

Increased Gene Duplication in Cichlid Adaptive Radiations, Lake Malawi, Africa

Heather E. Machado¹, Ginger Jui^{1,2}, Domino A. Joyce³ Christian R.L. Reilly^{1,4}, David H. Lunt³,
and Suzy C.P. Renn¹

¹Department of Biology, Reed College, Portland, OR 97202

²Department of Plant and Microbial Biology, University of California, Berkeley California
94720-3102

³Department of Biological Sciences, University of Hull, Kingston Upon Hull HU6 7RX, N
Humberside England

⁴Santa Catalina School, Monterey, CA 93940

Dr. Suzy C.P. Renn
Department of Biology, Reed College, Portland OR 97202 USA
Telephone: 503-517-7967
Fax: 503-777-7773
E-mail: renns@reed.edu

Abstract

Using the African Cichlid fish model, we test the hypothesis that gene duplication is a component of the genomic architecture for evolutionary adaptive radiation. Through a microarray-based technique of comparative genomic hybridization (aCGH), we show that for 3 Lake Malawi species (*Metriaclima estherae*, *Protomelas similis*, *Rhamphochromis chilingali*) from radiated lineages the number of duplicated genes is a three to four fold greater than for one species from a non-radiated lineage (*Astatotilapia tweddlei*). While further study is necessary to determine if this phenomenon is a general characteristic of adaptive radiation throughout the cichlids and in evolutionary radiations in general, the current results support the controversial claim that gene duplication can lead to phenotypic divergence and speciation.

Manuscript

Adaptive radiation, the simultaneous evolution of genetic and ecological diversity leading to the species proliferation in a lineage, is thought to result from divergent selection for resource specialization. Differential selection in heterogeneous environments can lead to adaptive radiation when there is a genetic basis for niche specialization (Dobzhansky 1937; Mayr 1963; Schluter 2000). An incredible series of adaptive radiations in response to the local physical, biological and social environment have produced the cichlid fishes of the African Great Lakes, which exhibit unprecedented inter-specific diversity in morphological and behavioral characteristics (Kocher 2004; Seehausen 2006). Importantly, each radiating lineage possesses close relatives that have not undergone radiations even when the ecological opportunity arose (Seehausen 2006). This ability to compare closely related lineages that have and that have not undergone evolutionary radiation provides a critical tool for identifying genomic features that promote, or correlate with, adaptive radiation.

Current genomic research supports the classic work of Ohno (1970) which proposed a prominent role for gene duplication in evolutionary expansion (e.g. primates: Fortna et al. 2004; Dumas et al. 2007; Marques-Bonet et al. 2009). According to this theory, the resultant extra gene copies are available for dosage effects, subfunctionalization, or neofunctionalization, and can contribute greatly to an organism's survival, fitness and adaptation. Given the high rate of gene duplication (similar point mutations: Lynch and Conery 2000), it should be considered as a process contributing to speciation. The recent and rapid speciation of the cichlid adaptive radiations allows us to test the hypothesis that there is an association between gene duplication and speciation.

In order to quantify duplicates, we use array-based comparative genomic hybridization (aCGH) to measure gene duplication for ~7000 genes (8679 array features) for three Lake Malawi species from lineages that have undergone radiation and one closely related riverine species not from a radiated lineage. Using this technique, gDNA from the species in question is competitively hybridized to a cDNA microarray with gDNA from the reference species, *Astatotilapia burtoni*, from which the microarray was produced. Genomic content is then relative to this reference, with greater hybridization to a feature on the array being caused by duplication at that locus (for review see Gresham, Dunham, and Botstein 2008). As the three species from the radiated lineages (*Metriaclima estherae*, *Protomelas similis*, *Rhamphochromis chilingali*) and the one non-radiated lineage (*Astatotilapia tweedlei*) are of similar phylogenetic distance to *A. burtoni*, there is no *a priori* reason to expect a substantially greater number of gene duplicates in any of these species (fig. 1).

In support of the hypothesized role for gene duplication in the process of rapid speciation, when quantifying the number of genes present in increased gene copy number relative to *A. burtoni*, we found three to four times more genes duplicated in the Lake Malawi species than in the non-radiated riverine species. According to stringent statistical analysis, of the 6923 genes tested, 39 were shown to be duplicated in *M. estherae*, 42 in *P. similis*, and 33 in *R. chilingali*, compared with only 10 in *A. tweedlei* (fig. 2, for details see supplementary table S1). Shared duplications were particularly prevalent among the three species from the Lake Malawi radiation, with seven duplications shared among all three, and 14 duplications shared between two of the three species

(fig. 2). This strong similarity in duplication profiles is visualized well by comparison of array hybridization ratios, which reflect actual genomic content resulting from different gene copy number in each species (fig. 3). Gene loss may also contribute to the observed phylogenetic patterns of genomic content. For example, two genes were found to have greater content in all four species relative to *A. burtoni*, possibly representing a reduction in copy number in the *A. burtoni* lineage.

The shared duplications among the radiated lineages are of great interest, as they likely occurred basally in the Lake Malawi radiations. During this early period of dramatic ecological specialization, it has been speculated that hybridization occurred, prior to development of the reproductive isolation present today (Seehausen 2004). The preponderance of shared duplications among the lake species may reflect this evolutionary history. Alternately, shared duplications may be due to a propensity of certain genes to duplicate. Among our current list of duplicated cichlid genes, several have been reported to have undergone duplication at other points in teleost history (cyp19: Chiang et al. 2001; MHC genes: Miller, Kaukinen, and Schulze 2002; PAC1: Cardoso et al. 2004; finTRIM: van der Aa et al. 2009).

Four loci found to be duplicated in various combinations of radiated and non-radiated cichlid lineages, were chosen for validation by quantitative PCR (qPCR). In each case, all species showing significant duplication in the microarray analysis also showed significantly higher genomic content than *A. burtoni* according to the qPCR analysis (fig. 4).

Few studies have tested, on a genome-wide scale, whether gene duplication provides the variation underlying natural ecological diversification (but see Dopman and Hartl 2007). Furthermore, only recently have studies begun to examine the pattern of gene copy number polymorphism across species in natural systems, beyond primates (e.g. Chen et al. 2008). Here we demonstrate that gene duplications are increased among rapidly radiating lineages compared to a non-radiating lineage for one comparison among the cichlid radiations. This model evolutionary system offers extraordinary ecological diversification and repeated radiations with which to further test the hypothesis of gene duplication associated with rapid speciation. Whether gene duplication is a cause or effect of adaptive radiation remains to be tested among other cichlid lineages as well as evolutionary model systems where radiated and non-radiated lineages are available. The pattern of increased gene duplication in radiated lineages supports the controversial claim that gene duplication can lead to phenotypic divergence and speciation.

Methods

The *A. burtoni* cDNA microarray was used (GEO platform GPL6416). Field-collected heterologous species samples were stored in ethanol. Genomic DNA, extracted according to a standard ProteinaseK/Phenol protocol, was size reduced by Hydroshear (Genome Solutions/Digilab) to 1 – 5 Kb. DNA (4µg) was fluorescently labeled with Alexa-Fluors conjugated dCTP by Klenow fragment polymerization (Invitrogen, Bio-Prime). Each species was hybridized twice, in dye swap, against a reference pool of *A. burtoni* genomic DNA from the inbred laboratory strain (Fernald and Hirata 1977). After a 16 hour hybridization (67.5°C, 3.4X SSC, 0.15% SDS, 1 mM DTT, Cot-1DNA), arrays were washed prior to scanning (Axon 4100B, Genepix).

Microarray data was filtered by omitting features with faint signal (<2 SD above background), lack of sequence information and known ribosomal content. Data was corrected for background intensity (“minimum”), and loess normalized within each array using 250 features expected to have sequence conservation across a range of fish species (Salzburger et al. 2008). This normalization corrects the skew introduced by sequence divergence under standard normalization that assumes of a mean Log intensity ratio of zero (van Hijum et al. 2008). "Duplicated genes" were identified based upon increased fluorescence according to the “lmFit” statistical model with “eBayes” correction and FDR adjustment for $P < 0.1$ significance level (Smyth 2004). The reported results are underestimates of duplication levels, due to the fact that diverged duplicates (Renn submitted) are less likely to be detected, and false positives may be introduced by mutations affecting gene structure (e.g. increased intron length).

Gene copy level was validated for four genes using qPCR (supplementary table S2). gDNA concentration was quantified with 1.5X SYBR Green I (Roche Applied Science) using a Nanodrop 3300 (Thermosavant). Triplicate qPCR reactions were run in an Opticon 2 (MJ Research) in 10 μ l reaction containing 0.75x SybrGreen, 1x Immomix (Biolabs), 200-500 nM primers and 0.2 ng sample DNA (95 °C- 10 min; 35 cycles of: 94 °C- 2 min, 60 °C- 20 sec, 72 °C- 15 sec, and 2 min extension). Copy number relative to *A. burtoni* was calculated as CT, the cycle number at the set threshold relative to the *A. burtoni* standard curve. Efficiency for each primer set was calculated with a dilution series for *A. burtoni* DNA and one test species (supplementary table S3).

Supplementary Material

Supplementary tables and figures S1, S2, and S3 are available at Molecular Biology and Evolution online (<http://www.mbe.oxfordjournals.org>).

Acknowledgements

Molly Shumer and Keith Karoly contributed and commented.

Fig. 1. Phylogenetic relationship of species used, maximum likelihood (D-loop sequence)

Fig. 2. Identified gene duplicates ($P < 0.1$ FDR)

Fig. 3. Genomic content for all 95 duplicated loci, relative to *A. burtoni*. Array features (rows) and species (columns) are clustered according to complete linkage of Euclidean distances. The vertical solid blue lines track hybridization ratio for each feature.

Fig. 4. qPCR validates gene copy number determined by aCGH. Abbreviations are genus and species initials.

** $P < 0.1$ FDR, * $P < 0.2$ FDR found by array analysis

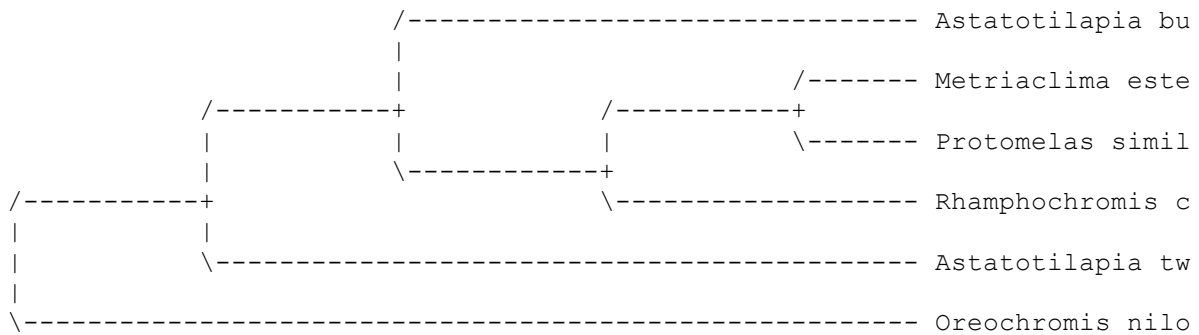


Fig. 1. Phylogenetic relationship of species analyzed according to D-loop sequences (PUT GENBANK IDS HERE) using *O. niloticus* as the outgroup in a maximum likelihood calculation employed by PAUP. (conducted by Heather)

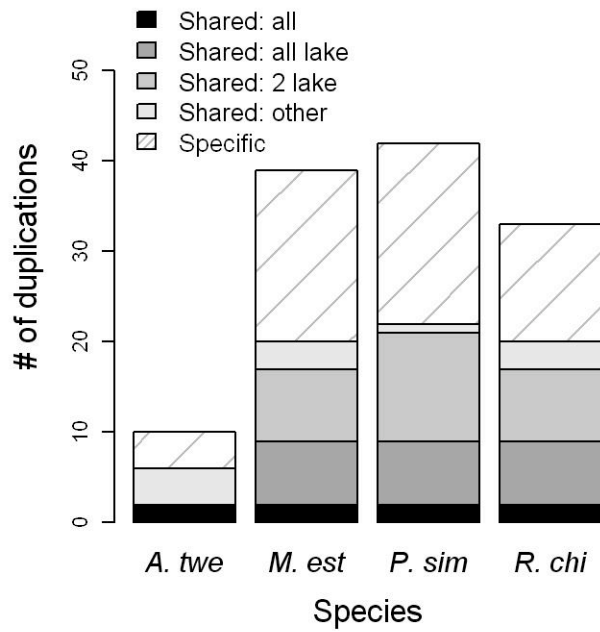


Fig. 2. Identified gene duplicates ($P < 0.1$ FDR)

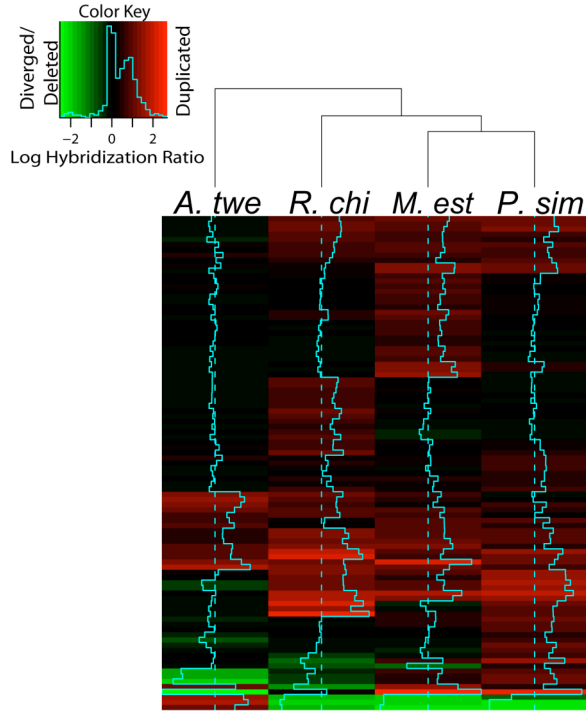


Fig. 3. Genomic content for all 95 duplicated loci, relative to *A. burtoni*. Array features (rows) and species (columns) are clustered according to complete linkage of Euclidean distances. The vertical solid blue lines track hybridization ratio for each feature.



Fig. 4. qPCR validates gene copy number determined by aCGH. Abbreviations are genus and species initials.

** $P < 0.1$ FDR, * $P < 0.2$ FDR found by array analysis

Supplementary Table S1: Hybridization ratios and P-values for the 95 genes found to be duplicated in any species.

GenBank	DFCI_TC	A_ratio	M_ratio	P_ratio	R_ratio	A_Pvalue	M_Pvalue	P_Pvalue	R_Pvalue	Duplicate	Best
CN470857	none	0.6598	0.6618	0.0935	0.9752	3.66E-02	8.75E-03	6.14E-01	1.28E-03	R	none
CN469913	none	0.2785	0.1515	0.3541	0.6363	1.61E-01	2.65E-01	2.75E-02	1.90E-03	R	none
CN470701	none	0.1155	0.3007	0.3579	0.6582	5.25E-01	4.63E-02	2.44E-02	1.42E-03	R	none

92 more lines in complete table to be submitted

Supplementary Table S2 Oligonucleotide primers used for qPCR designed against GenBank sequence available for microarray features.

GenBank Accession#	Primer Sequences	Best Hit Annotation	Predicted Length
DY626766	F: TCGGTCTCCTTAACCGGATG R: CTGAGTTTGGCTGCCCGTAA	None	193
DY627986	F: ACGAACACCCGAACGGAAAC R: GGTGCACGCACATGAACTGT	Hox gene cluster, contains hoxCa	222
DY631898	F: CGTCCCAGTGAGGATGAGGA R: TGATGCTGATCGGTTGATGC	MHC class II antigen	161
DY632057	F: ATTACTGCGAGTGCCGTCCA R: CTGCGCCCTGAAAGAACAGA	Pituitary adenylate cyclase activating polypeptide receptor 1A	150

Supplementary Table S3. Percent primer efficiency based on 4 XX-fold template dilution for *A. burtoni* and one heterologous species.

	DY626766	DY631898	DY632057	DY627986
<i>A. burtoni</i> (%)	86	100	82	91
Heterolg. species (%)	74	104	82	78
Heterolg. species (name)	<i>P. similis</i>	<i>M. estherae</i>	<i>R. chilingali</i>	<i>A. tweddlei</i>

References

- Cardoso, J. C. R., D. M. Power, G. Elgar, and M. S. Clark. 2004. Duplicated receptors for VIP and PACAP (VPAC(1)R and PAC(1)R) in a teleost fish, *Fugu rubripes*. *Journal of Molecular Endocrinology* **33**:411-428.
- Chen, Z. Z., C. H. C. Cheng, J. F. Zhang, L. X. Cao, L. Chen, L. H. Zhou, Y. D. Jin, H. Ye, C. Deng, Z. H. Dai, Q. H. Xu, P. Hu, S. H. Sun, Y. Shen, and L. B. Chen. 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America* **105**:12944-12949.
- Chiang, E. F. L., Y. L. Yan, G. G. Yann, J. Postlethwait, and B. C. Chung. 2001. Two Cyp19 (P450 aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. *Molecular Biology and Evolution* **18**:542-550.
- Dobzhansky, T. 1937. *Genetics of the evolutionary process*. Columbia University Press.
- Dopman, E. B., and D. L. Hartl. 2007. A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* **104**:19920-19925.
- Dumas, L., Y. H. Kim, A. Karimpour-Fard, M. Cox, J. Hopkins, J. R. Pollack, and J. M. Sikela. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Research* **17**:1266-1277.
- Fernald, R. D., and N. R. Hirata. 1977. Field study of haplochromis-burtoni habitats and cohabitants. *Environmental Biology of Fishes* **2**:299-308.
- Fortna, A., Y. Kim, E. MacLaren, K. Marshall, G. Hahn, L. Meltesen, M. Brenton, R. Hink, S. Burgers, T. Hernandez-Boussard, A. Karimpour-Fard, D. Glueck, L. McGavran, R. Berry, J. Pollack, and J. M. Sikela. 2004. Lineage-specific gene duplication and loss in human and great ape evolution. *PLOS BIOLOGY* **2**:937-954.
- Gresham, D., M. J. Dunham, and D. Botstein. 2008. Comparing whole genomes using DNA microarrays. *Nature Reviews Genetics* **9**:291-302.
- Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics* **5**:288-298.
- Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151-1155.
- Marques-Bonet, T., J. M. Kidd, M. Ventura, T. A. Graves, Z. Cheng, L. W. Hillier, Z. S. Jiang, C. Baker, R. Malfavon-Borja, L. A. Fulton, C. Alkan, G. Aksay, S. Girirajan, P. Siswara, L. Chen, M. F. Cardone, A. Navarro, E. R. Mardis, R. K. Wilson, and E. E. Eichler. 2009. A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**:877-881.
- Mayr, E. 1963. *Animal species and evolution*. Belknap Press of Harvard University Press, Cambridge, Massachusetts.
- Miller, K. M., K. H. Kaukinen, and A. D. Schulze. 2002. Expansion and contraction of major histocompatibility complex genes: a teleostean example. *Immunogenetics* **53**:941-963.
- Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag.
- Salzburger, W., S. C. P. Renn, D. Steinke, I. Braasch, H. A. Hofmann, and A. Meyer. 2008. Annotation of expressed sequence tags for the east African cichlid fish *Astatotilapia burtoni* and evolutionary analyses of cichlid ORFs. *BMC Genomics* **9**:1-14.
- Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.

- Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B-Biological Sciences* **273**:1987-1998.
- Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**:198-207.
- Smyth, G. K. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**:1-26.
- van der Aa, L. M., J. P. Levraud, M. Yahmi, E. Lauret, V. Briolat, P. Herbomel, A. Benmansour, and P. Boudinot. 2009. A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. *Bmc Biology* **7**.
- van Hijum, S., R. J. S. Baerends, A. L. Zomer, H. A. Karsens, V. Martin-Requena, O. Trelles, J. Kok, and O. P. Kuipers. 2008. Supervised Lowess normalization of comparative genome hybridization data - application to lactococcal strain comparisons. *Bmc Bioinformatics* **9**.