

# Reconstructing the evolutionary history of *microcephalin*, a gene controlling human brain size

Patrick D. Evans<sup>1,2</sup>, Jeffrey R. Anderson<sup>1</sup>, Eric J. Vallender<sup>1,2</sup>, Sun Shim Choi<sup>1</sup> and Bruce T. Lahn<sup>1,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute, Department of Human Genetics and <sup>2</sup>Committee on Genetics, University of Chicago, Chicago, IL 60637, USA

Received February 24, 2004; Revised and Accepted March 25, 2004

The defining process in the evolution of primates and particularly humans is the dramatic expansion of the brain. While many types of genes could potentially contribute to this process, genes that specifically regulate brain size during development may be especially relevant. Here, we examine the evolution of the *microcephalin* gene, whose null mutation in humans causes primary microcephaly, a congenital defect characterized by severe reductions in brain size without other gross abnormalities. We show that the evolution of *microcephalin*'s protein sequence is highly accelerated throughout the lineage from simian ancestors to humans and chimpanzees, with the most pronounced acceleration seen in the early periods of this lineage. We further demonstrate that this accelerated evolution is coupled with signatures of positive selection. Statistical analysis suggests that about 45 advantageous amino acid changes in *microcephalin* might have fixed during the 25–30 million years of evolution from early simian progenitors to modern humans. These observations support the notion that the molecular evolution of *microcephalin* may have contributed to brain expansion in the simian lineage leading to humans. We have recently shown that *ASPM*, another gene linked to primary microcephaly, experienced strong positive selection in the ape lineage leading to humans. We therefore propose that genes regulating brain size during development may have the general propensity to contribute to brain evolution in primates and particularly humans.

## INTRODUCTION

As a species, *Homo sapiens* is distinguished by its highly advanced mental capacity (1). A key biological basis for this is believed to be the enormous brain size of humans relative to other extant taxa (2–4). This is particularly true for the cerebral cortex, the region of the brain most prominently involved in higher cognitive functions (5,6). Indeed, there is a general correlation across taxa between the level of encephalization and cognitive complexity (7), which has led researchers to utilize anatomical dimensions of the brain, especially the size of the cerebral cortex, as proxies for cognitive abilities of extant and extinct species. More recently, discussions of brain evolution have moved from the traditional realms of anatomy, physiology and behavior to the domains of genes and genome organization (8,9). Of particular interest to researchers is the identification of specific genes whose evolution at the DNA

sequence level may have contributed to the evolutionary expansion of the brain.

Genes that underlie the developmental defect known as microcephaly may be especially relevant to this endeavor. Microcephaly (meaning small head) is a congenital abnormality characterized by the severe underdevelopment of the brain, particularly the cerebral cortex (10,11). Clinically, microcephaly is usually defined as a head circumference below the population mean by three standard deviations or more. Primary (or true) microcephaly is a subclass of microcephaly in which affected individuals show significantly reduced brain size (and accompanying intellectual impairment) but are free of other gross neuropathologies or dysmorphic features. Brains of primary microcephaly patients typically have a volume of around 400 cm<sup>3</sup>, far smaller than the 1200–1600 cm<sup>3</sup> of a normal adult brain (12). Superficially speaking, primary microcephaly can be viewed as an example

\*To whom correspondence should be addressed. Tel: +1 7738344393; Fax: +1 7738348470; Email: blahn@genetics.uchicago.edu

of evolutionary retrogression whereby the brain dimensions of affected individuals revert to a level resembling that of the very early hominids (12,13). This raises the tantalizing possibility that genes implicated in primary microcephaly—which are clearly involved in regulating brain size during development—may also play a role in the expansion of the brain during evolution.

Thus far, six autosomal recessive loci, named *MCPHI* through *MCPH6*, have been linked to clinically indistinguishable forms of primary microcephaly (14–20). For two of these *MCPH* loci, the underlying genes have been identified by genetic linkage studies. One is *microcephalin*, which corresponds to *MCPHI* (21); the other is *ASPM* (abnormal spindle-like microcephaly associated) corresponding to *MCPH5* (22,23).

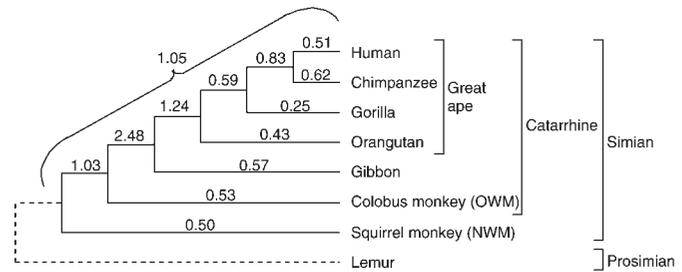
The human *microcephalin* gene spans 14 exons and has a deduced protein-coding region of roughly 2.5 kb (21). It contains three so-called BRCA1 C-terminal (BRCT) domains, one at its N-terminus and two at its C-terminus. This domain is found in the tumor suppressor gene *BRCA1* as well as multiple other eukaryotic genes, and is implicated in protein–protein and protein–DNA interactions (24). Apart from the BRCT domains, however, the biochemical function of *microcephalin* is unknown. Expression of *microcephalin* is found in a variety of human and mouse tissues (21). The most prominent expression is seen in the developing forebrain, within regions of active neurogenesis (i.e. the walls of lateral telencephalic ventricles). Such an expression pattern is consistent with the role of this gene in regulating brain size during development (21).

We have recently shown that the *ASPM* gene has experienced strong positive selection in the lineage from ape ancestors to humans (25). In this study, we examine the molecular evolution of *microcephalin* and present evidence that this gene displays strong signatures of adaptive evolution in the lineage from simian ancestors to humans and chimpanzees, particularly during early periods of this lineage. Thus, among the two primary microcephaly genes identified thus far, both are potentially implicated in the evolutionary enlargement of the brain in the primate lineage leading to humans.

## RESULTS

### Molecular evolution of *microcephalin*

We first focused on the evolution of *microcephalin* within primates. The entire coding sequence of this gene (~2.5 kb) was obtained from multiple primate species representing key positions of the primate evolutionary hierarchy, including human, chimpanzee, gorilla, orangutan, gibbon, colobus monkey (an Old World monkey), squirrel monkey (a New World monkey) and lemur (a prosimian). We constructed a phylogenetic tree with these sequences (Fig. 1). For each segment of the tree, we calculated the ratio of non-synonymous ( $K_a$ ) to synonymous ( $K_s$ ) substitution rates, which was used as a proxy for protein evolution rate (26). Close inspection of the tree revealed a prominent trend of accelerated protein evolution along the lineage from simian progenitors to humans and chimpanzees (Fig. 1). Indeed, of the seven phylogenetic segments within this lineage, all but one exhibited  $K_a/K_s$  ratios that were among the highest of the tree. In particular, the segments from simian progenitors to the last common



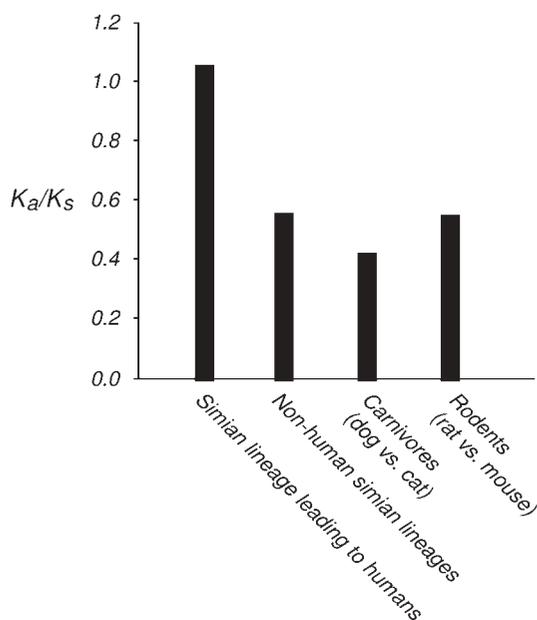
**Figure 1.** Evolution of *microcephalin* within primates. The  $K_a/K_s$  ratios of individual phylogenetic segments are indicated. The  $K_a/K_s$  of the entire lineage from simian progenitors to humans is also indicated. OWM: Old World monkey; NWM: New World monkey.

ancestors of great apes have  $K_a/K_s$  ratios that are consistently greater than 1, a signature of adaptive evolution (26) (see later discussions for further evidence of adaptive evolution).

Given that the evolutionary lineage leading from simian progenitors to humans was particularly relevant to the discussion of brain evolution, we considered this lineage separately from all the other primate branches that split off from it. The  $K_a/K_s$  ratio is about 1.05 for this lineage (referred henceforth as the simian lineage leading to humans), which is significantly higher than the  $K_a/K_s$  of 0.47 for all the other primate branches combined ( $P < 0.005$ ) (Figs 1 and 2). The  $K_a/K_s$  value of the simian lineage leading to humans is also significantly higher than that of any other primate branch considered individually, except for the chimpanzee branch. Even though these observations *per se* do not prove the presence of positive selection, it does argue convincingly that the evolution of *microcephalin*'s protein sequence in primates is characterized by a robust and statistically significant acceleration in the simian lineage leading to humans.

We next sought to compare the evolutionary rate of *microcephalin* in primates with that in other mammalian orders. *Microcephalin* sequence was obtained from four non-primate species representing two mammalian orders. These included dog and cat, representing carnivores, and rat and mouse, representing rodents. For each order, we compared *microcephalin* sequences between the representative pair of species to estimate the rate of protein evolution in that order. As shown in Figure 2, the evolutionary rate of *microcephalin* in the simian lineage leading to humans is significantly higher than in the non-primate orders ( $P$ -values were 0.0002 and 0.003 when the simian lineage was compared to carnivores and rodents, respectively). Based on these observations, we conclude that the dramatically accelerated evolution of *microcephalin* is a distinct feature in the simian lineage leading to humans relative to other mammalian taxa.

Bioinformatic searches of public databases failed to detect any convincing homologs of *microcephalin* in non-mammalian species (including some lower vertebrates, such as the Fugu fish, that have been extensively sequenced), even though there were a number of non-mammalian proteins that shared sequence homology with the BRCT domains of *microcephalin* (24). This suggests that *microcephalin* is either a newly emerged gene in the evolutionary lineage leading to mammals or it is an ancient gene that has undergone exceptionally rapid sequence evolution.



**Figure 2.** Evolution of *microcephalin* in various mammalian taxa. Species used to derive  $K_a/K_s$  ratios are indicated below the horizontal axis. 'Simian lineage leading to humans' refers to the lineage from simian progenitors to humans; 'Non-human simian lineages' refer to all the simian branches that split off from the lineage leading to humans.

### Signatures of adaptive evolution

We favor the interpretation that positive selection was responsible for the accelerated evolution of *microcephalin* along the simian lineage leading to humans. However, accelerated evolution could also result from relaxed functional constraint (i.e. the biochemical function of *microcephalin* has become less critical in the simian lineage leading to humans than in other taxa). While this scenario is possible in theory, we think that it is rather improbable in light of the fact that *microcephalin* plays an indispensable role in the proper development of the human brain (21). This functional argument notwithstanding, we sought for evolutionary signatures that could convincingly support the presence of positive selection. To this end, we performed several analyses.

The first analysis involved the McDonald–Kreitman test (27), which has been widely considered a stringent test for positive selection (25,28,29). The theory of this test rests on the null expectation that if a gene is evolving under a constant selective regime devoid of positive selection, then the ratio of non-synonymous to synonymous changes between species should be equivalent to that within a species. In case of a significant departure from this null expectation, i.e. a much higher non-synonymous to synonymous ratio between species than that within species, it is generally interpreted as the consequence of positive Darwinian selection that has driven the fixation of advantageous non-synonymous changes between species (28). To perform this test, we obtained polymorphism profiles across the *microcephalin* coding region from 27 unrelated individuals (54 haploid genomes) representing the world-wide diversity of human populations. As shown in Table 1, the non-synonymous to synonymous ratio in the simian lineage leading to humans is significantly higher than that within human populations ( $P < 0.02$ ). In this case, as in

**Table 1.** Numbers of non-synonymous ( $N$ ) and synonymous ( $S$ ) changes in *microcephalin*

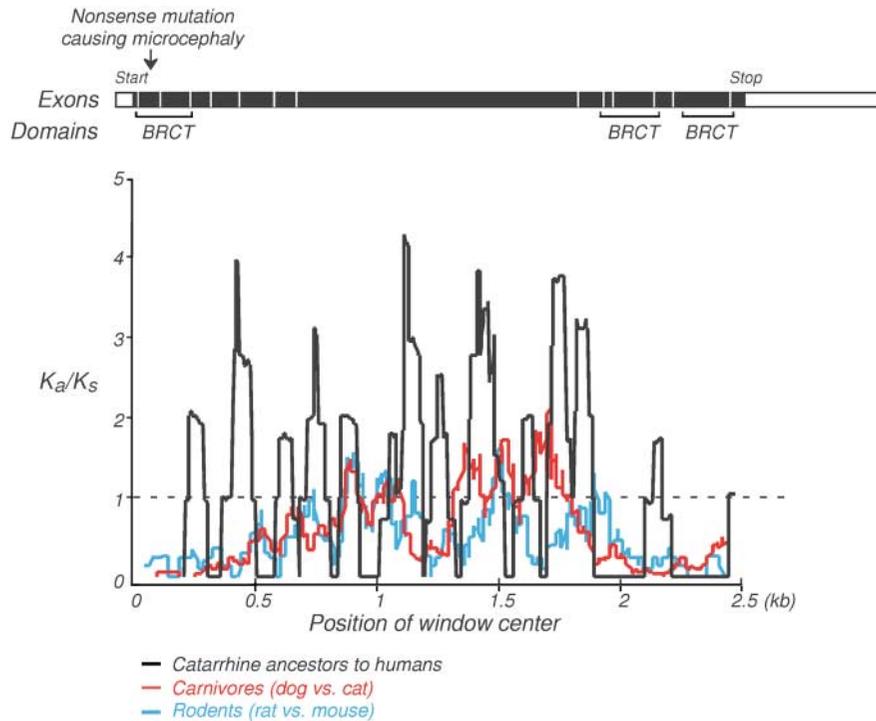
|  | $N$ | $S$ | $P$ -value |
|--|-----|-----|------------|
| Polymorphism within humans ( $n = 54$ )          | 11  | 11  |            |
| Divergence between simian progenitors and humans | 64  | 19  | 0.015      |

$n$  is the number of haploid human genomes sampled;  $P$ -value is calculated from Fisher's exact test.

others (25,29), the robust deviation from neutrality detected by the McDonald–Kreitman test argues that the accelerated evolution of *microcephalin* along the simian lineage leading to humans is the result of adaptation. Using the non-synonymous to synonymous ratio within humans as an estimator of selective constraint, we calculated that there was an excess of about 45 amino acid changes from simian progenitors to humans relative to that expected from the estimated level of constraint. Under the theoretical framework of the McDonald–Kreitman test (27), these excess non-synonymous substitutions are presumed to be the result of adaptation. Thus, about 45 advantageous amino acid replacements have fixed in *microcephalin* during the 25–30 million years of evolution from early simian ancestors to modern humans.

In the second analysis, we compared the ancestral catarrhine *microcephalin* sequence with the human sequence, and obtained a sliding-window  $K_a/K_s$  profile across the gene. In this profile, subregions of extreme peaks ( $K_a/K_s \gg 1$ ) and valleys ( $K_a/K_s \ll 1$ ) are evident throughout the gene (Fig. 3). Qualitatively, such a profile is in line with the presence of two types of subregions in the gene, those under strong selective constraint (and hence very low  $K_a/K_s$ ) and those under positive selection (and hence very high  $K_a/K_s$ ). Supporting this qualitative assessment, statistical tests showed that the topology of the  $K_a/K_s$  profile deviated significantly from the neutral expectation ( $P < 0.05$ ; see Materials and Methods). These results provide additional evidence for the adaptive nature of *microcephalin* evolution in the lineage leading to humans. Interestingly, the three most prominent  $K_a/K_s$  valleys corresponded to the three BRCT domains, arguing that these domains are under strong functional constraint. This is not surprising given that these domains are conserved across various BRCT-containing proteins of distantly related taxa. As a comparison, we also obtained  $K_a/K_s$  profiles for carnivores and rodents (Fig. 3). Even though there are moderate peaks in both carnivores and rodents where  $K_a/K_s > 1$ , these peaks are not statistically different from random fluctuation under neutrality, and therefore do not offer definitive proof of positive selection.

In the third analysis, we examined the statistical significance by which the number of amino acid changes between simian progenitors and humans deviated from the neutral expectation. To better appreciate the biochemical nature of the amino acid changes, we followed established convention to classify all amino acid changes according to their physicochemical characteristics into four categories: conservative, moderate, radical and very radical (29–31) (see Materials and Methods). For each category, the number of changes expected under selective neutrality was generated by computer simulation. We surmised



**Figure 3.** Sliding-window profile of  $K_a/K_s$  along the *microcephalin* coding region. Peaks above the dotted line have  $K_a/K_s$  values greater than 1, and therefore an excess of non-synonymous substitutions over the neutral expectation. Exons and domain structures of *microcephalin* are drawn above the  $K_a/K_s$  profile. Also indicated is the nonsense mutation in *microcephalin* known to cause primary microcephaly in humans (21).

that if positive selection was indeed at play then the conservative category would be the most likely to show an excess of changes relative to that expected under neutrality. This is because conservative changes are generally under the weakest purifying selection (32). Additionally, conservative changes have been postulated to be better substrates of positive selection (29) under the neo-Darwinian view that selection is likely to favor small steps over giant leaps (33). Comparison between observed and expected numbers of changes showed that, consistent with our expectation, there is a significant excess of conservative changes relative to the neutral expectation in the simian lineage leading to humans (compare the filled bar and the open bar in Figure 4, which represent the observed number of changes and the neutral expectation, respectively). This result resembles that found in male reproductive genes, which also show an excess of conservative amino acid changes while bearing other signatures of positive selection (29).

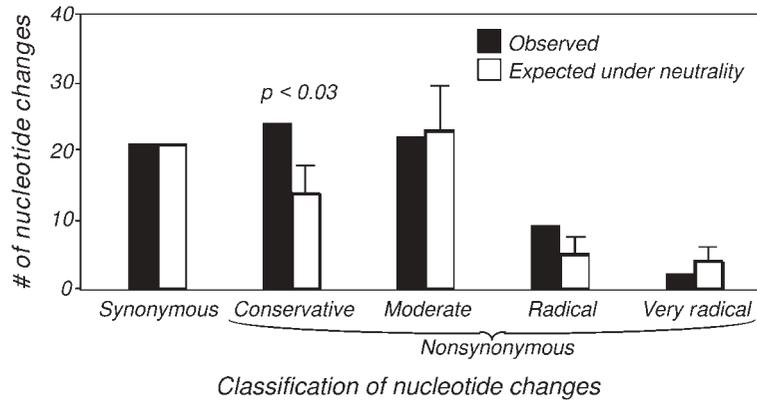
Collectively, the above studies indicate that the evolution of *microcephalin* is dramatically and specifically accelerated along the simian lineage leading to humans relative to other taxa, particularly during the early periods of this lineage. Furthermore, signatures of Darwinian selection were revealed by multiple analyses, supporting the notion that this acceleration is likely the result of adaptation.

## DISCUSSION

There is an inherent fascination in the genetic underpinnings of human evolution, especially with regard to the human

brain. We argue that genes controlling brain development are good candidates for exploring the genetic basis of brain evolution simply because evolutionary changes in these genes have the potential to alter the developmental outcome of brain morphology and function. Of particular interest to us are genes that underlie the congenital malformation known as primary microcephaly, which is characterized by severe reductions in cerebral cortical size without other gross abnormalities (10,11). Given the highly specific function of these genes in regulating brain size during development, it is reasonable to speculate that the same genes might also contribute to the expansion of the brain during evolution. Six autosomal recessive loci have been linked to primary microcephaly (14–20), and for two of these, the underlying genes have been identified. One of them, *ASPM*, was recently shown to bear robust signatures of adaptive evolution in the primate lineage leading to humans (25,34). In this study, we focused on the second gene, *microcephalin*, and showed that, like *ASPM*, this gene also exhibited strong signatures of adaptation during primate evolution.

A number of evolutionary parallels between *ASPM* and *microcephalin* are worthy of elaboration. First, the rate of protein evolution for both genes is dramatically accelerated in the primate lineage leading to humans relative to other primate branches and non-primate mammals (even though the exact rate distribution along the lineage may differ between these two genes; see later). Indeed, for either gene, it is not uncommon to find phylogenetic segments within the primate lineage leading to humans that exhibit  $K_a/K_s$  ratios greater than 1. Second, both genes show strong signatures of



**Figure 4.** Comparison of the observed number of non-synonymous changes in *microcephalin* with that expected under selective neutrality. All non-synonymous changes in *microcephalin* that occurred between the simian progenitors and humans were classified into four categories based on physicochemical properties of the resulting amino acid replacements (see Materials and Methods). The number of changes expected under neutrality was generated by computer simulation, with one standard deviation of the simulated result indicated by the error bar. The *P*-value is given when the observed number of changes significantly exceeds the simulated neutral expectation.

positive selection by the McDonald–Kreitman test (27). Based on this test, the frequency of advantageous amino acid changes could be estimated. This frequency is, on average, about one per 0.3–0.4 million years for *ASPM* in the human lineage since human–chimpanzee divergence (25), and one per 0.6–0.7 million years for *microcephalin* in the lineage from simian progenitors to humans. These are remarkably high rates of adaptive amino acid changes. Third, both genes exhibit  $K_a/K_s$  profiles along their coding regions that are consistent with portions of the gene being under strong purifying selection (i.e. very low  $K_a/K_s$ ) while other portions experiencing positive selection (i.e. very high  $K_a/K_s$ ). This argues that positive selection likely operated within particular functional motifs of these genes. Finally, for both genes, the accelerated evolution is not confined to the terminal human lineage post human–chimpanzee divergence. For *ASPM*, the most dramatic acceleration occurred in the lineage from ape ancestors to humans. For *microcephalin*, evolutionary acceleration took place in the lineage from simian progenitors to humans and chimpanzees, with the greatest acceleration observed in the early periods of this lineage (i.e. from simian progenitors to the last common ancestors of great apes). These observations argue that if *ASPM* and *microcephalin* indeed contributed to brain expansion in the lineage from ancestral primates to modern humans, then *ASPM* might be more involved in the late portion of this lineage whereas *microcephalin* might be more involved in the early portion. Indeed, there is no reason to expect that a gene important for brain evolution should only exert its influence in the terminal human branch post human–chimpanzee divergence. There is ample evidence to indicate that brain size and complexity has increased throughout the lineage from ancestral primates to modern humans (35,36), even though the increase is most dramatic in the last 3–4 million years of human evolution.

Given the strong evolutionary parallel between *ASPM* and *microcephalin*, we speculate that similar signatures of adaptive evolution may be found in additional genes implicated in primary microcephaly (or other forms of microcephaly). We further hypothesize that primary microcephaly genes

may belong to a class of evolutionarily ‘predisposed’ genes that, by virtue of their specific function in regulating brain size during development, may have a general propensity to play important roles in brain evolution. This hypothesis can be addressed with additional evolutionary studies of primary microcephaly genes as they become available.

*Microcephalin* is expressed in a variety of tissues (21). The most prominent expression is found in the neurogenic zone of the developing telencephalon, which is consistent with the role of this gene in regulating brain size during embryogenesis. The developmental function of *microcephalin* outside of the nervous system is apparently non-essential, because null mutations in humans do not appear to result in overt defects outside of the brain (21). Since the exact biochemical function of *microcephalin* is unknown, it is difficult to speculate on the nature by which the molecular evolution of *microcephalin* could give rise to phenotypic evolution of the brain. However, given the fact that the BRCT domains of *microcephalin* are found in proteins known to function in cell cycle regulation, such as the tumor suppressor *BRCA1* (24), it is possible that *microcephalin* is involved in regulating cell cycle during neurogenesis. If this is the case, then evolutionary changes in the protein sequence of *microcephalin* might alter the rate of cell division and/or differentiation of neural progenitor cells during neurogenesis, which in turn could change the size and morphology of the brain. These highly speculative possibilities could be tested when more detailed functional knowledge of *microcephalin* is obtained.

Curiously, the *BRCA1* gene has been shown to exhibit signatures of positive selection in the human and chimpanzee lineages after they diverged from each other (37,38). It may be a mere coincidence that both *microcephalin* and *BRCA1*—which share the BRCT domains in common—are subject to positive selection during primate evolution. Indeed, the BRCT domains in both genes are highly conserved and are themselves not subject to positive selection. However, it is also possible that *microcephalin* and *BRCA1* have similar functions in regulating cell cycle and are therefore subject to similar regimes of positive selection. Consistent with this possibility,

*BRCA1* knockout mice show profound defects in nervous system development such as failure of neural tube closure and severely retarded growth of the forebrain (39). These results indicate that *BRCA1*, like *microcephalin*, has a critical function in the proliferation and differentiation of neural progenitor cells (39), raising the possibility that positive selection on *BRCA1* was actually directed toward its activity in brain development rather than its function in tumor suppression.

## MATERIALS AND METHODS

### Sequence acquisition and analysis

The roughly 2.5 kb coding region of *microcephalin* was experimentally obtained from the following species: common chimpanzee (*Pan troglodytes*; accession nos AY552982–AY552995), lowland gorilla (*Gorilla gorilla*; accession nos AY552996–AY553009), orangutan (*Pongo pygmaeus*; accession nos AY553010–AY553023), white-handed gibbon (*Hylobates lar*; accession nos AY553024–AY553037), black-and-white colobus monkey (*Colobus guereza*; accession nos AY553038–AY553051), Bolivian squirrel monkey (*Saimiri boliviensis*; accession no. AY570949), ring-tailed lemur (*Lemur catta*; accession no. AY570945), domestic dog (*Canis familiaris*; accession no. AY570944) and domestic cat (*Felis catus*; accession no. AY570943). *Microcephalin* sequences of human (*Homo sapiens*; accession no. AX087870), mouse (*Mus musculus*; accession no. AY070216) and rat (*Rattus norvegicus*; obtained by piecing together genomic segments corresponding to exons retrieved from the University of Santa Cruz Genome Browser) were retrieved from public databases. For chimpanzee, gorilla, orangutan, gibbon and colobus monkey, genomic DNA was obtained from cell lines. For the rest of the species, total RNA was extracted from brain specimens, and primed by random hexamers or poly-T oligomers to synthesize first-strand cDNA. Standard PCR conditions were used to amplify *microcephalin* coding region from either genomic DNA (100–500 ng per 30  $\mu$ l reaction) or cDNA (0.5–5 ng per reaction). PCR primers were initially based on human and rodent *microcephalin* sequences until species-specific sequences were obtained. All sequencing was performed using standard dye-terminator chemistry on PCR products. Nucleotide sequences were aligned in-frame using the Megalign program in the DNASTAR software package (DNASTAR, Madison, WI, USA). In multiple-species alignments, the ancestral sequence was deduced by parsimony. In cases where single nucleotide polymorphisms were encountered, the ancestral allele was used for sequence comparison. The Diverge function from the Wisconsin Package version 10.2 (Accelrys Inc., San Diego, CA, USA) was used to analyze evolutionary divergence following a published method (40). To calculate the combined  $K_a/K_s$  ratio across multiple lineages, non-synonymous substitutions in all these lineages were added to calculate the combined  $K_a$ , and a similar procedure was used to calculate the combined  $K_s$ . For sliding-window analysis of the  $K_a/K_s$  ratio,  $K_a$  was calculated for a window size of 25 codons and a sliding increment of 1 codon, and  $K_s$  of the entire gene was used as denominator to avoid problems associated with stochastic variation that can lead to division by zero.

### Analysis of human polymorphism

Contiguous double-stranded sequences were obtained from 27 unrelated humans for the 2508 bp coding region of *microcephalin*. These individuals were chosen from the Coriell Human Variation Panels to represent a diverse selection of world-wide populations. They include four North Africans (Coriell numbers: 17378, 17379, 17380, 17383), four sub-Saharan Africans (17341, 17342, 17344, 17349), five Chinese (16654, 16688, 16689, 17014, 17016), four Russians (13820, 13838, 13852, 13911), two Basques (15884, 15885), two Iberians (17092, 17093), one Pacific Islander (17387), two Andeans (17301, 17302), two Southeast Asians (17081, 17082) and one Middle Easterner (17331). PCR primers were designed to amplify various *microcephalin* exons, and PCR products were sequenced on both strands using standard dye-terminator chemistry. Sequence chromatograms were aligned by the Sequencher software (Gene Codes Corporation, Ann Arbor, MI, USA). Polymorphisms were detected by direct visual inspection of sequence chromatograms.

### Statistical analysis

To assess the statistical significance that the  $K_a/K_s$  ratio of one lineage is distinct from that of another, the numbers of non-synonymous and synonymous substitutions were calculated for both lineages using a standard method (40). The resulting four values were placed in a  $2 \times 2$  contingency table and assessed for statistical significance by the two-tailed Fisher's exact test. To perform the McDonald–Kreitman test, the numbers of non-synonymous and synonymous changes were obtained for both interspecies divergence and within-species polymorphism. The resulting four values were placed in a  $2 \times 2$  table and assessed for statistical significance by the Fisher's exact test. A published method (41) was adapted to test whether the sliding-window profile of  $K_a/K_s$  is consistent with positive selection. First, the occurrence of point mutations in the coding sequence of *microcephalin* was simulated by assuming complete neutrality of all changes. When simulated synonymous changes reached the observed level, a sliding-window profile of  $K_a/K_s$  was generated from the simulated sequence. In windows where  $K_a/K_s$  ratios exceeded 1, the portions of  $K_a$  that exceeded neutral expectation were summed. The fraction of simulations in which this sum was equal to or greater than the sum observed for the real sequence was taken as the statistical significance that the observed  $K_a/K_s$  profile departs from the neutral expectation. Because this test considers the statistical significance of the overall topology of the  $K_a/K_s$  profile, it avoids the multiple testing problem associated with calculating the statistical significance of individual  $K_a/K_s$  peaks. A total of 10 000 simulations were performed for *microcephalin*. To assess the statistical significance that the observed non-synonymous changes exceeds that expected under selective neutrality, a previously published procedure was followed (29,42). Briefly, the neutral expectation was obtained by computer simulation, where nucleotide changes were randomly placed in the sequence using the observed synonymous transition-to-transversion ratio, till the number of synonymous changes reached the observed value. Data from a total of 10 000

replicas were then tallied, with non-synonymous changes classified according to Grantham's amino acid replacement distance matrix (30) into conservative (Grantham's distance  $\leq 50$ ), moderate (51–100), radical (101–150) or very radical ( $\geq 151$ ), as described previously (29,31). The fraction of simulation results that were at or exceeded the observed number of changes was used as the indicator of statistical significance.

## ACKNOWLEDGEMENTS

We thank L.G. Chemnick, A.R. Ryder and Yerkes Primate Center for providing tissue samples, and A. Di Rienzo, W. Dobyns, E.A. Grove, M. Kreitman, P.T. Schoenemann and H. Singh for technical input and/or helpful discussions. Supported in part by the Searle Scholarship and the Burroughs Wellcome Fund.

## REFERENCES

- Spuhler, J.N. (1959) *The Evolution of Man's Capacity for Culture*. Wayne State University Press, Detroit, MI.
- Brodmann, K. (1912) Ergebnisse über die vergleichende histologische lokalisation der grosshirnrinde mit besonderer berücksichtigung des stirmhirns. *Anat. Anz.*, **41** (suppl.), 157–216.
- Blinkov, S.M. and Glezer, I.J. (1968) *The Human Brain in Figures and Tables; A Quantitative Handbook*. Plenum Press, New York.
- Noback, C.R. and Montagna, W. (1970) *The Primate Brain*. Meredith Corporation, New York.
- Finlay, B.L. and Darlington, R.B. (1995) Linked regularities in the development and evolution of mammalian brains. *Science*, **268**, 1578–1584.
- Northcutt, R.G. and Kaas, J.H. (1995) The emergence and evolution of mammalian neocortex. *Trends Neurosci.*, **18**, 373–379.
- Williams, M.F. (2002) Primate encephalization and intelligence. *Med. Hypotheses*, **58**, 284–290.
- Olson, M.V. and Varki, A. (2003) Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.*, **4**, 20–28.
- Carroll, S.B. (2003) Genetics and the making of *Homo sapiens*. *Nature*, **422**, 849–857.
- Mochida, G.H. and Walsh, C.A. (2001) Molecular genetics of human microcephaly. *Curr. Opin. Neurol.*, **14**, 151–156.
- Dobyns, W.B. (2002) Primary microcephaly: new approaches for an old disorder. *Am. J. Hum. Genet.*, **112**, 315–317.
- Kumar, A., Markandaya, M. and Girimaji, S.C. (2002) Primary microcephaly: microcephalin and ASPM determine the size of the human brain. *J. Biosci.*, **27**, 629–632.
- Larson, C.S., Matter, R.M. and Gebo, D.L. (1998) *Human Origins: The Fossil Record*. Waveland Press, Prospect Heights.
- Jackson, A.P., McHale, D.P., Campbell, D.A., Jafri, H., Rashid, Y., Mannan, J., Karbani, G., Corry, P., Levene, M.I., Mueller, R.F. *et al.* (1998) Primary autosomal recessive microcephaly (MCPH1) maps to chromosome 8p22–pter. *Am. J. Hum. Genet.*, **63**, 541–546.
- Roberts, E., Jackson, A.P., Carradice, A.C., Deeble, V.J., Mannan, J., Rashid, Y., Jafri, H., McHale, D.P., Markham, A.F., Lench, N.J. *et al.* (1999) The second locus for autosomal recessive primary microcephaly (MCPH2) maps to chromosome 19q13.1–13.2. *Eur. J. Hum. Genet.*, **7**, 815–820.
- Moynihan, L., Jackson, A.P., Roberts, E., Karbani, G., Lewis, I., Corry, P., Turner, G., Mueller, R.F., Lench, N.J. and Woods, C.G. (2000) A third novel locus for primary autosomal recessive microcephaly maps to chromosome 9q34. *Am. J. Hum. Genet.*, **66**, 724–727.
- Jamieson, C.R., Govaerts, C. and Abramowicz, M.J. (1999) Primary autosomal recessive microcephaly: homozygosity mapping of MCPH4 to chromosome 15. *Am. J. Hum. Genet.*, **65**, 1465–1469.
- Jamieson, C.R., Fryns, J.P., Jacobs, J., Matthijs, G. and Abramowicz, M.J. (2000) Primary autosomal recessive microcephaly: MCPH5 maps to 1q25–q32. *Am. J. Hum. Genet.*, **67**, 1575–1577.
- Pattison, L., Crow, Y.J., Deeble, V.J., Jackson, A.P., Jafri, H., Rashid, Y., Roberts, E. and Woods, C.G. (2000) A fifth locus for primary autosomal recessive microcephaly maps to chromosome 1q31. *Am. J. Hum. Genet.*, **67**, 1578–1580.
- Leal, G.F., Roberts, E., Silva, E.O., Costa, S.M., Hampshire, D.J. and Woods, C.G. (2003) A novel locus for autosomal recessive primary microcephaly (MCPH6) maps to 13q12.2. *J. Med. Genet.*, **40**, 540–542.
- Jackson, A.P., Eastwood, H., Bell, S.M., Adu, J., Toomes, C., Carr, I.M., Roberts, E., Hampshire, D.J., Crow, Y.J., Mighell, A.J. *et al.* (2002) Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am. J. Hum. Genet.*, **71**, 136–142.
- Bond, J., Roberts, E., Mochida, G.H., Hampshire, D.J., Scott, S., Askham, J.M., Springell, K., Mahadevan, M., Crow, Y.J., Markham, A.F. *et al.* (2002) ASPM is a major determinant of cerebral cortical size. *Nat. Genet.*, **32**, 316–320.
- Bond, J., Scott, S., Hampshire, D.J., Springell, K., Corry, P., Abramowicz, M.J., Mochida, G.H., Hennekam, R.C., Maher, E.R., Fryns, J.P. *et al.* (2003) Protein-truncating mutations in ASPM cause variable reduction in brain size. *Am. J. Hum. Genet.*, **73**, 1170–1177.
- Huyton, T., Bates, P.A., Zhang, X., Sternberg, M.J. and Freemont, P.S. (2000) The BRCA1 C-terminal domain: structure and function. *Mutat. Res.*, **460**, 319–332.
- Evans, P.D., Anderson, J.R., Vallender, E.J., Gilbert, S.L., Malcom, C.M., Dorus, S. and Lahn, B.T. (2004) Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum. Mol. Genet.*, **13**, 489–494.
- Li, W.H. (1997) *Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- McDonald, J.H. and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, **351**, 652–654.
- Kreitman, M. (2000) Methods to detect selection in populations with applications to the human. *Annu. Rev. Genomics Hum. Genet.*, **1**, 539–559.
- Wyckoff, G.J., Wang, W. and Wu, C.I. (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature*, **403**, 304–309.
- Grantham, R. (1974) Amino acid difference formula to help explain molecular evolution. *Science*, **185**, 862–864.
- Li, W.H., Wu, C.-I. and Luo, C.C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.*, **2**, 150–174.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- Zhang, J. (2003) Evolution of the human ASPM gene, a major determinant of brain size. *Genetics*, **165**, 2063–2070.
- Jerison, J.H. (1973) *Evolution of the Brain and Intelligence*. Academic Press, New York.
- Walker, A., Falk, D., Smith, R. and Rickford, M. (1983) The skull of *Proconsul africanus*: reconstruction and cranial capacity. *Nature*, **305**, 525–527.
- Huttley, G.A., Easteal, S., Southey, M.C., Tesoriero, A., Giles, G.G., McCredie, M.R., Hopper, J.L. and Venter, D.J. (2000) Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Australian Breast Cancer Family Study. *Nat. Genet.*, **25**, 410–413.
- Fleming, M.A., Potter, J.D., Ramirez, C.J., Ostrander, G.K. and Ostrander, E.A. (2003) Understanding missense mutations in the BRCA1 gene: an evolutionary approach. *Proc. Natl Acad. Sci. USA*, **100**, 1151–1156.
- Gowen, L.C., Johnson, B.L., Latour, A.M., Sulik, K.K. and Koller, B.H. (1996) Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nat. Genet.*, **12**, 191–194.
- Li, W.H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.*, **36**, 96–99.
- Fares, M.A., Elena, S.F., Ortiz, J., Moya, A. and Barrio, E. (2002) A sliding window-based method to detect selective constraints in protein-coding genes and its application to RNA viruses. *J. Mol. Evol.*, **55**, 509–521.
- Choi, S.S. and Lahn, B.T. (2003) Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. *Genome Res.*, **13**, 2252–2259.