Objectives

Aim 1: Test the hypothesis that radiated lineages have increased gene duplicates. We will use aCGH to quantify gene duplicates in 15 cichlid species from 3 independent adaptive radiations and contrast results for each with a phylogenetically close, non-radiated lineage.

Aim 2: Expanded phylogenetic context for gene duplication rate and intra-specific CNV. We will characterize duplication rate and persistence time using qPCR across an extended phylogenetic range while also investigating individual variation within each species.

Aim 3: Sequence analysis to study the evolution of gene duplicates. We will use sequence analysis of paralogs and orthologs to investigate evolutionary history and modern fate of gene duplicates.

Aim 4: Test the hypothesis that sequence divergence of specific gene classes occurs in radiated lineages or is associated with adaptive phenotypes. We will use aCGH to identify highly diverged genes among the species addressed in Aim1.

Significance

Adaptive radiation, a prominent feature of evolution, results in remarkable phenotypic diversity. Similarly, gene duplication, a prominent feature of molecular evolution, results in remarkable evolutionary novelty and is thought to be responsible for the origin of 30-65% of all eukaryotic genes[4]. Although the adaptive potential for gene duplicates was recognized by early geneticists[5,6], Ohno's work[7] is the first synthesis of these ideas and description of the maintenance of gene duplicates contributing to adaptation. Now termed neofunctionalization, subfunctionalization and gene conservation (or dosage), these fates[8,9] are often modeled as the result of natural selection[8] and/or neutral evolution[10-12].

Among the East African cichlids, which exhibit remarkable morphological, ecological and behavioral diversity[13], there are lineages that have not undergone radiation even when the ecological opportunity arose[14], thus offering the opportunity to compare genomic architecture of radiated and non-radiated lineages. By collecting gene duplicate data on a genome-wide scale in multiple independent radiations, we aim to provide a rigorous test regarding the genomic architecture of adaptive radiation. Due to the recency of these radiations, and the array-based techniques, we will recover young gene duplicates whose characterization can address gaps in our understanding of the evolutionary processes that maintain gene duplicates. Answering these questions will also address the more fundamental problem of how genotype maps to phenotype.

The data generated by the macroevolutionary work proposed will make a significant contribution to ongoing cichlid genome sequencing projects[15], to the development of genomic tools, and will further studies concerning the microevolution of gene duplicates. Population genetic models of gene duplicate evolution are needed and the cichlid adaptive radiation provides an appropriate model system for empirical research at this level. We have primarily focused our attention at the level of species in order to conduct a rigorous, ambitious, genome-wide analysis that is achievable within the allotted timeframe.

In addition to the broad implications regarding the evolutionary processes and the general implications for the cichlid research community, the data collected will be directly relevant to the ongoing research programs of both PIs. These results will aid in the interpretation of the heterologous gene expression profiling performed by Dr. Renn and her students. Similarly, the data will be informative with regard to the process of speciation and patterns of gene exchange exhibited by different taxa in different ecological situations that are the focus of Dr. Lunt's research program. The Reed College students have demonstrated success with the aCGH, qPCR and cloning techniques that are proposed.

Background and Literature Review

Adaptive Radiation; Specifically the Cichlid Radiation

Adaptive radiation, the evolution of genetic and ecological diversity leading to species proliferation in a lineage, is classically viewed as the result of differential selection in heterogeneous environments[16-18]. Examples of such radiations include the Cambrian explosion of metazoans[19], the diversification of Darwin's finches in the Galapagos Islands[20], variations in amphipods and cottoid fish in Lake Baikal[21], the Caribbean anoles[22], the Hawaiian Silverswords[23] and the explosive speciation of the cichlid fishes in the African Great Lakes[24,25]. While fascinating unto itself, the rapid speciation associated with adaptive radiation also provides insight to a broad range of questions in evolutionary biology, relating to factors that play a role in speciation, such as genetic drift and population size[26,27], strength of selection[27] and gene flow with hybridization[28-30].

While cichlids can be found on several continents[31], the most dramatic assemblages, representing the majority of the morphological, ecological and behavioral diversity, are found among lacustrine radiations endemic to the large freshwater Lakes Malawi, and Victoria[24] and previously in lake Makgadikgadi[32]. As the product of an incredible series of adaptive radiations in response to local variation in the physical, biological and social environment[14,24,33], East African cichlid fish account for ~10% of the world's teleost fish. While all studies of African cichlid radiations agree that they are among the fastest and largest of the known vertebrate adaptive radiations[14,34,35], the genomic changes driving, or resulting from, this adaptive evolution are largely unknown. Importantly, some radiated lineages possesses a close relative that has not undergone radiation[14] (**Fig. 7**). The ability to compare closely related lineages that have and that have not undergone evolutionary radiation provides a critical tool to identify genomic features that promote, or correlate with, adaptive radiation.

The Genomic Architecture of Adaptive Radiation

Among the extant models of adaptive radiation we see common features that include evidence for parallel evolution, prevalent examples of convergence, roles for non-ecological processes and interesting exceptions of stasis[28]. While natural selection[36], sexual selection[37] hybridization[28], and drift[38] play roles in adaptive radiation, the apparent impunity to ecological forces witnessed by the lineages in stasis suggest that inherent genetic/genomic processes may also exist. Because the rate of speciation is higher in lineages whose precursors emerged from more ancient adaptive radiations, Seehausen[14] speculated that the propensity for speciation is attributable to "genomic properties that reflect a history of repeated episodes of radiation". If this is true, architectural differences should exist in genomic structure of radiated and non-radiated lineages. The proposed project uses modern genomic techniques to identify these architectural differences with regard to gene duplications and sequence divergence in order to address outstanding questions regarding the genetic basis of adaptive radiation.

The Role of Gene Duplication in Evolutionary Innovation

It is well established that gene duplication and the subsequent evolution of duplicates is an important source of genetic[39,40] and functional novelty[7,9]. Duplicated loci, which may arise through transposition[41], polyploidy[42,43], retrotransposition[44], and segmental[45] or tandem[46,47] duplications, are prevalent in genomes of organisms belonging to all three domains of life. Despite a high probability of loss due to drift[48] and silencing by degenerative mutations[49], 30-65% of all functional genes[4] are thought to have originated through the more rare event of fixation for gene duplicates[50]. Current genomic research (e.g. primates[51,52]; plants[53,54]; mammals[55]; genetic model organisms[49]) supports the classic work of Ohno[7] that proposed a prominent role for gene duplication in evolutionary expansion. The extra gene copies may have phenotypic effects that enhance survival, fitness and adaptation. For example, gene duplications are known to be involved in adaptive evolution in response to diet[56-59], chemical challenge[60,61], herbivore defense[62] and reproduction[40,63]. Such adaptations can allow diversification into new niches, as has been suggested for cold adaptation (Antarctic ice fish[64]; plants[65]) and novel metabolic processes (C-4 photosynthesis[66]; non-shivering thermogenesis[67]). Appreciation for the pervasive nature of inter-specific

gene duplication has been reinforced by intra-specific genome studies that identify dramatic gene copy number variation (CNV) between individuals (mouse[68]; human[69-71]; *Drosophila*[72]; *Arabidopsis*[73]). Among dog breeds, which offer a model for adaptive radiation despite being the product of artificial selection, recent work has identified breed-specific gene duplicates associated with breed-specific traits[74]. Importantly, between canines[75] and among dogs[74], a significant proportion of the gene duplications can be used to dissect evolutionary relationships despite the admixture prevalent among dog breeds - a situation to which the cichlid phylogeny is often likened. Within the cichlid radiation, a number of duplicate loci are known to be involved in functional/phenotypic divergence (pigmentation[76,77 78]; opsins[79-82]; sex-determination[83,84]; neurohormone systems[85]; hox genes[86])

The evolutionary fate of gene duplicates may increase fitness though different mechanisms that impact the evolutionary potential of a species. While there is high probability that any one gene duplicate will be rapidly lost by random genetic drift[87], a duplicate can also be retained as a by-product of non-adaptive evolution[10-12]. In the rare event that a functional duplicate rises in frequency, selective pressures may dominate. In some cases, multiple copies of the same gene are thought to have adaptive potential purely because these copies increase the amount of gene product[60,64,88] while gene sequence remains unchanged (conservation mechanism). Beyond such dosage effects, there can be adaptive nucleotide substitutions either during[89] or after[90] the fixation of the duplicated locus. Thus, the maintenance of gene duplication has long been thought to provide a faster track to adaptation and evolutionary novelty in various ways. Relaxation from pleiotropic constraint may allow the partitioning of previous gene function (subfunctionalization)[91,92] or the evolution of a novel function (neofunctionalization)[62,93,94] through major or minor sequence changes (e.g. opsins[95], olfactory receptors[96]). While possibly representing stages along a continuum for a single process[97], there is evidence for each fate.

Considering flies, worms and yeast, the estimated average duplication rate is 0.01 per gene per million years[49,98] and ranges from 0.0014 - 0.0039 among mammals[99]. Thus gene duplication has the potential to generate substantial molecular substrate for the origin of evolutionary novelty, as it occurs at a rate that is of the same order of magnitude as the rate of mutations per nucleotide site[100], such that an estimated 60 - 600 gene duplications per million years may be expected to arise between sister taxa[49].

Array Comparative Genomic Hybridization (aCGH) to Assay Genome Content

Originally championed for analysis and diagnosis of gene dosage and chromosomal aberrations underlying cancer[101-103], aCGH has been optimized for the identification of gene duplicates, which may be collapsed during the shotgun sequence assembly[104]. The thoroughly validated and standardized[53,102,103,105,106] aCGH technique relies on basic DNA chemistry. When using aCGH to compare genome content between two species, one of the gDNA samples is isolated from the species for which the microarray was constructed, and the other gDNA sample is isolated from a heterologous species of interest. Genomic regions that have been deleted or are highly diverged in the heterologous sample will fail to hybridize strongly to the array features resulting in a log ratio less than zero. Conversely, genomic regions that have been duplicated will hybridize at a ratio of 2:1 (or greater), resulting in a log ratio greater than zero. Through repetition with multiple heterologous species, a phylogenetic study can address genome content in a broad range of related organisms[1,2,70,107] without the need for full genome assembly, which is required for next-generation DNA sequencing techniques[52,108].

Beyond the assessment of copy number variation within a species[109] and between closely related lineages (*D. discoideum*[110]; *Drosophila*[111]; experimental evolution in yeast[112]; dog breeds[74]), the aCGH technique has been used to reveal genomic regions likely involved in an organism's ability to inhabit a specific environment (*Chlamydia trachomatis:* tissue specificity[113]; *Sinorhizobium meliloti:* root symbiont[114]; *Clostridium difficile:* host specificity[115]) or enact specific pathogenicity (*Vibrio cholerae*[116]; *Yersinia pesits*[117]; *Mycobacterium tuberculosis*[118,119]). The technique has also identified genomic duplications and deletions associated with population divergence and speciation in *Anopheles gambiae*[120,121] as well as

genomic regions that differentiate humans from other primate species[51,70,122]. With aCGH, it is also possible to identify rapidly evolving genes (*Paxillus involutus*[123]) and in some cases lend insight to phylogenetic relationships (*Shewanella*[124]; *Saccharomyces*[107,125]; *Salmonella*[126]).

While most studies rely only on presence or absence metrics, a few studies have suggested a roughly log-linear relationship between aCGH hybridization ratio and nucleotide identity [113,127,128]. Recently, we have conducted a detailed characterization this relationship[3] and investigated the interaction between sequence divergence and gene duplication. These efforts reveal the type of gene duplicates that are most successfully identified across diverged species by aCGH[2]. While these proof-of-concept experiments were conducted with Drosophilids , we have also initiated investigation of cichlid gene duplicates[1].

Summary

The African cichlid adaptive radiations constitute an ideal system to answer long-standing hypotheses about the role of gene duplication in phenotypic diversification and species proliferation. While it is well established that gene duplication and the subsequent evolution of duplicates is an important source of genetic and functional novelty, few studies have address this topic in a broad phylogenetic context beyond genome sequence analysis. We will directly test the association of gene duplicate number, duplication rate and retention of gene functional categories with the enhanced speciation and ecological diversification. Furthermore, by sequence analysis, we will test the hypothesis regarding the evolutionary history and modern fate of gene duplicates and address the degree to which positive selection is responsible for the evolution of gene duplicates. Not only does the proposed methodology ensure confident results, these techniques have been successfully employed by Reed College studdents. Hence, this work provides a golden opportunity to integrate of scientific inquiry with undergraduate research and curriculum.

Proposed Experiments and Methods

We propose to construct a Nimblegen 720K oligonucleotide microarray for our aCGH experiments. The development of a Nimblegen array platform will be a substantial improvement over the current cDNA platform[131] as it will provide a comprehensive dataset with state of the art technology in order to accurately interrogate all known cichlid genes.

All available cichlid sequence data will be taken into account for array design. In addition to the EST dataset[131,132], 454 transcriptome data (49K contigs, avg. length 950 bp) and anticipated genome sequence information is available for *A. burtoni*[15]. In collaboration with bioinformaticians at the Broad Institute supporting the International Cichlid Genome Consortium, these data will be used to design 720,000 Nimblegen oligonucleotide array probes (45-75 bp)[133,134]. The sequence data available for three Lake Victoria haplochromines[135-137], two Tanganyikan species (Salzburger, pers. comm.) and Tilapia (56K unique sequences; Kocher per. comm) will be taken into consideration along with the low coverage genomic sequence (0.1X) for five Malawi species (JGI; NCBI Trace)[138] in order to annotate probes and maximize representation of coding region sequence. With this technology, all known cichlid coding sequences will be represented by multiple probes as prescribed for accurate duplicate detection[139], and three arrays can be spotted on each glass slide.

Astatotilapia burtoni is a logical platform species for this study because it derives from a relatively deep East African haplochromine mtDNA lineage[140,141], is a clear sister group to the most recent radiation of the Lake Victoria superflock[140] and is among the species slated for full genome sequence[15].

We have specifically selected Nimblegen technology because it does not require core facility equipment beyond the scope of the Renn lab. Therefore, undergraduate students will have a comprehensive experience including sample preparation, labeling, hybridization and scanning in addition to data processing and analysis. There is a demonstrated high level of concordance between platforms and the greater customization of Nimblegen, making this platform a logical choice[142].

AIM 1: Test the Hypothesis that Radiated Lineages Have Increased Gene Duplicates.

Aim 1: Rationale for quantifying gene duplication

The diverse and rapid adaptive radiation of African Great Lake cichlids has occurred despite a rather unexceptional amount of genetic diversity in sequence. The fact that this tendency to speciate is not evenly distributed throughout the clade suggests the involvement of underlying genomic elements. Gene duplications provide one source of functionally diverse genomic elements that may lead to phenotypic diversity on which natural selection can act. Multiple independent comparisons between radiated lineages and their non-radiated relatives are important to identify such elements.

Aim 1: Preliminary data shows more gene duplicates in radiated lineages

Using aCGH on the current *A. burtoni* cDNA microarray[1] we find a 38% – 49% increase in gene duplicates for three species representing radiated lineages of Lake Malawi (*M. estherae*, *P. similis*, *R "chilingali"*) relative to the total number identified in a related species from a non-radiated lineage (*A. tweddlei*) (**Fig. 5**). While shared duplications were prevalent among the three Lake Malawi species (11% among all three and 11% between two), a large proportion (67%) of the duplications were lineage-specific.

Among our current list of duplicated cichlid genes, several have undergone duplication at other points in teleost history (MHC genes[143]; PAC1[144]; finTRIM[145]), and several others

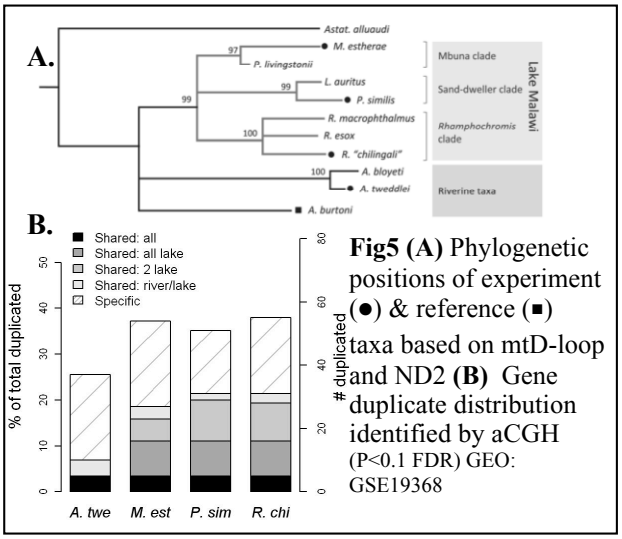

Fig5 (A) Phylogenetic positions of experiment (●) & reference (■) taxa based on mtD-loop and ND2 **(B)** Gene duplicate distribution identified by aCGH (P<0.1 FDR) GEO: GSE19368

are annotated with functions suggestive of an adaptive role. For example, FGF receptor activating protein (*M. estherae*) is a growth factor involved in angiogenesis, wound healing, and embryonic development, potentially leading to morphological diversity, as seen among cichlids. The finTRIM gene (*P. similis*) plays a role in immunity against viral infection, and is known to be the result of duplication retained by positive selection in zebrafish[145]. Several transcription factors, such as DAX1 (*P. similis*), which is a repressor of steroidogenesis and sex-determination[146], or CdaR (*M. estherae*) act as key regulator genes with the potential for cascading effects on gene expression and phenotypic variation.

These patterns, and the specific candidates, provide a tantalizing suggestion that gene duplication plays a functional role in these radiations. These duplicated genes may underlie adaptive phenotypes and be integral to the process of evolutionary adaptive radiation.

Aim 1: Experimental design for duplicate detection by aCGH

Cichlid tissue samples are available from multiple wild individuals of 5 different species within each of the 3 targeted independent African adaptive radiations (paleo-Magkadigkadi, Victoria and Malawi) as well as a cichlid species from a non-radiated lineage closely related to each radiation. DNA from each species (**Table 1**) will be competitively hybridized against a pooled sample of *A. burtoni* lab-strain gDNA on the new Nimblegen 720,000 oligo feature array, a design similar to that used for dog[74] , *Drosophila* strains[111] or primate[70] studies. A balanced reference design will include a pair of dye-swapped arrays for 3 biological replicates of each of the 18 species and the Tilapia outgroup[147]. This design allows a rigorous statistical analysis taking advantage of replicate probes per gene using Nimblescan and SignalMap software provided by Nimblegen in order to identify duplicated genes at the level of species and a less powerful test to uncover individual variation (expanded in Aim2) (114 hybridizations, 38 arrays).

	Radiation 1 - paleo-Magkadigkadi	Radiation 2 - Victoria	Radiation 3 - Malawi
Non-radiated	*T. demeusii*	*Astatotilapia bloyeti*	*Astatotilapia tweddlei*
Radiating taxa	*Serranochromis macrocephalus*	*Neochromis omnicaruleus*	*Dimidiochromis compressiceps*
	Sargochromis coulteri	*Para-labidochromis*	*Rhamphochromis esox*
	Chetia brevis	*Pundamilia pundamilia*	*Metriaclima emmiltos*
	Serranochromis robustus	*Pundamilia. azurea*	*Hemitilapia oxyrhynchus*
	Chetia brevicauda	*Lipochromis melanopterus*	*Copadichromis eucinostomus*

Table 1. Fish to be used in this experiment. If sample collection allows we will add a fourth radiation (Barombi Mbo) and non-radiated relative (*Saratherodon galilaeus*).

The number of duplicated loci vs. the number of duplicated genes will be estimated with rough synteny according to Tilapia BAC end sequences aligned to other fish species to identify multiple duplicated genes in close chromosomal proximity. Counting events rather than genes will allow us to test whether both the number of duplication events and the number of retained duplicated genes are different between radiated and non-radiated taxa. Furthermore, these analyses will identify chromosomal "hotspots" for gene duplication as have been shown to exist in other species[148-150]. Increased resolution will be possible when full Tilapia genome sequence is available.

Gene Ontology analysis[151] with traditional hypergeometric tests as well as rank-based methods (e.g Rank-GoStat[152]) will uncover trends among functional gene categories to suggest hypotheses about gene types or pathways that are most likely to be duplicated. Environment responsive categories are often found to be enriched, whereas certain metabolic processes are often found underrepresented for both intra- and inter-specific variation[111,153,154]. Annotations according to KEGG[155] or CoG[156] databases are also relevant. However, despite interesting results in genetic model organisms that have identified biases based on protein interaction network parameters[157,158], we do not feel that such detailed annotations are likely to translate accurately. Instead, we will test for over- or under-representation of duplicated genes among the identified cichlid gene expression modules[159-161].

Aim 1: Outcomes for gene duplicate quantification

These experiments will produce the primary data on which most of the tests in this proposal are based. In this aim, the data is interrogated to test the null hypothesis that the number of gene duplicates is the same between cichlid adaptive radiations and non-radiated relatives. aCGH data generated here will provide an extensive list of 'candidate loci' associated with morphological, behavioral and ecological diversification and speciation for subsequent investigation.

Aim 1: Further Considerations

Failure to reject the null hypothesis provides an equally interesting result. Gene duplication is heavily influenced by mutational bias and functional constraint [163] as demonstrated by the location and classes of genes that are seen to be affected [157,164-167]. These processes may temper the adaptive potential for duplication. As such, specific loci, rather than overall number, may influence the propensity for radiation.

Individual variation within a species is known to exist in humans[109,168] mice[169], and flies[111]. We expect individual variation in copy number within cichlids and have preliminary evidence that *A. burtoni* reveals copy number differences between the inbred lab-strain and recently captured wild individuals. Our experimental design will not confound intra-specific variation with inter-specific variation and will allow us to detect it here as well as by qPCR (Aim2).

AIM 2: Expanded Phylogenetic Context of Gene Duplication Rate and Intra-Specific CNV.

Aim 2: Rationale for expanding the study

The evolutionary significance of gene duplication depends on the rate at which duplicate genes arise and the processes by which they are retained[100]. Understanding the duplication rate and persistence time (half-life) will allow direct comparison to estimates from genome sequence surveys and aCGH studies that provide empirical data for theoretical models[49,99,112,170].

Though often detached in the biomedical literature, there is an obvious relationship between the identified lineage-specific gene duplicates among primates[51,52 99] and the current excitement concerning copy number variation (CNV) within humans[69-71] as it relates to disease susceptibility[171], behavior[172,173], and adaptation[174]. While several studies have investigated population structure for individual gene duplicates[56,169], few studies investigate intra-specific variation focused on lineage-specific duplications on a genome-wide scale (but see for *Drosophila*[111]).

Aim 2: Pilot data: Validation of qPCR duplicate detection

We selected 4 loci that exhibit different patterns of gene duplication among radiated (*M. estherae*, *P. similis*, *R "chilingali"*) and non-radiated lineages (*A. tweddlei*)[1] for validation by qPCR (**Fig. 6**). The strong agreement between these two techniques clearly demonstrates the ability to expand the phylogenetic study using qPCR. Importantly we show that for each primer, efficiency, as calculated with a four-fold dilution series, was similar for *A. burtoni* gDNA and for the heterologous species[1]. With only one exception (DY631898 in *M. estherae*), the two techniques agreed. The qPCR technique is routinely taught in courses and used in research by students at Reed College.

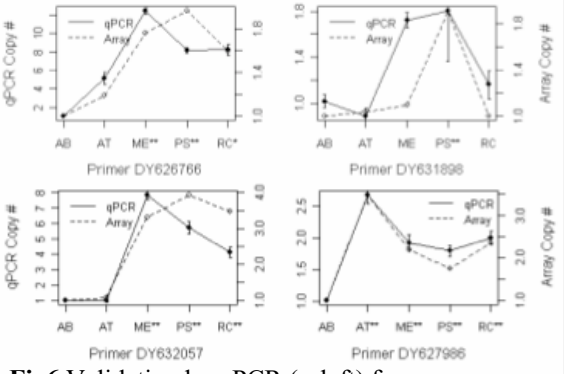

Fig6 Validation by qPCR (y-left) for gene copy number determined by aCGH (y-right). Abbreviations are genus and species initials.

Aim 2: Experimental methods to expand the phylogenetic range of gene duplicate information

Species and loci to be assayed: We will extend phylogenetic diversity by including species from each of the tribes in Lake Tanganyika as well as closely related non-radiated lineages[175] and increase the

resolution with additional species from the focal radiations (Aim1) to describe inter-specific gene duplicate patterns. We will survey10 individuals from each species to collect intra-specific CNV data. We will select loci that show an interesting pattern such as:

1) Duplications associated with a particular phenotype[176].
2) Duplications that are distributed across the radiation such that the inclusion of Tanganyikan species might resolve homoplasies.
3) Duplications that are species specific (in either a radiated or non-radiated lineage)
4) Loci that are unique to, and pervasive throughout, one of the 3 assayed radiated lineages

Appropriate species will be selected based upon the initial patterns observed for each loci (100's of samples available from Renn and Joyce).

Genomic content assayed by qPCR: gDNA concentration will be quantified with 1.5X SYBR Green I (Roche Applied Science) on a Nanodrop 3300 (Thermosavant). Primer design will be based on sequence alignment of cichlid and teleost sequence. Triplicate qPCR reactions (Opticon MJ Research) will be run in 10 µl containing 0.75x SybrGreen, 1x Immomix (Biolabs), 200-500 nM primers (with confirmed primer efficiency) and 0.2 ng sample gDNA (95 °C- 10 min; 35 cycles of: 94 °C- 2 min, 60 °C- 20 sec, 72 °C- 15 sec, and 2 min extension) on a DNA Engine Opticon Real-Time PCR System (Bio-Rad). Genomic content relative to *A. burtoni* labstock will be calculated as CT[177] standardized to *A. burtoni*.

The analysis of ancestral character (gene duplication) reconstruction on the phylogenetic tree (**Fig. 7**) will be implemented using parsimony, likelihood and Bayesian statistics. The Bayesian stochastic character-mapping approach[178,179] is the most flexible and powerful for our data since model-based character change can be determined along branches rather than just at nodes on the phylogeny. Visualization and data handling will be achieved in Mesquite[180]. We will then construct a character matrix of duplicated loci relative to Tilapia to conduct all pair-wise comparisons necessary for a relative rate test[181] in order to date the origins and persistence time of gene duplicates. This method tests whether a character (duplication) evolves at a constant rate in two lineages. Using known ages for the major nodes in the cichlid phylogeny that are based on the fossil record or geology[34,35,140,182] we will compare the number of duplicates between radiated and outgroup (Tilapia) versus non-radiated and outgroup. Likelihood models that rely on quantitative measures of gene duplicates [99,183] will also be addressed. These methods are complementary to the synonymous divergence (dS) approach used by Lynch [49] for genome data that can be applied to sequenced loci (Aim3). Rate and persistence times can be compared to estimates for other organisms[74,99].

Because segregation of ancestral polymorphisms is not an uncommon feature with regard to single nucleotide polymophisms (cichlids[184]; dogs[185]; silver swords[186,187]), by extending our analysis to Lake Tanganyika we can begin to evaluate whether this pattern holds for gene duplication. Sequence analysis (Aim3) will be necessary to conclusively differentiate homoplasy, loss and ancestral sorting of duplicates.

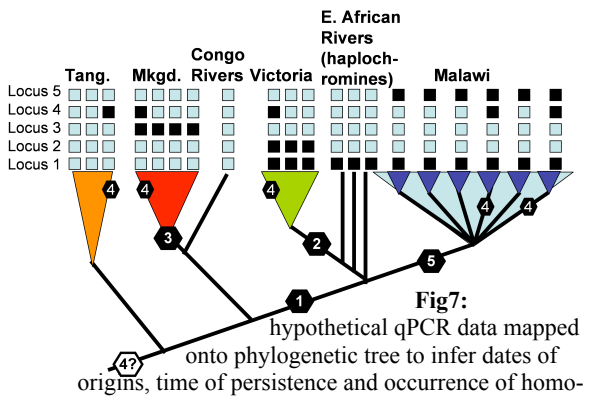

Fig7: hypothetical qPCR data mapped onto phylogenetic tree to infer dates of origins, time of persistence and occurrence of homo-plasy (solid "4") that can be distinguished from ancestral segregation (open "4?") by sequence (Aim3). Correlation with phenotype will also be tested.

Correlation between phenotypes and duplicates may indicate selective pressure or functional links that underlie the remarkable phenotypic diversity (for example: substrate preference[188]; trophic structures[189,190]; breeding strategy[13,191]; courtship behavior[192-194]; coloration[195-198]). Such differences in gene copy number between breeds are associated with breed-specific traits[74] in dogs. We will map duplicates against known phenotypes to generate these hypotheses.

There are three reasons to survey individual variation (intra-specific CNV).
First, the identification of gene duplicates that are fixed in a species indicates the potential to underlie species-specific traits. Second, we can ask whether individual variation differs between radiated and non-radiated lineages, as would be expected if an increased number of duplicates is due to increased rate of duplication rather than increased maintenance of duplicates. Third, such patterns are important because population genetic processes regulating intra-specific variation drive inter-specific variation such that the data collected here will contribute empirical data to further develop mathematical models describing the evolution of gene duplicates[199].

Aim 2: Outcomes

This work will provide detailed quantitative information on the scale, timing and distribution of cichlid gene duplication, providing novel information of broad significance in evolutionary genomics. Furthermore, while the population genetic models for the evolution of multigene families are still immature, this additional empirical evidence should provide insight to researchers developing models.

Aim 2: Further considerations and alternative approaches.

As genomic technology advances, we will continue to re-evaluate the cost vs. information gained by switching to a next-gen sequencing strategy[108,200]. No current technology surpasses aCGH for accurate estimates of gene duplicates on a genome-wide scale and while the data obtained in Aim1 could be used to design Nimblegen 380,000 arrays (12 arrays/slide) for multiplexing and population studies between several species, qPCR currently provides the most cost effective survey for a large number of individuals.

AIM 3: Sequence Analysis to Study the Evolution of Gene Duplicates.

Aim 3: Rationale for sequence analysis

Despite successful identification of many gene duplicates, the evolutionary mechanism responsible for their maintenance remains elusive[8]. Several studies suggest the relative importance of different evolutionary processes on gene duplicates during different phases of divergence[49,201,202]. DNA sequence allows insight to the evolutionary histories and modern fates of gene duplicates[8,203] in order to address these questions. Comparative analysis of orthologs and paralogs can suggest instances of positive natural selection that may have played a role in the diversification and speciation of the African cichlids.

Aim 3: Experimental design for sequencing duplicates.

To clone gene duplicates, degenerate primers, designed according to aligned cichlid and teleost sequence, will be used to amplify, clone and sequence ~20 duplicated loci from the species from Aim1 or Aim2, targeting loci that were fixed within species and avoiding large gene families or high copy number duplicates that would hinder our ability to sequence all paralogs and identify true orthologs. Again, we target loci that suggest convergent evolution (or assorted ancestral alleles), are species-specific, are unique to and pervasive throughout a lineage, or correlate with a phenotype. The number and distribution of species to be assayed will be tailored to the specific question for each locus. We will sequence ~15 colonies per transformation by standard procedures (~5,700 sequence reads) to identify paralogous sequences. From those sequences, paralog-specific primers can be designed for inverse PCR to clone surrounding sequence[204,205] and/or 5' and 3' RACE-PCR to clone the entire transcript. We have considerable experience with these techniques both in cichlids and a diverse range of other species. This strategy has successfully cloned recently duplicated genes in the soybean[206]

Ortholog and paralog sequences will be aligned with MUSCLE[207], and a maximum likelihood tree[208] will be built for each duplicate locus with PHYML[209,210] using appropriately estimated evolutionary models. For duplicates that occur in subsets of multiple distinct lineages, we will be able to test whether these instances represent convergent evolution or sorting of ancestral polymorphisms. Furthermore, sequence

data will be checked for frame shifts and stop codons that would indicate the presence of pseudogenes. Duplicates suspected of being pseudogenes will be omitted from further analyses.

Positive selective pressure on gene duplicates will be assessed with a codon-substitution model employed by the program PAML[211]. Likelihoods for each subtree when $d_N/d_S > 1$ (branch-site model A)[212] are compared to those for the neutral site model M1 where $d_N/d_S = 1$. By comparing twice the difference in log-likelihood values to a χ^2 distribution, a conservative measure for positive selection is obtained that is robust to the presence of pseudogenes[50,213]. This technique also results in a reduction in false positives compared with previous branch-site models [212]. Recent findings indicate low rates of gene conversion in duplicates[214], a potential source of false positives[215,216], further supporting a high true positive rate for these tests. Despite controversy regarding susceptibility to noise[217,218], statistical analysis of nonysnonomous mutations per nonsynonymous site to synonymous mutation per synonymous site d_N/d_S ratios (ω) still provides the most widely accepted and rigorous evidence for positive natural selection[50,219-222].

Aim 3: Outcomes

This DNA sequence information will allow insight into the evolutionary histories, describe modern fates and identify potentially adaptive roles for gene duplicates across the cichlid radiation.

Aim 3: Further Considerations

Substantial evidence suggests that expression patterns evolve rapidly for gene duplicates[223-228], likely influenced by sequence changes in 5' region[229], suggesting subfunctionalization. As such, gene expression profiling may augment the aCGH experiments proposed. Such a strategy would require an array designed to detect sequence differences among paralogs .The sequence information acquired in this aim is a necessary first step.

The historical hybridization among cichlids[28] means that lineages may not be completely independent, a violation of the assumptions underlying tests to detect signatures of selection[230]. While alternate comparative tests with different assumptions have been suggested, such as radical/conservative amino acid replacements[231], controversial[232-235] codon based methods[230], recent genetic robustness tests[236], and complex 3D-sliding window techniques[237,238], these have not been thoroughly validated and will not be considered until other studies provide such techniques with the necessary experimental validation.

Additional population level tests for adaptive evolution take advantage of the power gained by using sequence polymorphism data to look at levels of diversity in gene duplicates[239,240] or to compare non-synonymous to synonymous ratios between polymorphic and fixed differences[241-245] for investigation of paralogs[246] or orthologs[247,248]. Such tests may be applied in future studies that include a population level approach. We reserve sequencing of such intra-specific variation for future analysis.

Aim 4: Test the Hypothesis that Sequence Divergence of Specific Gene Classes Occurs in Radiated Lineages or is Associated with Adaptive Phenotypes.

Aim 4: Rationale for the identification of highly diverged genes

In aCGH experiments, variation in hybridization signal also reveals sequence divergence among orthologs for single copy genes[2]. Because the cichlid radiation has occurred recently, there is very low genomic divergence between species (**Fig. 4**). However, aCGH can detect the small number of genes that show greater than 5% sequence

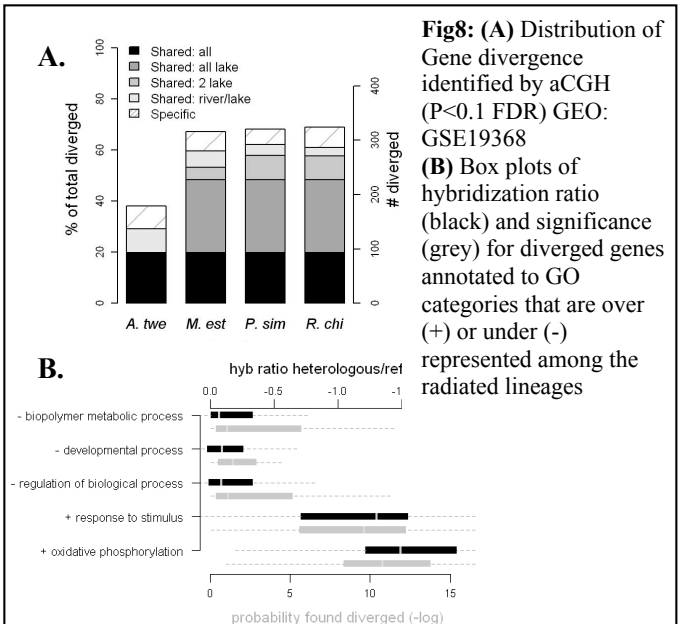

Fig8: (A) Distribution of Gene divergence identified by aCGH (P<0.1 FDR) GEO: GSE19368
(B) Box plots of hybridization ratio (black) and significance (grey) for diverged genes annotated to GO categories that are over (+) or under (-) represented among the radiated lineages

divergence[3,113,117,127] (**Fig. 8**). Identification of genes that are evolving more rapidly than expected due to phylogeny alone may indicate genes under directional, or disruptive selection, and that may underlie the remarkable phenotypic diversity among cihclids (for example; substrate preference[188]; trophic structures[189,190]; breeding strategy[13,191]; courtship behavior[192-194]; coloration[195-198]). The several apparent instances of convergent evolution of phenotypic traits[176] in these adaptive radiations provide an opportunity to identify divergent loci that underlie these traits. Using aCGH data generated in Aim1, we can identify diverged genes, map these against known phenotypes, and address biased divergence based on functional gene annotations.

Aim 4: Preliminary results show shared divergence

In our genome wide aCGH analysis of radiated and non-radiated cichlid species, we also identified highly diverged genes. The detection rate was congruent with our expectations based on analysis of available cichlid sequence by sequence alignment (**Fig. 4**). We find that the great majority of diverged genes are shared among all three radiated lineages, representing either ancestral sequence divergence predating the radiations, or loci that acquired mutations in parallel during the radiation process (**Fig. 8A**). This phylogenetic pattern differs from the largely lineage specific pattern observed for gene duplicates (**Fig. 5B**). We found GO categories that were overrepresented (directional selection or reduced constraint) and categories that were underrepresented (constraint) among the diverged genes (**Fig. 8B**).

Aim 4: Experimental design for microarray analysis

The aCGH experiments from Aim1 will be interrogated for highly diverged genes using our conserved gene normalization proven effective for cross-species analysis[1,249]. Statistical determination of biased hybridization (i.e. divergence) will be conducted in R using LIMMA[250] incorporating eBayes[251] and FDR correction and the Fisher test to combine p-values across multiple probes for a single gene[252].

Gene Ontology analysis[151] with traditional hyper-geometric tests will uncover trends among functional gene categories to suggest hypotheses about gene types or pathways under selection. Annotations according to KEGG[155] or CoG[156] databases and expression modules[159-161] will also be addressed.

Highly diverged genes will be mapped onto the phylogeny and phenotype to test for 1) association of phenotypic traits and sequence divergence in specific genes and 2) sub-branches of the phylogeny with rates of sequence divergence not expected by phylogenetic distance. These analyses will be conducted with permutation based tests[253], such as Mantel tests, comparing the distance matrices constructed for each array feature based on hybridization ratio against the phylogenetic distance matrix in a two-way Mantel test, or against a phenotype distance matrix in a three-way Mantel test controlling for phylogeny, using the program PASSaGE[254]. Mantel tests offer an alternative to the well-accepted technique of Phylogeneticaly Independent Contrasts methods (PIC)[255] and others like it[256,257], which assume a Brownian motion model that is not appropriate for the directional nature of the evolution of sequence divergence. In each Mantel test, significance will be assessed by comparing the *z*-statistic of the actual matrices to the *z*-statistics from 9999 random permutations to identify array features for which sequence divergence is greater or lesser than would be expected by phylogenetic distance, and features for which divergence is associated with phenotypic trait, thus potentially indicating the contribution of natural selection[258]. It should be noted that aCGH provides only a crude measure of sequence divergence and cannot be directly related to characteristics of sequence differences.

Aim 4: Outcomes for sequence divergence

These data will provide a genome-wide perspective for broad patterns of sequence divergence across the cichlid radiations.

Aim 4: Further Considerations for sequence analysis

While cloning of diverged genes is possible, we do not include such work in this aim given the difficulty in cloning exactly those genes that are most diverged. Furthermore, the aCGH approach will miss minor

changes in DNA sequence that could have large effect on phenotype (e.g. stickleback *Ectodysplasin* armor alleles differ by only 22 nucleotides over 16Kb[259]).

Possible Future Directions Include a Sequence Based Approach

With the ever-dropping cost, increasing scope, and availability to non-model organisms, it is tempting to jump to the new DNA sequencing technology because the aCGH approach provides only relative quantitative information. The development of sequence based techniques to detect gene duplicates using depth of reads[260] in combination with capture-array techniques to enrich the sample for DNA regions of interest[261,262] could be used for gene duplicate identification and description of sequence divergence for loci of interest. Such analyses might also provide richer data with which to identify parent/daughter relations[263], more definitive classification of neofunctionalization vs. subfunctionalization[203], mutations in enhancer regions[229] and signatures of selection in protein coding regions. Currently the cost and limitations on multiplexing make such approaches unfeasible. However, we will monitor the progress of technology and adapt our methods in each aim accordingly through frequent consultation with Broad Institute researchers who support the International Cichlid Genome Consortium.

Cited Literature

1. Machado HE, Jui G, Joyce DA, Reilly CRL, Lunt DH, Renn SCP (**in review**) Gene Duplication in an African Cichlid Adaptive Radiation. *Mol Biol Evol* http://dx.doi.org/NA
2. Machado HE, Renn SCP (**in review**) A critical assessment of cross-species aCGH for detection of gene duplicates. *Genome Res* http://dx.doi.org/NA
3. Renn SCP, Machado HE, Jones A, Soneji K, Kulathinal RJ, Hofmann HA (**in review**) Using comparative genomic hybridization to survey genomic sequence divergence across species: A proof-of-concept from Drosophila. *BMC Genomics* http://dx.doi.org/NA
4. Zhang JZ (**2003**) Evolution by gene duplication: an update. *Trends Ecol Evol* **18**:292-298. http://dx.doi.org/10.1016/s0169-5347(03)00033-8
5. Haldane JBS (**1933**) The part played by recurrent mutation in evolution. *Am Nat* **67**:5-19. http://dx.doi.org/10.1038/hdy.1979.80
6. Fisher RA (**1935**) The sheltering of lethals. *Am Nat* **69**:446-455. http://dx.doi.org/10.1086/280618
7. Ohno S. Evolution by Gene Duplication: Springer-Verlag; **1970**. isbn#: 0-04-575015-7
8. Hahn MW (**2009**) Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. *J Hered* **100**:605-617. http://dx.doi.org/10.1093/jhered/esp047
9. Taylor JS, Raes J (**2004**) Duplication and divergence: The evolution of new genes and old ideas. *Annu Rev Genet* **38**:615-643. http://dx.doi.org/10.1146/annurev.genet.38.072902.092831
10. Stoltzfus A (**1999**) On the possibility of constructive neutral evolution. *J Mol Evol* **49**:169-181. http://dx.doi.org/10.1554/04-543.1
11. Dykhuizen D, Hartl DL (**1980**) Selective Neutrality of 6pgd Allozymes in Escherichia-Coli and the Effects of Genetic Background. *Genetics* **96**:801-817.
12. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (**1999**) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**:1531-1545.
13. Barlow GW. Cichlid Fishes: Nature's Grand Experiment in Evolution: Harper Collins Publishers; **2000**. isbn#: 10: 0738205281
14. Seehausen O (**2006**) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B-Biological Sciences* **273**:1987-1998. http://dx.doi.org/10.1098/rspb.2006.3539
15. Consortium ICG (**2006**) Genetic Baisis of Vertebrate Diversity: the Cichlid Fish Model. *NHGRI White Paper* http://dx.doi.org/http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/CichlidGenomeSeq.pdf
16. Dobzhansky T. Genetics and the origin of species: Columbia University; **1937**. isbn#: 0-231-05475-0
17. Mayr E. Animal species and evolution: Belknap Press of Harvard University Press Cambridge; **1963**. isbn#: 10: 0674037502
18. Schluter D. The ecology of adaptive radiation: Oxford University Press; **2000**. isbn#: 10: 0198505221
19. Gould SJ. Wonderful Life: The Burgess Shale and the Nature of History: W.W. Norton; **1989**. isbn#: %(
20. Darwin C. The Origin of Species: Bantam Books; **1859**. isbn#: 10: 0451529065
21. Fryer G (**1990**) Comparative aspects of adaptive radiation and speciation in Lake Baikal and the great rift lakes of Africa. *Hydrobiologia* **211**:137-146. http://dx.doi.org/10.1086/397375
22. Losos JB, Jackman TR, Larson A, de Queiroz K, Rodriguez S (**1998**) Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**:2115-2118. http://dx.doi.org/10.1126/science.279.5359.2115
23. Baldwin BG, Sanderson MJ (**1998**) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc Natl Acad Sci U S A* **95**:9402-9406.

24. Fryer G, Iles TD. The cichlid fishes of the Great Lakes of Africa: Their biology and evolution: Oliver & Boyd, Croythorn House, 23 Ravelston Terrace, Edinburgh; **1972**. isbn#: 0876660308
25. Turner GF (**2007**) Adaptive radiation of cichlid fish. *Curr Biol* **17**:R827-R831.
26. Losos JB, Miles DB (**2002**) Testing the hypothesis that a clade has adaptively radiated: Iguanid lizard clades as a case study. *Am Nat* **160**:147-157. http://dx.doi.org/10.1086/341557
27. Eldredge N, Thompson JN, Brakefield PM, Gavrilets S, Jablonski D, Jackson JBC, Lenski RE, Lieberman BS, McPeek MA, Miller W (**2005**) The dynamics of evolutionary stasis. *Paleobiology* **31**:133-145. http://dx.doi.org/10.1666/0094-8373(2005)031[0133:TDOES]2.0.CO;2
28. Seehausen O (**2004**) Hybridization and adaptive radiation. *Trends Ecol Evol* **19**:198-207. http://dx.doi.org/10.1016/j.tree.2004.01.003
29. Grant PR, Grant BR. How and Why Species Multiply: The Radiaiton of Darwin's finches. Princeton, NJ: Princeton University Press; **2008**. isbn#: 10: 0691133603
30. Losos JB. Lizards in an Evolutionary Tree:Ecology and Adaptive Radiation of Anoles. Berkeley, CA: University of California Press; **2009**. isbn#: 9780520255913 0520255917
31. Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL (**2001**) How many species of cichlid fishes are there in African lakes? *Mol Ecol* **10**:793-806. http://dx.doi.org/10.1046/j.1365-294x.2001.01200.x
32. Joyce DA, Lunt DH, Bills R, Turner GF, Katongo C, Duftner N, Sturmbauer C, Seehausen O (**2005**) An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature* **435**:90-95. http://dx.doi.org/10.1038/nature03489
33. Kocher TD (**2004**) Adaptive evolution and explosive speciation: The cichlid fish model. *Nat Rev Genet* **5**:288-298. http://dx.doi.org/doi:10.1038/nrg1316
34. Sturmbauer C, Baric S, Salzburger W, Ruber L, Verheyen E (**2001**) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Mol Biol Evol* **18**:144-154. http://dx.doi.org/10.1098/rspb.2002.2188
35. Verheyen E, Salzburger W, Snoeks J, Meyer A (**2003**) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science* **300**:325-329. http://dx.doi.org/10.1126/science.1080699
36. Rundle HD, Nagel L, Boughman JW, Schluter D (**2000**) Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**:306-308. http://dx.doi.org/10.1126/science.287.5451.306
37. Salzburger W (**2009**) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Mol Ecol* **18**:169-185. http://dx.doi.org/10.1111/j.1365-294X.2008.03981.x
38. Templeton AR (**2008**) The reality and importance of founder speciation in evolution. *BioEssays* **30**:470-479. http://dx.doi.org/10.1002/bies.20745
39. Werth CR, Windham MD (**1991**) A Model for Divergent, Allopatric Speciation of Polyploid Pteridophytes Resulting from Silencing of Duplicate-Gene Expression. *Am Nat* **137**:515-526. http://dx.doi.org/10.1086/285180
40. Lynch M, Force AG (**2000**) The origin of interspecific genomic incompatibility via gene duplication. *Am Nat* **156**:590-605. http://dx.doi.org/10.1086/316992
41. Paques F, Haber JE (**1999**) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **63**:349-+. http://dx.doi.org/1092-2172/99/$04.00+0
42. Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC (**2006**) Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. *Trends Genet* **22**:597-602. http://dx.doi.org/10.1016/j.tig.2006.09.003
43. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (**2005**) Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A* **102**:5454-5459. http://dx.doi.org/10.1073/pnas.0501102102

44. Brosius J (**1991**) Retroposons - Seeds of Evolution. *Science* **251**:753-753. http://dx.doi.org/10.1126/science.1990437
45. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE (**2001**) Segmental duplications: Organization and impact within the current Human Genome Project assembly. *Genome Res* **11**:1005-1017. http://dx.doi.org/10.1101/gr.187101
46. Katju V, Lynch M (**2003**) The structure and early evolution of recently arisen gene duplicates in the Caenorhabditis elegans genome. *Genetics* **165**:1793-1803.
47. Nozawa M, Nei M (**2007**) Evolutionary dynamics of olfactory receptor genes in Drosophila species. *Proc Natl Acad Sci U S A* **104**:7122-7127. http://dx.doi.org/10.1073/pnas.0702133104
48. Lynch M, O'Hely M, Walsh B, Force A (**2001**) The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**:1789-1804. http://dx.doi.org/PMCID: PMC1461922
49. Lynch M, Conery JS (**2000**) The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151-1155. http://dx.doi.org/10.1126/science.290.5494.115
50. Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW (**2009**) Adaptive evolution of young gene duplicates in mammals. *Genome Res* **19**:859-867. http://dx.doi.org/10.1101/gr.085951.108
51. Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, Brenton M, Hink R, Burgers S, Hernandez-Boussard T and others (**2004**) Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol* **2**:937-954. http://dx.doi.org/10.1371/journal.pbio.0020207
52. Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang ZS, Baker C, Malfavon-Borja R, Fulton LA and others (**2009**) A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**:877-881. http://dx.doi.org/10.1038/nature07744
53. Fan CZ, Vibranovski MD, Chen Y, Long MY (**2007**) A microarray based genomic hybridization method for identification of new genes in plants: Case analyses of Arabidopsis and Oryza. *J Integr Plant Biol* **49**:915-926. http://dx.doi.org/10.1111/j.1672-9072.2007.00503.x
54. Flagel LE, Wendel JF (**2009**) Gene duplication and evolutionary novelty in plants. *New Phytol* **183**:557-564. http://dx.doi.org/10.1111/j.1469-8137.2009.02923.x
55. Carson AR, Scherer SW (**2009**) Identifying concerted evolution and gene conversion in mammalian gene pairs lasting over 100 million years. *BMC Evol Biol* **9**: http://dx.doi.org/10.1186/1471-2148-9-156
56. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R and others (**2007**) Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **39**:1256-1260. http://dx.doi.org/10.1038/ng2123
57. Zhang JZ, Rosenberg HF, Nei M (**1998**) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* **95**:3708-3713. http://dx.doi.org/10.1073/pnas.95.7.3708
58. Zhang JZ, Zhang YP, Rosenberg HF (**2002**) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet* **30**:411-415. http://dx.doi.org/10.1038/ng852
59. Fischer HM, Wheat CW, Heckel DG, Vogel H (**2008**) Evolutionary origins of a novel host plant detoxification gene in butterflies. *Mol Biol Evol* **25**:809-820. http://dx.doi.org/10.1093/molbev/msn014
60. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P and others (**2002**) A single P450 allele associated with insecticide resistance in Drosophila. *Science* **297**:2253-2256. http://dx.doi.org/10.1126/science.1074170
61. Labbe P, Berthomieu A, Berticat C, Alout H, Raymond M, Lenormand T, Weill M (**2007**) Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito Culex pipiens. *Mol Biol Evol* **24**:1056-1067. http://dx.doi.org/10.1093/molbev/msm025
62. Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J (**2006**) Positive selection driving diversification in plant secondary metabolism. *Proc Natl Acad Sci U S A* **103**:9118-9123. http://dx.doi.org/10.1073/pnas.0601738103

63. Ting CT, Tsaur SC, Sun S, Browne WE, Chen YC, Patel NH, Wu CL (**2004**) Gene duplication and speciation in Drosophila: Evidence from the Odysseus locus. *Proc Natl Acad Sci U S A* **101**:12232-12235. http://dx.doi.org/10.1073/pnas.0401975101
64. Chen ZZ, Cheng CHC, Zhang JF, Cao LX, Chen L, Zhou LH, Jin YD, Ye H, Deng C, Dai ZH and others (**2008**) Transcrintomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc Natl Acad Sci U S A* **105**:12944-12949. http://dx.doi.org/10.1073/pnas.0802432105
65. Sandve SR, Rudi H, Asp T, Rognli OA (**2008**) Tracking the evolution of a cold stress associated gene family in cold tolerant grasses. *BMC Evol Biol* **8**: http://dx.doi.org/10.1186/1471-2148-8-245
66. Monson RK (**2003**) Gene duplication, neofunctionalization, and the evolution of C4 photosynthesis. *Int J Plant Sci* **164**:S43-S54. http://dx.doi.org/10.1186/gb-2009-10-6-r68
67. Hughes DA, Jastroch M, Stoneking M, Klingenspor M (**2009**) Molecular evolution of UCP1 and the evolutionary history of mammalian non-shivering thermogenesis. *BMC Evol Biol* **9**: http://dx.doi.org/10.1186/1471-2148-9-4
68. Bailey JA, Church DM, Ventura M, Rocchi M, Eichler EE (**2004**) Analysis of segmental duplications and genome assembly in the mouse. *Genome Res* **14**:789-801. http://dx.doi.org/10.1101/gr.2238404
69. Bailey JA, Gu ZP, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (**2002**) Recent segmental duplications in the human genome. *Science* **297**:1003-1007. http://dx.doi.org/10.1126/science.1072047
70. Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, Sikela JM (**2007**) Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res* **17**:1266-1277. http://dx.doi.org/10.1101/gr.6557307
71. Feuk L, Carson AR, Scherer SW (**2006**) Structural variation in the human genome. *Nat Rev Genet* **7**:85-97. http://dx.doi.org/10.1038/nrg1767
72. Hahn MW, Han MV, Han SG (**2007**) Gene family evolution across 12 Drosophila genomes. *PLoS Genetics* **3**:2135-2146. http://dx.doi.org/10.1371/journal.pgen.0030197
73. Moore RC, Purugganan MD (**2005**) The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol* **8**:122-128. http://dx.doi.org/10.1016/j.pbi.2004.12.001
74. Chen WK, Swartz JD, Rush LJ, Alvarez CE (**2009**) Mapping DNA structural variation in dogs. *Genome Res* **19**:500-509. http://dx.doi.org/10.1101/gr.083741.108
75. Vali U, Brandstrom M, Johansson M, Ellegren H (**2008**) Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC Genet* **9**: http://dx.doi.org/10.1186/1471-2156-9-8
76. Watanabe M, Hiraide K, Okada N (**2007**) Functional diversification of kir7.1 in cichlids accelerated by gene duplication. *Gene* **399**:46-52. http://dx.doi.org/10.1016/j.gene.2007.04.024
77. Sugie A, Terai Y, Ota R, Okada N (**2004**) The evolution of genes for pigmentation in African cichlid fishes. *Gene* **343**:337-346. http://dx.doi.org/10.1016/j.gene.2004.09.019
78. Braasch I, Salzburger W, Meyer A, Ub (**2006**) Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. *Mol Biol Evol* **23**:1192-1202. http://dx.doi.org/10.1093/molbev/msk003
79. Carleton KL, Kocher TD (**2001**) Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Mol Biol Evol* **18**:1540-1550. http://dx.doi.org/10.1016/j.cub.2005.08.010
80. Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HDJ, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H and others (**2008**) Speciation through sensory drive in cichlid fish. *Nature* **455**:620-U23. http://dx.doi.org/10.1038/nature07285
81. Spady TC, Parry JWL, Robinson PR, Hunt DM, Bowmaker JK, Carleton KL, Df (**2006**) Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. *Mol Biol Evol* **23**:1538-1547. http://dx.doi.org/10.1093/molbev/msl014

82. Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M, Konijnendijk N, Mrosso HDJ and others (**2006**) Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol* **4**:2244-2251. http://dx.doi.org/10.1371/journal.pbio.0040433
83. Cnaani A, Kocher TD (**2008**) Sex-linked markers and microsatellite locus duplication in the cichlid species Oreochromis tanganicae. *Biology Letters* **4**:700-703. http://dx.doi.org/10.1098/rsbl.2008.0286
84. Shirak A, Golik M, Lee BY, Howe AE, Kocher TD, Hulata G, Ron M, Seroussi E (**2008**) Copy number variation of lipocalin family genes for male-specific proteins in tilapia and its association with gender. *Heredity* **101**:405-415. http://dx.doi.org/10.1038/hdy.2008.68
85. Chen CC, Fernald RD, Ff (**2006**) Distributions of two gonadotropin-releasing hormone receptor types in a cichlid fish suggest functional specialization. *J Comp Neurol* **495**:314-323. http://dx.doi.org/10.1002/cne.20827
86. Santini S, Bernardi G (**2005**) Organization and base composition of tilapia Hox genes: implications for the evolution of Hox clusters in fish. *Gene* **346**:51-61. http://dx.doi.org/10.1016/j.gene.2004.10.027
87. Nei M, Niimura Y, Nozawa M (**2008**) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* **9**:951-963. http://dx.doi.org/10.1038/nrg2480
88. Dean EJ, Davis JC, Davis RW, Petrov DA (**2008**) Pervasive and Persistent Redundancy among Duplicated Genes in Yeast. *Plos Genetics* **4**: http://dx.doi.org/10.1371/journal.pgen.1000113
89. Spofford JB (**1969**) Heterosis and Evolution of Duplications. *Am Nat* **103**:407-&. http://dx.doi.org/10.1086/282611
90. Hughes AL (**1994**) The Evolution of Functionally Novel Proteins after Gene Duplication. *Proceedings of the Royal Society of London Series B-Biological Sciences* **256**:119-124. http://dx.doi.org/10.1098/rspb.1994.0058
91. Bershtein S, Tawfik DS (**2008**) Ohno's Model Revisited: Measuring the Frequency of Potentially Adaptive Mutations under Various Mutational Drifts. *Mol Biol Evol* **25**:2311-2318. http://dx.doi.org/10.1093/molbev/msn174
92. Gibson TA, Goldberg DS (**2009**) Questioning the Ubiquity of Neofunctionalization. *Plos Computational Biology* **5**: http://dx.doi.org/10.1371/journal.pcbi.1000252
93. Escriva H, Bertrand S, Germain P, Robinson-Rechavi M, Umbhauer M, Cartry J, Duffraisse M, Holland L, Gronemeyer H, Laudet V (**2006**) Neofunctionalization in vertebrates: The example of retinoic acid receptors. *PLoS Genetics* **2**:955-965. http://dx.doi.org/10.1371/journal.pgen.0020102
94. Maston GA, Ruvolo M (**2002**) Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol Biol Evol* **19**:320-335.
95. Yokoyama S, Yokoyama R (**1996**) Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Annu Rev Ecol Syst* **27**:543-567. http://dx.doi.org/10.1146/annurev.ecolsys.27.1.543
96. Matsuo T (**2008**) Rapid evolution of two odorant-binding protein genes, obp57d and obp57e, in the Drosophila melanogaster species group. *Genetics* **178**:1061-1072. http://dx.doi.org/10.1534/genetics.107.079046
97. Des Marais DL, Rausher MD (**2008**) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**:762-U85. http://dx.doi.org/10.1038/nature07092
98. Lynch M, Conery JS (**2003**) The origins of genome complexity. *Science* **302**:1401-1404. http://dx.doi.org/10.1126/science.1089370
99. Hahn MW, Demuth JP, Han SG (**2007**) Accelerated rate of gene gain and loss in primates. *Genetics* **177**:1941-1949. http://dx.doi.org/10.1534/genetics.107.080077
100. Lynch M. The Origins of Genome Architecture. Sunderland, MA: Sinauer Associates; **2007**. isbn#: 978-0-87893-484-3

101. Pinkel D, Albertson DG (**2005**) Comparative genomic hybridization. *Annu Rev Genomics Hum Genet* **6**:331-354. http://dx.doi.org/10.1146/annurev.genom.6.080604.162140
102. Roukos DH (**2008**) Innovative genomic-based model for personalized treatment of gastric cancer: integrating current standards and new technologies. *Expert Review of Molecular Diagnostics* **8**:29-39. http://dx.doi.org/10.1586/14737159.8.1.29
103. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J and others (**2007**) The genomic landscapes of human breast and colorectal cancers. *Science* **318**:1108-1113. http://dx.doi.org/10.1126/science.1145720
104. Eichler EE (**2001**) Segmental duplications: What's missing, misassigned, and misassembled - and should we care? *Genome Res* **11**:653-656. http://dx.doi.org/10.1101/gr.188901
105. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod HL, Cheverud JM and others (**2007**) A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genetics* **3**: http://dx.doi.org/10.1371/journal.pgen.0030003
106. Gresham D, Dunham MJ, Botstein D (**2008**) Comparing whole genomes using DNA microarrays. *Nat Rev Genet* **9**:291-302. http://dx.doi.org/10.1038/nrg2335
107. Edwards-Ingram LC, Gent ME, Hoyle DC, Hayes A, Stateva LI, Oliver SG (**2004**) Comparative genomic hybridization provides new insights into the molecular taxonomy of the Saccharomyces sensu stricto complex. *Genome Res* **14**:1043-1051. http://dx.doi.org/10.1101/gr.2114704
108. Yoon ST, Xuan ZY, Makarov V, Ye K, Sebat J (**2009**) Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* **19**:1586-1592. http://dx.doi.org/10.1101/gr.092981.109
109. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen WW and others (**2006**) Global variation in copy number in the human genome. *Nature* **444**:444-454. http://dx.doi.org/10.1038/nature05329
110. Bloomfield G, Tanaka Y, Skelton J, Ivens A, Kay RR (**2008**) Widespread duplications in the genomes of laboratory stocks of Dictyostelium discoideum. *Genome Biology* **9**: http://dx.doi.org/10.1186/gb-2008-9-4-r75
111. Dopman EB, Hartl DL (**2007**) A portrait of copy-number polymorphism in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **104**:19920-19925. http://dx.doi.org/10.1073/pnas.0709888104
112. Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL and others (**2008**) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A* **105**:9272-9277. http://dx.doi.org/10.1073/pnas.0803466105
113. Brunelle BW, Nicholson TL, Stephens RS (**2004**) Microarray-based genomic surveying of gene polymorphisms in Chlamydia trachomatis. *Genome Biology* **5**:9. http://dx.doi.org/10.1186/gb-2004-5-6-r42
114. Giuntini E, Mengoni A, De Filippo C, Cavalieri D, Aubin-Horth N, Landry CR, Becker A, Bazzicalupo M (**2005**) Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of Sinorhizobium meliloti natural strains. *BMC Genomics* **6**: http://dx.doi.org/10.1186/1471-2164-6-158
115. Janvilisri T, Scaria J, Thompson AD, Nicholson A, Limbago BM, Arroyo LG, Songer JG, Grohn YT, Chang YF (**2009**) Microarray Identification of Clostridium difficile Core Components and Divergent Regions Associated with Host Origin. *J Bacteriol* **191**:3881-3891. http://dx.doi.org/10.1128/jb.00222-09
116. Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ (**2002**) Comparative genomic analysis of Vibrio cholerae: Genes that correlate with cholera endemic and pandemic disease. *Proc Natl Acad Sci U S A* **99**:1556-1561. http://dx.doi.org/10.1073/pnas.042667999
117. Hinchliffe SJ, Isherwood KE, Stabler RA, Prentice MB, Rakin A, Nichols RA, Oyston PCF, Hinds J, Titball RW, Wren BW (**2003**) Application of DNA microarrays to study the evolutionary

genomics of Yersinia pestis and Yersinia pseudotuberculosis. *Genome Res* **13**:2018-2029. http://dx.doi.org/10.1101/gr.1507303
118. Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N, Small PM (**2001**) Comparing genomes within the species Mycobacterium tuberculosis. *Genome Res* **11**:547-554. http://dx.doi.org/10.1101/gr166401
119. Zhou XHJ, Gibson G (**2004**) Cross-species comparison of genome-wide expression patterns. *Genome Biology* **5**: http://dx.doi.org/10.1186/gb-2004-5-7-232
120. Riehle MM, Markianos K, Niare O, Xu JN, Li J, Toure AM, Podiougou B, Oduol F, Diawara S, Diallo M and others (**2006**) Natural malaria infection in Anopheles gambiae is regulated by a single genomic control region. *Science* **312**:577-579. http://dx.doi.org/10.1126/science.1124153
121. Turner TL, Hahn MW, Nuzhdin SV (**2005**) Genomic islands of speciation in Anopheles gambiae. *PLoS Biol* **3**:1572-1578. http://dx.doi.org/10.1371/journal.pbio.0030285
122. Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, Pinkel D, Eichler EE (**2003**) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* **13**:347-357. http://dx.doi.org/10.1101/gr.1003303
123. Le Quere A, Eriksen KA, Rajashekar B, Schutzenbubel A, Canback B, Johansson T, Tunlid A (**2006**) Screening for rapidly evolving genes in the ectomycorrhizal fungus Paxillus involutus using cDNA microarrays. *Mol Ecol* **15**:535-550. http://dx.doi.org/10.1111/j.1365-294X.2005.02796
124. Murray AE, Lies D, Li G, Nealson K, Zhou J, Tiedje JM (**2001**) DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. *Proc Natl Acad Sci U S A* **98**:9853-9858. http://dx.doi.org/10.1073/pnas.171178898
125. Watanabe T, Murata Y, Oka S, Iwahashi H (**2004**) A new approach to species determination for yeast strains: DNA microarray-based comparative genomic hybridization using a yeast DNA microarray with 6000 genes. *Yeast* **21**:351-365. http://dx.doi.org/10.1002/yea.1103
126. Porwollik S, Wong RMY, McClelland M (**2002**) Evolutionary genomics of Salmonella: Gene acquisitions revealed by microarray analysis. *Proc Natl Acad Sci U S A* **99**:8956-8961. http://dx.doi.org/10.1073/pnas.122153699
127. Taboada EN, Acedillo RR, Luebbert CC, Findlay WA, Nash JHE (**2005**) A new approach for the analysis of bacterial microarray-based Comparative Genomic Hybridization: insights from an empirical study. *BMC Genomics* **6**:1-10. http://dx.doi.org/doi:10.1186/1471-2164-6-78
128. Dong J, Horvath S (**2007**) Understanding network concepts in modules. *BMC Systems Biology* **1**: http://dx.doi.org/10.1186/1752-0509-1-24
129. Kim CC, Joyce EA, Chan K, Falkow S (**2002**) Improved analytical methods for microarray-based genome-composition analysis. *Genome Biology* **3**:RESEARCH0065. http://dx.doi.org/10.1186/gb-2002-3-11-research0065
130. Dong YM, Glasner JD, Blattner FR, Triplett EW (**2001**) Genomic interspecies microarray hybridization: Rapid discovery of three thousand genes in the maize endophyte, Klebsiella pneumoniae 342, by microarray hybridization with Escherichia coli K-12 open reading frames. *Appl Environ Microbiol* **67**:1911-1921. http://dx.doi.org/10.1128/AEM.67.4.1911-1921.2001
131. Renn SCP, Aubin-Horth N, Hofmann HA (**2004**) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMCGenomics* **5**: http://dx.doi.org/10.1186/1471-2164-5-42
132. Salzburger W, Renn SCP, Steinke D, Braasch I, Hofmann HA, Meyer A (**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. http://dx.doi.org/10.1186/1471-2164-9-96
133. Flibotte S, Moerman DG (**2008**) Experimental analysis of oligonucleotide microarray design criteria to detect deletions by comparative genomic hybridization. *BMC Genomics* **9**: http://dx.doi.org/10.1186/1471-2164-9-497

134. Sharp AJ, Itsara A, Cheng Z, Alkan C, Schwartz S, Eichler EE (**2007**) Optimal design of oligonucleotide microarrays for measurement of DNA copy-number. *Hum Mol Genet* **16**:2770-2779. http://dx.doi.org/10.1093/hmg/ddm234
135. Kijimoto T, Watanabe M, Fujimura K, Nakazawa M, Murakami Y, Kuratani S, Kohara Y, Gojobori T, Okada N (**2005**) cimp1, a novel astacin family metalloproteinase gene from East African cichlids, is differentially expressed between species during growth. *Mol Biol Evol* **22**:1649-1660. http://dx.doi.org/10.1093/molbev/msi159
136. Watanabe M, Kobayashi N, Shin-I T, Horiike T, Tateno Y, Kohara Y, Okada N (**2004**) Extensive analysis of ORF sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock - Identity of EST sequences between Haplochromis chilotes and Haplochromis sp "Redtailsheller". *Gene* **343**:263-269.
137. Kobayashi N, Watanabe M, Horiike T, Kohara Y, Okada N (**2009**) Extensive analysis of EST sequences reveals that all cichlid species in Lake Victoria share almost identical transcript sets. *Gene* **441**:187-191. http://dx.doi.org/10.1016/j.gene.2008.11.023
138. Loh Y-HE, Katz LS, Mims MC, Kocher TD, Yi SV, Streelman JT (**2008**) Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biology* **9**:Unpaginated. http://dx.doi.org/10.1186/gb-2008-9-7-r113
139. Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, Tsang P, Curry B, Baird K, Meltzer PS and others (**2004**) Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci U S A* **101**:17765-17770. http://dx.doi.org/10.1073/pnas.0407979101
140. Koblmuller S, Sefc KM, Sturmbauer C (**2008**) The Lake Tanganyika cichlid species assemblage: recent advances in molecular phylogenetics. *Hydrobiologia* **615**:5-20. http://dx.doi.org/10.1007/s10750-008-9552-4
141. 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 Evol Biol* **5**:15. http://dx.doi.org/10.1186/1471-2148-5-17
142. Curtis C, Lynch AG, Dunning MJ, Spiteri I, Marioni JC, Hadfield J, Chin SF, Brenton JD, Tavare S, Caldas C (**2009**) The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* **10**:e588. http://dx.doi.org/10.1186/1471-2164-10-588
143. Miller KM, Kaukinen KH, Schulze AD (**2002**) Expansion and contraction of major histocompatibility complex genes: a teleostean example. *Immunogenetics* **53**:941-963. http://dx.doi.org/10.1007/s00251-001-0398-4
144. Cardoso JCR, Power DM, Elgar G, Clark MS (**2004**) Dupilicated receptors for VIP and PACAP (VPAC(1)R and PAC(1)R) in a teleost fish, Fugu rubripes. *J Mol Endocrinol* **33**:411-428. http://dx.doi.org/10.1677/jme.1.01575
145. van der Aa LM, Levraud JP, Yahmi M, Lauret E, Briolat V, Herbomel P, Benmansour A, Boudinot P (**2009**) A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. *BMC Biology* **7**: http://dx.doi.org/10.1186/1741-7007-7-7
146. Beuschlein F, Keegan CE, Bavers DL, Mutch C, Hutz JE, Shah S, Ulrich-Lai YM, Engeland WC, Jeffs B, Jameson JL and others (**2002**) SF-1, DAX-1, and ACD: Molecular determinants of adrenocortical growth and steroidogenesis. *Endocr Res* **28**:597-607.
147. Schwarzer J, Misof B, Tautz D, Schliewen UK (**2009**) The root of the East African cichlid radiations. *BMC Evol Biol* **9**: http://dx.doi.org/10.1186/1471-2148-9-186
148. Mefford HC, Eichler EE (**2009**) Duplication hotspots, rare genomic disorders, and common disease. *Curr Opin Genet Dev* **19**:196-204. http://dx.doi.org/10.1016/j.gde.2009.04.003
149. Bailey JA, Eichler EE (**2006**) Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet* **7**:552-564. http://dx.doi.org/10.1038/nrg1895

150. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R and others (**2005**) Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77**:78-88. http://dx.doi.org/10.1086/431652
151. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT and others (**2000**) Gene Ontology: tool for the unification of biology. *Nat Genet* **25**:25-29. http://dx.doi.org/10.1038/75556
152. Beissbarth T, Speed TP (**2004**) GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**:1464-1465. http://dx.doi.org/10.1093/bioinformatics/bth088
153. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV (**2002**) Selection in the evolution of gene duplications. *Genome Biol* **3**:RESEARCH0008.
154. Prachumwat A, Li WH (**2006**) Protein function, connectivity, and duplicability in yeast. *Mol Biol Evol* **23**:30-39. http://dx.doi.org/10.1093/molbev/msi249
155. Kanehisa M, Goto S (**2000**) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**:27-30. http://dx.doi.org/10.1093/nar/27.1.29
156. Tatusov RL, Galperin MY, Natale DA, Koonin EV (**2000**) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**:33-36.
157. Makino T, Hokamp K, McLysaght A (**2009**) The complex relationship of gene duplication and essentiality. *Trends Genet* **25**:152-155. http://dx.doi.org/10.1016/j.tig.2009.03.001
158. Korbel JO, Kim PM, Chen X, Urban AE, Weissman S, Snyder M, Gerstein MB (**2008**) The current excitement about copy-number variation: how it relates to gene duplications and protein families. *Curr Opin Struct Biol* **18**:366-374. http://dx.doi.org/10.1016/j.sbi.2008.02.005
159. Aubin-Horth N, Desjardins JK, Martei YM, Balshine S, Hofmann HA (**2007**) Masculinized dominant females in a cooperatively breeding species. *Mol Ecol* **16**:1349-1358. http://dx.doi.org/10.1111/j.1365-294X.2007.03249.x
160. Aubin-Horth N, Desjardins JK, Martei YM, Balshine S, Hofmann HA (**2005**) Dominance status, sex and gene expression: The case of the masculinized female brain. *Horm Behav* **48**:87-87.
161. Renn SCP, Aubin-Horth N, Hofmann HA (**2008**) Fish and chips: functional genomics of social plasticity in an African cichlid fish. *J Exp Biol* **211**:3041-3056. http://dx.doi.org/10.1242/jeb.018242
162. Clark TA, Townsend JP (**2007**) Quantifying variation in gene expression. *Mol Ecol* **16**:2613-2616. http://dx.doi.org/10.1111/j.1365-294X.2007.03354.x
163. Shakhnovich BE (**2006**) Relative contributions of structural designability and functional diversity in molecular evolution of duplicates. *Bioinformatics* **22**:E440-E445. http://dx.doi.org/10.1093/bioinformatics/btl211
164. Darai-Ramqvist E, Sandlund A, Muller S, Klein G, Imreh S, Kost-Alimova M (**2008**) Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions. *Genome Res* **18**:370-379. http://dx.doi.org/10.1101/gr.7010208
165. Li L, Huang YW, Xia XF, Sun ZR (**2006**) Preferential duplication in the sparse part of yeast protein interaction network. *Mol Biol Evol* **23**:2467-2473. http://dx.doi.org/10.1093/molbev/msl121
166. Liao BY, Zhang JZ (**2007**) Mouse duplicate genes are as essential as singletons. *Trends Genet* **23**:378-381. http://dx.doi.org/10.1016/j.tig.2007.05.006
167. Meisel RP (**2009**) Evolutionary Dynamics of Recently Duplicated Genes: Selective Constraints on Diverging Paralogs in the Drosophila pseudoobscura Genome. *J Mol Evol* **69**:81-93. http://dx.doi.org/10.1007/s00239-009-9254-1
168. Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA and others (**2008**) The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* **82**:685-695. http://dx.doi.org/10.1016/j.ajhg.2007.12.010
169. She XW, Cheng Z, Zollner S, Church DM, Eichler EE (**2008**) Mouse segmental duplication and copy number variation. *Nat Genet* **40**:909-914. http://dx.doi.org/10.1038/ng.172

170. Pan D, Zhang L (**2009**) An Atlas of the Speed of Copy Number Changes in Animal Gene Families and Its Implications. *PLoS One* **4**:Article No.: e7342. http://dx.doi.org/:10.1371/journal.pone.0007342
171. Zhang F, Gu WL, Hurles ME, Lupski JR (**2009**) Copy Number Variation in Human Health, Disease, and Evolution. *Annu Rev Genomics Hum Genet* **10**:451-481. http://dx.doi.org/10.1146/annurev.genom.9.081307.164217
172. Tam GWC, Redon R, Carter NP, Grant SGN (**2009**) The Role of DNA Copy Number Variation in Schizophrenia. *Biol Psychiatry* **66**:1005-1012. http://dx.doi.org/:10.1016/j.biopsych.2009.07.027
173. Merikangas AK, Corvin AP, Gallagher L (**2009**) Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet* **25**:536-544. http://dx.doi.org/10.1016/j.tig.2009.10.006
174. Perry AN, Grober MS (**2003**) A model for social control of sex change: interactions of behavior, neuropeptides, glucocorticoids, and sex steroids. *Horm Behav* **43**:31-38. http://dx.doi.org/10.1016/S0018-506X(02)00036-3
175. Clabaut C, Salzburger W, Meyer A (**2005**) Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: Nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. *J Mol Evol* **61**:666-681. http://dx.doi.org/10.1007/s00239-004-0217-2
176. Kocher TD, Conroy JA, McKaye KR, Stauffer JR (**1993**) Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Mol Phylogenet Evol* **2**:158-165. http://dx.doi.org/10.1006/mpev.1993.1016
177. Livak KJ, Schmittgen TD (**2001**) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**:402-408. http://dx.doi.org/10.1006/meth.2001.1262
178. Bollback JP (**2006**) SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**: http://dx.doi.org/10.1186/1471-2105-7-88
179. Nielsen R (**2002**) Mapping mutations on phylogenies. *Syst Biol* **51**:729-739. http://dx.doi.org/10.1080/10635150290102393
180. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 2.72 ::: http://mesquiteproject.org. 2009.
181. Gruar D, Li WH. Fundamentals of Molecular Evolution: Sinaure Assoc.; **2000**. isbn#: 0878932666
182. Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF (**2007**) Age of cichlids: New dates for ancient lake fish radiations. *Mol Biol Evol* **24**:1269-1282. http://dx.doi.org/10.1093/molbev/msm050
183. De Bie T, Cristianini N, Demuth JP, Hahn MW (**2006**) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269-1271. http://dx.doi.org/10.1093/bioinformatics/btl097
184. Swofford RW, Loh YHE, Streelman JT, Di Palma F, Lindblad-Toh K (**2009**) Deep segregation of single nucleotide polymorphisms (SNPs) in East African cichlids. *Integrative and Comparative Biology* **49**:E312-E312.
185. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC and others (**2005**) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**:803-819. http://dx.doi.org/10.1038/nature04338
186. Lawton-Rauh A (**2003**) Evolutionary dynamics of duplicated genes in plants. *Mol Phylogenet Evol* **29**:396-409. http://dx.doi.org/10.1016/j.ympev.2003.07.004
187. Lawton-Rauh A, Robichaux RH, Purugganan MD (**2007**) Diversity and divergence patterns in regulatory genes suggest differential gene flow in recently derived species of the Hawaiian

silversword alliance adaptive radiation (Asteraceae). *Mol Ecol* **16**:3995-4013. http://dx.doi.org/10.1111/j.1365-294X.2007.03445.x
188. Greenwood PH (**1980**) Toward a phyletic classification of the 'genus' Haplochromis (Pisces, Cichlidae) and related taxa. *Bulletin of the British Museum of Natural History and Zoology* **39**:1-101.
189. Liem KF, Osse JWM (**1975**) Biological Versatility, Evolution, and Food Resource Exploitation in African Cichlid Fishes. *Am Zool* **15**:427-454. http://dx.doi.org/10.1093/icb/15.2.427
190. Liem KF (**1980**) Adaptive Significance of Intraspecific and Interspecific Differences in the Feeding Repertoires of Cichlid Fishes. *Am Zool* **20**:295-314.
191. Koblmuller S, Salzburger W, Sturmbauer C (**2004**) Evolutionary relationships in the sand-dwelling cichlid lineage of lake Tanganyika suggest multiple colonization of rocky habitats and convergent origin of biparental mouthbrooding. *J Mol Evol* **58**:79-96. http://dx.doi.org/10.1007/s00239-003-2527-1
192. Oliveira RF, Ros AFH, Hirschenhauser K, Canario AVM (**2001**) Androgens and mating systems in fish: Intra- and inter-specific analysis. *Perspective in Comparative Endocrinology: Unity and Diversity* 985-993.
193. Pollen AA, Dobberfuhl AP, Scace J, Igulu MM, Renn SCP, Shumway CA, Hofmann HA (**2007**) Environmental complexity and social organization sculpt the brain in Lake Tanganyikan cichlid fish. *Brain Behav Evol* **70**:21-39. http://dx.doi.org/10.1159/000101067
194. Kidd MR, Kidd CE, Kocher TD (**2006**) Axes of differentiation in the bower-building cichlids of Lake Malawi. *Mol Ecol* **15**:459-478. http://dx.doi.org/10.1111/j.1365-294X.2005.02787.x
195. Duftner N, Sefc KM, Koblmuller S, Salzburger W, Taborsky M, Sturmbauer C (**2007**) Parallel evolution of facial stripe patterns in the Neolamprologus brichardilpulcher species complex endemic to Lake Tanganyika. *Mol Phylogenet Evol* **45**:706-715. http://dx.doi.org/10.1016/j.ympev.2007.08.001
196. Kidd MR, Kocher TD (**2005**) Patterns of ecological diversification within Pseudotropheus tropheops, a subgenus of rock dwelling cichlid from Lake Malawi. *Integrative and Comparative Biology* **45**:1025-1025.
197. Salzburger W, Braasch I, Meyer A (**2007**) Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. *BMC Biology* **5**: http://dx.doi.org/10.1186/1741-7007-5-51
198. Streelman JT, Kocher TD (**2000**) From phenotype to genotype. *Evolution & Development* **2**:166?173.
199. Kryazhimskiy S, Plotkin JB (**2008**) The Population Genetics of dN/dS. *PLoS Genetics* **4**: http://dx.doi.org/10.1371/journal.pgen.1000304
200. Turner DJ, Keane TM, Sudbery I, Adams DJ (**2009**) Next-generation sequencing of vertebrate experimental organisms. *Mamm Genome* **20**:327-338. http://dx.doi.org/10.1007/s00335-009-9187-4
201. Li WH, Gojobori T (**1983**) Rapid Evolution of Goat and Sheep Globin Genes Following Gene Duplication. *Mol Biol Evol* **1**:94-108.
202. Lynch M, Conery JS (**2003**) The evolutionary demography of duplicate genes. *J Struct Funct Genomics* **3**:35-44. http://dx.doi.org/PMID: 12836683
203. Hughes T, Liberles DA (**2007**) The pattern of evolution of smaller-scale gene duplicates in mammalian genomes is more consistent with neo-than subfunctionalisation. *J Mol Evol* **65**:574-588. http://dx.doi.org/10.1007/s00239-007-9041-9
204. Koh EGL, Lam K, Christoffels A, Erdmann MV, Brenner S, Venkatesh B (**2003**) Hox gene clusters in the Indonesian coelacanth, Latimeria menadoensis. *Proc Natl Acad Sci U S A* **100**:1084-1088. http://dx.doi.org/Hox gene clusters in the Indonesian coelacanth, Latimeria menadoensis
205. Ochman H, Gerber AS, Hartl DL (**1988**) Genetic Applications of an Inverse Polymerase Chain-Reaction. *Genetics* **120**:621-623.

206. Cai CM, Van K, Kim MY, Lee SH (**2008**) Sequence divergence of recently duplicate genes in soybean (Glycine max L. Merr.). *Genes & Genomics* **30**:271-281.
207. Edgar RC (**2004**) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792-1797. http://dx.doi.org/10.1093/nar/gkh340
208. Guindon S, Gascuel O (**2003**) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**:696-704. http://dx.doi.org/10.1080/10635150390235520
209. Guindon S, Delsuc F, Dufayard J-F, Gascuel O (**2009**) Estimating Maximum Likelihood Phylogenies with PhyML. *Bioinformatics for DNA Sequence Analysis* 113-137. http://dx.doi.org/:10.1007/978-1-59745-251-9_6
210. Guindon S, Lethiec F, Duroux P, Gascuel O (**2005**) PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **33**:W557-W559. http://dx.doi.org/10.1093/nar/gki352
211. Yang ZH (**2007**) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:1586-1591. http://dx.doi.org/10.1093/molbev/msm088
212. Zhang JZ, Nielsen R, Yang ZH (**2005**) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* **22**:2472-2479. http://dx.doi.org/10.1093/molbev/msi237
213. Wong WSW, Yang ZH, Goldman N, Nielsen R (**2004**) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**:1041-1051. http://dx.doi.org/10.1534/genetics.104.031153
214. McGrath CL, Casola C, Hahn MW (**2009**) Minimal Effect of Ectopic Gene Conversion Among Recent Duplicates in Four Mammalian Genomes. *Genetics* **182**:615-622. http://dx.doi.org/10.1534/genetics.109.101428
215. Casola C, Hahn MW (**2009**) Gene Conversion Among Paralogs Results in Moderate False Detection of Positive Selection Using Likelihood Methods. *J Mol Evol* **68**:679-687. http://dx.doi.org/10.1007/s00239-009-9241-6
216. Innan H (**2009**) Population genetic models of duplicated genes. *Genetica* **137**:19-37. http://dx.doi.org/10.1007/s10709-009-9355-1
217. Friedman R, Hughes AL (**2007**) Likelihood-ratio tests for positive selection of human and mouse duplicate genes reveal nonconservative and anomalous properties of widely used methods. *Mol Phylogenet Evol* **42**:388-393. http://dx.doi.org/10.1016/j.ympev.2006.07.015
218. Hughes AL, Friedman R (**2007**) The effect of branch lengths on phylogeny: An empirical study using highly conserved orthologs from mammalian genomes. *Mol Phylogenet Evol* **45**:81-88. http://dx.doi.org/10.1016/j.ympev.2007.04.022
219. Anisimova M, Liberles DA (**2007**) The quest for natural selection in the age of comparative genomics. *Heredity* **99**:567-579. http://dx.doi.org/10.1038/sj.hdy.6801052
220. Delport W, Scheffler K, Seoighe C (**2009**) Models of coding sequence evolution. *Brief Bioinform* **10**:97-109. http://dx.doi.org/10.1093/bib/bbn049
221. Kosakovsky Pond SL, Frost SDW (**2005**) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**:1208-22. http://dx.doi.org/10.1093/molbev/msi105
222. Studer RA, Penel S, Duret L, Robinson-Rechavi M (**2008**) Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. *Genome Res* **18**:1393-1402. http://dx.doi.org/10.1101/gr.076992.108
223. Li Z, Zhang H, Ge S, Gu XC, Gao G, Luo JC (**2009**) Expression pattern divergence of duplicated genes in rice. *BMC Bioinformatics* **10**: http://dx.doi.org/10.1186/1471-2105-10-s6-s8
224. Zou C, Lehti-Shiu MD, Thomashow M, Shiu SH (**2009**) Evolution of Stress-Regulated Gene Expression in Duplicate Genes of Arabidopsis thaliana. *PLoS Genetics* **5**: http://dx.doi.org/10.1371/journal.pgen.1000581

225. Karanth S, Lall SP, Denovan-Wright EM, Wright JM (**2009**) Differential transcriptional modulation of duplicated fatty acid-binding protein genes by dietary fatty acids in zebrafish (Danio rerio): evidence for subfunctionalization or neofunctionalization of duplicated genes. *BMC Evol Biol* **9**:Article No.: 219. http://dx.doi.org/:10.1186/1471-2148-9-219
226. Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA (**2009**) Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. *Genome Res* **19**:1404-1418. http://dx.doi.org/10.1101/gr.086827.108
227. Orozco LD, Cokus SJ, Ghazalpour A, Ingram-Drake L, Wang S, van Nas A, Che N, Araujo JA, Pellegrini M, Lusis AJ (**2009**) Copy number variation influences gene expression and metabolic traits in mice. *Hum Mol Genet* **18**:4118-4129. http://dx.doi.org/10.1093/hmg/ddp360
228. Mileyko Y, Joh RI, Weitz JS (**2008**) Small-scale copy number variation and large-scale changes in gene expression. *Proc Natl Acad Sci U S A* **105**:16659-16664. http://dx.doi.org/10.1073/pnas.0806239105
229. Kohn MH (**2008**) Rapid sequence divergence rates in the 5 prime regulatory regions of young Drosophila melanogaster duplicate gene pairs. *Genet Mol Biol* **31**:575-584. http://dx.doi.org/10.1590/S1415-47572008000300028
230. Plotkin JB, Dushoff J, Fraser HB (**2004**) Detecting selection using a single genome sequence of M-tuberculosis and P-falciparum. *Nature* **428**:942-945. http://dx.doi.org/10.1038/nature02458
231. Hanada K, Shiu SH, Li WH (**2007**) The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Mol Biol Evol* **24**:2235-2241. http://dx.doi.org/10.1093/molbev/msm152
232. Friedman R, Hughes AL (**2005**) Codon volatility as an indicator of positive selection: Data from eukaryotic genome comparisons. *Mol Biol Evol* **22**:542-546. http://dx.doi.org/10.1093/molbev/msi038
233. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N (**2005**) Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res* **15**:1153-1160. http://dx.doi.org/10.1101/gr.3567505
234. Pillai SK, Pond SLK, Woelk CH, Richman DD, Smith DM (**2005**) Codon volatility does not reflect selective pressure on the HIV-1 genome. *Virology* **336**:137-143. http://dx.doi.org/10.1016/j.virol.2005.03.014
235. Stoletzki N, Welch J, Hermisson J, Eyre-Walker A (**2005**) A dissection of volatility in yeast. *Mol Biol Evol* **22**:2022-2026. http://dx.doi.org/10.1093/molbev/msi192
236. Archetti M (**2009**) Genetic robustness at the codon level as a measure of selection. *Gene* **443**:64-69. http://dx.doi.org/10.1016/j.gene.2009.05.009
237. Liang H, Zhou WH, Landweber LF (**2006**) SWAKK: a web server for detecting positive selection in proteins using a sliding window substitution rate analysis. *Nucleic Acids Res* **34**:W382-W384. http://dx.doi.org/10.1093/nar/gkl272
238. Berglund AC, Wallner B, Elofsson A, Liberles DA (**2005**) Tertiary windowing to detect positive diversifying selection. *J Mol Evol* **60**:499-504. http://dx.doi.org/10.1007/s00239-004-0223-4
239. Moore RC, Purugganan MD (**2003**) The early stages of duplicate gene evolution. *Proc Natl Acad Sci U S A* **100**:15682-15687. http://dx.doi.org/10.1073/pnas.2535513100
240. Yi S, Charlesworth B (**2000**) A selective sweep associated with a recent gene transposition in Drosophila miranda. *Genetics* **156**:1753-1763. http://dx.doi.org/PMCID: PMC1461364
241. Arguello JR, Chen Y, Yang S, Wang W, Long M (**2006**) Origination of an X-linked testes chimeric gene by illegitimate recombination in Drosophila. *PLoS Genetics* **2**:745-754. http://dx.doi.org/10.1371/journal.pgen.0020077
242. Cirera S, Aguade M (**1998**) Molecular evolution of a duplication: The sex-peptide (Acp70A) gene region of Drosophila subobscura and Drosophila madeirensis. *Mol Biol Evol* **15**:988-996.
243. King LM (**1998**) The role of gene conversion in determining sequence variation and divergence in the Est-5 gene family in Drosophila pseudoobscura. *Genetics* **148**:305-315.

244. Long MY, Langley CH (**1993**) Natural-Selection and the Origin of Jingwei, a Chimeric Processed Functional Gene in Drosophila. *Science* **260**:91-95. http://dx.doi.org/10.1126/science.7682012
245. Thornton K, Long M (**2002**) Rapid divergence of gene duplicates on the Drosophila melanogaster X chromosome. *Mol Biol Evol* **19**:918-925.
246. Thornton K, Long M (**2005**) Excess of amino acid substitutions relative to polymorphism between X-linked duplications in Drosophila melanogaster. *Mol Biol Evol* **22**:273-284. http://dx.doi.org/10.1093/molbev/msi015
247. Betran E, Long M (**2003**) Dntf-2r, a young Drosophila retroposed gene with specific male expression under positive Darwinian selection. *Genetics* **164**:977-988.
248. Holloway AK, Begun DJ (**2004**) Molecular evolution and population genetics of duplicated accessory gland protein genes in Drosophila. *Mol Biol Evol* **21**:1625-1628. http://dx.doi.org/10.1093/molbev/msh195
249. van Hijum S, Baerends RJS, Zomer AL, Karsens HA, Martin-Requena V, Trelles O, Kok J, Kuipers OP (**2008**) Supervised Lowess normalization of comparative genome hybridization data - application to lactococcal strain comparisons. *BMC Bioinformatics* **9**: http://dx.doi.org/10.1186/1471-2105-9-93
250. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and Computational Biology Solutions using R and Bioconductor. New York: Springer; 2005. p 397-420.
251. Smyth GK (**2004**) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**:1-26. http://dx.doi.org/10.2202/1544-6115.1027
252. Sokal RR, Rohlf FJ. Biometry, The principles and practice of statistics in biological research. Third edition. New York: W.H. Freeman and Company; **1995**. isbn#: 10: 0716712547
253. Lapointe FJ, Garland T (**2001**) A generalized permutation model for the analysis of cross-species data. *J Classif* **18**:109-127. http://dx.doi.org/10.1007/s00357-001-0007-0
254. Rosenberg MS PASSaGE: a statistical package for spatial analysis. *Arizona State University*
255. Felsenstein J (**1986**) Phylogenies and the comparative method. *Am Nat* **125**:1-15. http://dx.doi.org/10.1086/284325
256. Garland T, Harvey PH, Ives AR (**1992**) Procedures for the Analysis of Comparative Data Using Phylogenetically Independent Contrasts. *Syst Biol* **41**:18-32. http://dx.doi.org/10.1093/sysbio/41.1.18
257. Garland T, Bennett AF, Rezende EL (**2005**) Phylogenetic approaches in comparative physiology. *J Exp Biol* **208**:3015-3035. http://dx.doi.org/10.1242/jeb.01745
258. Harmon LJ, Kolbe JJ, Cheverud JM, Losos JB (**2005**) Convergence and the multidimensional niche. *Evolution* **59**:409-421. http://dx.doi.org/10.1554/04-038
259. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM (**2005**) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**:1928-1933. http://dx.doi.org/10.1126/science.1107239
260. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O and others (**2009**) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* **41**:1061-U29. http://dx.doi.org/10.1038/ng.437
261. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song XZ, Richmond TA, Middle CM, Rodesch MJ, Packard CJ and others (**2007**) Direct selection of human genomic loci by microarray hybridization. *Nat Meth* **4**:903-905. http://dx.doi.org/10.1038/nmeth1111
262. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME (**2007**) Microarray-based genomic selection for high-throughput resequencing. *Nat Meth* **4**:907-909. http://dx.doi.org/10.1038/nmeth1109

263. Han MV, Hahn MW (**2009**) Identifying Parent-Daughter Relationships among Duplicated Genes. *Pacific Symposium on Biocomputing 2009* 114-125. http://dx.doi.org/10.1142/9789812836939_0012