

Comparative Genomics

Genomes of mice and men

Austin L Hughes

Heredity (2003) 90, 115–116. doi:10.1038/sj.hdy.6800222

The destinies of mice and men have been linked since sometime early in the Neolithic period. However, the newly published draft sequence of the mouse genome means that our unwelcome houseguest can now teach us more about ourselves than ever before (Mouse Genome Sequencing Consortium, 2002).

The house mouse (*Mus musculus*) first linked itself to humankind by learning to exploit a new ecological niche created when humans started to store grain. These hardy commensals have been with us ever since. Despite our best efforts to get rid of mice, we have continued to provide food and shelter for them, and indeed we have transported this originally Eurasian species around the globe.

The rise of modern biomedicine has given the mouse an opportunity to repay its debt to us by serving as the primary mammalian model in a wide variety of studies. Now this perennial pest of human households is poised to help us understand our own biology at the most fundamental level—that of the genome.

Biology is inevitably a comparative science because it is a historical science. We can only fully understand the molecular mechanisms underlying health and disease in our own species if we understand the evolutionary history that gave rise to those mechanisms. The best method we have to reconstruct the evolutionary past of any species is comparison with living relatives. The mouse genome will shed light on the human genome because this is the closest human relative whose genome has been sequenced so far. Furthermore, the human–mouse genomic comparison will shed important light on the general mechanisms of genome evolution because these are the two most closely related eukaryotic species from which genomic sequences are currently available.

One outcome of the human–mouse genome comparison is that it has made possible a refined estimate of the number of protein-coding genes in both genomes, which the authors (Mouse Genome Sequencing Consortium, 2002) place at around 30 000. This value is at the low end of the range previously

estimated on the basis of the first draft sequences of the human genome (International Human Genome Sequencing Consortium, 2001). It is of course less than a third as large as the ballpark estimate of 100 000 genes in the mammalian genome frequently cited in the pre-genomic era. Interestingly, less than 1% of the protein-coding genes in the mouse genome lack a detectable homologue in what is known of the human genome, while a similar fraction of human genes lack a known homologue in mouse (Mouse Genome Sequencing Consortium, 2002).

Comparison of genomes will help us to understand the molecular basis of biological differences between species. Some of the new data on the mouse indicate possible sources of such differences. For example, there are many gene families in which independent gene duplications have occurred in the mouse lineage after its separation from the primate lineage. Many of these families encode proteins involved in the reproductive, immune, or olfactory systems (Mouse Genome Sequencing Consortium, 2002), and it is plausible that these systems exhibit rodent-specific adaptations.

The pattern of nucleotide substitution between putatively orthologous human and mouse genes provides additional evidence of adaptive divergence. The ratio of nonsynonymous (amino acid altering) to synonymous substitution (called the K_A/K_S ratio) was particularly high in a small set of genes. A high K_A/K_S suggests the occurrence of positive Darwinian selection favoring changes at the amino-acid level (Hughes and Nei, 1988). Most of the genes with high K_A/K_S ratios encode proteins involved in the immune system, consistent with earlier results showing a high rate of amino-acid evolution in mammalian immune signaling proteins (Murphy, xxx).

Previous studies have suggested that mammalian genome evolution has involved the shuffling of large genomic segments containing syntenic groups of genes in which the gene order is largely conserved apart for a few local rearrangements (Nadeau and Taylor, 1984). The Mouse Genome Sequencing Consor-

tium (2002) found a total of 217 such conserved syntenic blocks between human and mouse. An unanswered question is whether the breaks between blocks occur at random or whether certain linkage groups are conserved in evolution because linkage of the genes involved is functionally important. Answering this question will require comparison of additional genomes from other orders of mammals.

Perhaps the most surprising discovery that emerges from the comparison of human and mouse genomes is that, in about 5% of nonoverlapping 50-bp windows spanning regions homologous between the two genomes, the rate of nucleotide substitution was lower than expected in the absence of selection (Mouse Genome Sequencing Consortium, 2002). This result suggests that about 5% of the genome is subject to selective constraints and thus is of functional importance in the two species. Protein-coding genes account for only about 1.5% of the genome in both species, so apparently 3.5% of the genome is functionally constrained despite not encoding proteins. A companion paper comparing human chromosome 21 with syntenic regions of the mouse genome provides additional support for the existence of a large number of conserved blocks that are not protein-coding (Dermitzakis *et al*, 2002).

The remaining selectively constrained regions may include non-protein-coding RNAs (which are notoriously difficult to identify using currently available software) as well as sequence elements important in the regulation of gene expression or in protein structure. Obviously, identifying the function of such conserved regions will play an important part in understanding our own genomic biology.

A frequently heard justification for genome sequencing projects is that they will shed light on the molecular mechanisms of human disease. Even in initial analyses, the mouse genome shows promise in this regard. The Mouse Genome Sequencing Consortium examined regions of the mouse genome homologous to 687 human disease genes. Of 7293 amino positions at which disease-associated variants have been reported in the human population, 90.3% show the same amino acid residue in the mouse as in the normal human sequence (Mouse Genome Sequencing Consortium, 2002). This high level of conservation is not surprising if the residues in question are functionally important, as indeed many of them must be if amino-acid replacements at these sites cause disease.

On the other hand, at 160 sites (2.2% of the disease-associated sites), the mouse sequence was the same as the human disease-associated sequence. Moreover, in 23 of these, there is documentation of a cause-and-effect relationship between the mutation and disease in humans (Mouse Genome Sequencing Consortium, 2002). This unexpected finding is important because it implies that the harmfulness of a mutation can depend on the biochemical context in which it occurs. A mutation that causes disease in humans may not have been harmful in the ancestor of the mouse because of other changes occurring in the rodent lineage

that served to buffer the mutation's effects.

As with any new genome sequence, the initial report of the mouse genome (Mouse Genome Sequencing Consortium, 2002), together with companion papers (Dermitzakis *et al*, 2002; The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I and II Team, 2002; Wade *et al*, 2002), only scratches the surface of the information made available to biologists through sequencing of the mouse genome. Comparison of human and mouse genomes is certain to yield important new insights in the near future as well as provide a rich source of testable hypoth-

eses for experimental biologists working in both rodent and primate systems. ■
Austin L Hughes is at the Department of Biological Sciences, University of South Carolina, Columbia, SC 29208 USA.

e-mail: austin@biol.sc.edu

Dermitzakis ET *et al* (2002). *Nature* **420**: 578–582.
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I and II Team (2002). *Nature* **420**: 563–578.
Hughes AL, Nei M (1988). *Nature* **335**: 167–170.
International Human Genome Sequencing Consortium (2001). *Nature* **409**: 860–921.
Mouse Genome Sequencing Consortium (2002). *Nature* **420**: 520–562.
Murphy PM (1993). *Cell* **72**: 823–826.
Nadeau JH, Taylor BA (1984). *Proc Natl Acad Sci USA* **81**: 814–816.
Wade CM *et al* (2002). *Nature* **420**: 574–578.

Evolutionary Genetics

Rose-colored goggles

Karen L Carleton and Thomas D Kocher

Heredity (2003) **90**, 116–117. doi:10.1038/sj.hdy.6800223

Evolutionary biologists are fascinated by how so many species of cichlid fishes have evolved in the Great Lakes of East Africa. It seems likely that visual cues played a key role in this explosive speciation process. In a recent paper, Terai *et al* (2002) provide compelling evidence from one of these amazing species flocks that selection is acting on a key vision gene, suggesting that for cichlids 'seeing is evolving'.

The mechanisms of speciation are difficult to unravel in the simplest systems. When hundreds of species arise quickly, within a small geographic area, the task becomes daunting. There are no more extreme examples than the Great Lakes Tanganyika, Malawi and Victoria, which each harbor hundreds of recently evolved cichlid species (Kornfield and Smith, 2000; Danley and Kocher, 2001).

Pre-mating isolation (where females will only mate with males of their own species) is key to the evolution and maintenance of so many cichlid species. Males display species-specific bright color patterns that differ between the most closely related taxa. Both in the laboratory and the field, females choose males based on these color patterns (Seehausen *et al*, 1997). Therefore, in their new study, Terai *et al* (2002) set out to examine a gene determining visual sensitivity in a broad array of cichlids from Lake Victoria.

Much is known about how visual sensitivity is controlled. The retina contains several types of cones. Each cone type has a unique visual pigment, which absorbs light in a different part of the spectrum (Figure 1). Each visual pigment is comprised of an opsin protein, wrapped around a vitamin-A-derived chromophore. The opsin genes fall into four broad spectral classes that arose early in vertebrate

evolution. In cichlids, these include very short-wavelength sensitive opsins (SWS1), short-wavelength sensitive opsins (SWS2), medium-wavelength sensitive opsins (RH2) and long-wavelength sensitive opsins (LWS) (Carleton and Kocher, 2001).

A cone's visual pigment can be tuned to maximally absorb different wavelengths of light in three ways (Bowmaker, 1995): (1) opsin genes from different spectral classes can be expressed, causing large shifts of 35–100 nm (Bowmaker, 1995; Carleton and Kocher, 2001), (2) the chromophore (or mixture of chromophores) used can differ in fish, shifting all the visual pigments simultaneously (with large shifts of 30–60 nm in the LWS pigments and small shifts of 5–10 nm in the SWS pigments, Harosi, 1994); or (3)

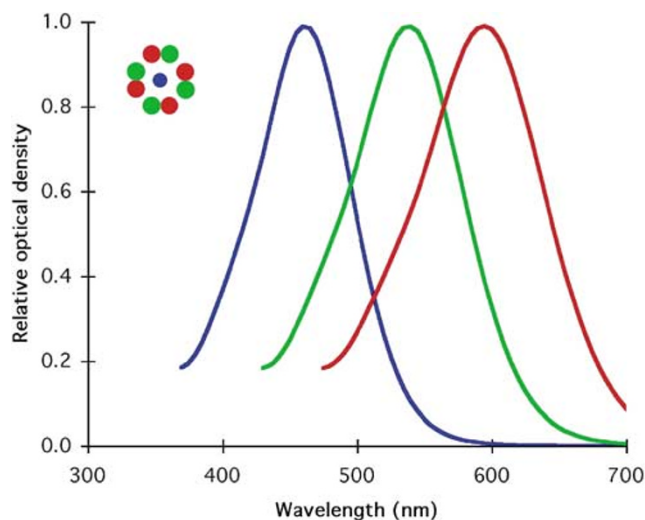


Figure 1 Visual pigment absorption spectra based on the microspectrophotometry of the Lake Victorian cichlid *Haplochromis pyrrhocephalus* (van der Meer and Bowmaker, 1995). The single cones (blue) contain a pigment with peak sensitivity at 462 nm, while the double cones contain pigments with peak sensitivity at 539 (green) and 595 (red) nm. The 595 nm pigment is coded for by the LWS opsin gene. The inset shows the square retinal mosaic typical of cichlids where single cones are surrounded by double cone pairs.