. *Molecular Biology and Evolution* 24: 1269-1282.

GLENN, T.C., SCHABLE, N.A. 2005 Isolating microsatellite DNA loci . *Molecular Evolution: producing the biochemical data, part B*. 395: 202-222.

KOCH, M., KOBLMULLER, S., SEFC, K.M., DUFTNER, N., KATONGO, C., STURMBAUER, C. 2007 Evolutionary history of the endemic Lake Tanganyika cichlid fish Tylochromis polylepis: a recent intruder to a mature adaptive radiation. *Journal of Zoological Systematics and Evolutionary Research*. 45: 64-71.

KOCHER, T.D. 2004 Adaptive evolution and explosive speciation: the cichlid fish model. Nature Reviews Genetics. 5: 288-298.

LEE, B.Y., LEE, W.J., STREELMAN, J.T., CARLETON, K.L., HOWE, A.E., HULATA, G., SLETTAN, A., STERN, J.E., TERAI, Y., KOCHER, T.D. 2005 A second-generation genetic linkage map of tilapia *Oreochromis sp.. Genetics* 170: 237-244.

LEE, W.J., KOCHER, T.D. 1996 Microsatellite DNA markers for genetic mapping in *Oreochromis niloticus*. *Journal of Fish Biology* 49: 169-171.

R DEVELOPMENT CORE TEAM 2006 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http: //www.R-project.org>

RICO, C., RICO, I., HEWITT, G. 1996 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of London Series B-Biological Sciences* 263: 549-557.

SALZBURGER, W., MACK, T., VERHEYEN, E., MEYER, A. 2005 Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology* 5: 1-15.

TAUTZ, D., TRICK, M., DOVER, G.A. 1986 Cryptic simplicity in DNA is a major source of genetic variation. *Nature*, 322: 652-656.

WEBSTER, M.T., HAGBERG, J. 2007 Is there evidence for convergent evolution around human microsatellites? *Molecular Biology and Evolution* 24: 1097-1100

WU, H., KERR, M.K., CUI, X., CHRUCHILL, G.A. 2003 MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments. The analysis of gene expression data: methods and software, New York: Springer.

ZANE, L., BARGELLONI, L., PATARNELLO, T. 2002 Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1-16

ZARDOYA, R., VOLLMER, D.M., CRADDOCK, C., STREELMAN, J.T., KARL, S., MEYER, A. 1996 Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes Pisces: Perciformes. *Proceedings of the Royal Society of London Series B-Biological Sciences* 263: 1589-1598.