Sequence Variation at Two Eosinophil-Associated Ribonuclease Loci in Humans

Jianzhi Zhang and Helene F. Rosenberg

Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892

Manuscript received May 30, 2000
Accepted for publication August 18, 2000

ABSTRACT

Host defense against invading pathogens is of great importance to the survival of higher organisms. We have been studying the evolution of mammalian eosinophil-associated ribonucleases (EARs), which are members of the ribonuclease A superfamily with known antipathogen activities. Earlier studies showed that positive selection promoted rapid diversification of paralogous EAR genes in both primates and rodents. Intraspecifically, however, it is unknown whether these genes also have divergent alleles. The recent discovery that the gene repertoire of the EAR family is much larger in rodents than in primates has led us to consider the possibility that primates maintain a large number of polymorphic alleles to compensate for a smaller gene repertoire. Here we present sequences of 2417 nucleotides at the two EAR loci, the eosinophil-derived neurotoxin (EDN, RNase 2) and eosinophil cationic protein (ECP, RNase 3), from >50 human individuals. Our data demonstrate that the nucleotide diversities (0.06–0.11%) at these loci are typical for human nuclear genes, thus permitting us to reject this polymorphism hypothesis. No significant departure from neutrality is noted and no signs of overdominant selection are observed. Similar patterns were observed in a preliminary study of chimpanzees. In summary, our results suggest that the antipathogen functions of the primate EARs are conserved after they are established and that these proteins are not currently undergoing rapid diversification in response to challenge from invading microorganisms.

ORIGINS of host defense systems in various animal and plant groups have attracted increasing attention in recent years partly because pathogens are recognized as important ecological factors in shaping the evolution of higher organisms (*e.g.*, Greenberg *et al.* 1995; Rosenberg and Dyer 1995; Meyers *et al.* 1998; Palacios *et al.* 1998; Zhang *et al.* 1998; Stahl *et al.* 1999). The central issue here is to understand at the molecular level how proteins with microbicidal activity originate and evolve under the selective pressures provided by invading pathogens. The eosinophil-associated ribonuclease (EAR) gene family is a good system for studying this problem because most gene members emerged within the past 50 million years, and some appeared as recently as 5 mya (Larson *et al.* 1996; Zhang *et al.* 1998, 2000), thus making it easier to study the evolutionary and molecular mechanisms underlying the functional specificity of individual genes. The primate EAR family includes only two functional members, eosinophil-derived neurotoxin (EDN or RNase 2) and eosinophil cationic protein (ECP or RNase 3), generated from a gene duplication event that occurred ~30 mya in the evolutionary lineage of hominoids and Old World monkeys (Hamann *et al.* 1990; Rosenberg *et al.* 1995; Zhang *et al.* 1998). Only one EAR gene (named EDN) has been found to exist in the genomes of New World monkeys and prosimians examined so far (Rosenberg *et al.* 1995; H. F. Rosenberg, unpublished results). In humans, the EDN and ECP genes are linked on chromosome 14q24–q31 (Hamann *et al.* 1990), and their protein products can be found in the large specific granules of eosinophilic leukocytes (reviewed in Rosenberg and Domachowske 1999). ECP has a cell-membrane-disruptive function that is likely to be responsible for its activity against bacteria and parasites *in vitro* (Young *et al.* 1986; Waters *et al.* 1987; Lehrer *et al.* 1989). EDN reduces the infectivity of both respiratory syncytial virus (RSV) and human immunodeficiency virus (HIV) *in vitro* (Domachowske *et al.* 1998c; Lee-Huang *et al.* 1999), an activity that is ribonuclease-dependent, at least in the case of RSV (Domachowske *et al.* 1998c). Sequence analysis revealed rapid evolution of EDN and ECP (Rosenberg *et al.* 1995), and positive Darwinian selection was detected in ECP during a relatively short period of evolutionary time following gene duplication, ending prior to the divergence of the hominoids from the Old World monkeys (Zhang *et al.* 1998). The EAR genes of rodents have also been studied. In contrast to the small size of the EAR family in primates, a survey of seven species of the rodent family Muridae shows that the average gene number per species is ~8, with one species having as many as 17 distinct genes (Larson *et al.* 1996; Batten *et al.* 1997; Singhania *et al.* 1999; Zhang *et al.* 2000). Furthermore, the rodent EAR family has undergone frequent gene duplications and massive gene deactivation, which has resulted in distinct or only partially overlapping gene inventories

Corresponding author: Helene F. Rosenberg, Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bldg. 10, Rm. 11N104, 9000 Rockville Pike, Bethesda, MD 20892. E-mail: hr2k@nih.gov

TABLE 1

Primers used in this study

EDN1f	5′ TCCAGAGTTTGGATCTAACCA
EDN2f	5′ CAGCTGCCCCTGAACCCCAG
EDN3f	5′ TTGTCTGCTTCTTCTGTTGG
EDN4f	5′ TTCAAACTGCAGGTATGCGC
EDN1r	5′ TCGTTGACTTACCCAGTCTC
EDN2r	5′ ATTGAGCCCAGGTAAACTGT
EDN3r	5′ GGTCTCGTCGTTGATCTCTG
EDN4r	5′ TAATGAAGACACACAGACACTTT
ECP1f	5′ CCCAGAGTCCAGATCCCACCG
ECP2f	5′ GTAGGGGAGACCAGCTGCCC
ECP3f	5′ AAATTTGTCTGCTTCTTCTG
ECP4f	5′ GGTGCACAGAATATTTCAAA
ECP1r	5′ TTGACTTACCCAGTCTCTGA
ECP2r	5′ AGCCCTCGTAAACTGTGGGG
ECP3r	5′ TCTCTGTTGTCACATGCAAC
ECP4r	5′ GTCACTAAATGACAGCAGAGC

EDN and ECP primers are derived from sequences with GenBank accession nos. X55987 and X16545, respectively.

in several closely related species (Zhang *et al.* 2000). Positive selection was also detected in the diversification of the paralogous genes in these rodent species. Interestingly, these evolutionary patterns of rapid gene birth and death, gene sorting, and positive selection are shared by a number of mammalian immune system genes such as the major histocompatibility complex (MHC), immunoglobulin (Ig), and T cell receptor genes. With this connection in mind, one naturally wonders how variable individual EAR genes are within a species—are they more similar to the highly polymorphic MHC genes or to the largely monomorphic Ig genes? The dramatic difference in size between rodent and primate EAR families has also prompted us to consider the possibility that the relatively small gene repertoire in primates may be compensated by highly polymorphic alleles. To answer these questions, we conducted an investigation of the DNA sequence variation at the EDN and ECP loci of humans and chimpanzees.

MATERIALS AND METHODS

Human genomic DNAs of normal individuals from different ethnic groups were purchased from Coriell Cell Repositories (Camden, NJ) or donated by Drs. David McDermott and Philip Murphy of the National Institutes of Health (McDermott *et al.* 1998). Genomic DNAs of six chimpanzees (*Pan troglodytes*) were purchased from BIOS Labs (New Haven, CT), Coriell Cell Repositories, or provided by Drs. Oliver Ryder and Leona Chemnick of San Diego Zoo. The EDN and ECP genes were amplified by polymerase chain reaction (PCR) with the primer sets EDN1f and EDN4r and ECP1f and ECP4r, respectively, which are derived from the published human sequences (Table 1). High-fidelity Taq was used to reduce the likelihood of PCR errors and the PCR reactions were performed under conditions recommended by the manufacturer (Life Technologies, Rockville, MD). PCR products were purified by the High Pure PCR product purification kit (Roche Molecular Biochemicals, Indianapolis) and were sequenced directly in both directions by the dideoxy chain termination method with dRhodamine dyes on Perkin-Elmer 377 automatic sequencers. The sequencing primers for the EDN gene were EDN1f-EDN4f and EDN1r-EDN4r, whereas those for the ECP genes were ECP1f-ECP4f and ECP1r-ECP4r (Table 1). Homozygotes and heterozygotes at individual sites were distinguished by their chromatographic patterns. Sequence variations that were found only in one chromosome (singletons) were reexamined so as to rule out the possibility of PCR errors. Most allelic haplotypes were determined by the approach of Clark (1990). For those individuals whose allelic haplotypes cannot be inferred by Clark's method, we subcloned PCR products into pCR4Blunt-TOPO vector (Invitrogen, Carlsbad, CA) and then sequenced individual clones.

Polymorphism data were analyzed according to Nei (1987), Hartl and Clark (1997), and Li (1997). Specifically, nucleotide diversity (π) was estimated following formula 10.6 of Nei (1987). The sampling variance of π, which is the error generated at the allele frequency survey, was estimated by (10.7) of Nei (1987). The stochastic variance of π was estimated by (10.10) of Nei (1987). Unless otherwise mentioned, the total variance is presented. Tajima's (1989) and Fu and Li's (1993) tests as implemented in DnaSP3.0 (Rozas and Rozas 1999) were used to test neutrality. McDonald and Kreitman's (1991) test was used to examine the homogeneity in evolutionary pattern between within-species variation and between-species divergence. The modified Nei-Gojobori method (Zhang *et al.* 1998) was used to compute rates of synonymous and nonsynonymous nucleotide substitutions.

RESULTS

Sequence variation at the EDN locus: We sequenced 1214 nucleotides at the EDN locus from 67 humans, including 20 African Americans, 30 Caucasians, 12 Asians, and 5 Native Americans. In addition, orthologous sequences from six unrelated chimpanzees were examined. Sequence variations observed among human individuals and between human and chimpanzee are depicted in Figure 1, along with the gene structure annotations. Nine polymorphic sites were found among the 134 human chromosomes. All of these variations are single nucleotide polymorphisms (SNPs) rather than insertions/deletions, and only two different nucleotides were observed at each of these sites. One commonly used parameter that measures the polymorphic level of a DNA region is nucleotide diversity (π), which is defined as the average number of differences per site between two alleles randomly chosen from the population. For the EDN data, π is estimated to be 0.00063 (Tables 2 and 3), which is within the normal range of diversity observed in a number of nuclear gene loci in humans (0.0005–0.0020; Li and Sadler 1991; Nickerson *et al.* 1998; Cargill *et al.* 1999; Halushka *et al.* 1999; Rieder *et al.* 1999). Among the five ethnic groups, African Americans have the highest diversity ($\pi = 0.00111$, Table 2), followed by Asians (0.00063), Native Americans (0.00046), and Caucasians (0.00018). This pattern is generally consistent with observations from other genes,

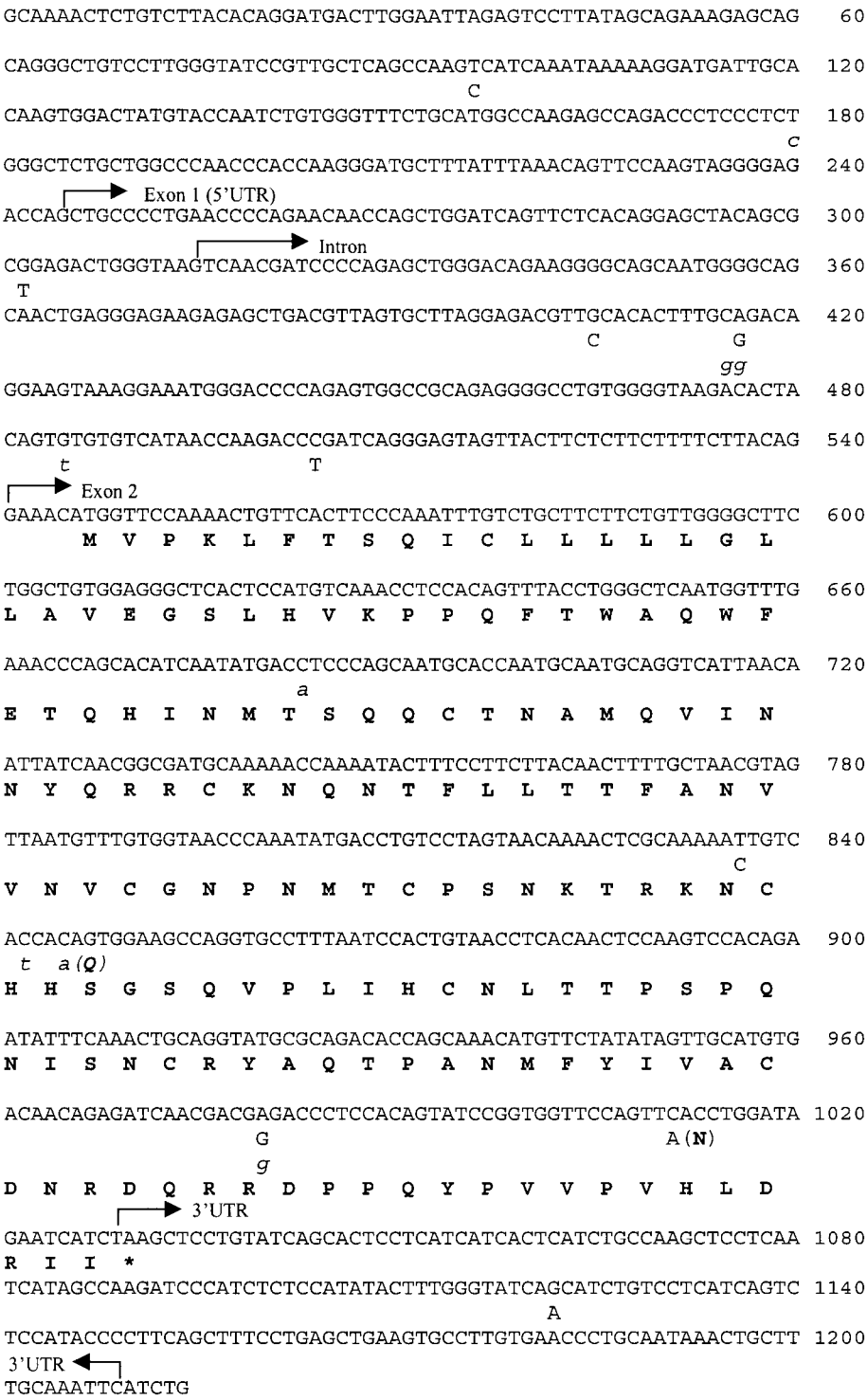

FIGURE 1.—DNA sequence variation at the EDN locus. Shown here is the human haplotype 1 sequence, with human polymorphisms listed in capital letters under the main sequence. The translated amino acids are shown by capital boldface letters. *, stop codon. The chimpanzee sequence (majority haplotype) is presented with italicized letters (lowercase for DNA and capital for protein) at those sites that differ from the human haplotype 1. The sequences have been deposited in GenBank with the following accession numbers: human EDN haplotypes 1–9, AF294007–AF294015; chimpanzee EDN haplotypes 1–3, AF294016–AF294018.

though the dramatically reduced diversity in Caucasians appears unusual. In fact, the nucleotide diversity in Caucasians is significantly smaller than any of the other three ethnic groups examined ($P < 0.01$).

Nine haplotypes were found in humans, and their relationships can be inferred by the parsimony principle (Figure 2). Haplotype X, which links some of the observed haplotypes, is missing from the samples examined. This may be due to small sample size and/or low frequency of this haplotype. Assuming random sampling, we estimated that haplotype X has a frequency of no more than 0.022 (95% confidence interval) in humans. Of the nine haplotypes, haplotypes 1 and 2 are found in all four ethnic groups with haplotype 1 being the predominant one in every population. Haplotypes 3, 4, 6, 8, and 9 are found only in African Ameri-

TABLE 2

EDN haplotypes in humans

Haplotypes	Nos. of copies in different populations				
	Af. Am.	Asian	Nat. Am.	Caucasian	Total (freq.)
1	24	12	5	53	94 (0.701)
2	3	11	5	6	25 (0.187)
3	3	0	0	0	3 (0.022)
4	1	0	0	0	1 (0.007)
5	0	0	0	1	1 (0.007)
6	1	0	0	0	1 (0.007)
7	4	1	0	0	5 (0.037)
8	3	0	0	0	3 (0.022)
9	1	0	0	0	1 (0.007)
Total:	40	24	10	60	134 (1.000)
$\pi \times 10{,}000$:	11.1 ± 2.0	6.3 ± 1.8	4.6 ± 0.6	1.8 ± 0.6	6.3 ± 1.0

The standard error of π is computed from sampling variance only, which does not include stochastic variance. Af. Am., African-American; Nat. Am., Native American; freq., frequency.

cans and haplotype 5 is restricted to Caucasians. Haplotype 7 is largely restricted to African Americans with only one copy found among the samples from Asians (Table 2). Recombination is not necessary to explain the observed haplotypes. That is, the minimal number of recombination events (D_m) is 0. Consistent with this result, no four-gametic types are found between any pairs of polymorphic sites, indicating that the polymorphisms are in complete linkage disequilibrium (Hudson and Kaplan 1985). The recombination rate C (per site) can also be estimated by Hudson's (1987) method from the variance in nucleotide difference among alleles, and it is estimated to be 8×10^{-7} for the EDN locus. With the limited number of polymorphisms as in the present data, the estimate of C may have a substantial variance (Hudson 1987). The relative rate of recombination to mutation per site is estimated to be 0.0013 for EDN.

Statistical tests of neutrality are applied to the above polymorphic data. Neither Tajima's test nor Fu and Li's test gives a significant result (Tajima's $D = -1.28$, $P > 0.10$; Fu and Li's $D^* = -1.72$, $P > 0.10$; Fu and Li's $F^* = -1.82$, $0.10 > P > 0.05$). However, the negative values of these statistics suggest relatively low nucleotide diversity giving the observed number of polymorphic sites. If not entirely due to stochastic error, this may be an indicator of background purifying selection, recent rapid spread of advantageous alleles in population (selective sweeps), or reduction in population size (bottlenecks).

TABLE 3

Sequence variation at the EDN locus

Sequence type	No. of sites	D^a	S^b	$\pi \times 10{,}000^c$
All sites	1214	8	9	6.3 ± 5.1
Coding	483	4	3	4.1 ± 5.7
Synonymous	110	3	2	16.6 ± 23.8
Nonsynonymous	373	1	1	0.4 ± 1.9
Noncoding	731	4	6	7.7 ± 6.9

[a] No. of human-chimpanzee differences.
[b] No. of polymorphic sites in humans.
[c] Nucleotide diversity in humans. Both stochastic and sampling variances are considered.

Six of the nine polymorphic sites involve transitional changes whereas the rest are transversions. Three polymorphisms are within the protein-coding sequence, with two representing synonymous changes and only one representing a nonsynonymous change. The other six polymorphisms are distributed in the promoter, intron, and untranslated mRNA (UTR) regions (Figure 1 and Table 3). Interestingly, the single nonsynonymous polymorphism is a C to A mutation (at site 1011) resulting in a His to Asn change at one of the three amino acid residues that form the catalytic site of the RNase. This change renders EDN nonfunctional as a ribonuclease, and therefore the mutant is presumably under strong purifying selection. Consistent with this prediction, only one copy of this allele is found among 134 chromosomes, suggesting that it may have been a recent mutation. The introns of EDN and ECP include transcriptional regulation sequences (EoTF and PU.1 sites) that are known to reduce gene expression when disrupted (Tiffany *et al.* 1996; Handen and Rosenberg 1997; Van Dijk *et al.* 1998), but none of the observed polymorphisms reside at these sites.

Among the 12 chromosomes of the six chimpanzees, only two polymorphisms were found (C → T at site 408 and C → T at site 828), and both are singletons. The nucleotide diversity is estimated to be 0.00027 ± 0.00030. One of the two polymorphisms is in the intron, while the other results in an Arg → Cys change in the protein sequence. The mature protein of EDN has eight

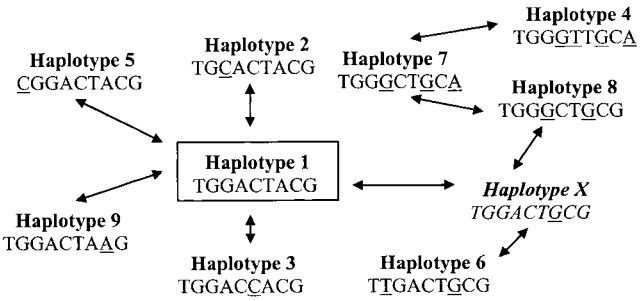

FIGURE 2.—Parsimonious relationships of the EDN haplotypes. The most frequently observed haplotype is boxed. Nucleotide differences from haplotype 1 are underlined. Haplotype *X* is hypothetical and was not observed in the sample examined. The nine polymorphic sites are at the positions 96, 302, 405, 416, 504, 836, 980, 1011, and 1122 of the sequence in Figure 1.

cysteines that form four disulfide bonds, which are of critical importance to the structure and function of EDN. An extra Cys in EDN could be harmful because it may interfere with protein folding.

Eight nucleotide differences are found between the most frequently observed haplotype sequences of chimpanzees and humans. Four of these substitutions occurred in the protein-coding region with three being synonymous and one nonsynonymous. Thus, the rate of nonsynonymous change per site is ~11% of that of synonymous change, suggesting that the gene is under purifying selection. Nevertheless, neutral evolution cannot be rejected statistically ($P = 0.052$, Fisher's test), apparently due to the small numbers of substitutions in this case. McDONALD and KREITMAN's (1991) test suggests that the evolutionary patterns in synonymous and nonsynonymous changes revealed from intraspecific and interspecific variations are compatible ($P > 0.50$). Availability of the chimpanzee sequences also makes it possible to infer the ancestral and descendant nucleotides at each of the nine polymorphic sites in humans and it appears that haplotype 8 is the ancestral form with regard to these sites. At other sites, there might have been mutations that occurred in the human lineage and are now fixed in populations.

Sequence variation at the ECP locus: We determined 1203 nucleotides at the ECP locus in 54 humans (21 African Americans, 16 Caucasians, 12 Asians, and 5 Native Americans) and six chimpanzees. Sequence variations observed are shown in Figure 3. Seven polymorphisms were found among the 108 human chromosomes, all of which are SNPs with two different nucleotides at each of these sites. The nucleotide diversity (π) is estimated to be 0.00113 (Tables 4 and 5). African Americans again have the highest diversity (0.00129, Table 4) among the four ethnic groups, followed by Caucasians (0.00101), Asians (0.00095), and Native Americans (0.00046).

Eight haplotypes were found in humans, and Figure 4 shows their parsimonious relationships. Haplotypes 2 and 4 are found in all four populations with haplotype 4 being the predominant one in the total population. Haplopytes 6, 7, and 8 are observed only in African Americans and haplotype 5 is restricted to African Americans and Asians. Haplotypes 1 and 3 are present in all populations except Native Americans (Table 4).

In the tests of neutrality, Tajima's test and Fu and Li's test give results in different directions (Tajima's $D = 0.042$, $P > 0.10$; Fu and Li's $D^* = -0.58$, $P > 0.10$; Fu and Li's $F^* = -0.44$, $P > 0.10$), but neither test rejects the null hypothesis of neutrality. Of the seven polymorphisms, three are transitions and four are transversions. Only one polymorphism resides in the coding region, and it results in an Arg/Thr polymorphism. Since the chimpanzee has Arg at this position, it can be inferred that Arg is the ancestral state and Thr appeared in humans after their separation from chimpanzees. Arginine residues have been considered to be important in ECP's antibacterial activity (ROSENBERG *et al.* 1989; ZHANG *et al.* 1998). However, the relatively high frequency (0.36) of Thr alleles (haplotypes 1, 3, and 7) in humans suggests that the difference in fitness between the Thr and Arg alleles is small. As the mature protein of ECP has 19 arginines, loss or gain of a few arginine residues will not have much impact on cationicity and thus may not ultimately alter function. Indeed, chimpanzee ECP has only 17 arginines.

Two copies of one polymorphism were found among the 12 chromosomes of six chimpanzees (T → C at site 1194), and it resides in the nontranscribed region that is 3′ to 3′ UTR. The nucleotide diversity is estimated to be 0.00025 ± 0.00029.

Compared with the human sequence (haplotype 4), the chimpanzee sequence (the majority haplotype) has two deletions of 1 and 3 nucleotides, respectively. In addition, 15 nucleotide differences are observed. Three of these differences are in the protein-coding region, and all are nonsynonymous changes. No heterogeneity between within-species variation and between-species divergence is detected in terms of the synonymous to nonsynonymous ratio ($P > 0.50$; McDonald and Kreitman test). Using the chimpanzee sequence, we predict that haplotype 5 represents the ancestral allele in humans with regard to the observed seven polymorphisms.

As in the case of EDN, one need not invoke recombination to explain the observed human ECP haplotypes, and no four-gametic types are found between any pairs of polymorphic sites, indicating that the polymorphisms are in complete linkage disequilibrium. The recombination rate C (per site) is estimated to be 0.1, and the relative rate of recombination to mutation is 94. These values are substantially greater than those for EDN, but more data are needed to test the statistical significance of the differences. Recombination between EDN and ECP can also be investigated, because both loci have been sequenced in 53 human individuals and the EDN-

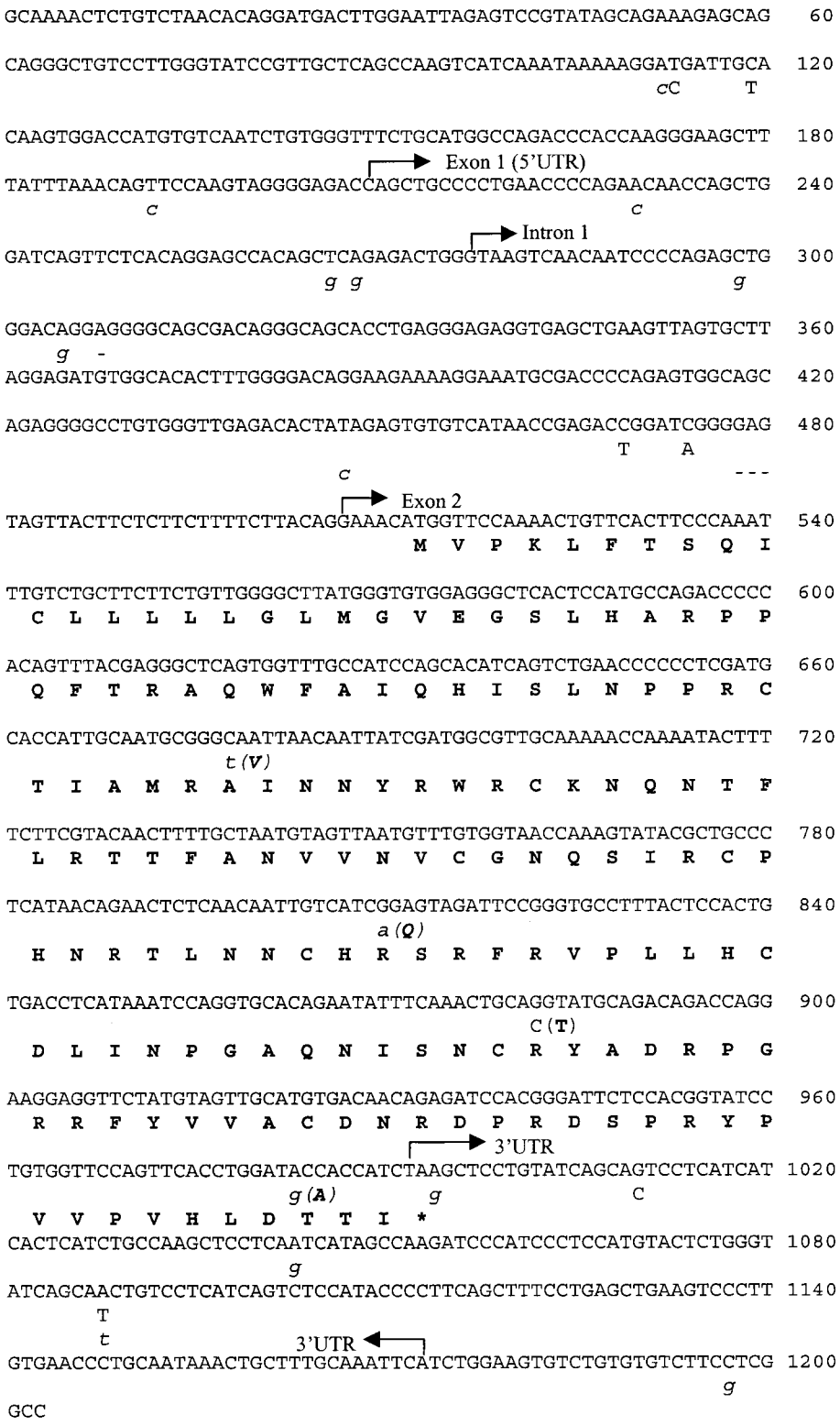

FIGURE 3.—DNA sequence variation at the ECP locus. Shown here is the human haplotype 4 sequence, with human polymorphisms listed in capital letters under the main sequence. The translated amino acids are shown by capital boldface letters. *, stop codon; -, deletion. The chimpanzee sequence (majority haplotype) is presented with italicized letters (lowercase for DNA and capital for protein) at those sites that show differences from the human haplotype 4. The sequences have been deposited in GenBank with the following accession numbers: human ECP haplotypes 1–8, AF294019–AF294026; chimpanzee ECP haplotypes 1–2, AF294027–AF294028.

ECP haplotypes can be inferred for those who are not heterozygous at both loci. Thus, we determined the EDN-ECP haplotypes of 36 individuals and this data set contains 7 and 5 polymorphisms in the EDN and ECP loci, respectively. Of the 35 pairs of polymorphic sites compared, 11 pairs have all four gametic types, and at least one recombination event (between sites 1122 of EDN and 474 of ECP) is required to explain the haplo-

TABLE 4

ECP haplotypes in humans

Haplotypes	Nos. of copies in different populations				
	Af. Am.	Asian	Nat. Am.	Caucasian	Total (freq.)
1	11	4	0	6	21 (0.194)
2	5	5	5	12	27 (0.250)
3	11	3	0	3	17 (0.157)
4	7	11	5	11	34 (0.315)
5	4	1	0	0	5 (0.046)
6	1	0	0	0	1 (0.009)
7	1	0	0	0	1 (0.009)
8	2	0	0	0	2 (0.019)
Total:	42	24	10	32	108 (1.000)
$\pi \times 10{,}000$:	12.9 ± 1.2	9.5 ± 1.4	4.6 ± 0.6	10.1 ± 1.2	11.3 ± 0.6

The standard error of π is computed from sampling variance only, which does not include stochastic variance. Af. Am., African-American; Nat. Am., Native American; freq., frequency.

types. This demonstrates that while EDN and ECP may be closely linked, recombination between loci did happen, even in the relatively recent evolutionary history of humans.

DISCUSSION

In this study, we examined intraspecific sequence variations at the human EDN and ECP loci. Our data showed that the nucleotide diversities at these loci are ~0.06–0.11%, which are typical for human nuclear genes (Li and Sadler 1991; Nickerson *et al.* 1998; Cargill *et al.* 1999; Halushka *et al.* 1999; Rieder *et al.* 1999), and that several statistical tests fail to reject neutrality (Kimura 1983) for the sequences generated. These results clearly demonstrate that human EDN and ECP genes are no more polymorphic than the average human gene and that no overdominant selection is involved in maintaining the different alleles. At the interspecific level, we noted earlier that the evolutionary pattern of the rodent EAR gene family is similar to those of many immune system gene families such as MHC, T cell receptor, and Ig (Zhang *et al.* 2000) as all four gene families are subject to evolution via rapid gene sorting and positive Darwinian selection (Zhang *et al.* 2000). When examined within species, it appears that the polymorphic levels of EDN and ECP are more similar to those of Ig genes, which have been described as largely monomorphic (Li and Hood 1995; Sasso *et al.* 1995). This is in contrast to what has been observed for some MHC loci, which are highly polymorphic (π = 4–7%; Nei and Hughes 1991), with overdominant or frequency-dependent selection apparently active in maintaining the divergent alleles (Hughes and Nei 1988).

One unexpected result from this investigation was the observation that nucleotide diversity among EDN alleles of Caucasians was significantly lower than that observed in other ethnic groups (Table 2). There are several possible explanations for this phenomenon. The first possibility is that our sample of Caucasians is so biased that only a regional population of Caucasians is effectively sampled. This, however, is unlikely because the Caucasian ECP locus shows a normal level of polymorphism and the EDN and ECP genes are linked. Second, recent population bottlenecks may have reduced the polymorphism, but this is also unlikely because if this were the case, the diversity observed at the ECP locus would also be reduced in the Caucasian population. Third, background selection against deleterious mutations may reduce the gene diversity, but it is hard to imagine why background selection on the same genes could be so different in different human populations. Fourth, a recent selective sweep of an advantageous allele in Caucasians may have reduced the gene diversity by quickly replacing other alleles. However, Caucasian EDNs are mostly of haplotype 1, which is also the majority in other populations, suggesting that it is a relatively old allele. Caucasians and Asians are more closely related genetically than they are related to Africans (*e.g.*, Nei and Roychoudhury 1993). Comparing

TABLE 5

Sequence variation at the ECP locus

Sequence type	No. of sites	D^a	S^b	$\pi \times 10{,}000^c$
All sites	1203	15	7	11.3 ± 7.7
Coding	480	3	1	9.7 ± 9.4
Synonymous	129	0	0	0
Nonsynonymous	351	3	1	13.3 ± 12.9
Noncoding	723	12	6	12.3 ± 9.5

[a] No. of human-chimpanzee differences.
[b] No. of polymorphic sites in humans.
[c] Nucleotide diversity in humans. Both stochastic and sampling variances are considered.

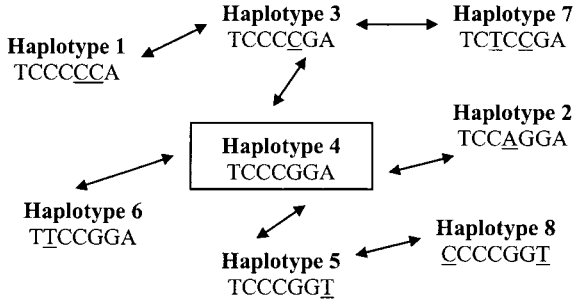

FIGURE 4.—Parsimonious relationships of the ECP haplotypes. The most frequently observed haplotype is boxed. Nucleotide differences from haplotype 4 are underlined. The seven polymorphic sites are at the positions 113, 119, 469, 474, 882, 1010, and 1088 of the sequence in Figure 2.

Asians with Caucasians, one can see that the frequency of haplotype 1 increases dramatically in Caucasians while that of haplotype 2 decreases. The difference between haplotypes 1 and 2 is an intronic G/C transversion at position 405, which is not in a region known to be involved in gene expression. According to Figure 2 and the prediction that haplotype 8 was the ancestral allele in humans, it follows that haplotype 2 appeared later than haplotype 1 (Figure 2). If this is the case, it is difficult to conceive of a mechanism whereby a selective sweep of haplotype 1 replaced haplotype 2. Of course, selection may not act directly at the EDN locus but at a closely linked site. In any case, although Tajima's D and Fu and Li's D^* and F^* statistics are all negative for the Caucasian EDN sequences, they are not significantly <0, suggesting that the present data are not sufficient for establishing either the background selection or the selective sweep hypothesis. As such, the cause of the reduced diversity at the EDN locus in Caucasians remains an open question.

Low nucleotide diversity ($\pi = 0.00025$–0.00027) was also found at both EDN and ECP loci in the chimpanzees examined. Previous studies at a number of loci of wild populations of chimpanzees showed that the nucleotide diversity in chimpanzees is often greater than that in humans, though exceptions do exist (see KAESSMANN *et al.* 1999 and references therein). In the present case, the substantially lower nucleotide diversity in chimpanzees than in humans ($P < 0.05$; randomization test) may be due to the limited sample size. In addition, it is possible that the six chimpanzees studied here were originally captured from a small geographic region, although they are not known to be related to one another.

As mentioned earlier, EDN and ECP genes emerged from gene duplication ~30 mya, and their respective antipathogen activities, such as they are currently understood, either originated or were substantially enhanced in a relatively short period of time after gene duplication (ROSENBERG and DYER 1995; ROSENBERG *et al.* 1995; DOMACHOWSKE *et al.* 1998a–c; ZHANG *et al.* 1998). In the case of ECP, directional Darwinian selection was detected after gene duplication, with the selection process resulting in a dramatic increase in the number of arginine residues (ZHANG *et al.* 1998). These positively charged arginine residues provide ECP with its distinct cationicity (pI = 11.4) and are likely to be responsible for ECP's membrane lytic effect and antibacterial activity (YOUNG *et al.* 1986; LEHRER *et al.* 1989). Cation-mediated membrane lysis is a nonspecific mechanism, which suggests that no species-specific interactions between ECP and its targets are involved. Our result demonstrating that the nucleotide diversity at the ECP locus is no higher than average suggests that there is no fitness advantage for the population to maintain many variant forms of ECP, which is consistent with our understanding of ECP's membrane lytic mechanism. However, as our understanding of eosinophil physiology and, by extension, of ECP, remains incomplete, we may ultimately need to reconsider the meaning of these findings.

Our lab has recently explored a novel function of EDN, namely, its ability to reduce the infectivity of RNA viruses of the family Paramyxoviridae for its target epithelial cells *in vitro* (DOMACHOWSKE *et al.* 1998c). While this activity is directly dependent on EDN's ribonuclease activity, its detailed molecular mechanism is far from clear. Evolutionary analysis of the gene sequences may shed some light on this process. If EDN's primary interaction is with the viral envelope, then a high mutation rate of RNA viruses and the "red queen" process (VAN VALEN 1973) would result in rapid evolution of EDN under positive selection. But, in fact, the ratio of the rates of nonsynonymous to synonymous substitution (0.4) falls well below 1.0 in comparisons between sequences of Old World monkeys and hominoids, failing to support this hypothesis. Rather, the unremarkable level of polymorphism at the EDN locus suggests that either EDN targets at only one particular strain or family of viruses (already shown to be unlikely, DOMACHOWSKE *et al.* 1998c; LEE-HUANG *et al.* 1999), or that somehow the variability of viral proteins has no effect on EDN's antiviral efficiency. Alternatively, EDN's antiviral activity may involve a primary interaction with one or more cellular targets, which remain more or less constant both within and among species; our most recent results stand in support of this latter hypothesis (J. B. DOMACHOWSKE and H. F. ROSENBERG, unpublished results).

As discussed earlier, the discovery of a dramatic difference in the size of the EAR family between rodents and primates has led to the hypothesis that the relatively small gene repertoire in primates may be compensated by highly polymorphic alleles. Our present study, however, rejects this hypothesis. Instead, our results imply that the ECPs and EDNs of primates use conservative mechanisms to fight against pathogens, and no extra variation is needed. In this respect, it is unclear what pathogens rodent EARs target and why rodents need more EAR genes than primates do. It is possible that

somewhat different mechanisms are involved in the host-defense process of rodent EARs, and more work is certainly needed to explore this possibility.

We are grateful to Drs. David McDermott, Philip Murphy, Leona Chemnick, and Oliver Ryder for providing some of the human and chimpanzee DNA samples used in this study and to Dr. Kimberly Dyer for excellent technical assistance and discussion.

LITERATURE CITED

Batten, D., K. D. Dyer, J. B. Domachowske and H. F. Rosenberg, 1997 Molecular cloning of four novel murine ribonuclease genes: unusual expansion within the ribonuclease A gene family. Nucleic Acids Res. **25:** 4235–4239.

Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie *et al.*, 1999 Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat. Genet. **22:** 231–238.

Clark, A. G., 1990 Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. **7:** 111–122.

Domachowske, J. B., C. A. Bonville, K. D. Dyer and H. F. Rosenberg, 1998a Evolution of antiviral activity in the ribonuclease A gene superfamily: evidence for a specific interaction between eosinophil-derived neurotoxin (EDN/RNase 2) and respiratory syncytial virus. Nucleic Acids Res. **26:** 5327–5332.

Domachowske, J. B., K. D. Dyer, A. G. Adams, T. L. Leto and H. F. Rosenberg, 1998b Eosinophil cationic protein/RNase 3 is another RNase A-family ribonuclease with direct antiviral activity. Nucleic Acids Res. **26:** 3358–3363.

Domachowske, J. B., K. D. Dyer, C. A. Bonville and H. F. Rosenberg, 1998c Recombinant human eosinophil-derived neurotoxin/RNase 2 functions as an effective antiviral agent against respiratory syncytial virus. J. Infect. Dis. **177:** 1458–1464.

Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Greenberg, A. S., D. Avila, M. Hughes, A. Hughes, E. C. McKinney *et al.*, 1995 A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature **374:** 168–173.

Halushka, M. K., J. B. Fan, K. Bentley, L. Hsie, N. Shen *et al.*, 1999 Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat. Genet. **22:** 239–247.

Hamann, K. J., R. M. Ten, D. A. Loegering, R. B. Jenkins, M. T. Heise *et al.*, 1990 Structure and chromosome localization of the human eosinophil-derived neurotoxin and eosinophil cationic protein genes: evidence for intronless coding sequences in the ribonuclease gene superfamily. Genomics **7:** 535–546.

Handen, J. S., and H. F. Rosenberg, 1997 Intronic enhancer activity of the eosinophil-derived neurotoxin (RNS2) and eosinophil cationic protein (RNS3) genes is mediated by an NFAT-1 consensus binding sequence. J. Biol. Chem. **272:** 1665–1669.

Hartl, D. L., and A. G. Clark, 1997 *Principles of Population Genetics*, Ed. 3. Sinauer Associates, Sunderland, MA.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

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.

Kaessmann, H., V. Wiebe and S. Paabo, 1999 Extensive nuclear DNA sequence diversity among chimpanzees. Science **286:** 1159–1162.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.

Larson, K. A., E. V. Olson, B. J. Madden, G. J. Gleich, N. A. Lee *et al.*, 1996 Two highly homologous ribonuclease genes expressed in mouse eosinophils identify a larger subgroup of the mammalian ribonuclease superfamily. Proc. Natl. Acad. Sci. USA **93:** 12370–12375.

Lee-Huang, S., P. L. Huang, Y. Sun, P. L. Huang, H. F. Kung *et al.*, 1999 Lysozyme and RNases as anti-HIV components in beta-core preparations of human chorionic gonadotropin. Proc. Natl. Acad. Sci. USA **96:** 2678–2681.

Lehrer, R. I., D. Szklarek, A. Barton, T. Ganz, K. J. Hamann *et al.*, 1989 Antibacterial properties of eosinophil major basic protein and eosinophil cationic protein. J. Immunol. **142:** 4428–4434.

Li, H., and L. Hood, 1995 Multiplex genotype determination at a DNA sequence polymorphism cluster in the human immunoglobulin heavy-chain region. Genomics **26:** 199–206.

Li, W. H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Li, W.-H., and L. A. Sadler, 1991 Low nucleotide diversity in man. Genetics **129:** 513–523.

McDermott, D. H., P. A. Zimmerman, F. Guignard, C. A. Kleeberger, S. F. Leitman *et al.*, 1998 CCR5 promoter polymorphism and HIV-1 disease progression. Multicenter AIDS Cohort Study (MACS). Lancet **352:** 866–870.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature **351:** 652–654.

Meyers, B. C., K. A. Shen, P. Rohani, B. S. Gaut and R. W. Michelmore, 1998 Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell **10:** 1833–1846.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., and A. L. Hughes, 1991 Polymorphism and evolution of the major histocompatibility complex loci in mammals, pp. 222–247 in *Evolution at the Molecular Level*, edited by R. Selander, A. Clark and T. Whittom. Sinauer Associates, Sunderland, MA.

Nei, M., and A. K. Roychoudhury, 1993 Evolutionary relationships of human populations on a global scale. Mol. Biol. Evol. **10:** 927–943.

Nickerson, D. A., S. L. Taylor, K. M. Weiss, A. G. Clark, R. G. Hutchinson *et al.*, 1998 DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat. Genet. **19:** 233–240.

Palacios, E., L. Digilio, H. M. McClure, Z. Chen, P. A. Marx *et al.*, 1998 Parallel evolution of CCR5-null phenotypes in humans and in a natural host of simian immunodeficiency virus. Curr. Biol. **8:** 943–946.

Rieder, M. J., S. L. Taylor, A. G. Clark and D. A. Nickerson, 1999 Sequence variation in the human angiotensin converting enzyme. Nat. Genet. **22:** 59–62.

Rosenberg, H. F., and J. B. Domachowske, 1999 Eosinophils, ribonucleases and host defense: solving the puzzle. Immunol. Res. **20:** 261–274.

Rosenberg, H. F., and K. D. Dyer, 1995 Eosinophil cationic protein and eosinophil-derived neurotoxin. Evolution of novel function in a primate ribonuclease gene family. J. Biol. Chem. **270:** 21539–21544.

Rosenberg, H. F., S. J. Ackerman and D. G. Tenen, 1989 Human eosinophil cationic protein. Molecular cloning of a cytotoxin and helminthotoxin with ribonuclease activity. J. Exp. Med. **170:** 163–176.

Rosenberg, H. F., K. D. Dyer, H. L. Tiffany and M. Gonzalez, 1995 Rapid evolution of a unique family of primate ribonuclease genes. Nat. Genet. **10:** 219–223.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Sasso, E. H., J. H. Buckner and L. A. Suzuki, 1995 Ethnic differences of polymorphism of an immunoglobulin VH3 gene. J. Clin. Invest. **96:** 1591–1600.

Singhania, N. A., K. D. Dyer, J. Zhang, M. S. Deming, C. A. Bonville *et al.*, 1999 Rapid evolution of the ribonuclease A superfamily: adaptive expansion of independent gene clusters in rats and mice. J. Mol. Evol. **49:** 721–728.

Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman and J. Bergelson, 1999 Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature **400:** 667–671.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tiffany, H. L., J. S. Handen and H. F. Rosenberg, 1996 Enhanced expression of the eosinophil-derived neurotoxin ribonuclease

(RNS2) gene requires interaction between the promoter and intron. J. Biol. Chem. **271:** 12387–12393.

Van Dijk, T. B., E. Caldenhoven, J. A. Raaijmakers, J. W. Lammers, L. Koenderman *et al.*, 1998 The role of transcription factor PU.1 in the activity of the intronic enhancer of the eosinophil-derived neurotoxin (RNS2) gene. Blood **91:** 2126–2132.

Van Valen, L., 1973 A new evolutionary law. Evol. Theory **1:** 1–30.

Waters, L. S., J. Taverne, P. C. Tai, C. J. Spry, G. A. Targett *et al.*, 1987 Killing of Plasmodium falciparum by eosinophil secretory products. Infect. Immun. **55:** 877–881.

Young, J. D. E., C. G. B. Peterson, P. Venge and G. J. Cohn, 1986 Mechanism of membrane damage mediated by human eosinophil cationic protein. Nature **321:** 613–616.

Zhang, J., H. F. Rosenberg and M. Nei, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95:** 3708–3713.

Zhang, J., K. D. Dyer and H. F. Rosenberg, 2000 Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. Proc. Natl. Acad. Sci. USA **97:** 4701–4706.

Communicating editor: S. Yokoyama