

Evolution of the Human *ASPM* Gene, a Major Determinant of Brain Size

Jianzhi Zhang¹

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109

Manuscript received July 7, 2003

Accepted for publication August 20, 2003

ABSTRACT

The size of human brain tripled over a period of ~2 million years (MY) that ended 0.2–0.4 MY ago. This evolutionary expansion is believed to be important to the emergence of human language and other high-order cognitive functions, yet its genetic basis remains unknown. An evolutionary analysis of genes controlling brain development may shed light on it. *ASPM* (abnormal spindle-like microcephaly associated) is one of such genes, as nonsense mutations lead to primary microcephaly, a human disease characterized by a 70% reduction in brain size. Here I provide evidence suggesting that human *ASPM* went through an episode of accelerated sequence evolution by positive Darwinian selection after the split of humans and chimpanzees but before the separation of modern non-Africans from Africans. Because positive selection acts on a gene only when the gene function is altered and the organismal fitness is increased, my results suggest that adaptive functional modifications occurred in human *ASPM* and that it may be a major genetic component underlying the evolution of the human brain.

AMONG mammals, humans have an exceptionally big brain relative to their body size. For example, in comparison with chimpanzees, the brain weight of humans is 250% greater while the body is only 20% heavier (McHENRY 1994). The dramatic evolutionary expansion of the human brain started from an average brain weight of 400–450 g ~2–2.5 million years (MY) ago and ended with a weight of ~1350–1450 g ~0.2–0.4 MY ago (McHENRY 1994; WOOD and COLLARD 1999). This process represents one of the most rapid morphological changes in evolution. It is generally believed that the brain expansion set the stage for the emergence of human language and other high-order cognitive functions and that it was caused by adaptive selection (DECAN 1992), yet the genetic basis of the expansion remains elusive. A study of human mutations that result in unusually small brains may help identify the genetic modifications that contributed to the human brain expansion. In this regard, primary microcephaly (small head) is of particular interest (MOCHIDA and WALSH 2001; BOND *et al.* 2002; KUMAR *et al.* 2002). Microcephaly is an autosomal recessive genetic disease with an incidence of 4–40 per million live births in western countries (MOCHIDA and WALSH 2001; KUMAR *et al.* 2002). It is defined as a head circumference >3 standard deviations below the population age-related mean, but with no associated malfunctions other than mild-to-moderate mental retardation (MOCHIDA and WALSH 2001; KUMAR *et al.* 2002).

The reduction in head circumference correlates with a markedly reduced brain size. Microcephaly is genetically heterogeneous, associated with mutations in at least five loci (MOCHIDA and WALSH 2001; KUMAR *et al.* 2002), one of which was recently identified and named *ASPM* (abnormal spindle-like microcephaly associated; BOND *et al.* 2002). Four different homozygous mutations in *ASPM* introducing premature stop codons were found to cosegregate with the disease in four respective families, while none of these mutations were found in 200 normal human chromosomes (BOND *et al.* 2002). Because the brain size of a typical microcephaly patient (430 g; MOCHIDA and WALSH 2001; KUMAR *et al.* 2002) is comparable with those of early hominids such as the 2.3- to 3.0-MY-old *Australopithecus africanus* (420 g; McHENRY 1994; WOOD and COLLARD 1999), I hypothesize that *ASPM* may be one of the genetic components underlying the human brain expansion. Signatures of accelerated evolution of *ASPM* under positive selection during human origins would strongly support my hypothesis, because the action of positive selection indicates a modification in gene function resulting in elevated organismal fitness (ZHANG *et al.* 2002). Below I provide population genetic and molecular evolutionary evidence for the operation of such adaptive selection on *ASPM*.

MATERIALS AND METHODS

Sequencing of *ASPM*: The human *ASPM* gene has 28 exons. All 28 exons were PCR amplified from genomic DNA samples of 14 human (*Homo sapiens*) individuals of different geographic origins (2 African Pygmies, 3 African Americans, 4 Europeans, 2 Southeast Asians, 1 Chinese, 1 Pacific islander, and 1 South American), using the high-fidelity Taq of In-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY367065–87.

¹Address for correspondence: Department of Ecology and Evolutionary Biology, University of Michigan, 3003 Natural Science Bldg., 830 N. University Ave., Ann Arbor, MI 48109. E-mail: jianzhi@umich.edu

vitrogen (Carlsbad, CA). The PCR products were then purified and sequenced in both directions. Polymorphisms that appear only once (singletons) were confirmed by a second PCR-sequencing experiment. The human DNA samples were purchased from Coriell (Camden, NJ). I also amplified all 28 exons from one chimpanzee (*Pan troglodytes*) and one orangutan (*Pongo pygmaeus*) and sequenced the insert DNAs after the PCR products were cloned into the pCR4TOPO vector (Invitrogen). To trace the evolutionary origin of a large insertion/deletion in exon 18, two segments (Figure 1, I and II) of exon 18 were also amplified and directly sequenced from genomic DNAs of hyrax (*Procapra capensis*), sea lion (*Zalophus californianus*), seal (*Phoca vitulina*), wolverine (*Gulo gulo*), fox (*Alopex lagopus*), dog (*Canis familiaris*), bear (*Ursus maritimus*), cat (*Felis catus*), pig (*Sus scrofa*), cow (*Bos taurus*), whale (*Balaena mysticetus*), rhesus monkey (*Macaca mulatta*), owl monkey (*Aotus trivirgatus*), and hamster (*Cricetulus griseus*).

Data analysis: The d_N/d_S ratios (LI 1997; NEI and KUMAR 2000) in the human, chimpanzee, and orangutan branches of the *ASPM* gene tree (Figure 2) were estimated using the maximum-likelihood method (YANG 1997), under the model of unequal codon frequencies (CodonFreq = 3) and unequal rates of transitions and transversions. This model fits the data significantly better than those with fixed codon frequencies (CodonFreq = 2) or equal rates of transitions and transversions. The hypothesis of $d_N/d_S = 1$ for a given branch and the hypothesis of equal d_N/d_S ratios between two branches were tested using the likelihood-ratio test, as well as a nonlikelihood method based on inferred ancestral sequences (ZHANG *et al.* 1997, 1998). The *ASPM* sequence for the common ancestor of humans and chimpanzees was estimated using the Bayesian method (YANG 1997; ZHANG and NEI 1997). Population genetic analysis was conducted using DnaSP3.99 (ROZAS and ROZAS 1999). TAJIMA's (1989), FU and LI's (1993), and FAY and WU's (2000) tests were conducted using coalescent simulations under the assumption of no recombination across the gene. This assumption makes the tests more conservative. When recombination is considered, the *P* values become lower in Tajima's and Fu and Li's tests, but remain virtually the same in Fay and Wu's test. The recombination rate in the *ASPM* locus was estimated from KONG *et al.* (2002) to be $1.8 \text{ cM}/10^6$ nucleotides. Since the coding sequences analyzed here span 62×10^3 nucleotides of the genome, the recombination rate for the sequence is $r = 1.8 \times 10^{-2} \times 0.062 = 1.1 \times 10^{-3}$ recombination/generation. Assuming that the human effective population size is 10^4 (TAKAHATA *et al.* 1995; HARPENDING *et al.* 1998), the population recombination parameter ($4Nr$) equals $4 \times 10^4 \times 1.1 \times 10^{-3} = 44$. Gene trees of segment I and II sequences from various mammals were reconstructed by the neighbor-joining method (SAITOU and NEI 1987) with KIMURA's (1980) two-parameter distances. Use of other distances or other methods (NEI and KUMAR 2000) does not change the result. Two thousand bootstrap replications (FELSENSTEIN 1985) were used to test the reliability of the trees. The MEGA2 software (KUMAR *et al.* 2001) was used for the phylogenetic analysis.

Computer simulation: Computer simulation of *ASPM* evolution under pure neutrality was conducted following the procedure described in ZHANG and WEBB (2003). The rate that an open reading frame (ORF) becomes disrupted is mainly determined by the sequence of the ORF, the rate of point mutations, and the rate of insertion/deletion (indel) mutations. Here I used a point mutation rate of 1×10^{-9} /site/year, which was estimated from a genomic comparison between humans and chimpanzees (YI *et al.* 2002). The relative mutational frequencies among the four nucleotides have only a negligible effect on the simulation result and I assumed that they are equal. I used an indel mutation rate of 1.1×10^{-10} /

site/year, which was also estimated from a genomic comparison between the human and chimpanzee (BRITTEN 2002). I assumed that all indels with sizes that are multiples of three nucleotides do not disrupt an ORF. This simplifies the simulation but does not affect the results, because the majority of indels generated by mutations have small sizes (less than or equal to nine nucleotides; BRITTEN 2002). In the above genomic data analyzed, 19% of the total 1019 indels are of sizes that are multiples of three nucleotides. A simulation was then performed for 20,000 replications with the inferred *ASPM* coding sequence of the common ancestor of humans and chimpanzees. Under no functional constraints, the substitution rate is identical to the mutation rate and mutations are assumed to be random. An ORF is interrupted when an indel of a size that is not a multiple of three nucleotides or a nonsense point mutation occurs. I thus determined the $t_{1/2}$ for *ASPM*, or the time required for an intact ORF to be interrupted in half of the simulation replications. Under the above parameters, $t_{1/2} = 0.48$ MY. When point and indel mutation rates are both halved, $t_{1/2} = 0.97$ MY. The probability that *ASPM* retains its ORF after *T*MY of neutral evolution is $(1/2)^{T/t_{1/2}}$. This estimation of the rate of pseudogenization is conservative, because other deleterious events such as insertions of transposable elements and null mutations at promoter regions and splicing sites are not considered here. The computer program for the simulation is from ZHANG and WEBB (2003).

RESULTS

Elevation of d_N/d_S in the human *ASPM* lineage: Human *ASPM* has 28 coding exons, spanning 62 kb in chromosome 1p31 and encoding a huge protein of 3477 amino acids (Figure 1). I determined the entire coding sequences of *ASPM* from one human, one chimpanzee, and one orangutan, and compared them in the phylogenetic tree of the three species (Figure 2). The orangutan sequence is used as the outgroup for humans and chimpanzees so that nucleotide substitutions on the human and chimpanzee lineages can be separated. I did not sequence the gorilla because the gorilla sequence may not be appropriate as the outgroup due to incomplete lineage sorting (SATTA *et al.* 2000). Use of orangutan, a slightly more distant outgroup, solves this problem. A commonly used indicator of natural selection at the DNA sequence level is the ratio of the rate of nonsynonymous nucleotide substitution (d_N) to that of synonymous substitution (d_S). Most functional genes show $d_N/d_S < 1$, because a substantial proportion of nonsynonymous mutations are deleterious and are removed by purifying selection, whereas synonymous mutations are more or less neutral and are generally uninfluenced by selection. A gene may occasionally exhibit $d_N/d_S > 1$ when a large fraction of nonsynonymous mutations are advantageous and are driven to fixation by positive selection (LI 1997; NEI and KUMAR 2000). I estimated the d_N/d_S ratio for *ASPM* in each of the three tree branches (Figure 2), using a maximum-likelihood method, and found that d_N/d_S is lowest in the orangutan branch (0.43), higher in the chimpanzee branch (0.66), and highest in the human branch (1.03). The hypothesis of $d_N/d_S = 1$ is rejected for the orangutan branch ($P < 0.001$, likeli-

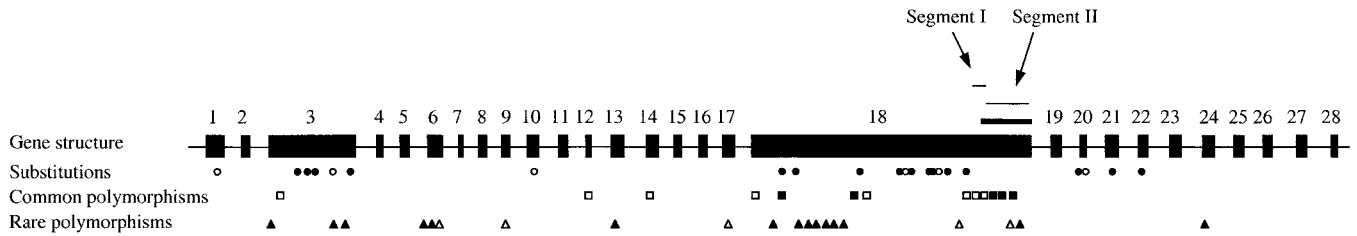


FIGURE 1.—Nucleotide polymorphisms and substitutions in the human *ASPM* gene. The exons are shown by solid boxes and introns by lines. The sizes of exons, but not introns, are drawn to scale. The thick horizontal line above exon 18 shows the 867-nucleotide fragment that is not found in the mouse *Aspm*, with the two thin lines above it showing the two segments amplified in various mammals (see text and Figure 3). Fixed nucleotide substitutions in the human lineage after the human-chimpanzee split are shown by circles, whereas common and rare polymorphisms within humans are shown by squares and triangles, respectively. Nonsynonymous changes are solid symbols and synonymous changes are open symbols.

hood-ratio test), but not for the other two branches, suggesting a difference in selection has occurred. Indeed, a test of the difference in d_N/d_S between the human and orangutan branches gives a marginally significant result ($P = 0.064$), but the difference between the chimpanzee and orangutan branches is not significant ($P = 0.29$), nor is the difference between human and chimpanzee branches ($P = 0.45$). Because the d_N/d_S ratio between the orangutan and mouse (*Mus musculus*) is also low (0.29), an increase of d_N/d_S in humans is more likely than a decrease in orangutans. The mouse sequence (GenBank accession no. AF533752) was not included in the phylogeny-based analysis as it is relatively distantly related to the ape sequences and contains multiple insertions and deletions, which would make the inference less reliable. Similar results are obtained when I first infer the *ASPM* sequence for the common ancestor of humans and chimpanzees and then estimate the d_N/d_S ratio by counting the numbers of synonymous and nonsynonymous nucleotide substitutions on each branch. For instance, this approach gives $d_N/d_S = 1.13$, 0.84, and 0.52, respectively, for the human, chimpanzee, and orangutan branches.

Complete functional relaxation does not adequately explain the elevation of d_N/d_S : Two hypotheses may explain the increase in d_N/d_S to 1.03 during the evolution of human *ASPM*. First, the functional constraints and purifying selection on *ASPM* may have been completely relaxed and many deleterious nonsynonymous mutations were fixed by random genetic drift. Alternatively, advantageous nonsynonymous substitutions un-

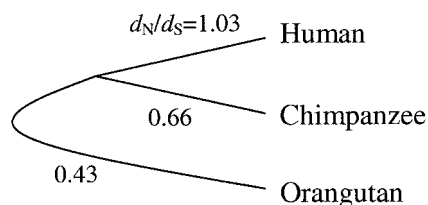


FIGURE 2.—The d_N/d_S ratios for the entire *ASPM* coding sequence in the human, chimpanzee, and orangutan lineages. The values were estimated from a likelihood analysis.

der positive selection occurred at some sites, while purifying selection acted at some other sites, resulting in an average d_N/d_S of ~ 1 . Under the first hypothesis, *ASPM* has been under pure neutral evolution since the human-chimpanzee separation ~ 6 – 7 MY ago (BRUNET *et al.* 2002). Using rates of single-nucleotide mutations and insertion/deletion mutations estimated from human-chimpanzee genomic comparisons (BRITTEN 2002; YI *et al.* 2002), I conducted a computer simulation of neutral evolution of *ASPM* (see MATERIALS AND METHODS). I found that the probability that *ASPM* retains its open reading frame after 6 MY of neutral evolution is extremely low (1.7×10^{-4}). Even when the above two mutation rates are both halved, the probability is still very small (0.014), suggesting that *ASPM* must have been under purifying selection. The fact that nonsense mutations in *ASPM* lead to microcephaly also demonstrates the presence of functional constraints on the gene. Thus, the hypothesis of complete relaxation of functional constraints and lack of purifying selection for the past 6–7 MY of human evolution is inconsistent with the data, and some sites in *ASPM* must have been subject to purifying selection ($d_N/d_S < 1$). This result would imply, although not prove, that some other sites are under positive selection ($d_N/d_S > 1$), so that the average d_N/d_S across the entire protein is ~ 1 . However, it is difficult to rule out the possibility of an incomplete functional relaxation in human *ASPM*, which can lead to a d_N/d_S ratio of ~ 1 when the number of substitutions is relatively small. A population genetic study may help resolve this question.

Signatures of purifying selection from population genetic data: The entire coding sequence of *ASPM* is determined from 14 human individuals of different geographic origins. A total of 33 single-nucleotide polymorphisms are found (Tables 1 and 2). The derived and ancestral alleles are inferred using the chimpanzee and orangutan sequences as outgroups. TAJIMA's (1989) and FU and LI's (1993) tests reveal slight departure of the data from the Wright-Fisher model of neutrality ($D = -1.29$, $P = 0.081$; $F = -1.76$, $P = 0.074$; Table 2). But FAY and WU's (2000) test, which is designed to detect recent

TABLE 1
DNA sequence variations in the *ASPM* gene of humans

Exon	Position	AP-1	AP-2	AA-1	AA-2	AA-3	EU-1	EU-2	EU-3	EU-4	PI-1	SA-1	SEA-1	SEA-2	CH-1
3	203	AA	AC	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
3	408	TT	CT	CC	TT	CC	CT	TT	TT	CT	TT	TT	TT	TT	TT
3	847	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
3	1010	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
6	45	AA	AT	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
6	120	CG	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
6	134	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
9	131	CC	CC	CC	CC	CC	CC	CC	CT	CC	CC	CC	CC	CC	CC
12	56	GG	GG	GG	GA	GG	GA	GA	GA	GA	GG	GA	GG	GG	GG
13	101	CC	CC	CT	CC	CT	CC	CC	CC	CC	CC	CC	CC	CC	CC
14	189	AA	AA	AT	AA	AT	AA	AA	AA	AT	AA	AA	AA	AA	AA
17	42	CC	CC	CA	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
18	384	AG	AA	AA	AG	AA	AA	AG	AG	AA	AG	AG	AA	AG	AA
18	731	GG	GG	GG	GG	GA	GG	GG	GG	GG	GG	GG	GG	GG	GG
18	998	CC	TC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
18	1018	CC	CT	CT	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
18	1120	CC	CC	CC	CC	CC	CC	CT	CC	CC	CC	CC	CC	CC	CC
18	1141	CC	CC	CC	CC	CG	CC	CC	CC	CC	CC	CC	CC	CC	CC
18	1303	GG	GA	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
18	1406	AA	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA
18	1514	CC	CT	CT	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
18	1756	TT	TT	CT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
18	1896	AA	AG	AA	AG	AA	AA	AA	AG	AA	AG	AA	AA	AG	AA
18	3486	TT	TT	TT	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT
18	3501	GG	GG	GA	GG	GA	GA	GG	GG	GA	GG	GG	GG	GG	GG
18	3540	GG	GG	GG	GG	GG	GG	GG	GG	GG	GA	GG	AA	GG	AA
18	3609	CT	CC	CC	CT	CC	CC	CT	CT	CC	CT	CT	CC	CT	CC
18	3619	AG	AA	AA	AG	AA	AA	AG	AG	AA	AG	AG	AA	AG	AA
18	3722	TT	CT	CT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
18	3852	AG	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
18	3874	CA	CC	CC	CA	CC	CC	CA	CA	CC	CA	CA	CC	CA	CC
18	4003	CT	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
24	137	AA	AG	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA

Shown here are the genotypes of the 14 humans at all polymorphic sites found in *ASPM*. For heterozygotes, the derived allele is shown after the ancestral allele. AP, African Pygmy; AA, African American; EU, European; PI, Pacific Islander; SA, South American; SEA, Southeast Asian; CH, Chinese.

selective sweeps, does not show a significant result ($H = -2.08$, $P = 0.21$). Thus, the negative D and F likely reflect recent population expansions and/or purifying background selection. A recent study suggested that negative D values may also be found under certain sampling schemes if there is fine-scale population differentiation (PTAK and PRZEWSKI 2002). When the synonymous and nonsynonymous sites were analyzed separately, I detected significant negative D and F values at nonsynonymous sites ($P < 0.05$; Table 2), but not at synonymous sites. H is not significant at either type of site. These results suggest that the nonsynonymous sites in human *ASPM* are subject to purifying selection. It should be mentioned that the recombination rate in the *ASPM* region is $\sim 1.8 \text{ cM}/10^6$ nucleotides (KONG *et al.* 2002), which translates into 1.1×10^{-3} recombination/meiosis for the sequences analyzed here. This relatively high recombination rate localizes signatures of selection to

a small region surrounding the selected sites. This might in part explain the above differences in the test results between synonymous and nonsynonymous sites.

Population genetic theory predicts that deleterious mutations do not reach high frequencies in populations, while neutral and advantageous mutations do. A comparison between rare and common polymorphisms may detect purifying selection of deleterious mutations (FAY *et al.* 2001). Fay *et al.* recommended a frequency of $\sim 10\%$ for the derived allele as a cutoff between rare and common polymorphisms (FAY *et al.* 2001, 2002). In the present sample of 28 chromosomes, derived alleles that appear one or two times are regarded as rare polymorphisms, and the rest are common. Because of the limited sample size, a truly rare allele may inadvertently appear more than twice in our sample and a truly common allele may inadvertently be regarded as rare. Using probability theory, I computed that the probability of

TABLE 2
Intra- and interspecific variations of the human *ASPM* gene

	Total	Synonymous	Nonsynonymous
No. of human chromosomes sequenced	28	28	28
No. of sites sequenced	10431	2972	7459
Nucleotide diversity per site π	0.00053	0.00103	0.00033
Watterson's θ_W per site	0.00081	0.00112	0.00069
Fay and Wu's θ_H per site	0.00077	0.00059	0.00080
Tajima's D	-1.29	-0.29	-1.84 ^a
Fu and Li's F	-1.76	-0.37	-2.40 ^a
Fay and Wu's H	-2.08	1.30	-3.38
No. of rare polymorphisms (<10% ^b)	20	5	15 ^c
No. of common polymorphisms (>10% ^b)	13	8	5
No. of fixed substitutions in the human branch	22	6	16 ^c

^a Five percent significant in comparison to the Wright-Fisher model of neutrality.

^b Use of a cutoff frequency of 5–15% gives similar results, although the significance level may change. See LI (1997) and NEI and KUMAR (2000) for definitions of π and θ_W .

^c Five percent significant in comparison with common polymorphisms.

the former error is <5% for an allele with frequency <3% and the probability of the latter error is <5% for an allele with frequency >20%. Thus, the present classification of rare and common alleles is expected to be relatively accurate. I observed that $n_r = 15$ nonsynonymous and $s_r = 5$ synonymous rare polymorphisms and $n_c = 5$ nonsynonymous and $s_c = 8$ synonymous common polymorphisms from the present data (Table 2; Figure 1). The ratio of n_c to n_r ($5/15 = 0.333$) is significantly lower than that of s_c to s_r ($8/5 = 1.6$; $\chi^2 = 4.41$, $P < 0.05$; Table 2). Since synonymous mutations are more or less neutral, the observed deficit of common nonsynonymous polymorphisms suggests that purifying selection has prevented the spread of nonsynonymous deleterious mutations. It is estimated by the likelihood method that there are $N = 7459$ and $S = 2972$ potentially nonsynonymous and synonymous sites in *ASPM*, respectively. Thus, for rare polymorphisms, there are $n_r/N = 15/7459 = 2.01 \times 10^{-3}$ polymorphisms/nonsynonymous site and $s_r/S = 5/2972 = 1.68 \times 10^{-3}$ /synonymous site. Their difference is statistically insignificant ($\chi^2 = 0.09$, $P > 0.5$). In contrast, for common polymorphisms, the number is significantly smaller per nonsynonymous site ($n_c/N = 5/7459 = 0.67 \times 10^{-3}$) than per synonymous site ($s_c/S = 8/2972 = 2.69 \times 10^{-3}$; $\chi^2 = 6.98$, $P < 0.01$), confirming that purifying selection has reduced the number of common nonsynonymous polymorphisms. This result also suggests the absence or rareness of advantageous nonsynonymous polymorphisms of *ASPM* that are currently segregating in humans, as such polymorphisms would predominantly show up as common polymorphisms and render n_c/N higher. This is consistent with the above result from Fay and Wu's test. The proportion of nonsynonymous polymorphisms not under purifying selection may be estimated by $(n_c/N)/(s_c/S) = (0.67 \times 10^{-3})/(2.69 \times 10^{-3}) = 0.25$ or by $(n_c/$

$s_c)/(n_r/s_r) = (5/8)/(15/5) = 0.21$. The two estimates are close to each other and to the d_N/d_S ratio between the mouse and orangutan (0.29). This indicates that human *ASPM* is currently under relatively strong purifying selection, and the strength of selection is comparable to or even greater than that in the long-term evolution of mammalian *ASPM*.

Comparison of polymorphism and divergence suggests past positive selection: Because both the synonymous and nonsynonymous common polymorphisms are largely neutral, comparing them with the fixed substitutions between humans and chimpanzees can reveal the signature of selection that has influenced the substitution processes (McDONALD and KREITMAN 1991; FAY *et al.* 2001, 2002; SMITH and EYRE-WALKER 2002). This comparison shows a significant excess of fixed nonsynonymous substitutions ($\chi^2 = 3.88$, $P < 0.05$, Table 2), suggesting that some nonsynonymous substitutions were fixed by positive selection. Because the expansion of brain size occurred in the human lineage after the human-chimpanzee split, it is more relevant to examine whether the human branch exhibits an excess of nonsynonymous substitutions. For this, the *ASPM* sequence of the common ancestor of humans and chimpanzees was inferred by the Bayesian method. Because the sequences considered are closely related, this inference is reliable, with the average posterior probability >0.999. Comparing the ancestral sequence with the polymorphic human sequences, I identified 16 nonsynonymous and 6 synonymous mutations that have been fixed in the human lineage (Table 2; Figure 1). Their ratio ($16/6 = 2.67$) is significantly greater than that for common polymorphisms ($n_c/s_c = 5/8 = 0.63$; $\chi^2 = 4.00$, $P < 0.05$). The number of neutral nonsynonymous substitutions may be estimated from the number of synonymous substitutions multiplied by n_c/s_c , which yielded $6 \times (5/$

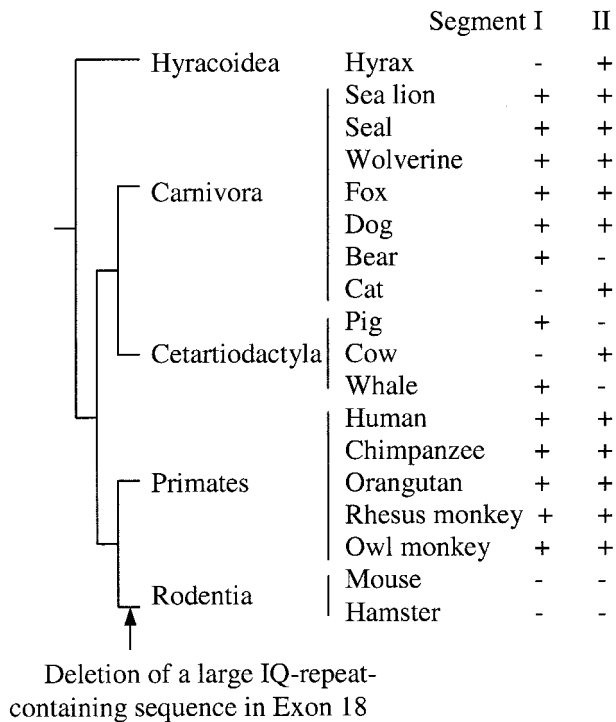


FIGURE 3.—Amplification of segments I and II of the 867-nucleotide IQ-repeat-containing fragment in exon 18 of orthologous *ASPM* genes of mammals. The two segments are shown in Figure 1. “+” represents a successful amplification and sequencing and “-” represents no amplification. Note that the entire coding sequence of the gene is known for human, chimpanzee, orangutan, and mouse. The phylogenetic relationships of the mammalian orders are depicted following MURPHY *et al.* (2001).

8) = 3.75 (FAY *et al.* 2001, 2002; SMITH and EYRE-WALKER 2002). The number of nonsynonymous substitutions unexplainable by neutral evolution is $16 - 3.75 = 12$, which may have been fixed by positive selection. It should be noted that a recent population expansion can cause an overestimate of the number of adaptive substitutions when slightly deleterious mutations are present. However, such overestimation is unlikely in the present case because the current effective population size of humans, even after the recent expansion, is still smaller than the long-term effective population size separating humans and chimpanzees and the effective population size of the common ancestor of humans and chimpanzees (TAKAHATA *et al.* 1995; CHEN and LI 2001; KAESSMANN *et al.* 2001; EYRE-WALKER 2002). It is interesting that there is no significant excess of nonsynonymous substitutions for either the chimpanzee or orangutan branches when the common polymorphisms and substitutions are compared ($P > 0.05$).

IQ repeats and brain size variation: Human *ASPM* contains multiple calmodulin-binding IQ repeats (BOND *et al.* 2002). In a comparison of putative orthologous *ASPM* genes from the human, mouse, fruit fly

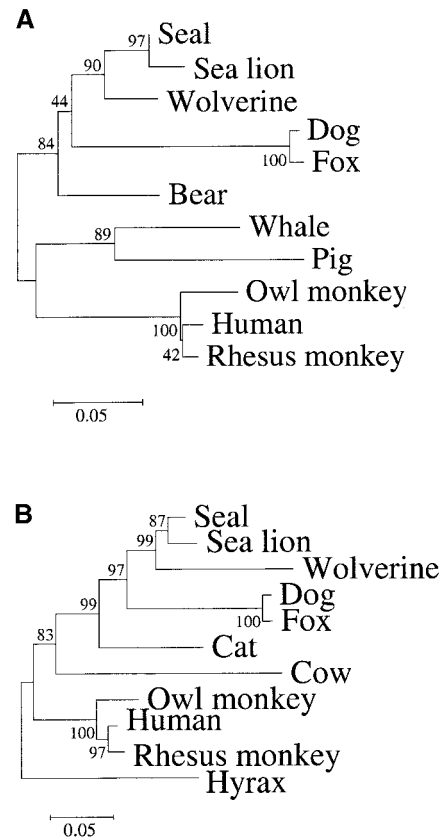


FIGURE 4.—Neighbor-joining trees of (A) segment I and (B) segment II sequences from *ASPM* exon 18 (see Figure 1). The aligned sequences of segment I have 155 nucleotide sites after the removal of primer-encoded regions, and those of segment II have 651 sites. Kimura’s two-parameter distances are used. Bootstrap percentages from 2000 replications are shown on interior nodes.

(*Drosophila melanogaster*), and nematode (*Caenorhabditis elegans*), BOND *et al.* (2002) noticed that organisms with larger brains have more IQ repeats, implying a possible relation of IQ repeats and brain size. In particular, the predominant difference between the human and mouse *ASPM* genes is a large IQ-repeat-encoding insertion of 867 nucleotides at the end of exon 18. However, my data showed no difference in the number of IQ repeats between human and chimpanzee *ASPM* sequences. To trace the origin of the large insertion in human *ASPM*, I amplified and sequenced from several mammals two DNA segments that cover most of the insertion (Figure 1). Segment I is of 212 nucleotides and segment II is of 706 nucleotides. One or both segments were obtained from species belonging to primates, Cetartiodactyla, Carnivora, and Hyracoidea, but not from mouse or hamster (Figure 3). Phylogenetic analyses were conducted to confirm that the obtained sequences are orthologous to the human sequence (Figure 4). While nonamplification of a sequence does not prove its nonexistence, the amplification of the orthologous sequence indicates its pres-

ence. From the recently established mammalian phylogeny (MURPHY *et al.* 2001), it can be inferred that the large human insertion was already present in the common ancestor of most placental mammals, but was deleted in mouse and possibly in other rodents (Figure 3). Thus, this IQ-repeat-containing sequence does not explain the brain size variation among many nonrodent mammals.

DISCUSSION

In the above, I provided evidence that advantageous amino acid substitutions unrelated to IQ repeats have been fixed by adaptive selection in human ASPM after the human-chimpanzee split, which strongly suggests that ASPM might be an important genetic component in the evolutionary expansion of human brain. The episode of positive selection on ASPM appears to have ended some time ago, as there is no evidence for positive selection on ASPM in current human populations; rather, relatively strong purifying selection is detected. Roughly, selective sweeps occurring in the past $0.5N$ generations may be detected (FAY and WU 2000), where N is the effective population size of humans and is thought to be $\sim 10,000$ (TAKAHATA *et al.* 1995; HARPENDING *et al.* 1998). That is, the positive selection detected in ASPM occurred some time between 6–7 and 0.1 MY ago ($0.5 \times 10,000$ generations $\times 20$ years/generation). The latter date coincides with the suggested time of migration of modern humans out of Africa (reviewed in CAVALLI-SFORZA and FELDMAN 2003). It is also interesting to note that although the precise time when positive selection acted on ASPM is difficult to pinpoint, my estimate is consistent with the current understanding that the human brain expansion took place between 2–2.5 and 0.2–0.4 MY ago (MCHENRY 1994; WOOD and COLLARD 1999). Furthermore, a selective sweep in human FOXP2, a gene involved in speech and language development, has been detected (ENARD *et al.* 2002; ZHANG *et al.* 2002). This sweep was estimated to have occurred no earlier than 0.1–0.2 MY ago (ENARD *et al.* 2002; ZHANG *et al.* 2002). That is, the adaptive evolution of FOXP2 post-dated that of ASPM, consistent with the common belief that a big brain may be a prerequisite for language (DECAN 1992).

Studies of ASPM in model organisms can help us understand how it impacts brain size. The mouse *Aspm* is highly expressed in the embryonic brain, particularly during cerebral cortical neurogenesis (BOND *et al.* 2002). The fruit fly ortholog *asp* is involved in organizing and binding together microtubules at the spindle poles and in forming the central mitotic spindle (GONZALEZ *et al.* 1990; WAKEFIELD *et al.* 2001). Mutations in *asp* cause dividing neuroblasts to arrest in metaphase, resulting in reduced central nervous system development (WAKEFIELD *et al.* 2001). The amino acid substitutions in human ASPM are located in exons 3, 18, 20, 21, and 22

(Figure 1), which encode a putative microtubule-binding domain and an IQ calmodulin-binding domain (BOND *et al.* 2002). These features suggest that the adaptive substitutions in human ASPM might be related to the regulation of mitosis in the nervous system, which can be tested in the future by functional assays of human ASPM as well as a laboratory-reconstructed ASPM protein of the common ancestor of humans and chimpanzees.

I thank David Webb for technical assistance and Douglas Futuyma, Priscilla Tucker, and David Webb for valuable comments on an earlier version of the manuscript. This work was supported by a start-up fund of the University of Michigan and a research grant from the National Institutes of Health (GM67030).

LITERATURE CITED

- BOND, J., E. ROBERTS, G. H. MOCHIDA, D. J. HAMPSHIRE, S. SCOTT *et al.*, 2002 ASPM is a major determinant of cerebral cortical size. *Nat. Genet.* **32**: 316–320.
- BRITTEN, R. J., 2002 Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci. USA* **99**: 13633–13635.
- BRUNET, M., F. GUY, D. PILBEAM, H. T. MACKAYE, A. LIKIUS *et al.*, 2002 A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**: 145–151.
- CAVALLI-SFORZA, L. L., and M. W. FELDMAN, 2003 The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* **33** (Suppl.): 266–275.
- CHEN, F. C., and W. H. LI, 2001 Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**: 444–456.
- DECAN, T. W., 1992 Biological aspects of language, pp. 128–133 in *The Cambridge Encyclopedia of Human Evolution*, edited by S. JONES, R. MARTIN and D. PILBEAM. Cambridge University Press, Cambridge, UK.
- ENARD, W., M. PRZEWORSKI, S. E. FISHER, C. S. LAI, V. WIEBE *et al.*, 2002 Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**: 869–872.
- EYRE-WALKER, A., 2002 Changing effective population size and the McDonald-Kreitman test. *Genetics* **162**: 2017–2024.
- FAY, J. C., and C.-I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- FAY, J. C., G. J. WYCKOFF and C.-I. WU, 2001 Positive and negative selection on the human genome. *Genetics* **158**: 1227–1234.
- FAY, J. C., G. J. WYCKOFF and C.-I. WU, 2002 Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**: 1024–1026.
- FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
- FU, Y. X., and W. H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GONZALEZ, C., R. D. SAUNDERS, J. CASAL, I. MOLINA, M. CARMENA *et al.*, 1990 Mutations at the *asp* locus of *Drosophila* lead to multiple free centrosomes in syncytial embryos, but restrict centrosome duplication in larval neuroblasts. *J. Cell Sci.* **96**: 605–616.
- HARPENDING, H. C., M. A. BATZER, M. GURVEN, L. B. JORDE, A. R. ROGERS *et al.*, 1998 Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* **95**: 1961–1967.
- KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- KONG, A., D. F. GUDBJARTSSON, J. SAINZ, G. M. JONSDOTTIR, S. A. GUDJONSSON *et al.*, 2002 A high-resolution recombination map of the human genome. *Nat. Genet.* **31**: 241–247.
- KAESSMANN, H., V. WIEBE, G. WEISS and S. PAABO, 2001 Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nat. Genet.* **27**: 155–156.
- KUMAR, A., M. MARKANDAYA and S. C. GIRIMAJI, 2002 Primary micro-

- cephaly: microcephalin and ASPM determine the size of the human brain. *J. Biosci.* **27**: 629–632.
- KUMAR, S., K. TAMURA, I. B. JAKOBSEN and M. NEI, 2001 MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.
- LI, W. H., 1997 *Molecular Evolution*. Sinauer, Sunderland, MA.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- McHENRY, H. M., 1994 Tempo and mode in human evolution. *Proc. Natl. Acad. Sci. USA* **91**: 6780–6786.
- MOCHIDA, G. H., and C. A. WALSH, 2001 Molecular genetics of human microcephaly. *Curr. Opin. Neurol.* **14**: 151–156.
- MURPHY, W. J., E. EIZIRIK, S. J. O'BRIEN, O. MADSEN, M. SCALLY *et al.*, 2001 Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**: 2348–2351.
- NEI, M., and S. KUMAR, 2000 *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- PTAK, S. E., and M. PRZEWORSKI, 2002 Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet.* **18**: 559–563.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SATTA, Y., J. KLEIN and N. TAKAHATA, 2000 DNA archives and our nearest relative: the trichotomy problem revisited. *Mol. Phylogenet. Evol.* **14**: 259–275.
- SMITH, N. G., and A. EYRE-WALKER, 2002 Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAKAHATA, N., Y. SATTA and J. KLEIN, 1995 Divergence time and population size in the lineage leading to modern humans. *Theor. Popul. Biol.* **48**: 198–221.
- WAKEFIELD, J. G., S. BONACCORSI and M. GATTI, 2001 The drosophila protein *asp* is involved in microtubule organization during spindle formation and cytokinesis. *J. Cell Biol.* **153**: 637–648.
- WOOD, B., and M. COLLARD, 1999 The human genus. *Science* **284**: 65–71.
- YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- YI, S., D. L. ELLSWORTH and W. H. LI, 2002 Slow molecular clocks in Old World monkeys, apes, and humans. *Mol. Biol. Evol.* **19**: 2191–2198.
- ZHANG, J., and M. NEI, 1997 Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.* **44** (Suppl. 1): S139–S146.
- ZHANG, J., and D. M. WEBB, 2003 Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc. Natl. Acad. Sci. USA* **100**: 8337–8341.
- ZHANG, J., S. KUMAR and M. NEI, 1997 Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Mol. Biol. Evol.* **14**: 1335–1338.
- ZHANG, J., H. F. ROSENBERG and M. NEI, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**: 3708–3713.
- ZHANG, J., D. M. WEBB and O. PODLAHA, 2002 Accelerated protein evolution and origins of human-specific features: *Foxp2* as an example. *Genetics* **162**: 1825–1835.

Communicating editor: S. YOKOYAMA