Human Molecular Genetics, 2004, Vol. 13, No. 11 ***1131–1137***
DOI: 10.1093/hmg/ddh127
Advance Access published on March 31, 2004

Molecular evolution of *microcephalin*, a gene determining human brain size

Yin-qiu Wang[1] and Bing Su[1,2,*]

[1]Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, The Chinese Academy of Sciences (CAS), Kunming, China and [2]Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati, OH, USA

Received February 4, 2004; Revised and Accepted March 25, 2004 DDBJ/EMBL/GenBank accession nos[†]

Microcephalin **gene is one of the major players in regulating human brain development. It was reported that truncated mutations in this gene can cause primary microcephaly in humans with a brain size comparable with that of early hominids. We studied the molecular evolution of *microcephalin* by sequencing the coding region of *microcephalin* gene in humans and 12 representative non-human primate species covering great apes, lesser apes, Old World monkeys and New World monkeys. Our results showed that *microcephalin* is highly polymorphic in human populations. We observed 22 substitutions in the coding region of *microcephalin* gene in human populations, with 15 of them causing amino acid changes. The neutrality tests and phylogenetic analysis indicated that the rich sequence variations of *microcephalin* in humans are likely caused by the combination of recent population expansion and Darwinian positive selection. The synonymous/non-synonymous analyses in primates revealed positive selection on *microcephalin* during the origin of the last common ancestor of humans and great apes, which coincides with the drastic brain enlargement from lesser apes to great apes. The codon-based neutrality test also indicated the signal of positive selection on five individual amino acid sites of *microcephalin*, which may contribute to brain enlargement during primate evolution and human origin.**

INTRODUCTION

The enlarged brain and highly developed cognitive skills are the most significant characteristics that set us apart from our relatives, the non-human primates. The brains of modern humans are more than 20 times larger than those of Old World monkeys, and more than four times larger than those of great apes (1). This evolutionary expansion is believed to be crucial to the highly developed cognitive abilities in humans, yet its genetic basis remains unsolved. Recent studies on autosomal recessive primary microcephaly (MCPH) have revealed genes responsible for human brain development (2,3). MCPH is a rare human brain disorder, which is defined as a head circumference more than 3 SD below the age and sex corrected mean, and with the absence of other syndromic features or significant neurological deficits (2–4). The brain size of MCPH patients is comparable with that of early hominids (5), an interesting evolutionary parallel leading to the proposal of an atavistic disorder for MCPH (6). Studies on worldwide microcephaly families revealed genetic heterogeneity of MCPH, and six loci have been mapped onto six different human chromosomes (7,8). Recently, two genes (*microcephalin* and *ASPM*) were identified to be responsible for MCPH, and the mutations observed in the affected individuals resulted in truncated gene products (2,3).

It has become common wisdom that the large brain of humans is the consequence of adaptive evolution, i.e. positive Darwinian selection (9). Because the cost of maintaining a large brain is so great, it is intrinsically unlikely that large brains will evolve merely because they can. Large brains will evolve only when the selection factor in their favor is sufficient to overcome the steep cost gradient (9). According to the so-called 'Social Brain Hypothesis', the ever-increasing group size and complication of social structure among primates are the driving forces leading to brain enlargement during primate evolution, especially during the origin of our own species (9). Hence, genes involved in human brain

*To whom correspondence should be addressed at: Center for Genome Information, University of Cincinnati, 3223 Eden Avenue, Cincinnati, OH 45267, USA. Tel: +1 5135586678; Fax: +1 5135584397; Email: bing.su@uc.edu
[†]AY307773–AY307854, AY505571–AY506356 and AY533018.

development are the potential targets of positive selection for functionally more sophisticated brains. The molecular signature of positive selection can be identified through comparative sequence analysis between humans and non-human primates. At the molecular level, when the positive selection of a gene is strong enough to rise above the background of neutral evolution or overcome other forms of selection, the rate of non-synonymous nucleotide substitution (K_a) may exceed that of synonymous substitution (K_s), i.e. $K_a > K_s$ or $K_a/K_s > 1$ (10). The recent study on *FOXP2* gene is a good example of how positive selection leads to sequence modification of a gene responsible for human speech and language ability (11,12).

Recent reports have suggested that the human *ASPM* (abnormal spindle-like microcephaly associated) gene went through an episode of accelerated sequence evolution due to positive selection in the human lineage (12,13). However, the evolutionary history of the *microcephalin* gene and its role in brain enlargement during human evolution remains unknown. This gene is located on human chromosome 8p23, encoding a BRCT (BRCA1 C-terminal) domain containing protein. It has 14 exons, spanning a length of 241 kb in the human genome, and encodes a protein of 834 amino acids (Fig. 1). The mutation reported in microcephaly patients created a premature stop codon leading to a truncated product of *microcephalin* (3). Expression study of human fetal tissues showed that it is expressed in fetal brain and other fetal tissues (3). *In situ* hybridization analysis on fetal mouse confirmed high levels of expression in the developing forebrain (3). There are three BRCT domains in *microcephalin*, which are engaged in DNA–protein and protein–protein interactions (3). As BRCT domains usually exist in proteins controlling cell cycles and DNA damage repairing (14), the function of *microcephalin* was suggested to be involved in cell cycle and apoptosis regulation during neurogenesis (3).

In this study, we sequenced the coding region of *microcephalin* gene in human populations and 12 non-human primate species in order to understand its contribution to brain enlargement during the course of human evolution, and to test whether selection acted on shaping up the mutation pattern of this gene in human populations.

RESULTS AND DISCUSSION

A total of 7828 bp was sequenced in 72 samples from humans and non-human primates, covering exons 2–14 (2507 bp) and the flanking intron segments of the *microcephalin* gene (see Supplementary Material for aligned DNA sequences). In the human populations, we observed 62 polymorphic sites, with 22 of them located in the 2507 bp coding region. The gene diversity (π) of the entire region sequenced is 0.00156 (Table 1), which falls into the commonly observed range of genetic variations in human populations (15). However, among the 22 polymorphic sites in the coding region, 15 of them are non-synonymous (amino acid changing) substitutions, indicating a high level of protein polymorphism of *microcephalin* in human populations.

When the human sequences were aligned with those of non-human primate species, a total of 1172 variant sites were observed (including those in human populations), with 315 of them located in the coding regions. There are 179 amino acid variant sites (21.5%, 179/834) in all the species tested, an indication of rapid amino acid substitutions of *microcephalin* during primate evolution. For example, the average non-synonymous substitution rate (K_a) between human and rhesus monkey is 0.05, which is higher than most of the genes reported (16). Among the 30 genes analyzed by Wyckoff *et al.* (16) only four genes have K_a values (human versus rhesus monkey) greater than 0.05. For the 15 amino acid variant sites observed in human populations, 12 of them are only polymorphic in humans while they are conserved in all the non-human primates species tested. This pattern was unexpected considering the functional role of *microcephalin* in human brain development (3). The gene diversities (π and θ_w) of humans and chimpanzees were calculated and listed in Table 1. The gene diversity (π) of humans is slightly larger than that of chimpanzees when both the coding and non-coding regions were considered, and the difference became more prominent when only the coding region was compared, where the diversity of humans is more than two times higher than that of the chimpanzees (0.00189 versus 0.00081) (Table 1). This observation is contradictory to the commonly accepted notion that chimpanzees generally show about four times higher diversity than humans because of their larger effective population size and longer evolutionary time (17). There are several possible driving forces that could cause the unusual high level of sequence variations in human populations, e.g. recent population expansion under neutral evolution, relaxation of negative selection and positive selection (18).

We conducted neutrality tests in human and chimpanzee populations (Table 1). Among the different methods employed, the Fu's *Fs*-test (19) based on haplotype diversity showed deviation from neutral expectation in human populations ($P < 0.001$) (Table 1). It is also significant for the chimpanzee population when only the coding region was considered ($P = 0.021$). Considering the influence of recombination on haplotype diversity in large genes like *microcephalin* (241 kb), we repeated the *Fs*-test by analyzing different domains of the gene separately. The Fu's *Fs*-test was still significant ($P < 0.05$) in human populations for the BRCT and structure domains of *microcephalin* (Table 2), but not significant in chimpanzees (data not shown). The significance of Fu's *Fs*-test in humans can be explained by either a recent population expansion or genetic hitch-hiking (positive selection) (19). The recent human population expansion has been documented using varied genetic markers (20,21), and it is likely that it would contribute to the excess of recent mutations of *microcephalin* in humans. However, when we looked into the details of the sequence variations in human populations, population expansion does not seem to be the only source causing the observed variation pattern. Among the 15 amino acid polymorphic sites observed in human populations, 11 of them are rare mutations with allele frequency lower than 10%, which is concordant with the expectation of recent population expansion (19). But at the other four sites, the mutations are prevalent in humans with frequencies

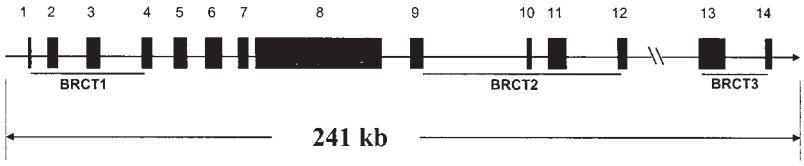

Figure 1. The schematic map of the *microcephalin* gene.

Table 1. Summary of sequence variations of *microcephalin* gene in human and chimpanzee populations

	Human			Chimpanzee		
	Total	Coding	Non-coding	Total	Coding	Non-coding
Sample size	82	82	82	18	18	18
Haplotypes	51	35	43	12	9	12
Length	7131	2483	4648	7236	2483	4753
Polymorphic sites	62	22	40	37	10	27
Amino acid polymorphic sites	15	15	\	6	6	\
Singletons	18	5	13	17	7	10
π	0.00156	0.00189	0.00139	0.00124	0.00081	0.00147
θ_w	0.00178	0.00178	0.00177	0.00149	0.00117	0.00166
Tajima's *D*	−0.391	0.188	−0.691	−0.679	−1.124	−0.452
Fu and Li's *D*	−1.063	−0.208	−1.390	−0.821	−1.799	−0.330
Fu and Li's *F*	−0.951	−0.073	−1.338	−0.905	−1.858	−0.423
Fay and Wu's *H*	−6.557	−2.576	−3.981	−5.804	−1.412	−4.392
Fu's *Fs*	−24.749*	−20.072*	−25.858*	−1.077	−3.697**	−1.961

*$P < 0.001$, **$P = 0.021$.
The sample size refers to the number of chromosomes sequenced. The human samples include 11 Africans, 10 Europeans and 20 Chinese.

ranging from 19.5 to 72% (sites 306, 681, 760 and 827; Table 3). For example, at site 827, the derived allele became the dominant one (72%) in human populations while the ancestral allele is well-conserved in the entire non-human primate species tested (Table 3 and Fig. 2), which cannot be explained solely by recent population expansion under neutral evolution (random genetic drift). The Fay and Wu's *H*-test (22) was developed to detect excess of high-frequency alleles due to positive selection, and it was marginally significant for humans ($P = 0.068$) (27), but not significant for chimpanzees ($P = 0.153$). Therefore, it is likely that both recent population expansion and positive selection contributed to the sequence variations of *microcephalin* in current human populations though the signal of positive selection is weak (marginally significant) due to the influence of recent population expansion. In addition, compared with non-human primates, there are seven fixed amino acid substitutions that are specific to humans (Fig. 2), and six of them are conserved in apes and Old World monkeys. The functional consequence of the human-specific amino acid substitutions is yet to be understood.

Based on the well-established phylogeny of primates (23), we calculated the K_a/K_s ratios of different evolutionary lineages following the Pamilo–Bianchi–Li's method (24,25) (Fig. 3). The one-tailed *Z*-tests showed that most of the lineages follow the expectations of either neutral evolution ($K_a/K_s = 1$) or negative selection ($K_a/K_s < 1$). The K_a/K_s ratio of the human lineage is 0.51, which is not significantly deviated from neutral expectation (Fig. 3). This is likely caused by recently accumulated synonymous mutations in humans due to population expansion.

In addition, the K_a/K_s ratio between nodes D and C (the common ancestor of humans and great apes) is exceptionally large compared with the other evolutionary lineages and the *Z*-test is marginally significant ($K_a/K_s = 3.18$, $P = 0.094$) (Fig. 3). This observation indicated possible adaptive sequence changes of *microcephalin* during the origin of the common ancestor of humans and great apes that occurred 14 million years ago (23), which coincides with the drastic brain enlargement during primate evolution from lesser apes to great apes (1).

Interestingly, we observed two deletions in exon 8 resulting in the absence of four (EESS) and three (PQQ) amino acids. These two deletions are shared by subtribe Hominina (human, chimpanzee and gorilla) and subfamily Homininae (human and apes) (23), respectively (Fig. 2). We sequenced exon 8 of one New World monkey species (common woolly monkey) and no deletions were observed, confirming therefore, that they are indeed deletions that occurred in apes and humans, not insertions obtained by the Old World monkeys. According to the species divergence proposed by Goodman *et al.* (23), these two deletions originated about 7 and 18 million years ago, respectively. As the amino acid sequences of these two segments are conserved in all the New World and Old World monkey species tested, they might have functional consequence in primate brain evolution.

Table 2. Gene diversity and neutrality tests of gene domains in human populations

	Exon 8	BRCT1	BRCT2	BRCT3
Sample size	92	82	82	82
Haplotypes	17	16	13	12
Length	1701	957	1292	594
Polymorphic sites	17	17	8	8
Singletons	5	6	2	4
π	0.00164	0.00359	0.00083	0.00275
θ_w	0.00196	0.00378	0.00124	0.00271
Tajima's *D*	−0.474	−0.146	−0.828	0.038
Fu and Li's *D*	−0.795	−1.528	−0.292	−1.855
Fu and Li's *F*	−0.808	−1.222	−0.558	−1.439
Fu's *Fs*	−4.227*	−2.430*	−7.682*	−3.646*

*$P < 0.05$.
All the 46 human samples were sequenced for exon 8. For BRCT1–3, 41 samples (11 Africans, 10 Europeans and 20 Chinese) were sequenced.

Table 3. The distribution of amino acid variations of *microcephalin* in human populations. The ancestral alleles of the variant sites in humans were determined by comparing with the alleles of the non-human primate species

Amino acid polymorphic site	Allele frequency			
	Asian ($n = 40$)	African ($n = 22$)	European ($n = 20$)	Total ($n = 82$)
Asn^{094}	0.025	0	0	0.012
Arg^{163}	0	0.273	0.100	0.098
Thr^{204}	0	0.136	0	0.037
Val^{256}	0	0.091	0	0.024
His^{280}	0	0.182	0	0.049
Ile^{296}	0.025	0.182	0.100	0.085
His^{306}	0.800	0.136	0.850	0.634
Cys^{323}	0	0	0.050	0.012
Pro^{397}	0.075	0	0	0.037
Met^{491}	0	0.045	0	0.012
Val^{586}	0.025	0	0	0.012
Asn^{681}	0.225	0.227	0.100	0.195
Ala^{760}	0.225	0.591	0.650	0.427
His^{790}	0	0.045	0	0.012
Pro^{827}	0.775	0.591	0.750	0.720

We conducted a codon-based neutrality test developed by Yang *et al.* (26) to detect positive selection at individual amino acid sites of *microcephalin.* The result showed that the K_a/K_s values of five sites, 218Y ($P = 0.008$), 424E ($P = 0.023$), 485R ($P = 0.013$), 760A ($P = 0.001$) and 827P ($P < 0.001$), were deviated from neutral expectations (Fig. 2). The 424E, 218Y and 485R are mutations shared by humans, chimpanzees and gorillas, confirming the proposed positive selection of the common ancestor of these species in the phylogenetic analysis. The 760A and 827P are located within and next to the BRCT3 domains. The deviation from neutrality at these two sites supports the influence of positive selection on human populations (Fig. 2). We also used the parsimony-based method developed by Suzuki and Gojobori (27), and none of the five sites showed significant deviation from neutral expectation (data not shown). The parsimony-based method is more conservative than the likelihood method developed by Yang *et al.* (26–28). Hence, functional studies are needed to test the biological effects of these amino acid variations.

In summary, our study showed that there are accelerated amino acid substitutions of *microcephalin* gene in primates, especially in human populations, possibly due to the combined influence of recent population expansion and positive selection. During the origin of the last common ancestor of humans and great apes, there were adaptive sequence changes in *microcephalin* paralleling the drastic brain enlargement from lesser apes to great apes.

MATERIALS AND METHODS

Samples

We sequenced 46 human individuals from the major continental populations, including Africans (14 individuals), Europeans (12 individuals) and Asians (20 Chinese individuals). To define the genetic root of *microcephalin*, we also sequenced 12 non-human primate species reflecting a 40 million-year history of primate evolution (23). The non-human primate panel include three great ape species (nine chimpanzees—*Pan troglodytes*, four gorillas—*Gorilla gorilla* and three orangutans—*Pango pygmaeus*), which diverged from humans about 4–14 million years ago, two lesser ape species (one white-browed gibbon—*Hylobates hoolock* and one white-cheeked gibbon—*Hylobates leucogenys*), which diverged from humans about 18 million years ago, six Old World monkey species (two rhesus monkeys—*Macaca mulatta*, one red guenon—*Erythrocebus patas*, one grey leaf monkey—*Trachypithecus phayrei*, one black leaf monkey—*Trachypithecus francoisi*, one douc langur—*Pygathrix nemaeus* and one Yunnan golden monkey—*Rhinopithecus bieti*), which diverged from humans about 25 million years ago and one New World monkey species (one common woolly monkey—*Lagothrix lagotricha*), which diverged from humans about 40 million years ago (23). The human sample IDs are Africans: AB1, AB2, AB4, AB6–8, AB11, AB13–15, AP86G, AP18G, AP115, AF19; Europeans: E150–154, E160–164, EBer5, EBer62; Chinese: KBC01, KHC01–12, KHC26–31, KMC01. All the DNA samples were from collections in Kunming Cell Bank of CAS, Kunming Blood Center and Shanghai National Genome Center in China.

PCR and sequencing

Thirteen fragments of *microcephalin* gene were amplified by PCR and sequenced in humans and 11 non-human primate species (excluding the common woolly monkey) covering exons 2–14 (Fig. 1). The exon 1 (22 bp) of *microcephalin* was not sequenced due to the difficulty of PCR amplification. For the common woolly monkey, only exon 8 was sequenced in order to determine the ancestry of two segment deletions observed in humans and apes. Universal primers for all the species were designed by comparing the published sequences

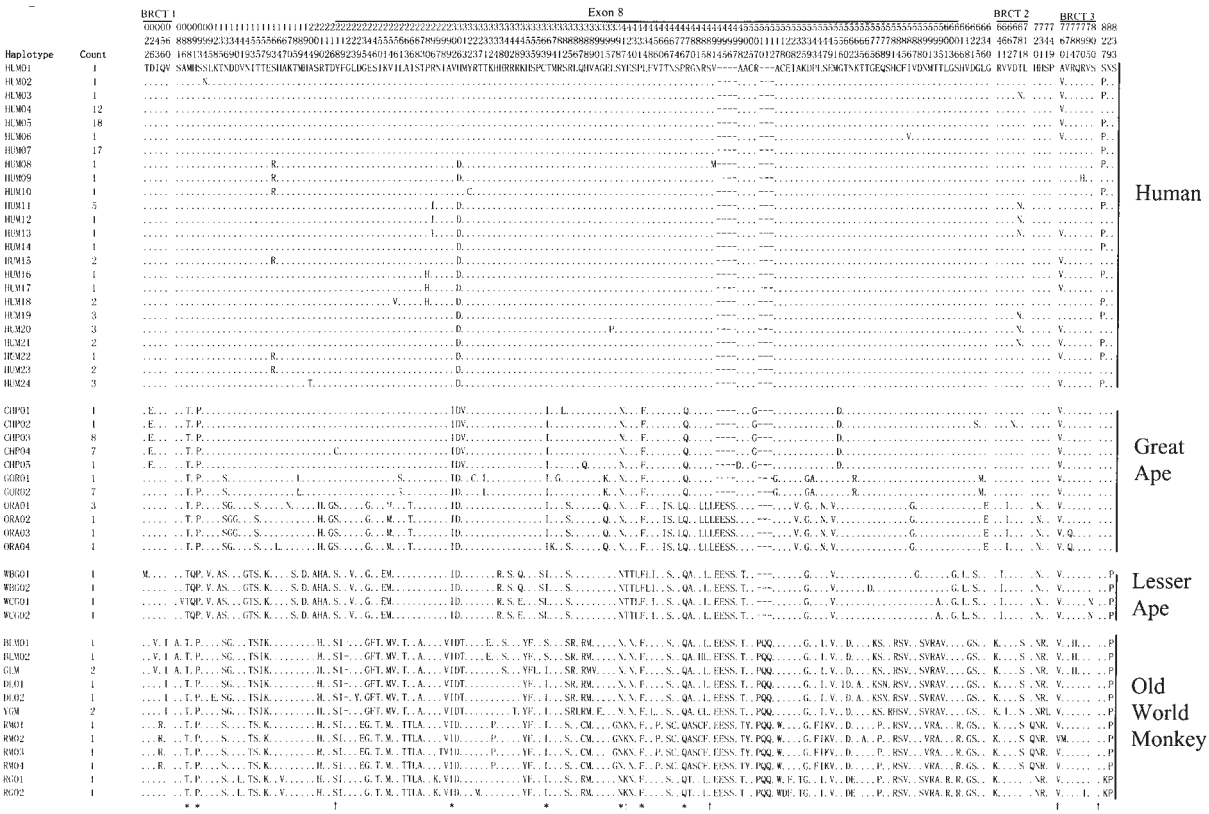

Figure 2. Alignment of the amino acid haplotypes of *microcephalin* in humans and non-human primates. The asterisks denote the sites fixed in human populations. The arrows denote the site under positive selection indicated by Yang's test (26). HUM—human, CHP—chimpanzee, GOR—gorilla, ORA—orangutan, WBG—white-browed gibbon, WCG—white-cheeked gibbon, BLM—black leaf monkey, GLM—grey leaf monkey, DL—douc langur, RM—rhesus monkey, YGM—Yunnan golden monkey, RG—red guenon.

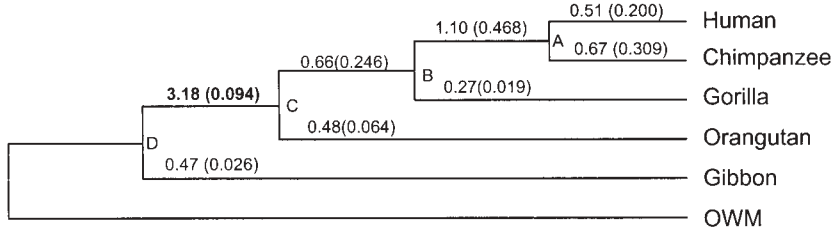

Figure 3. The K_a/K_s ratios of different evolutionary lineages in primates. Only the coding sequences were used for K_a/K_s calculation with Pamilo–Bianchi–Li's method (24,25). The one-tailed *Z*-test was used for detecting deviation of the K_a/K_s ratios from neutral expectation ($K_a/K_s = 1$) (30), and the *P*-values were shown in parentheses. The labels A–D are inferred internal nodes, and OWM stands for Old World monkey.

of human and mouse (3). The primer sequences are listed in Table 4. Sequencing were performed in both directions with forward and reverse primers using the BigDye Terminator sequencing kit on an ABI 3100 automated sequencer.

Data analysis

Sequence data was edited and aligned using DNASTAR (DNASTAR, Inc.) and manually inspected and confirmed. The DNA haplotypes of humans and great apes whose sequences have more than two heterozygous mutations were inferred using PHASE (29). The haplotypes of other non-human primates were determined by randomly separating the sequences of the two chromosomal copies of the gene when more than two heterozygous sites were observed. For phylogenetic analysis, the neighbor joining (NJ) tree was constructed using MEGA2 (14). The synonymous and non-synonymous substitution rates (K_s and K_a) were calculated based on Pamilo–Bianchi–Li's method, in which the transitional/transversional substitution bias was taken into account (24,25). The one-tailed *Z*-test was used to detect deviation of the K_a/K_s ratios from neutrality (30). The DnaSP3.99

Table 4. The PCR and sequencing primer sequences in this study

Name	Sequence (5′–3′)	Name	Sequence (5′–3′)
E2F	TGCTCGGCTAATTTTTTGTATC	E8F	GAAAGTTGAATATAGAATAATTTAAACC
E2R	TTGAGTTCTGAAAGTCCTACTGTTC	E8R	CATGTTTATTTAGCGAAGGTGACT
E2Fn	TCTCAAACCCCTGACTTCGTG	E8-3	TTTATTTAGCGAAGGTGACTTG
E2Rn	CTTACCCATTGCTTCGTCCAG	E8-4	GCAGCCTCGGCATGATAG
E2Fn2	GCAGGGGATGCTGGAATTTCA	E8-5	ACAAGCATATTTGAAATGTCTG
E2Rn2	TTCACACTTTGGCTCTTTTTTCTCC	E8-6	AGAGAAGTATCGTTTGTCTCCTAC
E3F	GCAGGGCTCCAACTCTATTCTT	E8-7	GGTCCTCAATGGTGTAAGAA
E3R	TCACATTCCCTCTTGCTTCCTA	E9F	AGTTTGACGAGTGCGTAAGGTAT
E3Fn	TAACTGGAACAGATATGTTTTAAGC	E9R	GGTTTATGTTTCATTGACCCAGA
E4F	ATTGTACTTCCTTTACCTGCCTCT	E10F	AAAAGTGGCTGACGAGTCTACAA
E4R	GCACCTAAATGTTTGCTTGTCTC	E10R	TCAAGGCTAAGTCGCTCACATT
E4Fn	TCAATTTCTTTACGTCCCTTCTACC	E10Fn	GTTGACACACTTGGAGGTCTGC
E4Rn	CTTGTCTCTTGGCACCTTTTAGTC	E11F	GACAGCCTGTGGGCTTGTCTT
E5F	CCGTGGTTACATGCTTGGTTC	E11R	CTTGCGTTTCTTTCAGCATCTTT
E5R	CGTGGCTCTATTATGGCTCCC	E11Rn	CTCTCTACCCTCCCCGTCCTC
E6F	GATTTGGGTTTGGATTTGAGGT	E12F	TCTGGGAAGGAATTTCAAGTGT
E6R	AAACAAAGCCAGCCATGAAATA	E12R	GGGCATAATTGTGCTTGACTG
E6Fn	GGCAGTTGCTCAGCAGATGTG	E12Fn	GCAACATGAAGATTCTGAAGGGAC
E6Rn	AGGTTGCAGTGAGCCCAGATC	E12Rn	GCAAGATATAAGACTACCTGCCCAAG
E6Fn2	GGACCATGCTGATCCCACTTC	E13F	AGCATGGCAGCCTTACATTCAG
E6Rn2	TATCAGGAAACATTAAAATATCTAGGC	E13R	TAACAGCAAGGAGCGGGAATG
E7F	AGGACTTCCTGCTGGCTTCAT	E14F	CCCTGCATCTAATGGGACATG
E7R	ACTAGGGCTCCCACATTACAAAC	E14R	GTGGCTTGTCTCCAGCATGAG
E7Fn	GGTAACTATGATGAGCATCCTTGAAG	E14R1	TTCTCATTTGAACATCCAAAGACAG
E7Rn	GTATGTGCGTAACATGGAACTAACTG	E14Fn	AACAGGTTATCACTTTCCTGTGTGG
E7Rn2	ATTTCCTTAGCTCTGACCTCAGTTC	E14Rn	GGTGGAGGTTGCAGTGAGCC
I7F	GCAGCAGCTTAGCATTACTCTCAG	E14Rn2	TCAGAAGTCATCTTGTATACGCCAG
I7R	ATCCTGAGTTTCCACAAAGATCATC		

program was used to calculate gene diversity and conduct neutrality tests, including π, θ_w, Tajima's *D*-test, Fu and Li's *D*-test, Fu and Li's *F*-test (31–36). The Fu's *Fs*-test and Fay and Wu's *H*-test were performed using the Neutrality Test Program (37,38). These tests were developed to detect whether the pattern of diversity in a population are consistent with the hypothesis of neutrality. If the statistics of the neutrality test is significantly deviated from the neutral expectation, we assume that selection and/or population history (e.g. recent expansion) are responsible for the observed diversity pattern (31–38). Using PAML (36), the ancestral sequences of the internal nodes of the phylogenetic tree were inferred and the substitution patterns were compared among different evolutionary lineages. Based on Yang's method (26), the individual codon-based substitution model was used to test positive selection at individual amino acid sites and the model used is M3 ($K = 3$) (26). The parsimony-based method for detecting positive selection at single amino acid sites was also conducted (28).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We thank Dr Fengtang Yang and Dr Wei Huang for generously providing DNA samples. We also thank Xiao-na Fan for technical assistance in this study. We are grateful to Dr Li Jin who critically reviewed this paper. This project was supported by grants from the Chinese Academy of Science and National Natural Science Foundation of China (grant no. KSCX2-SW-121).

REFERENCES

1. Parker, S.T. and McKinney, M.L., (1999) *Origin of Intelligence.* The Johns Hopkins University Press, Baltimore, MD, pp. 313–345.
2. 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.
3. 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.
4. Ross, J.J. and Frias, J.L. (1977) Microcephaly. In Congenital malformations of the brain and skull, Part 1. In Vinken, P.J. and Bruyn, G.W. (eds), *Handbook of Clinical Neurology*. Elsevier Holland Biomedical Press, Amsterdam, Vol. 30, pp. 507–524.
5. Wood, B. and Collard, M. (1999) The human genus. *Science*, **284**, 65–71.
6. Mochida, G.H. and Walsh, C.A. (2001) Molecular genetics of human microcephaly. *Curr. Opin. Neurol.*, **14**, 151–156.
7. Roberts, E., Hampshire, D.J., Pattison, L., Springell, K., Jafri, H., Corry, P., Mannon, J., Rashid, Y., Crow, Y., Bond, J. and Woods, C.G. (2002) Autosomal recessive primary microcephaly, an analysis of locus heterogeneity and phenotypic variation. *J. Med. Genet.*, **39**, 718–721.
8. 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.
9. Dunbar, R.I.M. (1998) The social brain hypothesis. *Evol. Anthropol.*, **6**, 178–190.

10. Li, W.H. (1997) *Molecular Evolution.* Sinauer Associates Press, Sunderland, MA.
11. Enard, W.M., Przeworski, S.E., Fisher, C.S.L., Lai, V., Wiebe, T., Kitano, A.P., Monaco and Pääbo, S. (2002) Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature*, **418**, 869–872.
12. Zhang, J. (2003) Evolution of the human *ASPM* gene, a major determinant of brain size. *Genetics*, **165**, 2063–2070.
13. 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.
14. Trevor, H., Bates, P.A., Zhang, X., Sternberg, M.J. and Freemont, P.S. (2000) The BRCA1 C-terminal domain: structure and function. *Mut. Res.*, **460**, 319–332.
15. Reich, D.E., Schaffner, S.F., Daly, M.J., McVean, G., Mullikin, J.C., Higgins, J.M. Richter, D.J., Lander, E.S. and Altshuler, D. (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat. Genet.*, **32**, 135–142.
16. 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.
17. Kaessmann, H., Wiebe, V. and Pääbo, S. (1999) Extensive nuclear DNA sequence diversity among chimpanzees. *Science*, **286**, 1159–1162.
18. Li, W.H. (1997) *Molecular Evolution.* Sinauer Associates Press, Sunderland, MA, pp. 237–268.
19. Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.
20. Reich, D.E. and Goldstein, D.B. (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proc. Natl Acad. Sci. USA*, **95**, 8119–8123.
21. Shriver, M., Jin, L., Ferrell, R.E. and Deka, R. (1997) Microsatellite data support an early population expansion in Africa. *Genome Res.*, **7**, 586–591.
22. Fay J.C. and Wu C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.
23. Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J., Gunnell, G. and Groves, C.P. (1998) Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.*, **9**, 585–598.
24. Pamilo, P. and Bianchi, N.O. (1993) Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. *Mol. Biol. Evol.*, **10**, 271–281.
25. Li, W.H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.*, **36**, 96–99.
26. Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A.M.K. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431–449.
27. Suzuki, Y. and Gojobori, T. (1999) A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.*, **16**, 1315–1328.
28. Suzuki, Y. and Nei, M. (2002) Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.*, **19**, 1865–1869.
29. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.
30. Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001) *MEGA2, molecular evolutionary genetics analysis software*. Arizona State University, Tempe, AZ.
31. Rozas, J. and Rozas, R. (1999) *DnaSP* version 3, an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.
32. Watterson, G.A. (1975) On the number of segregating sites in genetic models without recombination. *Theor. Pop. Biol.*, **7**, 256–276.
33. Hudson, R.R., Kreitman, M. and Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.
34. Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
35. Fu, Y.X. and Li, W.H. (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.
36. Schneider, S., Roessli, D. and Excoffie, L. (2000) *Arlequin*, *version 2.000, a software for population genetics data analysis*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.
37. Li, H.P. and Fu, Y.X. (2004) Neutrality Test, a novel software for performing tests of neutrality. *Bioinformatics*, in press.
38. Yang, Z. (1997) PAML, a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.*, **15**, 555–556.