Accelerated Evolution of Conserved Noncoding Sequences in Humans

Shyam Prabhakar,[1,2]* James P. Noonan,[1,2]* Svante Pääbo,[3] Edward M. Rubin[1,2]†

Traits that distinguish humans from other primates originated in human-specific DNA sequence changes. To investigate whether gene regulatory or other functional noncoding elements in the human genome bear the signature of accelerated evolution, we determined the occurrence of human-specific substitutions in 110,549 conserved noncoding sequences (CNSs) previously identified by whole-genome sequence comparisons (*1*).

To identify CNSs that accelerated in the human lineage, we developed a test statistic that evaluated the likelihood of observing the configuration of human-specific substitutions present in a given CNS. We assigned each CNS a human-acceleration *P* value based on the probability of observing a configuration of equal or smaller likelihood under the null model of constrained evolution (*1*). We identified 992 elements with a significant excess of human-specific substitutions at $P \leq 0.005$, 79% more than we would expect to see by chance (Fig. 1A and table S1).

To ascertain whether accelerated CNSs disproportionately occur near genes with particular functions, we obtained the Gene Ontology (GO) annotations of the closest neighboring RefSeq gene for all 110,549 CNSs and assigned those annotations to each CNS. Our method explicitly accounted for the unequal distribution of CNSs among GO categories. We then sought to identify GO terms with a significant excess of accelerated CNSs after correcting for variation in statistical power across CNSs (*1*). *P* value thresholds were set to adjust for multiple testing.

The cellular component GO term most significantly enriched in accelerated CNSs was basal lamina (Fig. 1B and table S1). Of the 12 accelerated CNSs in this category, 9 were associated with the dystrophin-associated glycoprotein complex, disruptions of which cause muscle and neuronal diseases (*2*, *3*). Cell adhesion was the only biological process displaying a significant excess of CNSs accelerated in human (Fig. 1B and table S1). Many of these CNSs were associated with genes involved in neuronal cell adhesion, such as cadherins and protocadherins, contactins, and neuroligins. To quantify this observation, we constructed a composite neuronal adhesion GO term by intersecting GO cell adhesion genes with genes annotated in the Entrez Gene database as having evidence of neuronal function. We chose this approach because neuronal adhesion genes are poorly annotated in GO. We found a significant excess of accelerated CNSs neighboring genes with both GO cell adhesion and Entrez Gene neuronal annotations [$P = 0.0069$, one-tailed power-adjusted exact test (*1*) (Fig. 1C and table S1)]. However, when these overlapping accelerated CNSs were removed from the analysis, the number of accelerated CNSs with only GO cell adhesion or Entrez Gene neuronal function annotations was not significantly greater than expected. Thus, the strongest signal of human-specific noncoding sequence evolution that we detected was near genes specifically involved in neuronal cell adhesion.

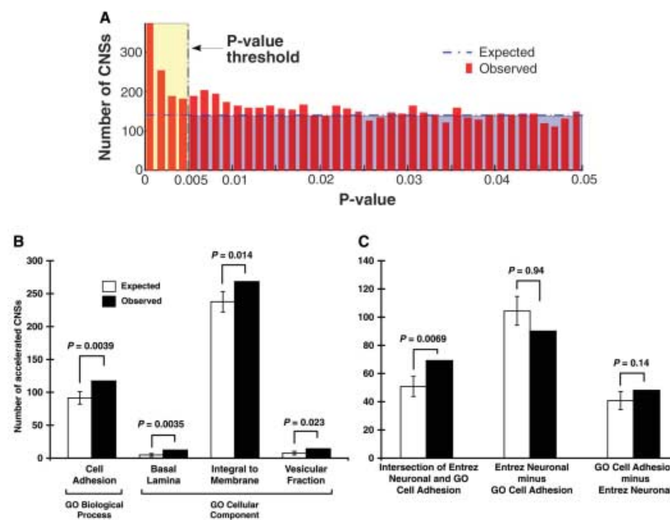

Fig. 1. (**A**) Numbers of human-specific substitution observed versus the uniform distribution expected by random chance. (**B**) GO biological process and cellular component terms enriched in accelerated CNSs. Enrichment *P* values were calculated with the power-adjusted exact test (one-tailed) (*1*). (**C**) Human-accelerated CNSs are disproportionately associated with genes functioning specifically in neuronal cell adhesion. Error bars indicate 1σ.

To determine whether the pattern of CNS acceleration in humans was recapitulated in other lineages, we identified accelerated CNSs in chimpanzee and mouse (*1*). We observed 1050 accelerated CNSs in chimpanzee, only 34 of which were also accelerated in human, indicating a general lack of overlap (table S1). However, CNSs accelerated in chimpanzee were also significantly enriched near neuronal cell adhesion genes (expected = 54, observed = 70, $P = 0.018$). This suggests independent accelerated evolution of neuronal cell adhesion functions in the human and chimpanzee lineages. For the 4607 CNSs accelerated in mouse, we did not detect any enrichment near neuronal cell adhesion genes ($P = 0.99$).

It is unlikely that acceleration of neuronal adhesion CNSs in humans and chimpanzees resulted in the same neuronal phenotypes, because the CNSs accelerated in the two lineages are largely disjoint and would therefore have had different consequences for brain development and cognitive function. Our findings do suggest that cis-regulatory and other noncoding changes may have contributed to the modifications in brain development and function that gave rise to uniquely human cognitive traits.

References and Notes

1. Materials and methods are available as supporting material on *Science* Online.
2. I. Dalkilic, L. M. Kunkel, *Curr. Opin. Genet. Dev.* **13**, 231 (2003).
3. M. P. Moizard *et al.*, *Eur. J. Hum. Genet.* **8**, 552 (2000).
4. Supported by grant HL066681 to the Berkeley PGA, NIH Programs for Genomic Application, funded by the National Heart, Lung, and Blood Institute, and by the director, Office of Science, Office of Basic Energy Sciences, of the U.S. DOE under Contract No. DE-AC02-05CH11231. J.P.N. was supported by NIH National Research Service Award fellowship 1-F32-GM074367.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5800/786/DC1
Materials and Methods
Table S1

1 June 2006; accepted 14 August 2006
10.1126/science.1130738

[1]U.S. Department of Energy (DOE) Joint Genome Institute, Walnut Creek, CA 94598, USA. [2]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: emrubin@lbl.gov

Downloaded from on March 24, 2016