Correspondence

Understanding the Degradation of Hominid Gene Control

Peter D. Keightley, Martin J. Lercher, Adam Eyre-Walker

Recently, two groups have examined the level of sequence constraint in noncoding DNA flanking mammalian genes, and appear to have found conflicting results. By comparing 500-bp blocks in mice and rats, we found that mean nucleotide divergence within 2 kb of the start and stop codons of protein-coding genes is substantially lower than that of introns, and decreases when approaching the coding sequence [1]. If nucleotide changes within introns are largely free from selection, this implies that noncoding blocks close to genes evolved under selective constraints, presumably because they contain gene expression control regions. In contrast, we find that upstream sequences in hominids do not evolve slower than introns, while downstream regions are under about half of the constraint seen in murids [1].

By analysing a similar set of noncoding DNA sequences, Bush and Lahn also found that the mean level of selective constraints in upstream regions between humans and chimpanzees is very low. However, their slightly more complex main analysis was to search for 16-bp sequences within upstream regions that are strongly conserved between humans, mice, and either dogs or chickens. They then examined the divergence between humans and chimpanzees at the flanking nucleotides, finding substantially reduced divergence compared with the genomic mean. This demonstrated selective constraints at certain upstream sequences in hominids. An analogous analysis of mouse–rat sequences showed that the selective constraints are about twice as strong in murids as in hominids [2].

These two findings—on one hand, a near absence of selective constraints in blocks upstream of hominid genes [1], and on the other, evidence for strong selective constraints in these regions [2]—appear to contradict each other. How can we square the two sets of results? The answer is rather simple—windows with high conservation scores are relatively rare, and they contribute little to the mean calculated over 500-bp windows (unfortunately, Bush and Lahn do not tell us the fraction of 5′ alignments within high conservation scores).

Bush and Lahn also suggest that the apparent discrepancy "likely results from the fact that in large 500-bp blocks, functional elements that are under constraint are mixed with large sections of nonfunctional DNA, which are not under constraint" [2]. We believe that this interpretation, while formally correct, obscures important and interesting information that can be gained from combining the two studies. Some sequences outside the conserved 16-mers identified by Bush and Lahn are also likely to be functional, since the same 500-bp regions (largely "nonfunctional" according to Bush and Lahn) show strong evidence of evolutionary constraints between mice and rats [1].

Bush and Lahn also note that constraint, on either side of conserved windows, is greater in murids than in hominids. However, they observe a much smaller difference than that seen in our analysis. This is deceptive because by concentrating attention on regions that are conserved between humans, mice, and dogs, they ignore the fact that there might be many more highly conserved regions in murids than there are in hominids.

In summary, there is no conflict between our results and those of Bush and Lahn; they concentrate their attention on a preselected subset of the sites we considered and so have a different perspective on the problem. What is clear from both studies is that there is a qualitative difference in the level of conservation in the 5′ flanking sequences between murids and hominids. We have argued that this is likely to be due to the fixation of slightly deleterious mutations in hominids that are otherwise selectively eliminated in rodents. Differences in constraints between hominids and murids demonstrate that the overwhelming majority of changes at upstream regulatory sites have only small effects on fitness. This has counterintuitive consequences: to obtain a comprehensive list of human regulatory sites, it might be better to examine conservation in murid rather than hominid genomes. ∎

Peter D. Keightley (keightley_pcb2006@spambob.net)
University of Edinburgh
Edinburgh, United Kingdom

Martin J. Lercher
University of Bath,
Bath, United Kingdom

Adam Eyre-Walker
University of Sussex
Brighton, United Kingdom

References

1. Keightley PD, Lercher MJ, Eyre-Walker A (2005) Evidence for widespread degradation of gene control regions in hominid genomes. PLoS Biol 3: DOI: 10.1371/journal.pbio.0030042
2. Bush EC, Lahn BT (2005) Selective constraint on noncoding regions of hominid genomes. PLoS Comput Biol 1: DOI: 10.1371/journal.pcbi.0010073

Authors' Reply

In their letter responding to our recent paper in *PLoS Computational Biology* [1,2], Keightley et al. provide a clear summary of the similarities and differences between the method used in their study [3] and that which was used in ours. They correctly point out that our study supports their conclusion that compared with rodents there has been an increase in sequence divergence rate in hominid noncoding sequences upstream of genes. They are also correct to say that the two studies are looking at slightly different populations of upstream noncoding sites. In their study, they calculate divergence in large blocks that include many different kinds of sites. Among these are nonfunctional sites, sites conserved

among primates, and sites conserved among all mammals. As a result, their method can be thought of as broad but low resolution. In contrast, our method considers only sites that are likely to be conserved among all mammals, making it more restricted but higher resolution.

Our difference in focus allows us to make an important clarification of their earlier results. We find that despite the overall increase in divergence rate in hominid noncoding regions, significant constraint remains at some sites. In their letter, Keightley et al. acknowledge this point, but argue that such sites are likely to be relatively rare. To respond to this, we can calculate frequency values for different conservation scores from Table S1 of our paper. Windows with a score of 13 or higher constitute 2.7% of the total. (Sites next to these have an average hominid divergence of 0.0086, which is significantly constrained compared with the genome-wide average divergence rate of approximately 0.012.) This means that an average 10-kb upstream noncoding region would have hundreds of bases of this type. This is not a trivial number, and suggests that there are many highly conserved noncoding sites in hominids.

On the other hand, we agree that the method of Keightley et al. includes many sites that we ignore, and may reveal things that our method misses. These sites include functional sites that are not conserved in mouse or dog. Such sites might show an especially high divergence rate in hominids. It would be very interesting to quantitate the hominid divergence rate specifically at such sites, and compare it with the corresponding divergence rate in other mammals.

We would also like to take this opportunity to bring up a cautionary note that applies equally to both studies. Comparing human–chimpanzee divergence with mouse–rat divergence raises a number of complex technical issues because human–chimpanzee divergence is more than one order of magnitude smaller. Such issues include back mutations and varying contributions of polymorphisms and sequencing errors. To extend the work by Keightley et al. and our group, a "cleaner" future study might be to compare hominids with two closely related rodents (or other

mammals) whose divergence is on par with human–chimpanzee divergence. It would be ideal to look at several such species pairs with varying population sizes, which may help one to assess whether the difference in divergence rate between hominids and rodents can be attributed to smaller historical population size in hominids.

Finally, whereas relaxation of selective constraint is a favored explanation for the higher divergence rate in hominids, it is by no means the only explanation. In the longer term, we also look forward to studies that quantitatively address the extent to which the higher hominid divergence rate is due to relaxation of functional constraint, positive selection, or other—as of yet poorly characterized—selective forces such as compensatory mutations. ∎

Eliot C. Bush

Bruce T. Lahn, (blahn@bsd.uchicago.edu)
University of Chicago
Chicago, Illinois, United States of America

References

1. Keightley PD, Lercher MJ, Eyre-Walker A (2006) Understanding the degradation of hominid gene control. PLoS Comput Biol 2: DOI: 10.1371/journal.pcbi.0020019
2. Bush EC, Lahn BT (2005) Selective constraint on noncoding regions of hominid genomes. PLoS Comput Biol 1: DOI: 10.1371/journal.pcbi.0010073
3. Keightley PD, Lercher MJ, Eyre-Walker A (2005) Evidence for widespread degradation of gene control regions in hominid genomes. PLoS Biol 3: DOI: 10.1371/journal.pbio.0030042