Recent Origin of a Hominoid-Specific Splice Form of Neuropsin, a Gene Involved in Learning and Memory

Yi Li,‡§ Ya-ping Qian,† Xiao-jing Yu,* Yin-qiu Wang,* Ding-gui Dong,‖ Wei Sun,‖ Run-mei Ma,¶ and Bing Su*†*

*Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming, Yunnan, China; †Center for Genome Information, Department of Environmental Health, University of Cincinnati; ‡Department of Chemistry, Qujing Teachers University, Qujing, Yunnan, China; §Graduate School of the Chinese Academy of Sciences, Beijing, China; ‖The First People's Hospital of Qujing, Qujing, Yunnan, China; and ¶The First Affiliated Hospital of Kunming Medical College, Kunming, Yunnan, China

Neuropsin is a secreted-type serine protease involved in learning and memory. The type II splice form of neuropsin is abundantly expressed in the human brain but not in the mouse brain. We sequenced the type II–spliced region of neuropsin gene in humans and representative nonhuman primate species. Our comparative sequence analysis showed that only the hominoid species (humans and apes) have the intact open reading frame of the type II splice form, indicating that the type II neuropsin originated recently in the primate lineage about 18 MYA. Expression analysis using RT-PCR detected abundant expression of the type II form in the frontal lobe of the adult human brain, but no expression was detected in the brains of lesser apes and Old World monkeys, indicating that the type II form of neuropsin only became functional in recent time, and it might contribute to the progressive change of cognitive abilities during primate evolution.

Introduction

Neuropsin is one of the secreted-type serine proteases that are considered essential in many aspects of neuronal activities (also called kallikrein 8, or KLK8 [Mitsui et al. 1999]). Mouse study showed that it is mainly expressed in the hippocampal pyramidal neurons and involved in hippocampal plasticity and pathogenesis that underlie some aspects of declarative memory encoding and recall (Chen et al. 1995; Suzuki et al. 1995; Komai et al. 2000; Davies et al. 2001). Ontogeny study on mouse also implicated its involvement in embryonic cell differentiation and neurite outgrowth and fasciculation (Hirata et al. 2001; Oka et al. 2002). Human study reported two types of human neuropsin cDNAs in the brain with one of them (type I) homologous to the mouse counterpart but the other (type II) only expressed in humans (Mitsui et al. 1999). There are six exons in the neuropsin gene, and the first exon is nontranslational. The alternative splicing site in humans is located in the intron region between exon 2 and exon 3 (between exon 1 and 2 in Mitsui et al. [1999]). The type I and type II neuropsin encode proteins with 260 and 305 amino acids, respectively. The type II neuropsin is identical with type I in amino acid sequences except that it carries extra 45 amino acids at the N-terminus of the protein because of alternative splicing (Mitsui et al. 1999). Expression analysis in humans showed that both type I and type II are expressed at similar level in multiple fetal tissues, including brains. However, in human adults, the type II form becomes the preferentially expressed form (major form) in the cerebral cortex and hippocampus, whereas the expression of type I decreases dramatically, implying the importance of the type II form to the normal function of the adult human brain (Mitsui et al. 1999).

Key words: neuropsin, hominoid, alternative splicing, cognition, molecular evolution.

E-mail: sub@mail.kiz.ac.cn.

Mol. Biol. Evol. 21(11):2111–2115. 2004
doi:10.1093/molbev/msh220
Advance Access publication July 28, 2004

It is widely accepted that human brain is the product of adaptive evolution; that is, Darwinian positive selection that favors larger and functionally more sophisticated brains during primate evolution. The large size and superior cognitive ability of the human brain is one of the most characteristic phenotypes that set us apart from our close relatives, the nonhuman primates. It is believed that new genes may emerge during the speciation process, leading to new phenotypes and/or new functions during evolution. Alternative splicing is one of the known mechanisms in creating new genes (Long et al. 2003). This gene-creating strategy is frequently used in the brain, where neurons are found to be rich in regulated alternative splicing events (Grabowski 1998). Hence, identifying species-specific or lineage-specific alternative splice forms will shed light on the mechanism of functional evolution of the human brain.

To trace the origin of the type II splice form of neuropsin, we conducted a comparative sequence analysis in humans and nonhuman primate species. Our results showed that the origin of the type II neuropsin is a relatively recent event that can be traced back to the ancestor of the hominoid species (humans and apes) about 18 MYA. The abundant expression of the type II form was observed in the human brain, but no expression was detected in the lesser ape and Old World monkey species, implying that the functionalization of the type II form in the brain may occur more recently during primate evolution.

The GenBank accession numbers of DNA sequences reported herein are AY563055 to AY563092.

Material and Methods

Samples

For comparative sequence analysis, we sampled 11 human individuals (Han Chinese) and 13 nonhuman primate species, including great apes (nine chimpanzees, *Pan troglodytes*; four gorillas, *Gorilla gorilla*; and three orangutans, *Pongo pygmaeus*), lesser apes (one white-cheeked

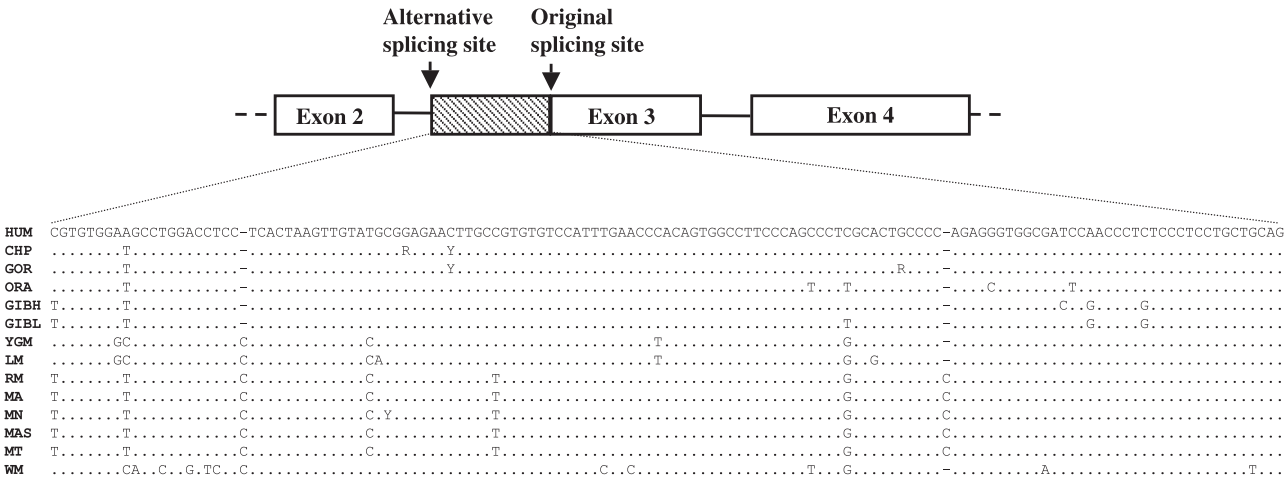

FIG. 1.—The structure of the neuropsin gene and the sequence alignment of the type II–spliced region in humans and nonhuman primate species tested. Hum = human, CHP = chimpanzee, GOR = gorilla, ORA = orangutan, GIBH = white-browed gibbon, GIBL = white-cheeked gibbon, YGM = Yunnan golden monkey, LM = Douc langurs, RM = rhesus monkey, MA = stump-tailed macaque, MN = pigtailed macaque, MAS = Assamese macaque, MT = Tibetan macaque, WM = white-shouldered marmoset. The highlighted sites indicate the indels. The "Y" and "R" indicate the within species polymorphisms, C/T and A/G, respectively.

gibbon, *Hylobates leucogenys* and one white-browed gibbon, *Hylobates hoolock*), Old World monkeys (one rhesus monkey, *Macaca mulatta*; one Assamese macaque, *Macaca assamensis*; one Tibetan macaque, *Macaca thibetana*; one pigtailed macaque, *Macaca nemestrina*; one stump-tailed macaque, *Macaca arctoides*; one Douc langurs, *Pygathrix nemaeus*; and one Yunnan snub-nosed monkey, *Rhinopithecus bieti*), and New World monkey (one white-shouldered marmoset, *Callitrix humeralifer*). The DNA samples were from the collections in Kunming Cell Bank of CAS, Kunming Blood Center, Coriell Cell Repositories (http://coriell.umdnj.edu), and Stanford University.

For gene expression analysis, we sampled brain tissues from humans, white-browed gibbons, rhesus monkeys, and Yunnan golden monkeys. They were from the collections of Kunming Cell Bank and local hospitals. The rhesus monkey was anesthetized before sacrifice. The human protocol was approved by the local hospitals. All the brain samples were immediately stored in liquid nitrogen after sampling. The RNA extraction tests indicated that none of the samples showed visible RNA degradation (data not shown).

Sequencing and RT-PCR

For PCR and sequencing, two sets of primers were designed to cover exon 3 and its flanking introns where the alternative splicing site locates. The primer sequences are as follows:

NP5S 5′-CCTCTCCATTTCCCAGACCTC-3′
NP5A 5′-CCACTCACGGTTTTTTACAGTG-3′
NP6S 5′-CTGTGGCGGTGTCCTTGTAG-3′
NP6A 5′-CCTGAGACGGTGCACTTCTG-3′)

The ABI-3100 automatic sequencer was used for sequencing.

The total RNAs were extracted from brain tissues using TRIzol (Invitrogen). The RNA samples were treated with DNase (Invitrogen) to remove possible genomic DNA contamination. The one-step RT-PCR kit (Qiagen) was used in detecting the expression of neuropsin following the standard protocol. The primers used for RT-PCR were the modified version based on Mitsui et al. (1999).

A type II–specific primer was designed to detect expression of this form in the brain. The primer is located in the type II–spliced region (5′-GTTGTATGCGGAGAACTTGCC-3′). To confirm the expression of different splice forms of neuropsin, the RT-PCR products were cloned into the pMD18-T vector (Takala, Inc.) and sequenced.

Data Analysis

The sequence data were edited and aligned by DNASTAR (DNASTAR, Inc.) and manually inspected and confirmed. The MEGA version 2.0 program (Kumar et al. 2001) was used to calculate the rates of synonymous (*Ks*) and nonsynonymous (*Ka*) substitutions of the coding region following Nei and Gojobori (1986), and the Pamilo-Bianchi-Li method (Li 1993; Pamilo and Bianchi 1993). For phylogenetic analysis, the ANC-GENE program developed by Zhang, Rosenberg, and Nei (1998) was used to infer ancestral sequences. To test the uniformity of mutation rates in different primate lineages, the relative rate test developed by Tajima (1993) was conducted.

Results

We sequenced exon 3 and the flanking introns of neuropsin gene in humans and 13 nonhuman primate species, including three great ape species (chimpanzee, gorilla, and orangutan), two lesser ape species, seven Old World monkey species, and one New World monkey species (see *Materials and Methods* for detailed sample description). When the sequences were aligned, an extra C at site 22 was observed in all the Old World and New World monkey species tested but not in the apes and humans (fig. 1). This extra C locates at the upstream of the

type II–spliced region (site 383 to 519 [see Supplementary Material online]) and would cause a frame shift leading to truncated proteins (premature stop codon) if the type II splice form existed in the Old World and New World monkey species. There is another C insertion at site 100 in the rhesus monkey, and this insertion is conserved in the five macaque species tested (fig. 1). Hence, for Old World and New World monkey species, the type II–spliced region is more likely a part of the intron sequence instead of coding sequence. In other words, the type II neuropsin is unlikely to generate functional proteins in Old World and New World monkeys. Within the type II–spliced region, compared with humans, there are no indels observed in the ape species; therefore, normal translations would be expected. Because the extra C at site 22 is conserved in both Old World and New World monkeys, a C deletion must have occurred in the ancestor of the hominoids (human and apes), which resulted in an intact open reading frame for the type II splice form. The marked difference between the hominoid and monkey species in the type II–spliced region suggested that the emergence of the type II neuropsin probably occurred in the ancestor of the hominoids about 18 to 20 MYA (Goodman et al. 1998).

To test whether the type II splice form are in fact expressed in the hominoid species, we analyzed the expression of neuropsin in human, lesser ape (white-browed gibbon,) and two Old World monkey species (rhesus monkey and Yunnan snub-nosed monkey). The frontal lobe and the hippocampus regions of the brains were analyzed in the three nonhuman primate species, and the fetal and adult human brain samples (frontal lobe) were used as control. The RT-PCR result showed that the type I form was expressed in all the samples of all the species tested, but the type II splice form was only visible in the adult human brain sample (fig. 2). The expression of the type II form in the adult human brain was confirmed by cloning and sequencing of the RT-PCR product. The expression level of the type II form in the adult human brain is less than that of the type I form. For the fetal human sample, only the type I form was detected. This observation in humans is different from the previous report, in which a much higher expression level of the type II form was reported in both the fetal and the adult human brains (Mitsui et al. 1999). This discrepancy might be caused by expression variations of different brain regions and/or different developmental stages in humans. The type II form was not detected in the gibbon or the two Old World monkeys. We conducted another RT-PCR analysis using the type II–specific primer, and, again, only the adult human brain showed positive expression of the type II form (data not shown). However, when the RT-PCR products were used as templates for a secondary PCR amplification, the type II form became visible for all the nonhuman primate samples (confirmed by sequencing), which is likely the result of background (not regulated) splicing. Therefore, the expression analysis demonstrated a sharp expression difference of the type II form between humans and the nonhuman primate species. It is likely that the origin of a functional type II form could be even more recent ($<$ 18 Myr) (Goodman et al. 1998) than the estimation based on the genomic DNA sequence data because the lesser ape does not express the type II form in the brain, although it has an intact open reading frame. Because the great ape species were not tested for expression in this study because of difficulty in obtaining brain samples, the question of when the type II form became functional is yet to be answered.

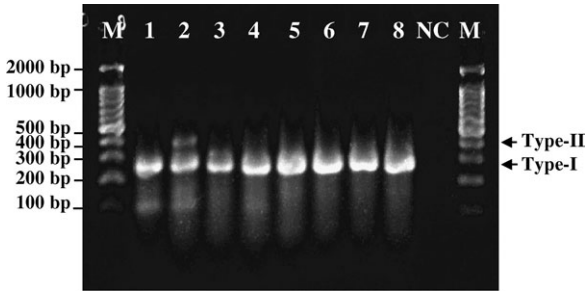

FIG. 2.—The electromorph of the RT-PCR result. The expected product sizes of type I and type II are 271 bp and 406 bp, respectively. The housekeeping gene, actin, was used as control, and no expression difference was observed in the samples tested (data not shown). The sample IDs are M = molecular standard, 1 = fetal human brain (frontal lobe), 2 = adult human brain (frontal lobe), 3 = white-browed gibbon brain (hippocampus), 4 = white-browed gibbon brain (frontal lobe), 5 = Yunnan golden monkey brain (hippocampus), 6 = Yunnan golden monkey brain (frontal lobe), 7 = rhesus monkey brain (hippocampus), 8 = rhesus monkey brain (frontal lobe), and NC = negative control.

When the coding region was translated into protein sequences, only two amino acid replacements were observed among humans, chimpanzees, and gorillas, with one of them polymorphic in gorillas, indicating strong functional constraint of neuropsin in these species (fig. 3). In the 11 human individuals (22 chromosomes) sequenced (845 bp), two polymorphic sites were observed in the intron region and no variations were detected in the coding region, including the type II–spliced region. The lack of sequence variations was also observed in the nine chimpanzees and four gorillas tested, again indicating strong functional constraint of neuropsin in these species. Therefore, the type II form may be functional in both chimpanzees and gorillas. In contrast, there are many substitutions in the orangutans and gibbons, especially within the type II–spliced region. There are seven substitutions between humans and orangutans in the coding region, with five of them located in the type II–spliced region. There are five substitutions between humans and the two gibbon species in the coding region, and four of them are located in the type II–spliced region.

The accelerated sequence substitution in orangutans and gibbons was confirmed by the relative rate test (Tajima 1993) (table 1), in which the orangutan and gibbon lineages showed significantly higher substitution rates than those of humans, chimpanzees, and gorillas (table 1). The relative rate tests were significant when only the intron region was considered but were not significant for the coding region, including the type II–spliced region (table 1). To further test whether the type II splice form is functional in orangutans and gibbons, we conducted a neutrality test by comparing the rates of synonymous and nonsynonymous substitution (*Ks* and

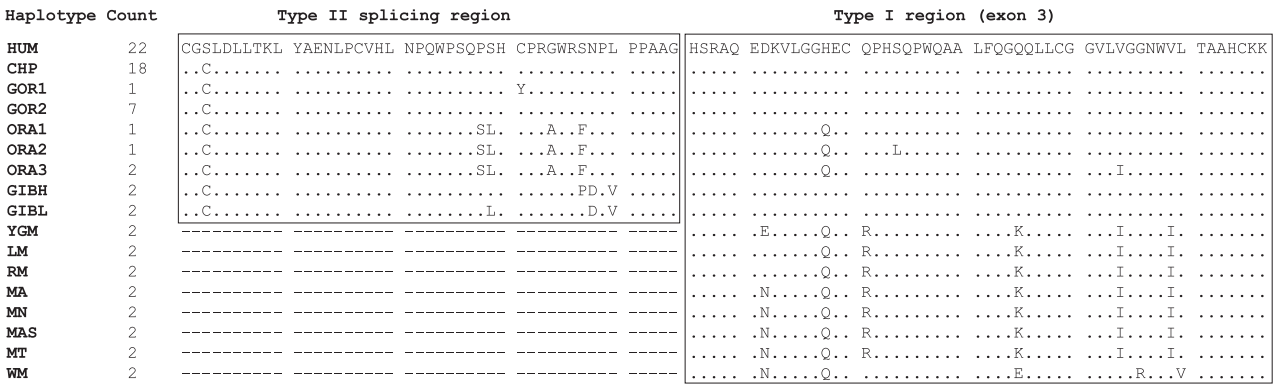

FIG. 3.—The alignment of putative amino acid sequences in humans and nonhuman primates.

Ka). An equal rate would be expected for *Ks* and *Ka* when the segment undergoing testing is selectively neutral, whereas negative and positive selection would result in *Ka*/*Ks* ratios significantly deviated from 1. We calculated the *Ka*/*Ks* ratios of different lineages in primates using the method developed by Nei and Gojobori (1986) (fig. 4). Compared with the other primate lineages, the orangutan and gibbon lineages have large *Ka*/*Ks* ratios for the entire coding region (2.56 and 1.81, respectively), but they are not significant when the one-tailed Fisher's exact tests were conducted ($P > 0.1$ [fig. 4]). In addition, using PAML (Yang 1997), we compared the *Ka*/*Ks* ratios among different primate lineages, and no significant difference was observed when either the entire coding region or only the type II–spliced region was considered, respectively ($P > 0.05$). Similar result was obtained when using Pamilo-Bianchi-Li's method (Li 1993; Pamilo and Bianchi 1993). Hence, the lack of sequence conservation in orangutans and gibbons may be caused by a higher neutral mutation rate rather than natural selection, and both of them may not have a functional type II form, which is consistent with the expression result of gibbon. Therefore, the origin of a functional type II form could be less than 18 MYA.

In humans, there is one amino acid substitution, from cysteine to serine (C to S), which is located in the type II–spliced region. This mutation is fixed in humans but conserved in all the ape species. The functional consequence of this amino acid change in humans is yet to be understood.

Discussion

Alternative splicing is a common phenomenon in the eukaryotic genomes. In humans, less than 40% of genes were suggested to have multiple splice forms (Mironov et al. 1999; Stamm 2002; Zhu et al. 2003). A recent study comparing human and mouse genome showed that alternative splicing is associated with a large increase in frequency of recent exon creation and/or loss (Modrek and Lee 2003). The newly recruited exons are usually not conserved and likely the targets of Darwinian positive selection in forming new proteins (Modrek and Lee 2003). The origin of type II neuropsin in primates seems to follow the pattern proposed in the aforementioned study. The type II splice form recruited a 135-bp fragment from the intron region. After the new splice form occurred in the ancestor of hominoids about 18 MYA, it began to accumulate mutations in the spliced region (previous intron sequence). The most crucial mutation is the deletion of a C nucleotide in the spliced region that makes the open reading frame intact and adds 45 amino acids to the original proteins. Whether type I and type II neuropsin are functionally different is not known. However, according to the expression study in humans, the type II form may acquire a new function within the brain, where it is preferentially expressed during human adulthood (Mitsui et al. 1999).

In view of evolution, alternative splicing is a relatively safe strategy in creating new genes/proteins. The newly created splice forms are not exposed to strong selection pressures because the old forms can still function normally

Table 1
The Result of Relative Rate Tests

Species	Outgroup	Entire Region (845 bp)	Intron (548 bp)	Type II Splicing Region (135 bp)	Exon 3 (160 bp)
Human-Gibbon	Rhesus monkey	0.041*	0.041*	1.000	0.317
Human-Orangutan	Rhesus monkey	0.041*	0.041*	0.317	0.317
Chimp-Gibbon	Rhesus monkey	0.034*	0.033*	1.000	0.317
Chimp-Orangutan	Rhesus monkey	0.034*	0.033*	0.317	0.317
Gorilla-Gibbon	Rhesus monkey	0.059	0.061	1.000	0.317
Gorilla-Orangutan	Rhesus monkey	0.059	0.061	0.317	0.317
Rhesus-Gibbon	Marmoset	0.696	0.662	0.655	0.705
Rhesus-Orangutan	Marmoset	0.895	0.655	0.705	0.180

NOTE.—Data are from Tajima (1993). * $P < 0.05$.

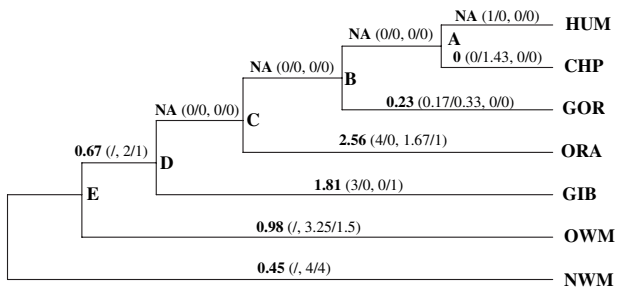

FIG. 4.—The *Ka*/*Ks* values of different evolutionary lineages in primates (Nei and Gojobori 1986). The phylogeny was based on Goodman et al. (1998). The A to E refer to the internal nodes (ancestral sequences) of the phylogenetic tree. The numbers in parentheses refer to synonymous and nonsynonymous mutations observed in the type II–spliced region and exon 3, respectively. The *Ka*/*Ks* ratios of the Old World and New World monkey (OWM and NWM) lineages were calculated using only the exon 3 sequences. NA = not applicable.

(Modrek and Lee 2003). Considering the relatively rich alternative splicing events observed in the brain (Grabowski 1998), this strategy may play an important role in shaping up brain structure and function, especially in primates as brain enlargement and progressive change of cognitive abilities is the most prominent character of primate evolution.

Acknowledgments

We are grateful for the technical help of Xiao-Na Fan. We thank Dr. Peter Underhill for generously providing the DNA samples of great apes. This work was supported by grants from the Chinese Academy of Sciences (KSCX2-SW-121), the National Science Foundation of China (NSFC), and the Natural Science Foundation of Yunnan Province, China.

Literature Cited

Chen, Z. L., S. Yoshida, K. Kato, Y. Momota, J. Suzuki, T. Tanaka, J. Ito, H. Nishino, S. Aimoto, and H. Kiyama. 1995. Expression and activity-dependent changes of a novel limbic-serine protease gene in the hippocampus. J. Neurosci. **15**:5088–5097.

Davies, B., I. R. Kearns, J. Ure, C. H. Davies, and R. Lathe. 2001. Loss of hippocampal serine protease BSP1/neuropsin predisposes to global seizure activity. J. Neurosci. **21**:6993–7000.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, C. P. Groves. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9**:585–598.

Grabowski, P. J. 1998. Splicing regulation in neurons: tinkering with cell-specific control. Cell **92**:709–712.

Hirata, A., S. Yoshida, N. Inoue et al. (12 co-authors). 2001. Abnormalities of synapses and neurons in the hippocampus of neuropsin-deficient mice. Mol. Cell. Neurosci. **17**:600–610.

Komai, S., T. Matsuyama, K. Matsumoto, K. Kato, M. Koubayashi, K. Imamura, S. Yoshida, S. Ugawa, and S. Shiosaka. 2000. Neuropsin regulates an early phase of schaffer-collateral long-term potentiation in the murine hippocampus. Eur. J. Neurosci. **12**:1479–1486.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics. **17**:1244–1245.

Li, W. H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**:96–99.

Long, M.Y., E. Betran, K. Thornton, and W. Wang. 2003. The origin of new genes: glimpses from the young and old. Nat. Rev. Genet. **4**:866–875.

Mironov, A. A., J. W. Fickett, and M. S. Gelfand. 1999. Frequent alternative splicing of human genes. Genome Res. **9**:1288–1293.

Mitsui, S., N. Tsuruoka, K. Yamashiro, H. Nakazato, and N. Yamaguchi. 1999. A novel form of human neuropsin, a brain-related serine protease, is generated by alternative splicing and is expressed preferentially in human adult brain. Eur. J. Biochem. **260**:627–634.

Modrek, B., and C. J. Lee. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat. Genet. **34**:176–180.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

Oka, T., M. Akisada, A. Okabe, K. Sakurai, S. Shiosaka, and K. Kato. 2002. Extracellular serine protease neuropsin (KLK8) modulates neurite outgrowth and fasciculation of mouse hippocampal neurons in culture. Neurosci. Lett. **321**:141–144.

Pamilo, P., and N. O. Bianchi. 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol. Biol. Evol. **10**:271–281.

Stamm, S. 2002. Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. Hum. Mol. Genet. **11**:2409–2416.

Suzuki, J., S. Yoshida, Z. L. Chen, Y. Momota, K. Kato, A. Hirata, and S. Shiosaka. 1995. Ontogeny of neuropsin mRNA expression in the mouse brain. Neurosci. Res. **23**:345–351.

Tajima, F. 1993. Simple methods for testing molecular clock hypothesis. Genetics **135**:599–607.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **15**:555–556.

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.

Zhu, J., J. Shendure, R. D. Mitra, and G. M. Church. 2003. Single molecule profiling of alternative pre-mRNA splicing. Science **301**:836–838.

Lisa Matisoo-Smith, Associate Editor

Accepted July 19, 2004