Cell

Brain Evolution and Uniqueness in the Human Genome

Jordan P. Amadio[1,2] and Christopher A. Walsh[1,*]
[1]Division of Genetics, Children's Hospital Boston, Howard Hughes Medical Institute, Beth Israel Deaconess Medical Center, and Broad Institute of MIT and Harvard, Boston, MA 02115, USA
[2]Harvard-Massachusetts Institute of Technology Division of Health Sciences and Technology, Boston, MA 02115, USA
*Contact: cwalsh@bidmc.harvard.edu
DOI 10.1016/j.cell.2006.09.007

Despite an ever-expanding database of sequenced mammalian genomes to be mined for clues, the emergence of the unique human brain remains an evolutionary enigma. In their new study, Pollard et al. (2006) trawl the human genome and those of other mammals in search of short conserved DNA elements that show extremely rapid evolution only in humans. As they report in a recent issue of *Nature*, their scan yielded a gene for a novel noncoding RNA that adopts a human-specific structure and may regulate neurodevelopment.

The completion of the human and chimpanzee genome sequences has provided opportunities for studying the evolution of *Homo sapiens* in remarkably stimulating and systematic ways. How did our uniquely human features originate? How was the capacity for language, reason, and culture encoded in genes during evolution? Recent efforts in comparative genomics have begun to address such questions. Using a comparative genomics approach, Haussler and colleagues (Pollard et al., 2006) have now performed a genome-wide scan for regions highly conserved across mammalian genomes that appear to have undergone a sudden and rapid evolution in the human lineage.

Molecular evolution occurs by several mechanisms. Sometimes, new genes can arise or vanish spontaneously—though this turns out to be surprisingly rare. More commonly, nonsynonymous changes in an amino acid coding sequence confer new biochemical properties upon a protein, altering its function. Alternatively, evolution can preserve a protein's amino acid sequence but dramatically alter the spatial and temporal expression of the gene encoding that protein. Previous attempts to elucidate genome evolution across species have focused on either the addition or deletion of genes (Chimpanzee Sequencing and Analysis Consortium, 2005), because they are easy to recognize, or on changes in amino acid sequences (Nielsen et al., 2005), because the statistical methods to test for natural selection acting on coding regions are well established (for reviews, see Sabeti et al., 2006 and Yang and Bielawski, 2000). These approaches have identified several brain-related genes (as well as many genes expressed in other tissues) that have undergone surprisingly rapid evolution in the primate lineage leading to humans. Examples include *FOXP2* (Lai et al., 2001), which encodes a homeodomain protein essential for normal human speech; *ASPM* (Zhang, 2003) and *Microcephalin* (Evans et al., 2004), which are essential for normal cerebral cortical size in humans; and *AHI1* (Ferland et al., 2004), which regulates axon guidance. Such evidence of "positive" selection, where a gene's amino acid sequence has changed more than would be predicted by a neutral model, implies that the gene's alterations are associated with increased evolutionary success. It is now widely agreed, however, that changes in noncoding segments of the genome—regulatory elements, splicing signals, RNA genes, and so forth—can also be extremely important. Yet noncoding DNA, despite comprising most of the genome, is much more difficult to systematically assess for evidence of adaptation; there are no well-established statistical tests for evaluating selection in noncoding regions, and the functional effects of changes to noncoding DNA are rarely obvious.

Pollard et al. (2006) devised a way to scan the genome for selection without explicitly discriminating between coding and noncoding regions. By comparing genomic DNA sequences from several nonhuman mammals, they identified 34,498 mammalian conserved regions (short regions of DNA displaying marked conservation among mammals). Among these otherwise highly conserved regions, they described and ranked 49 "human accelerated regions" (HARs) displaying a significantly accelerated nucleotide substitution rate in humans.

The authors then investigated in detail the highest ranked of the HARs: *HAR1*, a previously uncharacterized 118 bp region in the last band of chromosome 20q. The *HAR1* sequence is well conserved across amniotes, differing at only two nucleotide positions between chickens and nonhuman primates, but it shows an estimated 18 fixed substitutions during the short evolutionary time that separates humans from the ancestor that they share with chimpanzees. It is transcribed as part of two overlapping genes: *HAR1F*, which contains

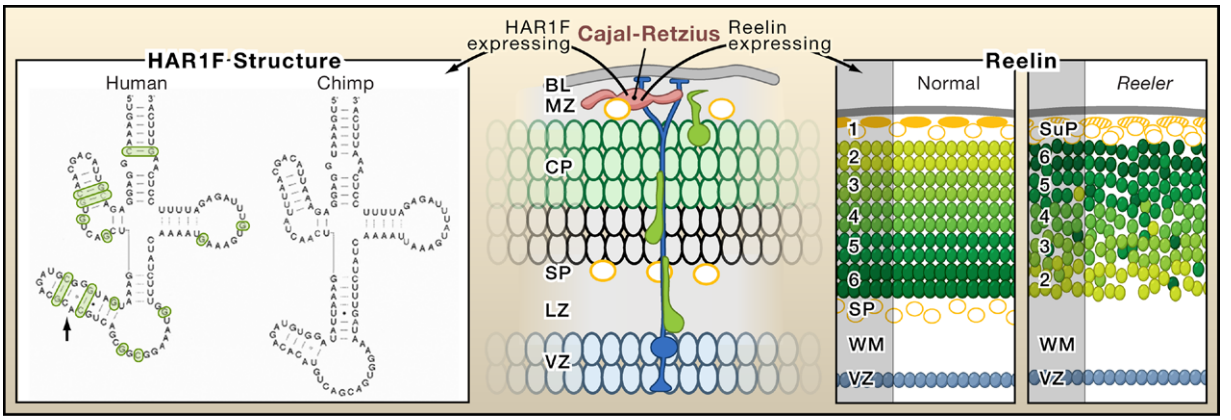

Figure 1. *HAR1F* and Cortical Development

(Left) The *HAR1* region, which lies within a putative noncoding RNA gene, has incorporated 18 human-specific fixed nucleotide substitutions (highlighted) since the divergence of humans and chimpanzees from their common ancestor less than 7 million years ago. The predicted secondary structure of this region of the forward RNA transcript (*HAR1F*) is shown for both human and chimpanzee. In the human structure, which appears to be unique among mammals, one RNA helix is selectively elongated.

(Middle) A representation of the developing neocortex is shown. Neurons (solid green) migrate along the radial glia (solid blue) that span the entire cortical wall from the ventricular zone (VZ) to the basal lamina (BL). Neurons migrate from the ventricular zone through the fiber rich-intermediate zone (IZ) into the developing cortical plate (CP). Cajal-Retzius cells (solid red), which are found in the marginal zone (MZ), express both *HAR1F* and reelin. Although the function of *HAR1* in neuronal development is not known, reelin has been implicated in orchestrating the correct layering of neurons in the cortical plate.

(Right) In wild-type mice, neurons migrate into the cortical plate and form six well-defined layers (green), which overly the subplate (SP) and a band of white matter (WM). In mice lacking reelin, this layering appears disorganized or inverted. In such "reeler" mice, the cortical plate develops beneath the subplate (here called the superplate [SuP] because of its altered position). Future work may establish whether *HAR1* contributes to patterning or migration of cortical neurons.

HAR1 in its first exon, and *HAR1R*, an alternatively spliced gene of two or three exons that contains *HAR1* in its final exon. With the exception of the *HAR1* segment, the *HAR1F* and *HAR1R* transcripts are poorly conserved across mammals, suggesting a disproportionate functional constraint on the *HAR1* segment itself.

Biologically speaking, then, what does the human genome's fastest evolving 118 bp stretch of DNA actually do? *HAR1F* and *HAR1R* do not encode proteins. Rather, the *HAR1* segment was found to form a stable RNA secondary structure. The computationally predicted structure, containing five RNA helices, was empirically supported by *HAR1*'s pattern of human-specific substitutions (Figure 1). Of these, ten substitutions constitute pairs of compensatory changes that preserve Watson-Crick pairing within the predicted helices. The authors used dimethyl sulfate probing and primer extension of in vitro-synthesized *HAR1F* RNAs to experimentally confirm the structure.

HAR1F and *HAR1R* lack homology to any known RNA in the Rfam RNA database and show no evidence of containing tRNA motifs or precursor miRNAs. Hence, they are noncoding RNA (ncRNA) genes. Several ncRNA regulators of RNA polymerase II transcription have recently been discovered, including at least one involved in neuronal differentiation (Goodrich and Kugel, 2006). Like proteins, some ncRNAs are known to catalyze reactions or modulate enzyme activity by folding into stable higher-order structures.

In fact, Pollard et al. (2006) demonstrated that human *HAR1F* RNA folds into a structure found only in humans, wherein one of the helices is extended at the expense of its neighbor. This suggests that human *HAR1F* RNA adopts a conformation appreciably distinct from that of the *HAR1F* RNA belonging to the common human/chimpanzee ancestor. Interestingly, all 18 of the human-specific *HAR1* substitutions are from A or T nucleotides to G or C—in effect, upgrades from weak to strong base pairing. Curiously, this weak-to-strong substitution bias in *HAR1* extends over 1.2 kb, a region far larger than *HAR1* itself. Such changes, which also appear to characterize the HARs as a group, undoubtedly serve to strengthen RNA helices against dissociation and may promote enhanced expression or stability. Nevertheless, the weak-to-strong bias in *HAR1* nucleotide substitutions remains unexplained.

The work of Pollard et al. (2006) is notable in part because the discovery of the fast-evolving *HAR1* led the authors fortuitously—though perhaps not unexpectedly—to neurodevelopment. Human embryonic brain sections showed strong *HAR1F* expression, but no detectable *HAR1R* expression, between 7 and 19 weeks of gestation—a critical period for cortical neuron migration and fate specification. Moreover, *HAR1F* RNA appears to be coexpressed with the cortical patterning protein reelin. Reelin is a specific marker of the Cajal-Retzius neurons inhabiting a part of the cortex that is especially well developed in humans (the subpial granular layer, or SGL). A similar expression pattern was observed in cynomolgus macaque embryos, sug-

gesting some conserved function for *HAR1F* in cortical development over at least the past 25 million years—a function that may have been susceptible to recent sequence changes in the *HAR1F* gene. Reelin acts to specify the layering of the cerebral cortex, and although defects in the reelin pathway have been linked to a number of neurological and psychiatric disorders, it is not yet known whether reelin acts differently in humans compared with nonhuman primates (Olson and Walsh, 2002).

Given the sequence overlap and tissue coexpression of *HAR1F* and *HAR1R*, the authors propose the possibility of antisense regulation between the two transcripts. For instance, the data are consistent with a model in which *HAR1R* is expressed later in development to downregulate *HAR1F*. It is interesting that, in humans, *HAR1F* expression is generally higher than *HAR1R* expression. By contrast, in the mouse, real-time PCR analysis of fetal and adult tissues showed similar levels of the two transcripts. This finding suggests that the regulation of *HAR1F* and *HAR1R* transcripts has itself been tweaked during evolution.

Of the 48 other HARs, nearly a quarter are adjacent to a known gene involved in neurodevelopment. Two are in coding regions: *HAR23* lies within a hypothetical protein called MGC27016, and *HAR38* lies within ITPR1, an inositol 1,4,5-triphosphate receptor that is essential for normal brain function. Of the 23 noncoding HARs predicted to form RNA secondary structures, 8 contain one or more compensatory substitutions supporting the putative structure. Elucidation of these other HARs will provide rich fodder for future studies. Most interesting of all will be further experiments on *HAR1*, which are needed to determine the true extent of its role in brain evolution.

Whole-genome studies like that of Pollard et al. (2006) illustrate the fact that we now tend to think of the human genome in finite terms: Our 20,000–25,000 genes can be compared to the residents of an average-sized town, each with names and addresses, all speaking a language for which the essential grammatical elements are understood. The brain, on the other hand, still evokes celestial metaphors, with its vast numbers of neurons compounded by the 10,000-fold complexity of the interconnections between them. The universal language of neurons, if it exists, has yet to be decoded. As we celebrate the centenary of the first Nobel prizes for neuroscience—to Camillo Golgi and Santiago Ramón y Cajal for their studies on the structure of the nervous system—we are reminded that we really do not understand much about the essential anatomical and functional differences that distinguish our brain from those of our primate relatives. But we do know that those differences were begotten solely from changes to a shared ancestor's DNA. Ultimately, discerning the origins of our uniquely human neurobiology will require integrating our knowledge of genomic structure and nervous-system structure into a functional whole.

ACKNOWLEDGMENTS

C.A.W. is supported by grants from the NINDS and by the Marilyn and James Simons Foundation and the NLM Foundation C.A.W. is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

Chimpanzee Sequencing and Analysis Consortium (2005). Nature *437*, 69–87.

Evans, P.D., Anderson, J.R., Vallender, E.J., Choi, S.S., and Lahn, B.T. (2004). Hum. Mol. Genet. *13*, 1139–1145.

Ferland, R.J., Eyaid, W., Collura, R.V., Tully, L.D., Hill, R.S., Al-Nouri, D., Al-Rumayyan, A., Topcu, M., Gascon, G., Bodell, A., et al. (2004). Nat. Genet. *36*, 1008–1013.

Goodrich, J.A., and Kugel, J.F. (2006). Nat. Rev. Mol. Cell Biol. *7*, 612–616.

Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). Nature *413*, 519–523.

Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fledel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., et al. (2005). PLoS Biol. *3*, e170.

Olson, E.C., and Walsh, C.A. (2002). Curr. Opin. Genet. Dev. *12*, 320–327.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). Nature, in press. Published online August 16, 2006. 10.1038/nature05113.

Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., and Lander, E.S. (2006). Science *312*, 1614–1620.

Yang, Z., and Bielawski, J.P. (2000). Trends Ecol. Evol. *15*, 496–503.

Zhang, J. (2003). Genetics *165*, 2063–2070.