

Methodology article

Open Access

Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray

Susan CP Renn[†], Nadia Aubin-Horth^{*†} and Hans A Hofmann

Address: Harvard University, Bauer Center for Genomics Research, 7 Divinity Avenue, Cambridge, MA, 02138, USA

Email: Susan CP Renn - SRenn@cgr.harvard.edu; Nadia Aubin-Horth* - Naubin-Horth@cgr.harvard.edu;

Hans A Hofmann - Hans@cgr.harvard.edu

* Corresponding author †Equal contributors

Published: 06 July 2004

Received: 16 June 2004

BMC Genomics 2004, **5**:42 doi:10.1186/1471-2164-5-42

Accepted: 06 July 2004

This article is available from: <http://www.biomedcentral.com/1471-2164/5/42>

© 2004 Renn et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Unravelling the path from genotype to phenotype, as it is influenced by an organism's environment, is one of the central goals in biology. Gene expression profiling by means of microarrays has become very prominent in this endeavour, although resources exist only for relatively few model systems. As genomics has matured into a comparative research program, expression profiling now also provides a powerful tool for non-traditional model systems to elucidate the molecular basis of complex traits.

Results: Here we present a microarray constructed with ~4500 features, derived from a brain-specific cDNA library for the African cichlid fish *Astatotilapia burtoni* (Perciformes). Heterologous hybridization, targeting RNA to an array constructed for a different species, is used for eight different fish species. We quantified the concordance in gene expression profiles across these species (number of genes and fold-changes). Although most robust when target RNA is derived from closely related species (<10 MA divergence time), our results showed consistent profiles for other closely related taxa (~65 MA divergence time) and, to a lesser extent, even very distantly related species (>200 MA divergence time).

Conclusion: This strategy overcomes some of the restrictions imposed on model systems that are of importance for evolutionary and ecological studies, but for which only limited sequence information is available. Our work validates the use of expression profiling for functional genomics within a comparative framework and provides a foundation for the molecular and cellular analysis of complex traits in a wide range of organisms.

Background

The expression activities of all the genes represented in an organism's genome at any given time constitute a complex phenotype that is closely connected with, but not solely dependent upon, the genotype. In fact, gene expression profiles represent the primary level of integration between environmental factors and the genome, providing the basis for protein synthesis which ultimately guides the implementation of complex macro-phenotypes such as

morphology and behaviour. Therefore, by comparing gene expression profiles of different strains, populations, or even species, one can directly study the molecular basis of phenotypic variation. This comparative approach has recently been employed in expression profiling of model organisms such as *Drosophila* and *S. cerevisiae*, unveiling surprising patterns of sex-driven evolution and providing insights into the genetics of population-level variation in adaptive traits ([1-5]). Similarly, across-species

comparisons of expression profiles between human and non-human primates have been used to test hypotheses about the functional complexity of the human brain ([6-8]).

Although it has become increasingly clear that genomics research benefits from such a comparative outlook, whole genome sequences are available for only a few of the less traditional model systems (e.g. Fugu ([9]; *Anopheles* [10]; dog [11]; honey bee [12]). Given the current DNA microarray technologies (long or short oligonucleotide arrays and cDNA arrays), which allow simultaneous monitoring of thousands of genes, the PCR amplified cDNA array is the most accessible for studies in non-model organisms (but see [6] for oligonucleotide strategies). The construction of a cDNA microarray is not limited by the need for probe design based on advanced bioinformatics analysis of genome sequences only available for genetic "model organisms" (e.g., [12-14]). Due to the length of the probes (which are often full-length), it has been suggested that cDNA microarrays can also be used in heterologous hybridizations across strains and even relatively closely related species as long as sequence divergence is limited for a given gene [15]. This approach has been used successfully, if only occasionally, to study the molecular basis of traits not present in traditional model species (e.g. hibernation, [16]). Heterologous hybridization offers a promising approach to study molecular mechanisms in species for which a wealth of ecological data as well as natural phenotypic and genetic variation are already available. As research areas such as ecophysiology ([17-19]) as well as ecology and evolution ([20-22]) are now merging with functional genomics ([23]; reviewed in [24]) the technique of heterologous hybridization will become more prominent. While the feasibility of this technique has been indirectly suggested, [15] its potential has not been systematically tested using a biologically meaningful approach over a wide range of species.

In this paper we quantify the utility of a microarray constructed from an African cichlid fish cDNA library to study expression profiling in other fish species through the use of heterologous hybridization. We not only test RNA hybridization from closely related cichlids but also from more distantly related species of teleost fishes, a group that represents more than half of living vertebrates [25]. Our results demonstrate that heterologous microarray hybridization can yield biologically meaningful data even in relatively distantly related species and establish this technique as a tool for comparative functional genomics in organisms not previously open to an integrative molecular analysis.

The rapid, recent, and repeated radiation of the cichlid fishes in East Africa's Great Lakes has produced a system

especially amenable to this approach. Each of the Great Lakes -Victoria, Tanganyika, and Malawi- boasts hundreds of phenotypically diverse endemic cichlid species, evolved from one or a few common ancestors in very recent evolutionary history ([26-30]) and thus sharing highly similar genomic sequence. The astonishing variation in phenotype at the morphological and behavioural level ([31-38]), both within and among cichlid species has likely contributed to the evolutionary success of this group ([39,40]). Therefore, through heterologous hybridization, a genomic approach can be used to uncover the molecular basis and the evolution of these complex traits.

Results

Characterization of the cichlid fish cDNA array

As our study presents the first use of this cDNA microarray to determine gene expression in the brain of multiple cichlids species, we first characterized this tool. This array was originally built using anonymous clones from a brain cDNA library for the African cichlid *Astatotilapia burtoni*. Sequencing is being done in parallel with initial experiments. The gel electrophoretic analysis of PCR products performed during array construction identified reactions that either had failed to produce a product or had produced more than one product. From this we concluded that 4570 amplified cDNAs were reliably represented on the array. For these, the initial 5' sequencing returned sequence for eighty-six percent (3936) of the clones (Genbank CN468542 - CN472211 for clones > 150 bp). Sixty-four percent (2462) of the EST's represented unique sequences ("singletons") and the remaining 37% could be aligned to 575 EST clusters, predicting that up to 3000 unique genes may be represented on this array. Approximately 40% of the different genes showed significant homology to proteins predicted from the Fugu genome ([41], Salzburger et al., in prep.)

Self-hybridization experiments, in which two samples of the platform species *A. burtoni* genomic DNA were competitively hybridized, revealed that 93% (mean = 4264 + SD 229; n = 6) of the features on the array could be hybridized at a fluorescent intensity at least two standard deviations above mean background intensity. Signal intensities for the 635 nm and 532 nm channels were highly correlated in these experiments (r = 0.991 same DNA isolation, r = 0.941 same animal but different DNA isolation, r = 0.974 two different animals, all p < 0.0001) indicating that technical variability - due to differences in DNA source, isolation, and fluorescent labelling - was very low. These results suggest that the majority of the *A. burtoni* library spots will provide reliable data.

Self-hybridization experiments, where two samples of RNA from a pool of several *A. burtoni* brains were competitively hybridized, revealed that ~94% of the features on

the array (4316+SD 431; n = 10) represent genes that are expressed at detectable levels using a whole brain strategy. Again, signal intensities for the 635 nm and 532 nm channels were highly correlated in these experiments ($r = 0.974 - 0.997$ all $p < 0.0001$) indicating that technical variability due to Cy3 or Cy5 labelling is very low.

Establishing a reference profile in *A. burtoni*

In order to evaluate the sensitivity and consistency of heterologous hybridization for several species on this array platform, we first devised a direct comparison experiment to maximize expression differences within *A. burtoni*, the species from which the array was built. These results provide a reference for the evaluation of the heterologous hybridization results presented below. As a large percentage of a genome's protein-coding genes are expressed in the brain ([42,43]), we expected this microarray to be useful for expression profiling in other tissues. Therefore, brain-derived RNA was competitively hybridized against mixed-tissue RNA (skin, muscle, blood vessels) extracted from the same individual. The distinct nature of these tissues predicts a dramatic difference in the expression of a large number of genes. Four replicates from each of two individual *A. burtoni* (one adult and one juvenile) were used to identify consistent gene expression differences between the two tissues, independent of age and reproductive status. 88% of the *A. burtoni* microarray probes hybridized above the background cut-off (4165+SD 421, n = 8).

Analysis of each *A. burtoni* individual transcription profile revealed spots that showed significant up-regulation (in a wide range of fold-differences) in the brain as compared to mixed-muscle (n = 1146 and 920, respectively). Due to the variability among animals (adult and juvenile), the number of significantly regulated spots identified by increasing only technical replicates was greater than the number identified by the same absolute number of arrays involving biological replicates (data not shown). We identified 804 spots that showed consistent up-regulation in the brain of both *A. burtoni* individuals and used this core set as the reference for the heterologous hybridization experiments (Figure 1).

The inclusion of replicate spots in microarray construction provides a standard for internal control, and increased confidence in the estimation of differential regulation for these spots. Based on the sequence information (available for 656 of the 804 spots) we estimated that this reference set contains a total of 472 genes. 41 of these genes are represented by three or more cDNAs (173 spots total). We find greater than 75% concordant regulation across representative spots (at least three out of four clones or all three for those clusters which contain only three clones) for 27 of these genes. Concordance smaller

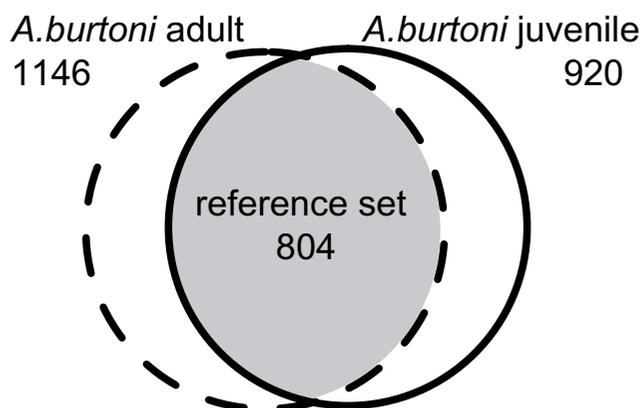


Figure 1
The core reference set for brain up-regulated gene expression. Venn diagram of spots up regulated in brain for two individual *A. burtoni*. Statistical analysis (see methods) of four replicates, including two dye-swaps for each animal predicts significant regulation ($p < 0.05$). The intersection of these results represents the core reference set.

than 75% could be caused by (i) hybridization failure (due to low probe concentration), (ii) improper assignment of clones to genes, (iii) sequencing errors, (iv) alternative splicing, (v) chance (false negatives).

Heterologous hybridization

We used the same brain vs. body intra-species experimental design to test the hypothesis that hybridization of heterologous RNA to the array can yield biologically meaningful results. We chose two other cichlid species endemic to Lake Tanganyika, *Enantiopus melanogenys* and *Neolamprologus brichardi*, and a more distantly related African cichlid, the Nile tilapia, *Oreochromis niloticus*. Based on their wide use in behavioural, evolutionary, developmental, and genetic studies, we also selected four more distantly related species that are not part of the order perciform [25]. Two poeciliid fish: platyfish (*Xiphophorus* sp. [44]) and guppy (*Poecilia reticulata* [45]; 65 MY divergence time [46]); one salmonid: Atlantic salmon (*Salmo salar* [47]), and one cyprinid: zebrafish (*Danio rerio* [48]; 200 MY divergence time [46]). We first quantified the extent of hybridization for a given species with a simple threshold cut-off, and then compared these heterologous expression profile results to the reference set of the 804 brain-enriched genes from *A. burtoni* in order to quantify the performance of heterologous hybridization for detection of hybridization, regulation, and biologically meaningful results.

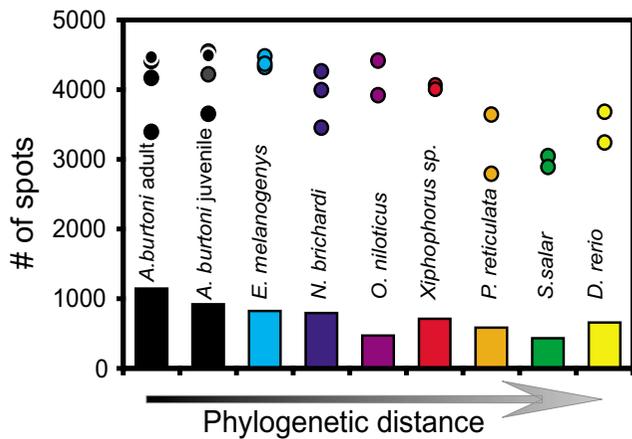


Figure 2
Detection of hybridization and regulation across phylogenetic distance. Heterologous hybridization of RNA from *A. burtoni* and seven other teleost species. The Y-axis defines the number of spots that hybridized above background (circles) for each hybridization experiment as well as the number of spots that showed significant ($p < 0.05$) up-regulation (bars) in brain RNA compared to mixed muscle RNA for each species. Colour coding of species is consistent throughout the manuscript

Detection of hybridization

Without regard for the identity of the array features, hybridization dynamics inform on the utility of the array. The number of spots hybridizing above threshold intensity (2 SD above background) was determined for each species (Figure 2, circles). While heterologous hybridization performed well for the two Tanganyikan cichlids and the closely related tilapia, a marginally significant decrease in the number of hybridized spots was seen for the more distantly related species (Kruskal-Wallis test: $\chi^2 = 13.35$; $p = 0.06$). We hypothesize that this is due to the fact that sequence divergence increases with phylogenetic distance. While detecting hybridization demonstrates the extent to which diverged sequences still give reliable hybridization signals across the entire array, more sophisticated analysis is required to quantify the power to detect expression differences in other species. For this reason, we quantified the detection of expression ratios by heterologous hybridization for tissue specific gene expression.

Correlation of expression profiles across species

To estimate the effect of evolutionary distance on our ability to obtain biologically significant results with heterologous hybridization, we performed a linear regression analysis between fold-change determined for *A. burtoni* and each of the other species. Regression coefficients indi-

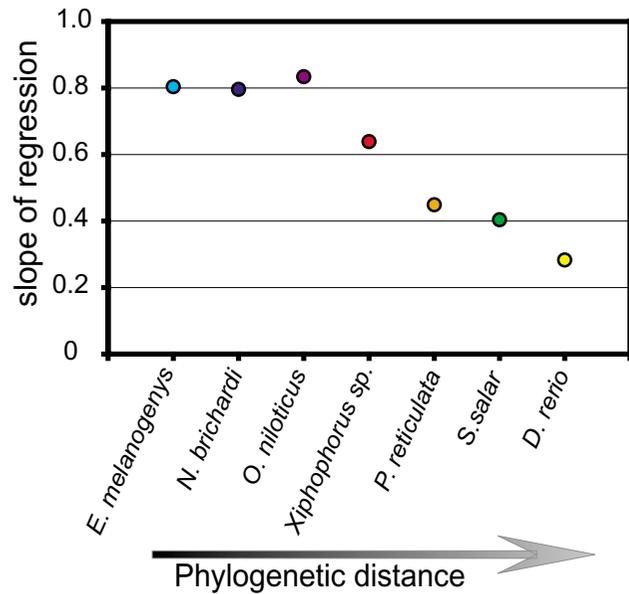


Figure 3
The detected magnitude of fold change decreases with phylogenetic distance. For all spots on the array, the fold change value from a combined analysis of all *A. burtoni* hybridizations was correlated independently with the fold change determined for each of the other species. The Y-axis denotes the slope of each regression, with circles colour coded as in previous figures.

cate how well the *A. burtoni* fold-changes account for the fold-change variation in the other species. The slope of the regression indicates the relative magnitude of gene regulation detected for the different species compared to *A. burtoni*. All regressions were highly significant ($p < 0.0001$). For the three cichlid species, the regression coefficients ranged between $r^2 = 0.58$ and $r^2 = 0.80$, demonstrating that most expression differences between brain and mixed-muscle tissues (58% to 80%) could be explained by corresponding expression differences observed in *A. burtoni* (data not shown). Likewise, slopes of the regression between these fold-changes were similar within cichlids (Figure 3) indicating that absolute fold-change values were very similar in magnitude between *A. burtoni* and the other members of this family.

In poeciliids, regression coefficients were $r^2 = 0.45$ (*Xiphophorus sp.*) and $r^2 = 0.37$ (*P. reticulata*). However, in the more distantly related taxa (salmonids and cyprinids) regression coefficients were $r^2 = 0.21$ and $r^2 = 0.11$, demonstrating that the *A. burtoni* data explained less of the fold-change variation. This reflects the fact that more spots showed different expression in these distantly

Table 1: Many genes are regulated in each species. Even in more distantly related fish species, the spots determined to be up-regulated in the brain encompass a great number of unique spots, and are not simply a few genes represented multiple times on the array (n refers to the number of array hybridizations). The estimated number unique genes up-regulated in brain is calculated independently for each species based on the sum of singleton ESTs and the number of genes predicted by clustered ESTs relative to the sequence information available. This result indicates that the redundancy rate for spots determined to be up-regulated in the brain is in agreement with the overall array redundancy and that differences in expression can be detected for mRNAs of low copy number.

Species	Total number of spots	No sequence available	Sequenced ESTs			Estimated number of unique genes
			Singletons	Clustered ESTs		
				Number of ESTs	Number of clusters	
<i>A. burtoni</i> (adult) (n = 4)	1146	203	530	413	189	874
<i>A. burtoni</i> (juvenile) (n = 4)	921	172	373	376	171	669
<i>N. brichardi</i> (n = 3)	795	145	302	348	157	561
<i>E. melanogenys</i> (n = 3)	824	152	320	352	156	584
<i>O. niloticus</i> (n = 2)	472	89	164	219	110	338
<i>Xiphophorus</i> sp. (n = 2)	585	126	244	215	121	465
<i>P. reticulata</i> (n = 2)	713	157	323	233	137	590
<i>S. salar</i> (n = 2)	433	99	197	137	83	363
<i>D. rerio</i> (n = 2)	658	141	329	188	130	584

related species compared to *A. burtoni*. The drop in the regression slope with phylogenetic distance (Figure 3) suggests that although gene expression in *A. burtoni* and the other species were in the same direction, the magnitude of change in expression that was detected was much lower in these species. The effect of phylogenetic distance on both our ability to detect subtle gene regulation and its magnitude shows that the sensitivity of our array is very good for cichlids and even species that are not members of the perciformes (e.g., platyfish and guppy), but is lower for distantly related species such as salmon and zebra fish.

Detection of regulated gene expression

For the Tanganyikan cichlid species, the number of spots that showed significant up-regulation in the brain (*E. melanogenys*: 824; *N. brichardi*: 795; $p < 0.05$) was similar to the number of spots observed for each *A. burtoni* (Figure 2, bars). However, the number of significantly regulated spots decreased with phylogenetic distance in more distantly related species (*O. niloticus*: 472; *Xiphophorus* sp.: 585; *P. reticulata*: 713; *S. salar*: 433; *D. rerio*: 658). Based upon the preliminary EST sequence data, we estimate that ~78% of these spots represent unique genes (Table 1). These encouraging results demonstrate the power to detect significant gene regulation using heterologous hybridization. However, in order to validate the comparative approach with this technique it is crucial to demonstrate the ability to quantify the degree to which a given spot is similarly regulated in the different species.

Detection of consistent gene regulation

Assuming that the reference set of 804 brain-enriched spots identified in *A. burtoni* represents genes that are always active in the brain independent of age, sex, and reproductive status, we would expect a large fraction of these spots to be similarly up-regulated in the brains of other fish species. Figure 4A shows the percentage of spots from the reference set that are also found to be up-regulated in the brains of each species. Within the cichlid species, 42% of the *A. burtoni* reference spots were also identified in all three of the other cichlid species (340 spots) (Figure 4B). An additional 26% (206 spots) were identified in the two more closely related Tanganyikan cichlids, *E. melanogenys* and *N. brichardi*, bringing the total to 68%. Within the non-cichlid species, 62% of the 804 *A. burtoni* reference spots were also identified in at least one species. The intersection of up-regulated brain spots for members of the poeciliid family (*Xiphophorus* sp. and *P. reticulata*) contained 42% (340 spots) of the reference set (Figure 4C). In the very distantly related species, we identify 27% (217 spots) and 29% (236 spots) for *S. salar* and *D. rerio*, respectively.

We investigated whether we could identify concordant regulation through heterologous hybridization experiments in the other species using the 27 genes selected above (> 75% concordant regulation across clones in *A. burtoni*). For the Tanganyikan cichlids, *N. brichardi* and *E. melanogenys*, 24 and 26 of these genes, respectively, passed the same stringent cut-off. *O. niloticus* showed 19 genes exceeding this threshold. As is expected from the data for

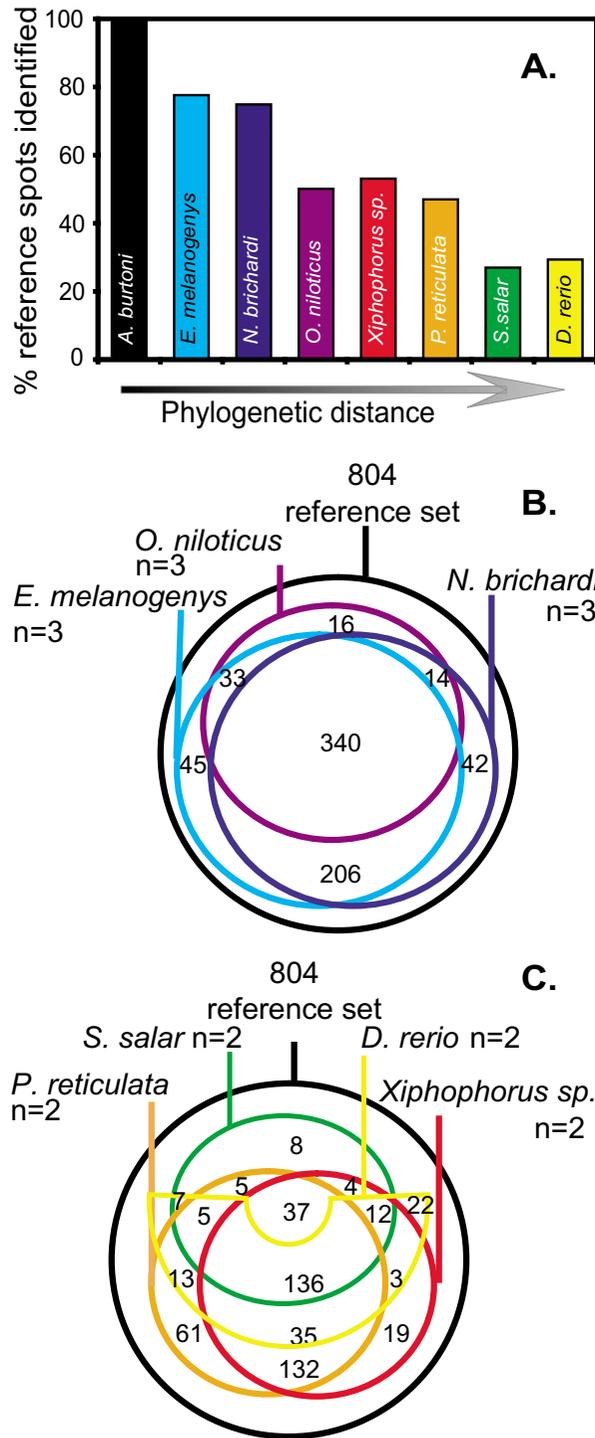


Figure 4

Detection of biologically meaningful gene regulation. (3A) Core reference spots showing significant up-regulation in the brain decreases with phylogenetic distance. All 804 spots were examined in each species. The Y-axis depicts the percentage of the core reference set spots that were identified as significantly up-regulated in the brain of each species. Venn diagrams depict the relationship of identified spots from the core reference set as re-identified in each species. The number of *A. burtoni* reference set spots that are also significantly up-regulated in the brains of (3B) other cichlids and (3C) distantly related fish are shown in the appropriately represented area. 108 and 302 spots are not re-identified in the cichlids or distantly related fish respectively (not shown).

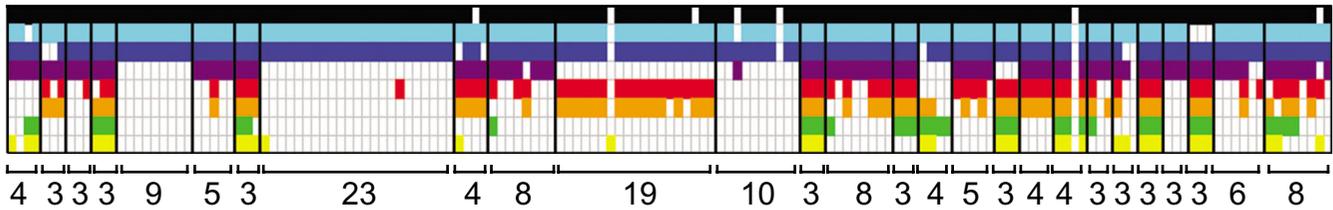


Figure 5
Concordant detection of regulation. All clones (vertical columns) of genes (outlined by brackets) that show > 75% concordant regulation in *A. burtoni* are represented. Each row represents the species used in this study (colour-coded as in Figure 1). Filled boxes represent ESTs significantly up-regulated ($p < 0.05$) in the brain. Numbers indicate the number of clones in each EST cluster.

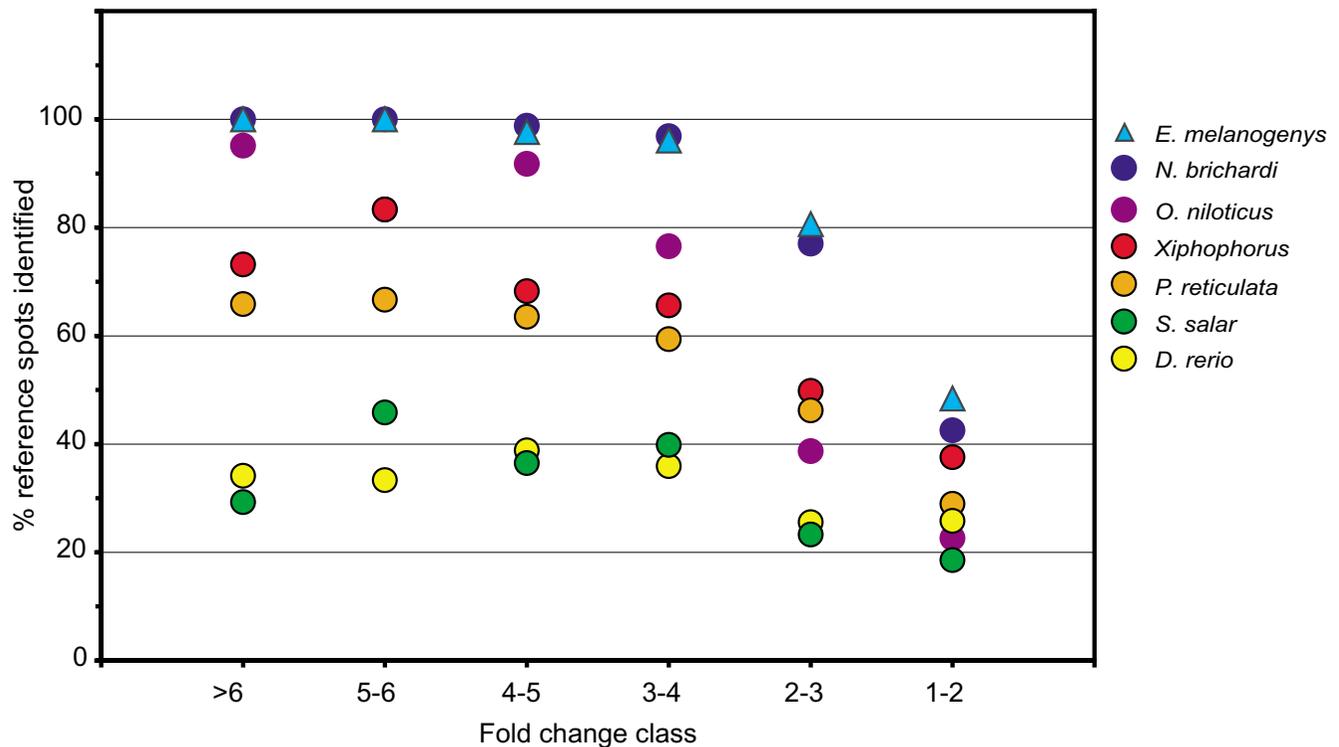


Figure 6
Spots of low fold change are under represented in heterologous hybridization. The 804 reference set spots are divided into 6 classes according to fold change (> 6.0 $n = 41$; 5.0 – 6.0 $n = 24$; 4.0 – 5.0 $n = 85$; 3.0 – 4.0 $n = 128$; 2.0 – 3.0 $n = 305$ and 1.0 – 2.0 $n = 221$). For each fold change class (x-axis) the percentage of spots in that class (y-axis) that are also identified ($p < 0.05$) in other species are depicted by colour coded symbols. The under representation of low fold change classes is most dramatic for more distantly related fish (note salmon, green and zebrafish, yellow).

the entire reference set, the number of brain-enriched genes that are re-identified in the more distantly related species decreases with phylogenetic distance (*Xiphophorus*: 12; *P. reticulata*: 11; *D. rerio*: 8; *S. salar*: 8). Notably, the

clones are not eliminated from the list at random but rather EST clusters tend to be either detected as brain up-regulated or not in a concordant fashion (Figure 5).

Detection of subtle gene regulation

To determine if the capacity to detect subtle gene regulation with this array declines in other cichlid species, the 804 reference spots were binned into six classes according to their average fold-change, as determined by a combined analysis of all eight *A. burtoni* replicates. The percentage of spots in each fold-change class that was also significantly up-regulated in the brain tissue in the other species is shown in Figure 6. Note that a spot is called "regulated" only if significant ($p < 0.05$) regardless of the magnitude of the fold-change in the heterologous hybridization. Spots that showed > 3.0 -fold change in *A. burtoni* also showed significant brain-specific regulation in the Tanganyikan cichlid species. Similarly, a large portion (42% to 81%) of small fold-change spots (1-fold to 3-fold) were also detected as significantly over-expressed in the brains of *E. melanogens* and *N. brichardi*. For heterologous hybridization with the more distantly related cichlid *O. niloticus* (ca. 10 million years divergence time), all fold-change classes were underrepresented, but similar to the closely related Tanganyikan cichlids, the detection of significant regulation was more drastically compromised for low fold-change classes. For heterologous hybridization in the poeciliid family (~65 million years divergence time), represented by *Xiphophorus* sp. and *P. reticulata*, spots regulated > 3.0 -fold in *A. burtoni* were fairly well represented (59% – 83%), while spots in the lower fold-change classes were relatively underrepresented. In more distantly related fish families, represented by *S. salar* and *D. rerio* (> 200 million years divergence time), brain-specific expression profiles were more uniformly and drastically compromised across all fold-change classes (18% – 45%). These data demonstrate that phylogenetic distance most dramatically affects detection of subtly regulated genes.

Discussion

In this paper we describe the first systematic analysis of heterologous microarray hybridizations across a range of vertebrate species. This work validates the use of expression profiling for functional genomics within a comparative framework and provides a foundation for the molecular and cellular analysis of complex traits at the organismal, population, and ecological levels [24]. We clearly show the utility of the array for heterologous hybridization across a range of fish species for which there is no other microarray available. We can detect array features (though reduced in number) that hybridize above background as well as spots that show tissue-specific regulation, many of which correspond to those regulated in *A. burtoni*.

The variation in brain-specific gene expression between individual fish of different Tanganyikan cichlid species is comparable to the variation observed between adult and

juvenile individuals of *A. burtoni*. The slight decrease in our ability to identify the *A. burtoni* genes of subtle regulation (i.e., low fold-change classes) in other cichlid species may be due to either the smaller number of replicates performed for these species or the increased individual variation in these fold-change classes. Alternatively, this result could also reflect real species-specific differences in gene expression. Even in distantly related species 26% – 53% of the significantly up-regulated *A. burtoni* genes could still be re-identified. In this analysis we assume that the increasing number of genes that failed to hybridize with increased phylogenetic distance was due mainly to sequence divergence. This assumption provides a conservative guideline regarding the utility of heterologous hybridization. Tissue-specific gene regulation is obviously not expected to be identical in all species. Therefore, it is possible that more than 26% – 53% of the array spots are informative for distantly related species. Heterologous hybridization experiments on any microarray are of limited use for genes that have undergone rapid evolutionary change in coding regions, large rearrangements, and duplication (e.g., functional divergence of paralogous genes). Our regression analysis across species demonstrates that gene regulation is robust and identifiable, although its magnitude decreases with phylogenetic distance. Our results suggest that with sound statistical analysis and additional replicates ([49-51]) even subtly regulated genes can be identified in the distantly related species. Given our results using species that have diverged more than 65 million years ago (guppies and platyfish), it is clear that this array will perform splendidly in the $> 12,000$ species within the large order Perciformes, to which cichlids belong (e.g., gouramis, mackerels, blennies, wrasses, bass, sunfish, perch, gobies, and damselfish).

Future detailed studies focusing on multiple species will benefit from inter-species genomic DNA hybridizations in order to determine spots that are most affected by sequence divergence [2]. Such experiments will differentiate between genes whose regulation is different (genomic-hybridization ratios equal to 1) and genes whose sequence has diverged considerably (genomic-hybridization ratios significantly different from 1). We explored this strategy by competitively hybridizing to the array *A. burtoni* genomic DNA against genomic DNA from either the Nile tilapia (ca. 10 million years divergence time) or the zebrafish (more than 200 million years of divergence). As we had previously determined which of the 804 reference spots were significantly regulated in either of these two species (see Figure 4), we divided the genomic hybridization results for the reference spots into two classes depending on whether they were also brain-enriched in the other species or not. Interestingly, the mean ratios of these two classes were not different in the

O. niloticus/*A. burtoni* genomic DNA hybridization (Student t-test: $t = -1.6$, $p = 0.1$). However, when genomic DNA from *A. burtoni* and the distantly related *D. rerio* was competitively hybridized to the array, we not only found many spots that hybridized preferentially with *A. burtoni* genomic DNA; we also found a significant difference for the mean hybridization ratios ($t = -9.4$, $p < 0.001$) between the two reference spot classes (i.e., those spots that did and those that did not show significant brain-specific regulation in *D. rerio*). These results suggest that the difference in gene regulation observed between *A. burtoni* and the Nile tilapia may be due to real functional differences while the small number of re-identified reference spots observed in zebrafish may be largely due to sequence divergence. Sequence divergence hinders accurate hybridization at these spots during heterologous hybridization experiments, indicating that these spots cannot be used for functional analysis within this species. In conclusion, genomic DNA hybridization experiments can be used to estimate the false negative rate for a within-species RNA experiment and may be essential for distinguishing between variation due to sequence divergence and variation due to transcript abundance in across-species RNA experiments. Two general rules can be derived from this analysis: First, identify the phylogenetically closest existing array platform; second, before initiating an extensive expression profiling experiment utilizing heterologous hybridization to any array, conduct a statistical analysis of genomic hybridization results. These steps will maximize the number of useful spots and assure the disqualification of those spots whose DNA hybridization ratios are significantly different from 1.

The great number of ecological, evolutionary, aquaculture and conservation studies in widely divergent fish species will be greatly enhanced by the development of genomic resources. Because natural variation is fundamentally polygenic and arises from complex interactions within the genome as well as with the environment, a multiple-gene approach to the study of phenotypic regulation will provide new insights. The combination of diverse ecological characteristics in African cichlid fishes and their ability to reproduce a full behavioural repertoire in captivity provides a powerful framework for studies both in the field and in the laboratory. Their astonishing phenotypic diversity, despite minimal genetic divergence, the result of a uniquely rapid and recent radiation (e.g. [52-54]), allows us to utilize a single cichlid microarray to study the more than 2000 different East African cichlid species. We foresee the utility of this array for examining natural variation of gene expression as it relates to phenotypic plasticity, adaptation, and speciation, and population studies central to organismal and evolutionary biology. Both within and across species this microarray can be used to study the molecular basis of species-specific characters such as jaw

morphology, male colour patterns, brain anatomy, reproduction, and behaviour, as well as the mechanisms underlying phenotypic plasticity, which may contribute to the rapid rate of speciation (reviewed in [55]).

While the cichlid fish cDNA microarray will greatly facilitate the comparative functional genomic approach for an important group of fishes, we expect that the results of our systematic heterologous hybridization studies presented here will encourage researchers in many fields to utilize existing cDNA arrays for diverse groups of teleosts and other taxa.

Conclusions

We have constructed a cDNA microarray with ~4500 features from a brain-specific cDNA library for the African cichlid fish *Astatotilapia burtoni*. We describe the first quantitative functional analysis of heterologous hybridization across a range of vertebrate species to a single cDNA microarray platform. We validate a genomic strategy that overcomes some of the restrictions imposed by systems for which only limited sequence information is available. Although most robust when sample RNA is derived from closely related cichlids, expression profiling results showed consistent hybridization for other closely related taxa (~65 million years divergence) and, to a lesser extent, even very distantly related species. This work represents a first step toward bringing genomics to bear in cichlids and other non-traditional model systems. Crucially, we demonstrate the feasibility of functional genomic studies in a comparative context for any organism.

Methods

Part I: Construction of a custom-made cDNA array

Library construction – A full-length, directional (EcoRI – XhoI) cDNA library was constructed in Lambda ZapII phage vector (Stratagene) with mRNA from *A. burtoni* brains (both sexes at all stages of development and social condition were included) and was generously provided by U. DeMarco and R. Fernald (Stanford University). The pBluescript phagemid, pBSIIsk, was excised from the Lambda ZAP vector, following protocol for transformation into XL1-Blue MRF' (*Stratagene*) *E. coli* strain for plating and picking.

Plating, selection, and amplification of bacterial colonies – Cells were plated on LB agar supplemented with ampicillin in 20 cm Q-bot trays (Genetix). 5755 Bacterial colonies were selected by the Q-bot (Genetix) and inoculated into 96-well plates with 150 μ l LB+amp glycerol for overnight growth at 37°C in a humid incubator.

Replicated plates (without glycerol), produced a working set of 58 plates for PCR amplification. Plasmid inserts

were amplified by colony PCR in Microseal 96-well plates (MJ Research) on MJ Tetrads (MJ Research) using custom vector primers for pBSIIsk- (CSVP2: TTCCCAGTCAC-GACGTTGTA AAA, 23mer, $T_m = 60.9^\circ\text{C}$; CSVP3: AAGCGCGCAATTAACCCTCACTA, 23mer, $T_m = 62.7^\circ\text{C}$). Reaction conditions were as follows: 1 × Taq Buffer + 2 mM MgCl₂; 0.25 mM dNTP mix; 0.18 μM each primer; 1.5 units FastStart Taq (Roche). Samples were denatured for 5 min at 95 °C followed by the 35 cycles of 95 °C for 45 sec, 60 °C for 20 sec, 72 °C for 3 min. Samples were then held at 72 °C for 5 min and stored at 4 °C. PCR products were visualized on 1% agarose gels and scored for strong, single product (4570 passes = 79.4%). The plates were purified by vacuum filtration to remove excess dNTPs and primers using the MultiScreen-PCR 96-well Filtration System (Millipore); re-suspended in MilliQ-grade water to an average estimated concentration of 100–200 ng/μl; transferred to Costar 96-well V bottom polypropylene storage plates (Corning); and dehydrated for storage. After all inserts had been amplified, the products were re-suspended in nuclease-free de-ionized water and compressed into a 384-well plate format without re-configuration using a BioMek FX liquid handling robot (Beckman Instruments) and sterile barrier tips (Beckman-Coulter). The plates were dehydrated for storage and re-hydrated in 10 μl of 3 × SSC for array printing.

Array production – All *A. burtoni* cDNA clones (including 1185 that failed the gel analysis above) and 120 control clones were spotted in duplicate arrays onto NaOH cleaned, poly-lysine (Sigma) coated slides using the 16-pin format on an OmniGrid-100 arrayer (GeneMachines). Yeast, *Arabidopsis*, mouse, and human clones were included as negative controls.

Slides were re-hydrated and UV cross-linked with 6000 mJ (Stratalinker). Slides were blocked with succinate anhydride, 1-Methyl-2 polypyrrolidinone and sodium borate, then denatured in boiling water and spun dry according to standard protocol [56]. Hydrated and blocked arrays were stored in light-proof containers in a desiccator until hybridization.

Part II: Characterization of cDNA array protocol and cross-species hybridizations

Fish species used – Male *A. burtoni*, *Enantiopus melanogenys* and *Neolamprologus brichardi* were randomly selected from a lab-reared stock. The Tilapia (*O. niloticus*) was obtained from aquaculture supplier. The other non-cichlid species were obtained from a local supplier (*Xiphophorus* sp., *Poecilia reticulata*) from the Harvard University zebrafish facility (*Danio rerio*) and the S.O. Conte Anadromous Fish Research Center (Atlantic salmon, *S. salar*).

Fish were killed with 0.03% tricaine methanesulfonate (Sigma) in accordance with the animal protocol (#22-22) approved by the Harvard University Institutional Animal Care & Use Committee, and brains and a mixture of "body tissues" containing muscle, skin, and bone, were dissected out immediately. The samples were minced in 1 ml of RNAlater solution (Ambion) and stored in 4 ° overnight followed by long term storage at -20 °C.

DNA extraction- Genomic DNA was isolated from mixed tissue. Approximately 100 mg of tissue was homogenized and digested in buffer solution (60 mM Tris, pH 8.0, 100 mM EDTA, 0.5% SDS) containing proteinase K (0.5 mg/ml) at 37 °C for 12 to 16 hours followed by phenol:chloroform:iso-amyl alcohol extraction (25:24:1) using the Phase Lock Gel light system (Eppendorf) for phase separation. Yield and quality was evaluated by gel analysis and standard spectrophotometric measurements.

DNA labelling- For each DNA probe 2 μg of genomic DNA was restriction-digested with Sau3aI (New England Biolabs) and labelled according to a standard Klenow protocol (Invitrogen) with direct incorporation of Cy3 or Cy5-dCTP (Amersham). Labelled DNA was purified and concentrated on a YM30 Amicon (Millipore) filter, salts were adjusted to 3XSSC and 1.5 % SDS. The denatured probe was applied beneath a lifter cover slip (Erie Scientific Corp.) and hybridized overnight in the dark at 65 °C in a humidified chamber (Telechem) submerged in a water bath. Excess probe was removed by rinsing in 2 × SSC 0.01 % SDS at 65 °C followed by two rinses at room temperature (1 × SSC and 0.2 × SSC) and centrifuged to dry.

RNA extraction – Total RNA was extracted from brains and mixed tissue of males according to a standard Trizol protocol (Invitrogen), following tissue homogenization (Tissue Tearor, Biospec Products). The RNA was analyzed for quantity and quality on the Bioanalyzer (Agilent) and a standard spectrophotometer (Agilent).

RNA labelling – Two μg of total RNA was labelled for each sample ([56] by first annealing primer in a 15 μl reaction with 1 μl of primer solution (5 μg/μl each poly dT 12–18 or 5 μg/μl each poly dT 12–18 with 5 μg/μl random hexamer oligonucleotides). Reverse transcription reactions were prepared on ice: 5 μl 5 × 1st strand buffer (Invitrogen); 2 μl 0.1 M DTT; 0.6 μl 50 × amino-allyl-dUTP/dNTP mix (2.5 mM each dATP, dCTP, dGTP, 1.5 mM dTTP (Invitrogen) and 10 mM amino-allyl dUTP (Sigma)); and 2 μl (200 U/μl) SuperScript II (Invitrogen), and then incubated at 42 °C for 2 hours. RNA was hydrolyzed, and the reaction was stopped by adding 10 μl of 1 N NaOH and 10 μl of 0.5 M EDTA and placed at 65 °C for 7 min. The reaction was neutralized with 25 μl of 1 M HEPES pH 7.5

(GIBCO BRL). The cDNA was then rinsed and concentrated on a YM-30 filter (Millipore). The dye-coupling reaction required adding 1.5 μ l of 1 M sodium bicarbonate pH 9.0 and the appropriate Cy3 or Cy5 CyDye Post-labeling reactive Dye Pack (Amersham) and placing it for 1 hour at room temperature in the dark. The labelled cDNA was then purified using a Qiagen PCR column, pooled with the appropriate sample for competitive hybridization and concentrated to 50 μ l on a YM 30 filter. The appropriate hybridization buffer conditions were achieved by adding 6 μ l 20 \times SSC (Gibco), 3 μ l poly (dA) poly(dT) (Sigma) and 0.96 μ l 1 M HEPES and 0.9 μ l 10% SDS to each combined labelled probe. Hybridizations and post-hybridization processing were performed as in the DNA hybridization procedure (see above). Note that Cy3 and Cy5 dyes were "swapped" between tissues when technical replicates were performed, such that brain RNA was labelled at least once in "green" (Cy3) and once in "red" (Cy5) in a given species to avoid gene-by-dye effects [1].

Analysis – Hybridized arrays were scanned with an Axon 4000B scanner (Axon Instruments) using Genepix 4.0 software (Axon Instruments) for initial spot finding. The data sets were filtered for spots flagged as "bad" because of irregularities on surface of array (dust, speckle, scratch). Intensity values of spots showing hybridization intensity two standard deviations above background intensity in both channels were used for spot counting and correlation analysis on technical replicates of *A. burtoni* genomic DNA.

Raw data from Genepix was imported into R and analyzed using the LIMMA library (*Linear Models for Microarray Data*, [57]) for within-array print-tip loess normalization of intensities, identification of statistically significant regulation (moderated t-statistics using empirical Bayes shrinkage of the standard errors), and calculation of average fold-changes. Background subtracted intensities from unflagged spots were used in normalization and model fitting. The normalized and fitted data of intensities, number of significantly regulated spots and fold change were used for all remaining intra- and inter-species analysis.

The raw and analyzed data for the 24 microarray experiments used in this study have been submitted to Gene Expression Omnibus (SERIES ID = GSE975, available online [58]). The ESTs representing the cDNAs on the microarray have been submitted to NCBI GenBank.

All correlations analyses were performed using Pearson correlation coefficient tests. Linear regression analyses were used to estimate the amount of variation in fold change observed in a heterologous hybridization that could be explained by the fold change observed in *A. bur-*

toni and estimate the slope of the relationship between these two variables.

Authors' contributions

SCPR and NAH jointly designed and carried out the laboratory manipulations, microarray experiments, and statistical analyses and drafted the manuscript. HAH conceived of the project, participated in its design and coordination and took part in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Uli DeMarco and Russell Fernald (Stanford University) for providing the cDNA library. We thank Sarah Annis, Claire Bailey, Christian Daly and Keith Morneau for expert technical assistance with the array construction and hybridization protocol, Christian R. Landry for bioinformatics support as well as Josiah Altschuler for animal husbandry. Colin Meiklejohn, Yuk Fai Leung, Ping Ma and Jon Wilkins provided valuable comments on earlier versions of the manuscript. This work was supported by a NIH National Research Service Award to SCPR, a FQRNT (Fonds québécois de la recherche sur la nature et les technologies) postdoctoral fellowship to NAH, and by the Bauer Center for Genomics Research.

References

- Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G: **The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster***. *Nat Genet* 2001, **29**:389-395.
- Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL: **Sex-dependent gene expression and evolution of the *Drosophila* transcriptome**. *Science* 2003, **300**:1742-1745.
- Meiklejohn CD, Parsch J, Ranz JM, Hartl DL: **Rapid evolution of male-biased gene expression in *Drosophila***. *Proc Natl Acad Sci U S A* 2003, **100**:9894-9899.
- Bochdanovits Z, van der Klis H, de Jong G: **Covariation of larval gene expression and adult body size in natural populations of *Drosophila melanogaster***. *Mol Biol Evol* 2003, **20**:1760-1766.
- Townsend JP, Cavalieri D, Hartl DL: **Population genetic variation in genome-wide gene expression**. *Mol Biol Evol* 2003, **20**:955-963.
- Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrouf RE, Paabo S: **Intra- and interspecific variation in primate gene expression patterns**. *Science* 2002, **296**:340-343.
- Gu J, Gu X: **Induced gene expression in human brain after the split from chimpanzee**. *Trends Genet* 2003, **19**:63-65.
- Nahon JL: **Birth of 'human-specific' genes during primate evolution**. *Genetica* 2003, **118**:193-208.
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels CA, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoeve F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes***. *Science* 2002, **297**:1301-1310.
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusser DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Lofth B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburg P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscuti D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ,

- Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL: **The genome sequence of the malaria mosquito *Anopheles gambiae***. *Science* 2002, **298**:129-149.
11. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC: **The dog genome: survey sequencing and comparative analysis**. *Science* 2003, **301**:1898-1903.
 12. Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE: **Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee**. *Genome Res* 2002, **12**:555-566.
 13. Whitfield CW, Cziko AM, Robinson GE: **Gene expression profiles in the brain predict behavior in individual honey bees**. *Science* 2003, **302**:296-299.
 14. Gracey AY, Troll JV, Somero GN: **Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis***. *Proc Natl Acad Sci U S A* 2001, **98**:1993-1998.
 15. Rise ML, von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR, Hunt P, Shukin R, Zelnik JA, Nelson C, Jones SR, Smailus DE, Jones SJ, Schein JE, Marra MA, Butterfield YS, Stott JM, Ng SH, Davidson WS, Koop BF: **Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics**. *Genome Res* 2004, **14**:478-490.
 16. Hittel D, Storey KB: **Differential expression of adipose- and heart-type fatty acid binding proteins in hibernating ground squirrels**. *Biochim Biophys Acta* 2001, **1522**:238-243.
 17. Gracey AY, Cossins AR: **Application of microarray technology in environmental and comparative physiology**. *Annu Rev Physiol* 2003, **65**:231-259.
 18. Hogstrand C, Balesaria S, Glover CN: **Application of genomics and proteomics for study of the integrated response to zinc exposure in a non-model fish species, the rainbow trout**. *Comp Biochem Physiol B Biochem Mol Biol* 2002, **133**:523-535.
 19. Ju Z, Dunham RA, Liu Z: **Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation**. *Mol Genet Genomics* 2002, **268**:87-95.
 20. Oleksiak MF, Kolell KJ, Crawford DL: **Utility of natural populations for microarray analyses: isolation of genes necessary for functional genomic studies**. *Mar Biotechnol (NY)* 2001, **3**:S203-11.
 21. Oleksiak MF, Churchill GA, Crawford DL: **Variation in gene expression within and among natural populations**. *Nat Genet* 2002, **32**:261-266.
 22. Nakazawa M, Cebria F, Mineta K, Ikeo K, Agata K, Gojobori T: **Search for the evolutionary origin of a brain: planarian brain characterized by microarray**. *Mol Biol Evol* 2003, **20**:784-791.
 23. Stearns SC, Magwene P: **The naturalist in a world of genomics**. *Am Nat* 2003, **161**:171-180.
 24. Feder ME, Mitchell-Olds T: **Evolutionary and ecological functional genomics**. *Nat Rev Genet* 2003, **4**:651-657.
 25. Helfman GS, Collette BB, Facey DE: **The diversity of fishes**. Malden, Blackwell science; 1997:528.
 26. Fryer G, Iles TD: **The Cichlid Fishes of the Great Lakes of Africa**. Edinburgh, Oliver & Boyd; 1972:641.
 27. Streelman JT, Zardoya R, Meyer A, Karl SA: **Multilocus phylogeny of cichlid fishes (Pisces: Perciformes): evolutionary comparison of microsatellite and single-copy nuclear loci**. *Mol Biol Evol* 1998, **15**:798-808.
 28. Sturmbauer C: **Explosive speciation in cichlid fishes of the African Great lakes: a dynamic model of adaptive radiation**. *Journal of Fish Biology* 1998, **53**:18-36.
 29. Albertson RC, Markert JA, Danley PD, Kocher TD: **Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa**. *Proc Natl Acad Sci U S A* 1999, **96**:5107-5110.
 30. Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M: **Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences**. *Mol Phylogenet Evol* 2003, **26**:121-138.
 31. Verheyen E, Salzburger W, Snoeks J, Meyer A: **Origin of the super-flock of cichlid fishes from Lake Victoria, East Africa**. *Science* 2003, **300**:325-329.
 32. Taborsky M: **Broodcare helpers in the cichlid fish *Lamprologus brichardi*: their costs and benefits**. *Animal Behavior* 1984, **32**:1236-1252.
 33. Liem KF: **Functional Morphology**. *Cichlid Fishes: Behaviour, ecology and evolution* Edited by: Keenleyside MHA. Dordrecht, Kluwer Academic Publishers; 1991.
 34. Meyer A: **Phylogenetic relationships and evolutionary processes in East African cichlid fishes**. *Evolutionary Ecology* 1993, **8**:279-284.
 35. Huysseune A: **Phenotypic plasticity in the lower pharyngeal jaw dentition of *Astatoreochromis allaudi* (Teleostei: Cichlidae)**. *Arch Oral Biol* 1995, **40**:1005-1040.
 36. Balshine-Earn S, Neat FC, Reid H, Taborsky M: **Paying to stay or paying to breed? Field evidence for direct benefits of helping behavior in a cooperatively breeding fish**. *Behavioral Ecology* 1998, **9**:432-438.
 37. Barlow GW: **The cichlid fishes: Nature's grand experiment in evolution**. Cambridge, Perseus Publishing; 2000:335.
 38. Hofmann HA, Fernald RD: **Social status controls somatostatin neuron size and growth**. *J Neurosci* 2000, **20**:4740-4744.
 39. Seehausen O, Witte F, Alphen JJM, Bouton N: **Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria**. *Journal of Fish Biology* 1998, **53**:37-55.
 40. Danley PD, Kocher TD: **Speciation in rapidly diverging systems: lessons from Lake Malawi**. *Mol Ecol* 2001, **10**:1075-1086.
 41. DOE Joint Genome Institute: **JGI Fugu v3.0**. <http://genome.jgi-psf.org/fugu6/fugu6.home.html>.
 42. Datson NA, van der Perk J, de Kloet ER, Vreugdenhil E: **Expression profile of 30,000 genes in rat hippocampus using SAGE**. *Hippocampus* 2001, **11**:430-444.
 43. Evans SJ, Datson NA, Kabbaj M, Thompson RC, Vreugdenhil E, De Kloet ER, Watson SJ, Akil H: **Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis. Serial Analysis of Gene Expression**. *Eur J Neurosci* 2002, **16**:409-413.
 44. Ryan MJ, Rosenthal GG: **Variation and selection in swordtails**. *Model Systems in Behavioral Ecology: Integrating Conceptual, Theoretical, and Empirical Approaches* Edited by: Dugatkin LA. Princeton, Princeton University Press; 2001:133-148.
 45. Reznick DA, Bryga H, Endler J: **Experimentally induced life-history evolution in a natural population**. *Nature* 1990, **346**:357-359.
 46. Carroll RL: **Vertebrate paleontology and evolution**. New York, W.H. Freeman and company; 1987:698.
 47. Stearns SC, Hendry AP: **Evolution illuminated: Salmon and their relatives**. Oxford, Oxford university press; 2003.
 48. Westerfield M: **The Zebrafish Book**. Eugene, University of Oregon Press; 1995.
 49. Churchill GA: **Fundamentals of experimental design for cDNA microarrays**. *Nat Genet* 2002, **32** Suppl:490-495.
 50. Cui X, Churchill GA: **Statistical tests for differential expression in cDNA microarray experiments**. *Genome Biol* 2003, **4**:210.
 51. Lee ML, Kuo FC, Whitmore GA, Sklar J: **Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations**. *Proc Natl Acad Sci U S A* 2000, **97**:9834-9839.
 52. Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C: **Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas**. *Syst Biol* 2002, **51**:113-135.
 53. Seehausen O: **Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14,600 year history for its cichlid species flock**. *Proc R Soc Lond B Biol Sci* 2002, **269**:491-497.
 54. Nishida M: **Phylogenetic relationships and evolution of Tanganyikan cichlids: A molecular perspective**. *Fish Communities in Lake Tanganyika* Edited by: Kawanabe H, Hori M and Nagoshi M. Kyoto, Kyoto University Press; 1997:1-23.
 55. Kocher TD: **Adaptive evolution and explosive speciation: the cichlid model**. *Nature Reviews Genetics* 2004, **5**:288-298.
 56. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J: **A concise guide to cDNA**

- microarray analysis.** *Biotechniques* 2000, **29**:548-50, 552-4, 556 passim.
57. Smyth GK, Yang YH, Speed T: **Statistical issues in cDNA microarray data analysis.** *Methods Mol Biol* 2003, **224**:111-136.
58. Gene Expression Omnibus: <http://www.ncbi.nlm.nih.gov/geo/query?acc=GSE975>.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

