

28. J. Savarino, M. Legrand, *J. Geophys. Res.* **103**, 8267 (1998).
29. P. D. Jones, M. E. Mann, *Rev. Geophys.* **42**, 2003RG000143 (2004).
30. A. Moberg, D. M. Sonechkin, K. Holmgren, N. M. Datsenko, W. Karlen, *Nature* **433**, 613 (2005).
31. M. O. Andreae, P. Merlet, *Global Biogeochem. Cycles* **15**, 955 (2001).
32. C. McEvedy, R. Jones, *Atlas of World Population History* (Penguin, London, 1978), pp. 368.
33. W. Denevan, *The Native Population of the Americas in 1492* (Univ. of Wisconsin Press, Madison, WI, 1992), pp. 386.
34. B. Glaser, L. Haumaier, G. Guggenberger, W. Zech, *Naturwissenschaften* **88**, 37 (2001).
35. C. MacFarling Meure, thesis, University of Melbourne (2004).
36. We thank the staff of the Australian Antarctic Program, especially Casey Station, for field support; A. Smith for firm-air sampling assistance; the Bureau of Meteorology (Australia) for Cape Grim archive-air collection assistance; R. Francey, P. Steele, C. Allison, and S. Coram at CSIRO for logistical and technical help; and especially B. Ruddiman and B. Allan for valuable discussions. Supported by NSF (grant no. OPP0087357); NOAA/Climate Modeling and Diagnostics Laboratory; NIWA, New Zea-

land (Foundation for Research Science and Technology grant no. C01X0204); and the Australian Government's Antarctic Climate and Ecosystems Cooperative Research Centre and CSIRO Atmospheric Research.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5741/1714/DC1

Materials and Methods  
References and Notes

23 May 2005; accepted 28 July 2005  
10.1126/science.1115193

## Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans

Patrick D. Evans,<sup>1,2</sup> Sandra L. Gilbert,<sup>1</sup> Nitzan Mekel-Bobrov,<sup>1,2</sup> Eric J. Vallender,<sup>1,2</sup> Jeffrey R. Anderson,<sup>1</sup> Leila M. Vaez-Azizi,<sup>1</sup> Sarah A. Tishkoff,<sup>4</sup> Richard R. Hudson,<sup>3</sup> Bruce T. Lahn<sup>1\*</sup>

The gene *Microcephalin* (*MCPH1*) regulates brain size and has evolved under strong positive selection in the human evolutionary lineage. We show that one genetic variant of *Microcephalin* in modern humans, which arose ~37,000 years ago, increased in frequency too rapidly to be compatible with neutral drift. This indicates that it has spread under strong positive selection, although the exact nature of the selection is unknown. The finding that an important brain gene has continued to evolve adaptively in anatomically modern humans suggests the ongoing evolutionary plasticity of the human brain. It also makes *Microcephalin* an attractive candidate locus for studying the genetics of human variation in brain-related phenotypes.

The most distinct trait of *Homo sapiens* is the exceptional size and complexity of the brain (1, 2). Several recent studies have linked specific genes to the evolution of the human brain (3–12). One of these is *Microcephalin* (7, 8); mutations in this gene cause primary microcephaly [MCPH; Online Mendelian Inheritance in Man (OMIM) accession 251200] (13, 14). MCPH is defined clinically as severe reductions in brain size coupled with mental retardation, but remarkably, an overall retention of normal brain structure and a lack of overt abnormalities outside of the nervous system (15–17). This led to the notion that the brains of MCPH patients function normally for their size and that genes underlying MCPH are specific developmental regulators of brain size (15–17).

*Microcephalin* is one of six known loci, named *MCPH1* through *MCPH6*, for which recessive mutations lead to MCPH (14, 18–23). For four of these, the underlying genes have been identified as *Microcephalin* (*MCPH1*), *CDK5RAP2* (*MCPH3*), *ASPM* (*MCPH5*), and

*CENPJ* (*MCPH6*) (14, 21, 23). Patients with loss-of-function mutations in *Microcephalin* have cranial capacities about 4 SD below the mean at birth. As adults, their typical brain size is around 400 cm<sup>3</sup> (whereas the normal range is 1200 to 1600 cm<sup>3</sup>), and the cerebral cortex is especially small (13, 14). *Microcephalin* is suggested to control the proliferation and/or differentiation of neuroblasts during neurogenesis. This postulate was consistent with several observations. First, mouse *Microcephalin* is expressed prominently in the proliferative zones of the embryonic brain (14). Second, the *Microcephalin* protein contains several copies of the BRCT domain that is found in cell cycle regulators, such as *BRCAl* (14, 24). Finally, cell culture studies indeed suggested a role of *Microcephalin* in regulating cell cycle (25–27).

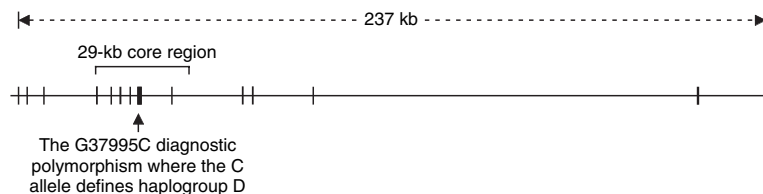
The finding that *Microcephalin* is a critical regulator of brain size spurred the hypothesis

that it might have played a role in brain evolution (16, 28). Consistent with this hypothesis, phylogenetic analysis of *Microcephalin* revealed signatures of strong positive selection in the lineage leading to humans (7, 8). Here, we examine the possibility that positive selection has continued to operate on this gene after the emergence of anatomically modern humans.

The human *Microcephalin* locus has 14 exons spanning about 236 kb on chromosome 8p23 (14) (Fig. 1). We previously sequenced all the exons in 27 humans (8). When re-analyzing the data, we noticed that one haplotype had a much higher frequency than the other haplotypes. Additionally, this haplotype differed consistently from the others at position 37995 of the genomic sequence (counting from the start codon) or position 940 of the open reading frame. This polymorphism falls in exon 8 and changes amino acid residue 314 from an ancestral aspartate to a histidine. (This polymorphism is described as G37995C with G denoting the ancestral allele.)

To investigate whether positive selection has acted on the high-frequency haplotype, we resequenced 23.4 kb of a 29-kb region centered around the G37995C polymorphism (Fig. 1). Sequencing was performed on a panel of 89 individuals from the Coriell Institute, which broadly represents human diversity (see SOM). To assign the ancestral state of polymorphisms, we also sequenced the common chimpanzee. Several GC-rich segments were not sequenced because of technical difficulties. The resulting sequence data contained 220 polymorphic sites, including 213 single-nucleotide polymorphisms (SNPs) and 7 insertion/deletion polymorphisms (indels) (table S1).

Haplotypes were inferred using the PHASE 2.1 program (29, 30). A total of 86 haplotypes



**Fig. 1.** Genomic structure of the human *Microcephalin* gene. The region sequenced in the 89-individual Coriell panel is bracketed.

<sup>1</sup>Howard Hughes Medical Institute, Department of Human Genetics, <sup>2</sup>Committee on Genetics, <sup>3</sup>Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. <sup>4</sup>Department of Biology, University of Maryland, College Park, MD 20742, USA.

\*To whom correspondence should be addressed. E-mail: blahn@bsd.uchicago.edu

were identified along with their frequencies (Fig. 2 and table S2). One haplotype, denoted 49, had a much higher frequency than the other haplotypes. It had the derived C allele at the G37995C SNP site and corresponded to the high-frequency haplotype in the aforementioned exon-only polymorphism survey (8). In the Coriell panel, haplotype 49 had a frequency of 33% (59 out of 178 chromosomes) and is found in all the populations sampled in the panel. The remaining 85 haplotypes varied in frequency from 0.6 to 6.2% (1 to 11 chromosomes).

Positive selection on an allele can increase the frequency of the haplotype bearing the allele while maintaining extended linkage disequilibrium (LD) around that allele (31–36). Our data on haplotype 49 are consistent with these signatures of selection. We formally tested the statistical significance of positive selection using the previously established coalescent model (37, 38). Given the slight uncertainty in haplotype inference, we considered only the 18 individuals in the Coriell panel who are homozygous for haplotype 49 (table S1).

By simulation, we calculated the probability of obtaining 18 or more individuals (out of 89) who are homozygous for a single haplotype across a region of 220 segregating sites under neutral evolution. Here, recombination and gene conversion rates were set to values previously established for the *Microcephalin* locus (39), and a demographic model with a severe bottleneck followed by exponential growth was assumed (see SOM). Prior studies have shown that the bottleneck specified here is likely to be much more stringent than that associated with the real demographic history of human populations (40, 41); thus, the test is conservative (38). Under these parameters, the probability of obtaining 18 homozygotes out of 89 is highly significant ( $P = 0$  based on 5,000,000 replicates).

We then tested several additional demographic models, including (i) constant size, (ii) very ancient expansion, (iii) very recent expansion, (iv) repeated severe bottlenecks with subsequent expansion, and (v) population structure with between two and five subpopulations (see SOM). All produced exceedingly significant results. Even though the exact demographic history of humans is yet to be defined, our tests are highly significant under a broad range of demographic scenarios, which furthers the argument that the statistical significance is unlikely to be altered by reasonable variations in the supposed human demography. We also tested the significance of the inferred haplotype data (i.e., the significance of having 59 copies of haplotype 49 among 178 chromosomes), which similarly produced highly significant results. These data strongly suggest that haplotype 49 was driven to high frequency by positive selection. However, our data do not address whether the positive selection is

frequency-dependent selection, heterozygote advantage, or simple additive positive selection.

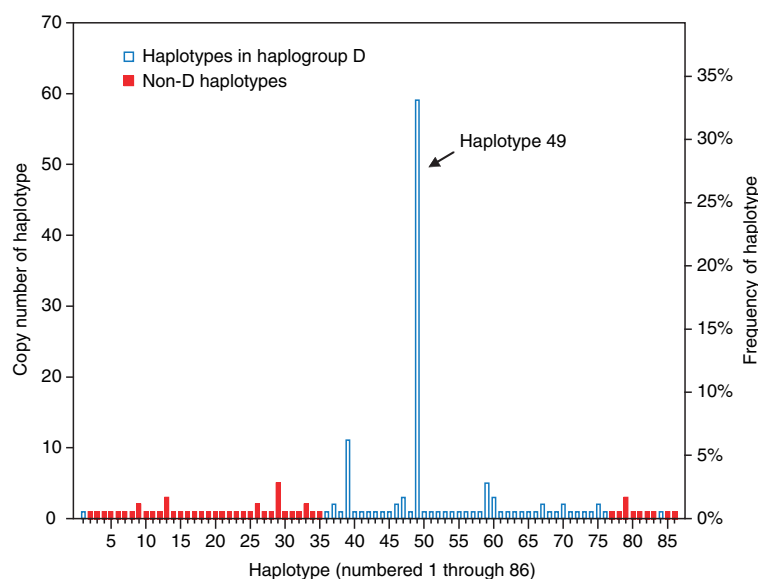
Using the G37995C polymorphism as a diagnostic site, we divided all the haplotypes into two groups: those that carry the derived C allele and those that carry the ancestral G allele. We designated the former group as haplogroup D (where D stands for “derived”). It includes 43 haplotypes that together have a 70% frequency in the Coriell panel, and haplotype 49 is the predominant member (table S2). Although the derived C allele at the G37995C site only provides an operational definition for haplogroup D, several observations make evident that haplogroup D is systematically different from the non-D haplotypes. First, this haplogroup consists exclusively of haplotype 49 or its minor variants, whereas non-D haplotypes show much greater sequence divergence from haplogroup D chromosomes. This greater divergence is because haplogroup D and non-D haplotypes have multiple fixed differences relative to each other in addition to G37995C (table S2). The only exceptions are a few recombinant haplotypes between D and non-D chromosomes (discussed below). Second, for sites that are polymorphic within haplogroup D chromosomes (excluding recombinants between D and non-D chromosomes), the non-D chromosomes are invariably monomorphic for the ancestral alleles. These data indicate that haplogroup D constitutes a genealogical clade of closely related haplotypes that is altogether separate from the more distantly related non-D haplotypes (again, excluding recombinants between D and non-D chromosomes, which represent mixed genealogies).

Collectively, the above observations support an evolutionary scenario with two aspects.

First, haplotype 49 swept from a single copy to high frequency in a short period of time. Second, during the sweep, minor variants of haplotype 49 emerged through rare mutations and recombinations. These variants, together with haplotype 49, make up haplogroup D. Haplotype 49 evidently represents the most recent common ancestor (MRCA) of haplogroup D, because it consistently has the ancestral allele for the sites polymorphic within haplogroup D.

We next estimated the coalescence age (i.e., time to MRCA) of haplogroup D chromosomes in the Coriell panel. We used the average number of mutations from the MRCA of a haplogroup clade to its descendant lineages as a molecular clock for estimating the age of the clade (42, 43). This approach is known to be unbiased by demographic history (42). The age of haplogroup D was found to be ~37,000 years, with a 95% confidence interval of 14,000 to 60,000 years. In comparison, the coalescence age of all the chromosomes in the Coriell panel is about 1,700,000 years. The emergence of anatomically modern humans has been estimated to be 200,000 years before present (44). Haplogroup D is obviously much younger, which indicates that positive selection was at work in a period considerably postdating the emergence of anatomically modern humans in Africa. We note that the age of haplogroup D coincides with the introduction of anatomically modern humans into Europe about 40,000 years ago, as well as the dramatic shift in the archeological record indicative of modern human behavior, such as art and the use of symbolism (i.e., the “Upper Paleolithic revolution”) (45).

If haplogroup D indeed experienced a recent selective sweep, it should show low poly-



**Fig. 2.** Frequencies of 86 inferred *Microcephalin* haplotypes in the 89-individual Coriell panel. Haplotypes in haplogroup D are indicated by blue-edged bars; non-D haplotypes are indicated by solid red bars.

morphism and an excess of rare alleles (46). To confirm this, we calculated nucleotide diversity ( $\pi$ ) and Tajima's  $D$  for the 47 individuals who are homozygous for haplogroup D chromosomes, and we compared these values to those of the non-D chromosomes. The  $\pi$  value of the D chromosomes is lower, by a factor of 12, than that of the non-D chromosomes (0.000077 and 0.00092, respectively), even though the D chromosomes represent about 70% of the chromosomes in the panel. Tajima's  $D$ , which is a summary statistic for the frequency spectrum of alleles, is  $-2.3$  for haplogroup D (whereas it is  $-1.2$  for the non-D chromosomes). This strongly negative Tajima's  $D$  indicates a starlike genealogy for haplogroup D chromosomes (47). Thus, both summary statistics contrast sharply between D and non-D chromosomes and are consistent with the recent age and rapid expansion of haplogroup D. We note that these calculations do not provide a statistically stringent test of positive selection, because they are done on subsets of the genealogy. Nevertheless, they do

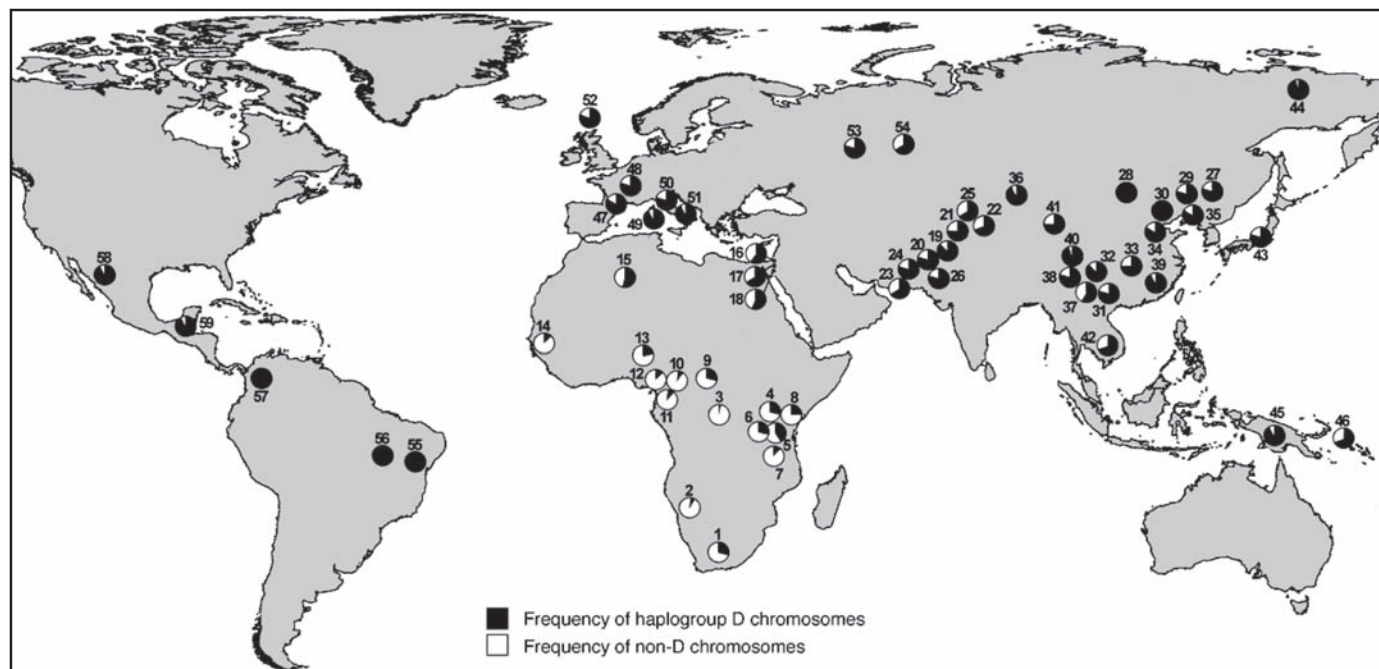
reveal qualitative signatures of positive selection that further corroborate the more stringent statistical tests described earlier.

Another sign of a positive selective sweep is extended LD around the selected allele. This is apparent in the region of *Microcephalin* investigated here, where haplogroup D chromosomes show near-complete LD across the entire region. The only exceptions are haplotypes 1, 68, and 84 (each found in a single copy in the Coriell panel), which are recombinants between D and non-D chromosomes as evidenced by recombination tracts (table S2). The remaining 121 copies of haplogroup D chromosomes show no evidence of recombination. By comparison, the non-D chromosomes do not display any significant LD across the region.

To probe the extent of LD beyond the 29-kb core region, we sequenced the Coriell panel for two segments of about 3 kb each, situated at the beginning and end of the gene separated from each other by about 235 kb. In these flanking regions, there is clear evidence

of LD decay from the core region, which supports the idea that selection has most likely operated on a site (or sites) around the core region. Our present data cannot resolve the exact site(s) of selection, and the G37995C nonsynonymous SNP used to define haplogroup D is just a candidate.

To obtain a more detailed frequency distribution of haplogroup D across the globe, we analyzed a much larger human population panel containing 1184 globally diverse individuals. We genotyped the diagnostic G37995C SNP in this panel to infer the frequency of haplogroup D chromosomes (Fig. 3). Geographic variation was observed, with sub-Saharan populations generally having lower frequencies than others. The statistic for genetic differentiation,  $F_{ST}$ , is 0.48 between sub-Saharan and others, which indicates strong differentiation (48) and is significantly higher than the genome average of 0.12 ( $P < 0.03$  based on previously established genomewide  $F_{ST}$  distribution) (49). Such population differentiation may reflect a Eurasian origin of haplogroup D, local adaptation, and/or



**Fig. 3.** Global frequencies of *Microcephalin* haplogroup D chromosomes (defined as having the derived C allele at the G37995C diagnostic SNP) in a panel of 1184 individuals. For each population, the country of origin, number of individuals sampled, and frequency of haplogroup D chromosomes are given (in parentheses) as follows: 1, Southeastern and Southwestern Bantu (South Africa, 8, 31.3%); 2, San (Namibia, 7, 7.1%); 3, Mbuti Pygmy (Democratic Republic of Congo, 15, 3.3%); 4, Masai (Tanzania, 27, 29.6%); 5, Sandawe (Tanzania, 32, 39.1%); 6, Burunge (Tanzania, 28, 30.4%); 7, Turu (Tanzania, 23, 15.2%); 8, Northeastern Bantu (Kenya, 12, 25%); 9, Biaka Pygmy (Central African Republic, 32, 26.6%); 10, Zime (Cameroon, 23, 8.7%); 11, Bakola Pygmy (Cameroon, 24, 10.4%); 12, Bamoun (Cameroon, 28, 17.9%); 13, Yoruba (Nigeria, 25, 24%); 14, Mandenka (Senegal, 24, 16.7%); 15, Mozabite [Algeria (Mzab region), 29, 53.5%]; 16, Druze [Israel (Carmel region), 44, 60.2%]; 17, Palestinian [Israel (Central), 40, 63.8%]; 18, Bedouin [Israel (Negev region), 44, 54.6%]; 19, Hazara (Pakistan, 20, 85%); 20, Balochi (Pakistan, 23, 78.3%); 21, Pathan (Pakistan, 23, 76.1%); 22, Burusho (Pakistan, 25, 66%); 23, Makrani (Pakistan, 24,

62.5%); 24, Brahui (Pakistan, 25, 78%); 25, Kalash (Pakistan, 24, 62.5%); 26, Sindhi (Pakistan, 25, 78%); 27, Hezhen (China, 9, 77.8%); 28, Mongola (China, 10, 100%); 29, Daur (China, 10, 85%); 30, Orogen (China, 10, 100%); 31, Miaozi (China, 9, 77.8%); 32, Yizu (China, 10, 85%); 33, Tujia (China, 10, 75%); 34, Han (China, 41, 82.9%); 35, Xibo (China, 9, 83.3%); 36, Uygur (China, 10, 90%); 37, Dai (China, 9, 55.6%); 38, Lahu (China, 10, 85%); 39, She (China, 9, 88.9%); 40, Naxi (China, 10, 95%); 41, Tu (China, 10, 75%); 42, Cambodian (Cambodia, 11, 72.7%); 43, Japanese (Japan, 27, 77.8%); 44, Yakut [Russia (Siberia region), 25, 98%]; 45, Papuan (New Guinea, 17, 91.2%); 46, NAN Melanesian (Bougainville, 18, 72.2%); 47, French Basque (France, 24, 83.3%); 48, French (France, 28, 78.6%); 49, Sardinian (Italy, 26, 90.4%); 50, North Italian [Italy (Bergamo region), 13, 76.9%]; 51, Tuscan (Italy, 8, 87.5%); 52, Orcadian (Orkney Islands, 16, 81.3%); 53, Russian (Russia, 24, 79.2%); 54, Adygei [Russia (Caucasus region), 15, 63.3%]; 55, Karitiana (Brazil, 21, 100%); 56, Surui (Brazil, 20, 100%); 57, Colombian (Colombia, 11, 100%); 58, Pima (Mexico, 25, 92%); 59, Maya (Mexico, 25, 92%).

demographic factors such a bottleneck associated with human migration out of Africa 50,000 to 100,000 years ago.

Previous studies have shown that *Microcephalin* is a specific regulator of brain size (13, 14) and that this gene has evolved under strong positive selection in the primate lineage leading to *Homo sapiens* (7, 8). Here, we present compelling evidence that *Microcephalin* has continued its trend of adaptive evolution beyond the emergence of anatomically modern humans. The specific function of *Microcephalin* in brain development makes it likely that selection has operated on the brain. Yet, it remains formally possible that an unrecognized function of *Microcephalin* outside of the brain is actually the substrate of selection. If selection indeed acted on a brain-related phenotype, there could be several possibilities, including brain size, cognition, personality, motor control, or susceptibility to neurological and/or psychiatric diseases. We hypothesize that D and non-D haplotypes have different effects on the proliferation of neural progenitor cells, which in turn leads to different phenotypic outcomes of the brain visible to selection.

References and Notes

1. J. N. Spuhler, *The Evolution of Man's Capacity for Culture* (Wayne State Univ. Press, Detroit, MI, 1959).
2. J. H. Jerison, *Evolution of the Brain and Intelligence* (Academic Press, New York, 1973).
3. W. Enard et al., *Nature* **418**, 869 (2002).

4. J. Zhang, *Genetics* **165**, 2063 (2003).
5. P. D. Evans et al., *Hum. Mol. Genet.* **13**, 489 (2004).
6. N. Kouprina et al., *PLoS Biol.* **2**, E126 (2004).
7. Y. Q. Wang, B. Su, *Hum. Mol. Genet.* **13**, 1131 (2004).
8. P. D. Evans, J. R. Anderson, E. J. Vallender, S. S. Choi, B. T. Lahn, *Hum. Mol. Genet.* **13**, 1139 (2004).
9. R. J. Ferland et al., *Nat. Genet.* **36**, 1008 (2004).
10. H. H. Stedman et al., *Nature* **428**, 415 (2004).
11. F. Burki, H. Kaessmann, *Nat. Genet.* **36**, 1061 (2004).
12. S. Dorus et al., *Cell* **119**, 1027 (2004).
13. A. P. Jackson et al., *Am. J. Hum. Genet.* **63**, 541 (1998).
14. A. P. Jackson et al., *Am. J. Hum. Genet.* **71**, 136 (2002).
15. W. B. Dobyns, *Am. J. Hum. Genet.* **112**, 315 (2002).
16. G. H. Mochida, C. A. Walsh, *Curr. Opin. Neurol.* **14**, 151 (2001).
17. C. G. Woods, J. Bond, W. Enard, *Am. J. Hum. Genet.* **76**, 717 (2005).
18. E. Roberts et al., *Eur. J. Hum. Genet.* **7**, 815 (1999).
19. L. Moynihan et al., *Am. J. Hum. Genet.* **66**, 724 (2000).
20. C. R. Jamieson, C. Govaerts, M. J. Abramowicz, *Am. J. Hum. Genet.* **65**, 1465 (1999).
21. J. Bond et al., *Nat. Genet.* **32**, 316 (2002).
22. G. F. Leal et al., *J. Med. Genet.* **40**, 540 (2003).
23. J. Bond et al., *Nat. Genet.* **37**, 353 (2005).
24. T. Huyton, P. A. Bates, X. Zhang, M. J. Sternberg, P. S. Freemont, *Mutat. Res.* **460**, 319 (2000).
25. S. Y. Lin, S. J. Eledge, *Cell* **113**, 881 (2003).
26. X. Xu, J. Lee, D. F. Stern, *J. Biol. Chem.* **279**, 34091 (2004).
27. M. Trimborn et al., *Am. J. Hum. Genet.* **75**, 261 (2004).
28. S. L. Gilbert, W. B. Dobyns, B. T. Lahn, *Nat. Rev. Genet.* **6**, 581 (2005).
29. M. Stephens, N. J. Smith, P. Donnelly, *Am. J. Hum. Genet.* **68**, 978 (2001).
30. M. Stephens, P. Donnelly, *Am. J. Hum. Genet.* **73**, 1162 (2003).
31. S. A. Tishkoff et al., *Science* **293**, 455 (2001).
32. P. C. Sabeti et al., *Nature* **419**, 832 (2002).
33. E. Wang et al., *Am. J. Hum. Genet.* **74**, 931 (2004).
34. T. Bersaglieri et al., *Am. J. Hum. Genet.* **74**, 1111 (2004).
35. E. E. Thompson et al., *Am. J. Hum. Genet.* **75**, 1059 (2004).
36. H. Stefansson et al., *Nat. Genet.* **37**, 129 (2005).

37. R. R. Hudson, *Oxf. Surv. Evol. Biol.* **7**, 1 (1990).
38. R. R. Hudson, *Bioinformatics* **18**, 337 (2002).
39. A. Kong et al., *Nat. Genet.* **31**, 241 (2002).
40. E. Zietkiewicz et al., *J. Mol. Evol.* **47**, 146 (1998).
41. H. Harpending, A. Rogers, *Annu. Rev. Genomics Hum. Genet.* **1**, 361 (2000).
42. H. Tang, D. O. Siegmund, P. Shen, P. J. Oefner, M. W. Feldman, *Genetics* **161**, 447 (2002).
43. R. Thomson, J. K. Pritchard, P. Shen, P. J. Oefner, M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7360 (2000).
44. I. McDougall, F. H. Brown, J. G. Fleagle, *Nature* **433**, 733 (2005).
45. R. G. Klein, *The Human Career: Human Biological and Cultural Origins* (Univ. of Chicago Press, Chicago, 1999).
46. M. Bamshad, S. P. Wooding, *Nat. Rev. Genet.* **4**, 99 (2003).
47. F. Tajima, *Genetics* **123**, 585 (1989).
48. S. Wright, *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1978).
49. J. M. Akey, G. Zhang, K. Zhang, L. Jin, M. D. Shriver, *Genome Res.* **12**, 1805 (2002).
50. We thank the Coriell Institute for Medical Research, the Centre d'Etude du Polymorphisme Humain (CEPH), and A. Froment for human DNA samples. We thank H. M. Cann, S. Dorus, E. E. Eichler, N. M. Pearson, A. Di Rienzo, M. Kreitman, and J. K. Pritchard for technical support and/or helpful discussions. Supported in part by the Searle Scholarship and the Burroughs Wellcome Career Award (to B.T.L.), and David and Lucile Packard Career Award, the Burroughs Wellcome Career Award, and NSF grant BCS-0196183 (to S.A.T.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5741/1717/DC1  
 Materials and Methods  
 Tables S1 and S2  
 References and Notes

18 April 2005; accepted 14 June 2005  
 10.1126/science.1113722

# Ongoing Adaptive Evolution of *ASPM*, a Brain Size Determinant in *Homo sapiens*

Nitzan Mekel-Bobrov,<sup>1,2</sup> Sandra L. Gilbert,<sup>1</sup> Patrick D. Evans,<sup>1,2</sup> Eric J. Vallender,<sup>1,2</sup> Jeffrey R. Anderson,<sup>1</sup> Richard R. Hudson,<sup>3</sup> Sarah A. Tishkoff,<sup>4</sup> Bruce T. Lahn<sup>1\*</sup>

The gene *ASPM* (*abnormal spindle-like microcephaly associated*) is a specific regulator of brain size, and its evolution in the lineage leading to *Homo sapiens* was driven by strong positive selection. Here, we show that one genetic variant of *ASPM* in humans arose merely about 5800 years ago and has since swept to high frequency under strong positive selection. These findings, especially the remarkably young age of the positively selected variant, suggest that the human brain is still undergoing rapid adaptive evolution.

Homozygous null mutations of *ASPM* cause primary microcephaly, a condition characterized by severely reduced brain size with otherwise normal neuroarchitecture (1). Studies

have suggested that *ASPM* may regulate neural stem cell proliferation and/or differentiation during brain development, possibly by mediating spindle assembly during cell division (1, 2). Phylogenetic analysis of *ASPM* has revealed strong positive selection in the primate lineage leading to *Homo sapiens* (3–5), especially in the past 6 million years of hominid evolution in which *ASPM* acquired about one advantageous amino acid change every 350,000 years (4). These data argue that *ASPM*

may have contributed to human brain evolution (3–6). Here, we investigate whether positive selection has continued to operate on *ASPM* since the emergence of anatomically modern humans.

Human *ASPM* has 28 exons with a 10,434–base pair open reading frame (1) (fig. S1). We resequenced the entire 62.1-kb genomic region of *ASPM* in samples from 90 ethnically diverse individuals obtained through the Coriell Institute and from a common chimpanzee (7). This revealed 166 polymorphic sites (table S1). Using established methodology (7), we identified 106 haplotypes. One haplotype, numbered 63, had an unusually high frequency of 21%, whereas the other haplotypes ranged from 0.56% to 3.3% (fig. S2). Moreover, this haplotype differed consistently from the others at multiple polymorphic sites (save for a few rare haplotypes that are minor mutational or recombinational variants of haplotype 63, as discussed later) (table S2). Two of these polymorphic sites are nonsynonymous, both in exon 18, and are denoted A44871G and C45126A (numbers indicate genomic positions from the start codon, and letters at the beginning and end indicate ancestral and derived alleles, respectively). These two sites reside in a region of the open reading frame that was shown previously to have experienced par-

<sup>1</sup>Howard Hughes Medical Institute, Department of Human Genetics, <sup>2</sup>Committee on Genetics, <sup>3</sup>Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. <sup>4</sup>Department of Biology, University of Maryland, College Park, MD 20742, USA.

\*To whom correspondence should be addressed. E-mail: blahn@bsd.uchicago.edu