Adaptive Evolution of the IgA Hinge Region in Primates

Kenta Sumiyama,† Naruya Saitou,† and Shintaroh Ueda**

*Department of Biological Sciences, Graduate School of Science, University of Tokyo; †Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima 411-8540, Japan

IgA is a major component that prevents the penetration of pathogenic bacteria into mucosal surfaces. The IgA antibody is cleaved at the IgA hinge region with high specificity by IgA-specific proteases produced by several pathogenic bacteria. We conducted a genomic sequence analysis of the IgA genes of a wide spectrum of primates, including the first intron and second exon, which consist of the hinge region and the CH2 domain, to find evidence of positive selection. Because the hinge region is quite small, we combined the largest collection of sequences that could be clearly aligned and evaluated the total numbers of synonymous and nonsynonymous substitutions on the phylogenetic tree. The nonsynonymous to synonymous substitution ratio (d_N/d_S test) showed that hominoids, Old World monkeys, and New World monkeys have d_N/d_S ratios of 5.4, 6.3, and 4.2, respectively. Fisher's exact probability tests showed statistical significance for the Old World monkey. Because the substitution rates of the flanking sequences are more or less similar to the synonymous rates of the hinge region, these high values of d_N/d_S should be the result of positive selection at the hinge region. Combining the high sequence variability in each population and the highly accelerated nonsynonymous substitution rates in the hinge region, we conclude that this unusual IgA evolution is a molecular evidence of adaptive evolution possibly caused by the host-parasite relationship.

Introduction

The IgA antibody is a major component of the secreted immune response and is known to inhibit colonization of pathogenic bacteria on the mucosal surface. The heavy chain constant region of IgA consists of three CH domains and a hinge region, which are encoded by an immunoglobulin Cα gene. The hinge region, located in the beginning of the CH2 exon, links Fab, including the antigen recognition sites, and Fc, including the effector sites, and can vary the distance between two antigen recognition sites with some flexibility. Hominoids have two Cα loci in their genomes, except for orangutan, whereas Old World monkeys have a single Cα gene (Kawamura, Omoto, and Ueda 1990 and references therein). Sequence comparison among the hominoid Cα genes reveals the hypervariability of the hinge region that might be caused by frequent remodeling associated with their reiterated structure (Kawamura, Omoto, and Ueda 1990). Because the IgA is a secreted antibody, this molecule may be subject to positive selection driven by bacterial proteolysis. It is known that the bacterial IgA-specific proteases inactivate the IgA molecule by cleaving it at the hinge region (Plaut et al. 1975; Kilian, Mestecky, and Schrohenloher 1979; Male 1979; Plaut 1983). As for IgA, we found by a survey based on PCR–single strand conformation polymorphism (SSCP) that the Cα hinge region of Old World monkeys had an unusually high polymorphism, although the hinge region is located within the constant region (Sumiyama et al. 1998).

In view of the evolutionary hypervariability among hominoids and the unusually high polymorphism of the Cα gene hinge region among Old World monkeys, there is a possibility that this gene has experienced positive selection during primate evolution. With this aim, we obtained 31 new Cα hinge sequences from Old World and New World monkeys and evaluated their d_N and d_S values by using a tree-based comparison method.

Key words: immunoglobulin Cα, IgA, hinge, positive selection.

Address for correspondence and reprints: Kenta Sumiyama, Department of Molecular, Cellular and Developmental Biology, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520. E-mail: kenta.sumiyama@yale.edu.

Mol. Biol. Evol. 19(7):1093–1099. 2002
© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Materials and Methods

Specimens

Nucleotide sequences of the following species were newly determined: patas monkey (*Erythrocebus patas*), Tibetan macaque (*Macaca thibetana*), stump-tailed macaque (*Macaca arctoides*), rhesus macaque (*Macaca mulatta*), lion-tailed macaque (*Macaca silenus*), pig-tailed macaque (*Macaca nemestrina*), Taiwan macaque (*Macaca cyclopis*), Assamese macaque (*Macaca assamensis*), crab-eating macaque (*Macaca fascicularis*), anubis baboon (*Papio anubis*), hamadryas baboon (*Papio hamadryas*), green monkey (*Cercopithecus aethiops*), black-and-white colobus (*Colobus polykomos*), banded leaf monkey (*Presbytis melalophos*), silvered leaf monkey (*Presbytis cristata*), obscurus leaf monkey (*Presbytis obscurus*), Javan leaf monkey (*Presbytis comata*), white-thighed leaf monkey (*Presbytis femoralis*), night monkey (*Aotus trivirgatus*), cotton-top tamarin (*Saginus oedipus*), white-lipped tamarin (*Saginus labiatus*), Geoffroy's spider monkey (*Ateles geoffroyi*), and long-haired spider monkey (*Ateles belzebuth*). Their total genomic DNAs were prepared from peripheral blood by standard procedure.

PCR Amplification and Fragment Analysis

The following PCR primer pair was used to amplify an approximately 400-bp DNA fragment involving the intron, hinge, and CH2 regions: 5′-CCA AGC TTC TAC ACG AAT CCC AGC CAG GAT GTG-3′ and 5′-CCG AAT TCT CCC ACT TGA GGG CGT CCA GGT GAA-3′. The PCR program was 35–40 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 2 min. A PCR-SSCP analysis was performed to check whether the PCR products contain single or multiple sequences, where

PCR products were run on nondenaturing gel containing 5% acrylamide, 5% glycerol, and 0.5× TBE buffer at 16°C constant temperature, as described before (Sumiyama et al. 1998). In the PCR-SSCP analysis, the following internal primers were used to obtain better resolution in gel: 5′-TCT GAC CAG TTC AGG CCA TCT C-3′ and 5′-GGC CGG CGC AGC GAC AGT-3′.

Sequence Determination

Nucleotide sequences of both strands were determined using the ABI Taq DyeDeoxy Terminator Cycle sequencing kit, according to the manufacturer's instructions, using an ABI PRISM 377 DNA sequencer. In some heterozygous cases, their nucleotide sequences were determined by using allele-specific sequencing primers.

Sequence Analyses

Hominoid sequences were taken from Kawamura et al. (1991). The Neighbor-joining (NJ) method (Saitou and Nei 1987) was used for tree reconstruction of the hinge-flanking region by using CLUSTAL W (Thompson, Higgins, and Gibson 1994), with ''exclude gap'' option. Maximum likelihood analysis was done by using the NucML program of the MOLPHY package (Adachi and Hasegawa 1996) with the HKY85 model in order to evaluate the reliability of tree topology. We used the PAUP 3.1.1. program (Illinois Natural History Survey) for the construction of the maximum parsimony (MP) tree for the hinge region. For the purpose of estimating the ancestral sequences on the tree internal nodes, we used the baseml program of the PAML package (Yang 1999) with the HKY85 model. Values of d_N and d_S were calculated for each phylogenetic tree branch between the given sequences using the computer program package ODEN (Ina 1994) that employs Nei and Gojobori's (1986) method. Window analysis was conducted using the wina program of Endo, Ikeo, and Gojobori (1996). Statistical tests for positive selection were conducted by using Fisher's exact probability test, where we set the null hypothesis that nonsynonymous substitutions per site and synonymous substitutions per site are equal (strictly neutral condition). Therefore, the ratio of unchanged nonsynonymous sites to unchanged synonymous sites was expected to be equal to that of the nonsynonymous changes to synonymous changes (e.g., Zhang, Kumar, and Nei 1997).

Results and Discussion

Window Analysis of Hominoid Genomic Sequences

We performed a window-based analysis on the published Cα sequences to study the nonsynonymous and synonymous substitution pattern for the entire coding region of the immunoglobulin Cα genes. Kawamura et al. (1991) previously determined the whole genomic sequences of the chimpanzee, gorilla, and orangutan Cα genes. The human genomic sequence (Takahashi et al. 1982; accession number is J00220) was also available. We first obtained pairwise distances of both synonymous (d_S) and nonsynonymous (d_N) substitutions in a default 20-codon window of the wina program. The hinge region of hominoids consists of 18 codons such that this default window size seems to be appropriate.

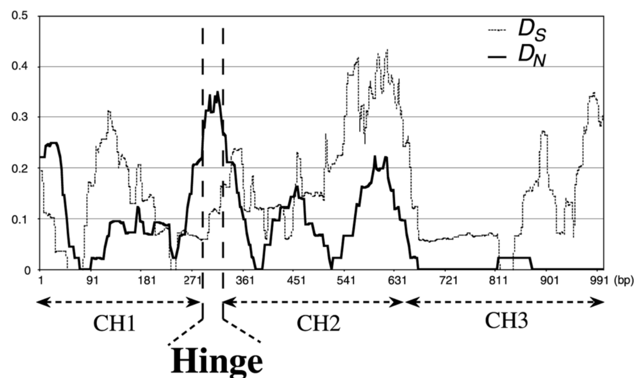

FIG. 1.—Window analysis of the hominoid immunoglobulin Cα1 genes. Vertical and horizontal axes indicate the total distance (D_T) value and the window position, respectively. D_S (dotted line) and D_N (thick line) designate D_T of synonymous and nonsynonymous substitutions, respectively. Sliding window is shifted by three bases (one codon). Window size is 20 codons. Note that the nucleotide position indicates the start point of each window.

We calculated the total distance (D_T) of the phylogenetic tree among species for d_S and d_N, according to the established tree topology—(((human, chimpanzee), gorilla), orangutan). Ancestral nucleotide sequences of all interior nodes were reconstructed by using the PAML program, and then the number of nucleotide substitutions on each branch was calculated. D_T is defined as the sum of all branches. D_T of each window was plotted against the first nucleotide position of each window along the entire coding sequence (fig. 1). The D_T of the nonsynonymous substitutions clearly exceeds that of the synonymous substitutions at the hinge region. This observation supports the idea that the hominoid hinge region has been evolving under some kind of positive selection. Besides the hinge region, both synonymous and nonsynonymous rates were raised at the end of the CH2 domain. We do not know the reason for this observation, however, it is possible that recombinations or gene conversions occurred and raised overall variation in this region.

Genomic Sequences of the IgA Gene of Nonhominoid Primates

In our previous report, we showed by PCR-SSCP DNA fragment analysis that some nonhominoid primates have very high polymorphism and high heterozygosity (Sumiyama et al. 1998). Here, we determined the hinge and flanking sequences of various nonhominoid primates to perform a more detailed sequence analysis. Newly determined immunoglobulin Cα gene sequences of nonhominoid primates are supplied as supplementary material, and the predicted hinge amino acid sequences are shown in table 1. Sequence data are available in DDBJ/EMBL/GenBank International Nucleotide Sequence Database (accession numbers AB013752–AB013790).

Table 1
Amino Acid Sequences of the Immunoglobulin Cα Hinge Region in Primates

Species	Amino Acid Sequence
Old World monkeys (Cercopithecinae)	
Rhesus macaque-2, crab-eating macaque-1	SETKPCL
Lion-tailed macaque, pit-tailed macaque-1, Taiwan macaque-2,[a] Assamese macaque-2,[a] rhesus macaque-4[a]	SQTKPCL
Rhesus macaque-3[a]	PQPKPCL
Tibetan macaque,[a] stump-tailed macaque[a]	PPPKPCL
Assamese macaque-1	LPPNSCL
Rheses macaque-1, crab-eating macaque-2[a]	PPPNSCL
Hamadryas baboon-1	PKPNSCL
Anubis baboon,[a] hamadryas baboon-2	SQPKSCL
Patas monkey	IPPKPCF
Green monkey	ASPKCPL
Hamadryas baboon-3	PSSKCPSLK
Rhesus macaque-5, Taiwan macaque-1, pig-tailed macaque-2	PPITPPCPS
Rhesus macaque-6[a]	IPPTPIPLPCPS
Old World monkeys (Colobinae)	
Black-and-white colobus	IPPRCPP
Silvered leaf monkey-1,	QPPCPPCCS
Silvered leaf monkey-2, obscurus leaf monkey, banded leaf monkey-1	RPPCPPCCS
Banded leaf monkey-2, white-thighed leaf monkey, Javan leaf monkey	HPPCYS
Hominoids	
Human Cα1	PSTPPTPSPSTPPTPSPS
Chimpanzee Cα1	PSTPCPPTPSTPPTPSPS
Gorilla Cα1	PSTPPTPSPSTPPTPSPP
Orangutan Cα	PRPTPTPSTPPCPPPS
Human Cα2, chimpanzee Cα2	PPPPP
Gorilla Cα2	PPSPP
Gibbon Cα2	PPPHP
New World monkeys	
Cotton-top tamarin	PPQPLCPCHS
Night monkey	PPECPPPCHP
Geoffroy's spider monkey	PPLPPCPCHS
Long-haired spider monkey	PPLPPPPCHS
White-lipped moustached tamarin	PPLPSCPCHS

[a] Only the hinge region is determined.

Our sequences begin at the end of exon 1, corresponding to the CH1 domain, and end at the middle of exon 2, corresponding to the hinge region and the CH2 domain. There is a gap located just before the exon 1 and intron 1 boundary. This 3-bp gap does not cause a nonsense mutation. Exon-intron boundary sequences are well conserved throughout the sequences. We did not observe any large deletion or insertion in the subsequent intron region, although there are several single substitutions and minor deletions, such that the multiple alignment of all intron sequences was straightforward. This is also the case in exon 2, other than the hinge region, and most substitutions in this region occurred at the third positions leading to synonymous substitutions. We did not find any large gap here. The boundary sequences between intron 1 and exon 2 are also well conserved.

In contrast to the flanking region, the hinge region is unusually hypervariable. Because of the complexity associated with the hinge region (different length, many substitutions, and partially reiterated structure; see table 1), a multiple alignment of all sequences could not be performed at once. Instead of making an overall multiple alignment in the hinge region, we show the alignment within the hinge to be the largest collection of sequences for which homology can be clearly assessed within Old World monkeys, New World monkeys, and hominoids (fig. 2). We use these unambiguously aligned sequences for subsequent analysis of the hinge region.

Phylogenetic Analysis of the Nonhypervariable Hinge-Flanking Sequences

To elucidate the basic phylogenetic relationship among sequences used here, we constructed an NJ tree using 343 bp hinge-flanking sequences with 1,000 bootstrap replications (fig. 3). It should be noted that these nonhinge region sequences were unambiguously aligned. In this NJ tree, Old World monkeys, hominoids, and New World monkeys are separated clearly. Interestingly, internal clusters in Old World monkeys do not necessarily correspond to that of the species relationship. For example, three rhesus macaque sequences (rhesus mac-1, -2, and -5) are located in different clusters. The same is true for two pig-tailed macaque sequences. These patterns suggest that an ancestral polymorphism may have been maintained in the extant Old World mon-

a

```
crab-1      5'  TCC GAA ACC AAG CCT TGC TTA 3'
                 S   E   T   K   P   C   L

lion        5'  TCC CAA ACC AAG CCT TGC TTA 3'
                 S   Q   T   K   P   C   L

rhesus-3    5'  CCC CAA CCC AAG CCT TGC TTA 3'
                 P   Q   P   K   P   C   L

tibetan     5'  CCC CCA CCC AAG CCT TGC TTA 3'
                 P   P   P   K   P   C   L

assam-1     5'  CTC CCA CCC AAC TCT TGC TTA 3'
                 L   P   P   N   S   C   L

rhesus-1    5'  CCC CCA CCC AAC TCT TGC TTA 3'
                 P   P   P   N   S   C   L

hamad-1     5'  CCC AAA CCC AAC TCT TGC TTA 3'
                 P   K   P   N   S   C   L

hamad-2     5'  TCC CAA CCC AAG TCT TGC TTA 3'
                 S   Q   P   K   S   C   L

patas       5'  ATC CCA CCC AAG CCT TGC TTC 3'
                 I   P   P   K   P   C   F

green       5'  GCC TCA CCC AAA TGT CCC TTA 3'
                 A   S   P   K   C   P   L
```

b

```
gspider     5'  CCT CCC CTG CCC CCA TGC CCA TGT CAC TCA 3'
                 P   P   L   P   P   C   P   C   H   S

lspider     5'  CCT CCC CTG CCC CCA CCC CCA TGT CAC TCA 3'
                 P   P   L   P   P   P   P   C   H   S

night       5'  CCT CCC GAG TGC CCA CCC CCA TGT CAC CCA 3'
                 P   P   E   C   P   P   P   C   H   P

wtamarin    5'  CCT CCC CTG CCT TCA TGC CCA TGT CAC TCA 3'
                 P   P   L   P   S   C   P   C   H   S

ctamarin    5'  CCT CCC CAG CCC CTA TGC CCA TGT CAC TCA 3'
                 P   P   Q   P   L   C   P   C   H   S
```

c

```
hsa1  5'  CCC TCA ACT CCA CCT ACC CCA TCT CCC TCA ACT CCA CCT ACC CCA TCT CCC TCA 3'
           P   S   T   P   P   T   P   S   P   S   T   P   P   T   P   S   P   S

ptr1  5'  CCC TCA ACT CCA TGT CCC CCA ACT CCC TCA ACT CCA CCT ACC CCA TCT CCC TCC 3'
           P   S   T   P   C   P   P   T   P   S   T   P   P   T   P   S   P   S

ggo1  5'  CCC TCA ACT CCA CCT ACC CCA TCT CCC TCA ACT CCA CCT ACC CCA TCT CCC CCA 3'
           P   S   T   P   P   T   P   S   P   S   T   P   P   T   P   S   P   P

ppo   5'  CCC --- --- CGA CCT ACC CCA ACT CCC TCA ACT CCA CCT TGC CCA CCT CCC TCA 3'
           P           R   P   T   P   T   P   S   T   P   P   C   P   P   P   S
```

FIG. 2.—The nucleotide sequences and their predicted translations of the immunoglobulin Cα genes in the hinge region of the largest collection of unambiguously aligned sequences from (*a*) Old World monkeys, (*b*) hominoids, and (*c*) New World monkeys. The full-length sequence alignment from the end of the exon 1 to the middle of the exon 2 region will be shown at the MBE website as supplementary data. To save space, names for each sequence are shortened, such as crab (crab-eating macaque), lion (lion-tailed macaque), rhesus (rhesus macaque), tibetan (Tibetan macaque), assam (Assamese macaque), hamad (hamadryas baboon), patas (patas monkey), green (green monkey), gspider (Geoffroy's spider monkey), lspider (long-haired spider monkey), night (night monkey), wtamarin (white-lipped moustached tamarin), ctamarin (cotton-top tamarin), hsa1 (human Cα1), ptr1 (chimpanzee Cα1), ggo1 (gorilla Cα1), and ppo (orangutan Cα).

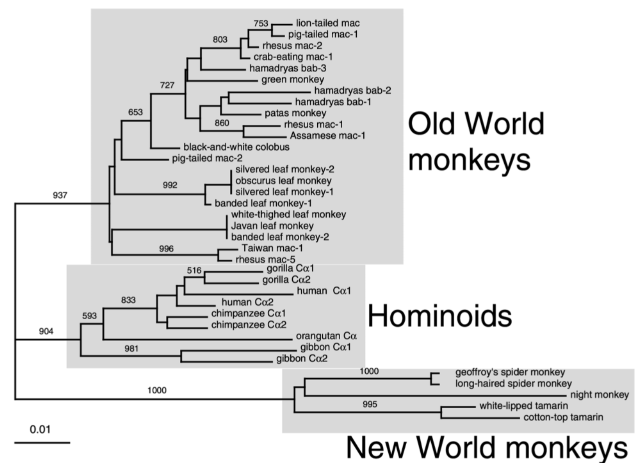

FIG. 3.—A phylogenetic tree reconstructed for the Cα hinge-flanking sequences by using the NJ method. The bootstrap values out of 1,000 replications are indicated where their values are larger than 500. Macaque and baboon are abbreviated as mac and bab, respectively.

keys (trans-species polymorphism). This is indicative of overdominant selection (Klein et al. 1993).

Because the bootstrap probabilities for the internal branches were not high in figure 3, it is difficult to conclude this with certainty. Hence, we compared this unusual topology (original tree: ORTR) with the species tree (SPTR) using the NucML program of the MOLPHY package. For simplification purpose, we compared only four sequences: rhesus macaque-1, rhesus macaque-5, patas monkey, and human as an apparent outgroup. ORTR (log likelihood = −687.6, and RELL-BP = 0.93) was clearly favored to SPTR (log likelihood = −698.1, with standard error 7.2, and RELL-BP = 0.053). We also compared pig-tailed macaque-1, pig-tailed macaque-2, African green monkey, and human and obtained similar results. ORTR (log likelihood = −686.8, and RELL-BP = 0.93) was favored to SPTR (log likelihood = −696.0, with standard error 6.3, and RELL-BP = 0.033).

We previously reported a very high extent of heterozygosity in the hinge region in macaques (Sumiyama et al. 1998), which is another indication of overdominant selection. Considering these results together, existence of overdominant selection is quite probable in the Cα gene hinge region for Old World monkeys.

When hominoids are considered, special attention is necessary because they have two Cα loci (Kawamura, Saitou, and Ueda 1992). Cα1 and Cα2 are positioned close to each other in the phylogenetic tree in every hominoid species. This is the result of increased similarity caused by a gene conversion event between the two duplicated Cα loci (Kawamura, Saitou, and Ueda 1992). In other primates, there is only one locus for the Cα gene; thus, they are free from this gene conversion problem.

Phylogeny and Substitution Pattern of the Hypervariable Hinge Sequences

Because the hinge region itself is small and hypervariable, it was necessary to use methods different from those used for the hinge-flanking regions. First, we constructed MP trees for those nucleotide sequences corresponding to 10 Old World monkey hinge regions shown in figure 2*a*. Ten equally parsimonious trees were obtained (fig. 4*a*). Nine of the ten MP trees contained only nonsynonymous substitutions. Only one MP tree (MPTREE_5) contained one synonymous substitution, as shown with a diamond in figure 4*a*.

Second, we constructed a phylogenetic network from the same data set (fig. 4*b*). The phylogenetic network is a summarization of all possible phylogenies and is more informative than the simple consensus tree because it can show the existence of parallel substitutions or recombinations (or both) with parallelogram. In the phylogenetic network, only one possible synonymous substitution is observed in a triangle. Parallelograms shown in this network were probably caused by parallel substitutions.

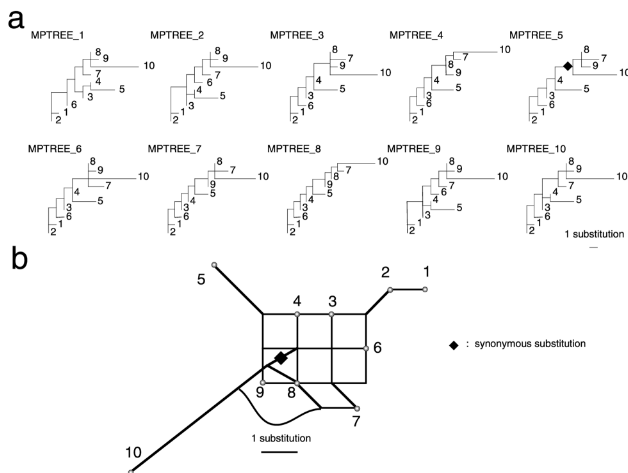

FIG. 4.—Phylogenetic relationship and nucleotide substitution patterns among immunoglobulin Cα genes-alleles of Old World monkeys. *a*, All MP trees (10) of Old World monkeys (fig. 2*a*) calculated by PAUP 3.1.1. Operational taxonomic units (OTUs) are as folows—1: rhesus macaque-2, crab-eating macaque-1; 2: lion-tailed macaque, pig-tailed macaque-1, Taiwan macaque-2, Assamese macaque-2, rhesus macaque-4; 3: rhesus macaque-3; 4: Tibetan macaque, stump-tailed macaque; 5: patas monkey; 6: anubis baboon, hamadryas baboon-2; 7: hamadryas baboon-1; 8: rhesus macaque-1, crab-eating macaque-2; 9: Assamese macaque-1; 10: green monkey. A synonymous substitution is indicated by a diamond. *b*, Phylogenetic network for Old World monkeys. Parallelograms designate existence of incompatible sites. A curve in the network indicates an incompatible site, which cannot be drawn in three dimensions. All 10 MP trees in figure 4*a* are included in this network. A synonymous substitution is designated by a diamond.

To further examine our observations, we computed Fisher's exact tests (table 2). Note that the commonly used *t*-test is not suitable for very low expected values of substitutions. The results are listed in table 2. Note that the nonsynonymous and synonymous substitution reconstruction may be biased by the codon usage and the GC%, it is observed that $GC_3 = 0.795$ and ENc (effective number of codon) = 39.75 for the entire crab-eating macaque Cα-coding sequence (X53705). Nevertheless, the null hypothesis of equal expectation of nonsynonymous and synonymous substitutions was rejected for all MP trees. We also conducted the same analysis on hominoids and New World monkeys. They also show high nonsynonymous to synonymous ratios, but these were not statistically significant (see table 2).

To test whether a high ratio of nonsynonymous substitution to synonymous substitution is caused by an increased nonsynonymous substitution rate or the suppression of the synonymous rate, substitutions of the flanking intron were counted and compared with the hinge nonsynoymous substitutions. We reconstructed the tree and ancestral sequence on each node using the PAML program and then calculated the total substitution number for both intron and hinge regions, using Assamese mac-1 and -2, crab-eating mac-1 and -2, hamadryas bab-1 and -2, patas monkey, and African green monkey sequences. The results reveal 23 nonsynonymous changes and 127 invariant sites in total in the hinge region. In the intron region, there are 18 site changes and a total of 2,122 invariant sites. This is highly significant ($P < 0.00000001$) by Fisher's exact test. Synonymous changes in the hinge region are 1 and 59 for the invariant sites. Fisher's exact test cannot rule out the null hypothesis of the equal substitution rate ($P = 0.409$). If we assume that the intron shows a neutral rate of evolution, these data suggest that only the nonsynonymous substitution rate has increased in the hinge region. Together with the high ratio of nonsynonymous to synonymous substitutions in the hinge region, we conclude that positive selection operates in the hinge region.

Table 2
Numbers of Nonsynonymous Substitutions and Synonymous Substitutions in all Maximum Parsimony Trees of Each Group, Their $\Sigma d_N/\Sigma d_S$ Values, and Fisher's Exact Probabilities

		Nonsynonymous	Synonymous	P	$\Sigma d_N/\Sigma d_S$
Old World monkeys					
MPTREE 1–4, 6–10	Change	19	0	0.0045	—
	No change	157	55		
MPTREE 5	Change	18	1	0.034	6.3
	No change	158	54		
Hominoids	Change	10	1	0.070	5.4
	No change	130	69		
New World monkeys	Change	10	1	0.125	4.2
	No change	76	33		

NOTE.—Substitution numbers are based on parsimony estimates so that it could have some error.

Other Supports for Positive Selection: IgA Sequence Variation in Other Mammals and Vast IgA Protease Variation in Pathogenic Bacteria

We found that the IgA hinge region showed evidence of adaptive evolution in primates. Furthermore, some other mammalian species also show hypervariability in their IgA hinge regions, although nonsynonymous-synonymous substitution ratio tests have not been conducted because of alignment difficulties (Osborne et al. 1988; Burnett et al. 1989; Brown and Butler 1994). Our results suggest that adaptive evolution of the IgA hinge is likely to have occurred in many mammalian species. IgA is the most prevalent immunoglobulin in secretions, and it plays a crucial role for the prevention of pathogenic bacterial penetration of the mucosal surface. Because IgA is an indispensable factor for mammalian species, the bacterial IgA protease could be a strong selective pressure on the IgA gene.

There is a possibility of coevolution between host IgA and proteases of the pathogenic bacteria. IgA proteases of pathogenic bacteria consist of several different classes. They are found in a wide variety of pathogenic bacteria, including *Haemophilus*, *Neisseria*, and *Streptococcus* (Plaut et al. 1975; Kilian, Mestecky, and Schrohenloher 1979; Male 1979). These proteases are highly specific enzymes that cleave certain peptide bonds of the IgA hinge region. Interestingly, their catalytic mechanisms and gene structures are completely different among bacteria (Mortensen and Kilian 1984; Bachovchin et al. 1990; Gilbert, Plaut, and Wright 1991;

Lomholt, Poulsen, and Kilian 1995), such as thiol-activated protease from *Bacteroides melaninogenicus*, serine-type proteases from *Haemophilus influenzae* and *Neisseria gonorrhoeae*, and metalloprotease from *Streptococcus sanguis*. Such wide variations among IgA proteases from different origins with the same substrate specificity strongly suggest functional convergence in IgA protease evolution. This convergent evolution of the IgA proteases of pathogenic bacteria indicates the importance of inactivation of the host IgA to retain their pathogenicity. This fact indicates strong host-parasite interaction between host IgA and parasite protease, which results in positive selection pressure on both sides.

Conclusions

The d_N/d_S value is a powerful indicator for positive selection with which several genes evolving under positive selection have so far been identified (Hughes and Nei 1988, 1989; Tanaka and Nei 1989; Messier and Stewart 1997). Here, we used a tree-based multiple sequence comparative method in order to test the smaller region. We showed that nonsynonymous substitutions in the hinge were significantly larger than synonymous substitutions in the hinge and flanking intron substitutions. This strongly suggests the existence of positive selection in the IgA hinge region of primates. This selection could be a result of the adaptation against pathogenic parasites where an increasing variety in the IgA hinge sequence is requisite for the secretory immune system.

Supplementary Materials

The full version of the alignment data will be shown at the MBE website (www.molbiolevol.org).

Acknowledgments

This study was supported by Grants-in-Aids for Scientific Research from the Ministry of Education, Science, Sports, and Culture, Japan, NIG Cooperative Research Program, and the Cooperation Research Program of Primate Research Institute, Kyoto University. We thank Drs. T. Ishida and O. Takenaka for providing samples of crab-eating macaque and leaf monkeys, respectively.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1996. Computer science monographs, No. 28. MOLPHY. Version 2.3. Programs for molecular phylogenetics based on maximum likelihood. Institute of Statistical Mathematics, Tokyo.

BACHOVCHIN, W. W., A. G. PLAUTO, G. R. FLENTKE, M. LYNCH, and C. A. KETTNER. 1990. Inhibition of IgA1 proteinases from *Neisseria gonorrhoeae* and *Hemophilus influenzae* by peptide prolyl boronic acids. J. Biol. Chem. **265**:3738–3743.

BROWN, W. R., and J. E. BUTLER. 1994. Characterization of a C alpha gene of swine. Mol. Immunol. **31**:633–642.

BURNETT, R. C., W. C. HANLY, S. K. ZHAI, and K. L. KNIGHT. 1989. The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes. EMBO J. **8**: 4041–4047.

ENDO, T., K. IKEO, and T. GOJOBORI. 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13**:685–690.

GILBERT, J. V., A. G. PLAUT, and A. WRIGHT. 1991. Analysis of the immunoglobulin A protease gene of *Streptococcus sanguis*. Infect. Immun. **59**:7–17.

HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167–170.

———. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Positive Darwinian selection observed at the variable-region genes of immunoglobulins. Proc. Natl. Acad. Sci. USA **86**:958–962.

INA, Y. 1994. ODEN: a program package for molecular evolutionary analysis and database search of DNA and amino acid sequences. Comput. Appl. Biosci. **10**:11–12.

KAWAMURA, S., K. OMOTO, and S. UEDA. 1990. Evolutionary hypervariability in the hinge region of the immunoglobulin alpha gene. J. Mol. Biol. **215**:201–206.

KAWAMURA, S., N. SAITOU, and S. UEDA. 1992. Concerted evolution of the primate immunoglobulin alpha–gene through gene conversion. J. Biol. Chem. **267**:7359–7367.

KAWAMURA, S., H. TANABE, Y. WATANABE, K. KUROSAKI, N. SAITOU, and S. UEDA. 1991. Evolutionary rate of immunoglobulin alpha noncoding region is greater in hominoids than in Old World monkeys. Mol. Biol. Evol. **8**:743–752.

KILIAN, M., J. MESTECKY, and R. E. SCHROHENLOHER. 1979. Pathogenic species of the genus *Haemophilus* and *Streptococcus pneumoniae* produce immunoglobulin A1 protease. Infect. Immun. **26**:143–149.

KLEIN, J., Y. SATTA, C. O'HUIGIN, and N. TAKAHATA. 1993. The molecular descent of the major histocompatibility complex. Annu. Rev. Immunol. **11**:269–295.

LOMHOLT, H., K. POULSEN, and M. KILIAN. 1995. Comparative characterization of the iga gene encoding IgA1 protease in *Neisseria meningitidis*, *Neisseria gonorrhoeae* and *Haemophilus influenzae*. Mol. Microbiol. **15**:495–506.

MALE, C. J. 1979. Immunoglobulin A1 protease production by *Haemophilus influenzae* and *Streptococcus pneumoniae*. Infect. Immun. **26**:254–261.

MESSIER, W., and C. STEWART. 1997. Episodic adaptive evolution of primate lysozymes. Nature **385**:151–154.

MORTENSEN, S. B., and M. KILIAN. 1984. Purification and characterization of an immunoglobulin A1 protease from *Bacteroides melaninogenicus*. Infect. Immun. **45**:550–557.

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

OSBORNE, B. A., T. E. GOLDE, R. L. SCHWARTZ, and S. RUDIKOFF. 1988. Evolution of the IgA heavy chain gene in the genus Mus. Genetics **119**:925–931.

PLAUT, A. G. 1983. The IgA1 proteases of pathogenic bacteria. Annu. Rev. Microbiol. **37**:603–622.

PLAUT, A. G., J. V. GILBERT, M. S. ARLENSTEIN, and J. D. CAPRA. 1975. *Neisseria gonorrhoeae* and *Neisseria meningitidis*: extracellular enzyme cleaves human immunoglobulin A. Science **190**:1103–1105.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SUMIYAMA, K., S. KAWAMURA, O. TAKENAKA, and S. UEDA. 1998. A high sequence variety in the immunoglobulin C alpha hinge region among Old World monkeys. Anthropol. Sci. **106**:31–39.

TAKAHASHI, N., S. UEDA, M. OBATA, T. NIKAIDO, S. NAKAI, and T. HONJO. 1982. Structure of human immunoglobulin

gamma genes: implications for evolution of a gene family. Cell **29**:671–679.

TANAKA, T., and M. NEI. 1989. Positive darwinian selection observed at the variable-region genes of immunoglobulins. Mol. Biol. Evol. **6**:447–459.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

YANG, Z. 1999. PAML: phylogenetic analysis by maximum likelihood. Version 2. University College London, U.K.

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.

RODNEY HONEYCUTT, reviewing editor

Accepted February 15, 2002