Comparative Genomics and Evolution of the Alpha-Defensin Multigene Family in Primates

Sabyasachi Das,*[,1] Nikolas Nikolaidis,[2] Hiroki Goto,[3] Chelsea McCallister,[2] Jianxu Li,[1] Masayuki Hirano,[1] and Max D. Cooper*[,1]

[1]Department of Pathology and Laboratory Medicine, Emory Vaccine Center, School of Medicine, Emory University

[2]Department of Biological Science, California State University, Fullerton

[3]Department of Biology and Center for Comparative Genomics and Bioinformatics, Pennsylvania State University–University Park

***Corresponding author**: E-mail: max.cooper@emory.edu; sdas8@emory.edu.

Associate editor: Yoko Satta

Abstract

Defensin genes encode small cationic antimicrobial peptides that form an important part of the innate immune system. They are divided into three families, alpha (α), beta (β), and theta (θ), according to arrangement of the disulfide bonding pattern between cysteine residues. Considering the functional importance of defensins, investigators have studied the evolution and the genomic organization of defensin genes. However, these studies have been restricted mainly to β-defensins. To understand the evolutionary dynamics of α-defensin genes among primates, we identified the α-defensin repertoires in human, chimpanzee, orangutan, macaque, and marmoset. The α-defensin genes in primates can be classified into three phylogenetic classes (class I, II, and III). The presence of all three classes in the marmoset indicates that their divergence occurred before the separation of New World and Old World monkeys. Comparative analysis of the α-defensin genomic clusters suggests that the makeup of the α-defensin gene repertoires between primates is quite different, as their genes have undergone dramatic birth-and-death evolution. Analysis of the encoded peptides of the α-defensin genes indicates that despite the overall high level of sequence divergence, certain amino acid residues or motifs are conserved within and between the three phylogenetic classes. The evolution of α-defensins in primates, therefore, appears to be governed by two opposing evolutionary forces. One force stabilizes specific amino acid residues and motifs to preserve the functional and structural integrity of the molecules and the other diversifies the sequences generating molecules with a wide range of activities against a large number of pathogens.

Key words: alpha-defensin gene, primate evolution, innate immunity, comparative genomics, positive selection, birth-and-death evolution.

Introduction

Innate immunity is an evolutionarily ancient defense strategy used by multiple species to defend against pathogens. A large number of antimicrobial peptides are known to be involved in the innate immunity. Among them defensins constitute one of the most important families of antimicrobial peptides because of their capacity to enhance phagocytosis and the production of proinflammatory cytokines, promote neutrophil recruitment, suppress anti-inflammatory mediators, and regulate complement activation (Yang et al. 2002; Zhang et al. 2002; Ganz 2003; Kim et al. 2005; Wehkamp et al. 2005; Presicce et al. 2009). The defensins are engaged in host defenses against a broad spectrum of pathogens (i.e. bacteria, fungi, and viruses) by either interacting directly with the pathogens, or acting in concert with other components of the immune system (Ganz 2003; Klotman and Chang 2006; Menendez and Brett Finlay 2007). In mammals and especially in primates, in addition to their natural antimicrobial activity, defensins participate in other physiological processes such as sperm protection, immune recognition, and cell signaling (Yang et al. 2002; Yudin et al. 2005). It has also been proposed that defensins may function as a link between innate and adaptive immune responses (Yang et al. 1999).

Mammalian defensins are small cationic peptides containing three pairs of intramolecular disulfide bonds mediated by six conserved cysteines (Ganz and Lehrer 1994; Ganz 2003). On the basis of the disulfide bonding pattern and the position of the cysteines, defensins are divided into alpha (α), beta (β), and theta (θ) families (Ganz 2003), although functional θ-defensin genes are found only in the rhesus macaque (*Macaca mulatta*) and olive baboon (*Papio anubis*) (Garcia et al. 2008; Tran et al. 2008). The α- and β-defensins form a triple-stranded β-sheet structure stabilized by disulfide bonds, whereas θ-defensins are structurally unrelated to the α and β families (Tang et al. 1999; Selsted 2004). All functional α-defensin genes are expressed as prepropeptides (Valore and Ganz 1992). The mature peptide results from sequential removal of the signal peptide and prosegment giving rise to a mature, tridisulfide-containing peptide (Michaelson et al. 1992).

Recent studies have shown that many physiological and morphological characters are generally controlled by genes belonging to multigene families (i.e., immunoglobulin, T-cell receptor, major histocompatibility complex, histone, ubiquitin, and olfactory receptor genes). Therefore, detailed investigation of the evolution of multigene families is an important step toward understanding the evolution of

phenotypic characters. It has been shown that several multigene families are subject to birth-and-death evolution and that the rates of gene gain and gene loss vary considerably between families (Ota and Nei 1994; Nei et al. 1997; Su and Nei 2001; Niimura and Nei 2005; Das et al. 2008a). Due to rapid birth-and-death evolution, the number of functional genes may be quite different between closely related species or even between individuals of the same species (Nei 2007; Nozawa et al. 2007; Das et al. 2008b; Das 2009). Considering the functional importance of defensins, investigators have studied the genomic organization and the evolution of defensin genes in several vertebrate species. However, these studies are largely restricted to β-defensins (Boniotto et al. 2003; Morrison et al. 2003; Semple et al. 2003; Xiao et al. 2004; Semple et al. 2005; Hollox and Armour 2008) with the exception of only a few studies on the α-defensin family (Patil et al. 2004; Lynn and Bradley 2007). Fortunately, the draft genome sequences of several primate species with greater than 5× coverage are available. These sequences allow us to carry out genome-wide comparisons of the α-defensin clusters in primates. Here, we present the complete repertoire of α-defensin genes in human, chimpanzee, orangutan, macaque, and marmoset. This analysis aims to provide a better understanding into the general pattern of the evolutionary processes that have shaped the differences in α-defensin repertoire between primate species and new insights into the evolutionary changes of the functional activities of α-defensin genes.

Materials and Methods

Identification of Alpha-Defensin Genes

The procedure of the retrieval of functional and nonfunctional α-defensin genes is drawn as a flow chart in supplementary figure S1 (Supplementary Material online). To identify all the α-defensin genes, we performed a two-round TBlastN search against the draft genome sequences of human (*Homo sapiens*; assembly: GRCh37, February 2009), chimpanzee (*Pan troglodytes*, assembly: CHIMP 2.1, March 2006), orangutan (*Pongo pygmaeus*, assembly: PPYG2, September 2007), macaque (*Macaca mulatta*, assembly: MMUL 1.0, February 2006), and marmoset (*Callithrix jacchus*, assembly: Callithrix_jacchus-3.2, February 2009) from Ensembl Genome Browser. In the first round, the amino acid sequences of five functional α-defensin genes from human available at the RefSeq protein database (accession numbers: NP_001916, NP_001917, NP_004075, NP_005208, and NP_066290) were used as queries. These five α-defensin sequences in the query data set align to the same genomic regions because they are similar to one another. For this reason, we retrieved only nonoverlapping genomic sequences that produced alignments with the lowest *E* values. Taking into account the alignment with the query α-defensin genes, we categorized the retrieved sequences into potential functional genes if they contained a start codon, aligned with query sequence without any frameshifts or premature stop codons, and encoded for six conserved cysteine residues in the mature peptide region. Other sequences were regarded as α-defensin pseudogenes. In the second round, the procedure was repeated by using the α-defensin genes identified in the first round to identify additional α-defensin genes. Intron–exon boundaries were identified with reference to the open reading frame. For each species after collecting all nonoverlapping sequences, we annotated the entire α-defensin cluster according to their genomic positions.

Phylogenetic Analysis

The translated amino acid sequences from identified α-defensin genes were aligned using ClustalW program. To ensure codon-to-codon alignment, the nucleotide alignments were retrieved from the amino acid sequences. The evolutionary distances were computed using the maximum composite likelihood method (Tamura et al. 2004) after elimination of nucleotide sequence alignment gaps only in pairwise sequence comparisons (pairwise deletion option). The phylogenetic trees were constructed by 1) neighbor joining (NJ) (Saitou and Nei 1987) and 2) maximum parsimony (MP) (Eck and Dayhoff 1966) methods using the MEGA4.0 program (Tamura et al. 2007). The reliability of the trees was assessed by bootstrap resampling of 1000 replications. One mouse α-defensin sequence (accession no. NM_010031) was used as outgroup.

Determination of Orthologous Sequences

Because α-defensin coding sequences are short and evolve rapidly (Patil et al. 2004), to determine true orthologous relationships, we have used three different methods: 1) phylogenetic analysis with the combination of any two species under study (i.e., human–chimpanzee, human–orangutan, chimpanzee–macaque, orangutan–macaque, macaque–marmoset, etc.); 2) reciprocal Blast best hits and 3) comparison of the repetitive elements flanking the 5′ and 3′ sides of α-defensin genes. The repetitive elements of the entire α-defensin cluster were identified using the CENSOR software tool (Kohany et al. 2006).

Tests for Positive Selection

To detect positive selection, the CODEML program as implemented in PAML was used to calculate the codon substitution models for heterogeneous selection pressure at amino acid positions (Yang 2007; Yang et al. 2000, 2005). The models used in this study were M0, M1a, M2a, M7, M8, and M8a. M1a (nearly neutral), M7 (beta), and M8 (beta and $\omega = 1$) were null models that did not support $\omega > 1$, whereas the alternative models M2a (positive selection) and M8a (beta and ω) have an additional class that allowed $\omega > 1$. Using these null and alternative models, three likelihood ratio tests (LRT) were carried out: 1) M1a versus M2a, 2) M7 versus M8, and 3) M8a versus M8. For LRTs, twice the log-likelihood difference, $2\Delta l = 2(l1 - l0)$ was compared with a χ^2 distribution to test whether the null model could be rejected, where $l1$ and $l0$ were the log likelihood for the alternative model and the null model, respectively. Moreover, naive empirical Bayes analysis was employed to calculate the posterior

Table 1. Number of α-Defensin Genes in the Five Primate Species.

	Species	Func	Pseu	Part	Total
	Human	6	7	0	13
	Chimpanzee	6 (1)	6	1	14
Hominidae	Orangutan	7	7	1	15
Old World monkey	Macaque	8 (1)	8	1	18
New World monkey	Marmoset	12	2	0	14

NOTE.—Due to incompleteness of few genomic regions, the genes located on these regions are represented as partial. The number in the parenthesis is the number of partial genes, which are designated as functional on the basis of 100% unique identities with the deposited nucleotide sequences in NCBI (see text). Func, functional genes; Pseu, pseudogenes; Part, partial genes.

probability that each site belonged to a particular site class. Sites with high posterior probability of belonging to the site class of $\omega > 1$ were inferred to be under positive selection.

Structural Analysis

To understand the structural characteristics of α-defensin peptides, we predicted the folding pattern of several primate sequences using the SWISS-MODEL (http://swissmodel.expasy.org) (Arnold et al. 2006) and PHYRE (http://www.sbg.bio.ic.ac.uk/~phyre) (Bennett-Lovsey et al. 2008) web servers. Pairwise structural alignments and structural superimposition were performed using the DaliLite (http://www.ebi.ac.uk/Tools/dalilite) (Holm and Park 2000) web server. Electrostatic potential was calculated using the PBEQ-Solver (Jo et al. 2008) web server. All figures were generated with PyMOL (DeLano Scientific; http://pymol.org). The hydrophobicity score was calculated using the Kyte–Doolittle scale (Kyte and Doolittle 1982).

Results

Number of α-Defensin Genes in Five Primate Species

Our homology searches (see Materials and Methods) detected 41 functional and 30 nonfunctional α-defensin genes in the genomes of the five primate species under study (table 1). The α-defensin genes in human, chimpanzee, orangutan, and macaque are located in the subtelomeric region of chromosome 8 of the respective species (supplementary fig. S2, Supplementary Material online). The chromosomal location for marmoset α-defensin genes is not available due to incompleteness of the genome assembly. However, most (12 of 14) of the α-defensin genes in marmoset are located in a single genomic region (contig 333). All identified α-defensin genes in five primate species are listed in supplementary table S1 (Supplementary Material online). In all species, with the exception of human, the numbers of functional genes given in table 1 are the minimum estimates because some genomic regions of the α-defensin cluster are incomplete in the draft genome sequences. It is possible that the partial genes can be annotated as functional genes when more complete versions of the genome sequences will be available. In this study, we have found five partial genes (two from chimpanzee, one from orangutan, and two from macaque). These genes contain either the entire first exon or the entire second exon and some portion of the intron. We used these partial sequences as queries in similarity searches against the nucleotide database of NCBI to find out whether these partial α-defensin sequences are functional or nonfunctional genes. We found that two of five identified partial α-defensin genes, chimpanzee Ptr 1 and macaque Mmu 18 (see supplementary table S1, Supplementary Material online) exhibit 100% sequence identity with two nucleotide sequences. The sequence exhibiting 100% identity with Ptr 1 is a full-length messenger RNA sequence (accession number AY746440), whereas for Mmu 18 the 100% identical sequence is part of a BAC clone (accession number AC202726). The retrieved α-defensin gene (100% identical to the nucleotide sequence of Mmu 18) from this BAC clone is a full-length gene. Therefore, we decided to regard Ptr 1 and Mmu 18 sequences as functional genes in this study. As shown in table 1, the proportions of total numbers of functional and nonfunctional α-defensin genes are nearly the same for hominids (human, chimpanzee, and orangutan) and Old World monkeys (macaque). By contrast, the total number of functional α-defensin genes in marmoset (member of New World monkeys) is considerably higher than that of catarrhine primate species (hominids and Old World monkeys). It is also noticeable that marmoset showed the lowest number of pseudogenes from all the species studied.

Phylogenetic Relationships of α-Defensin Genes

To examine the evolutionary relationships between the α-defensin genes, we constructed phylogenetic trees for the data set of 41 functional and 22 nonfunctional α-defensin genes using the NJ and MP methods. We excluded three partial genes and eight pseudogenes because they were truncated and they had highly diverged sequences, respectively. In fig. 1, we present condensed phylogenetic trees at the 50% bootstrap consensus value level. The tree topologies produced by both methods (figs. 1*a* and 1*b*) are nearly the same and classify the α-defensin genes in primates into three major phylogenetic classes (classes I, II, and III). In both phylogenetic trees, the branch leading to class I is supported by >90% bootstrap values. For classes II and III, although the bootstrap support is relatively low (>60% and <80%), both classes are reproduced by NJ and MP trees. The presence of α-defensin genes from all five primate species in all three phylogenetic classes suggests that the separation between the classes I, II, and III genes occurred before the divergence between New World and Old World monkeys.

The number of α-defensin genes varies considerably among the three classes (table 2). In all species, the largest numbers of genes belong to class III, with the exception of macaque. In the latter species, the highest number of genes is found in class I. In hominids, class II contains one functional gene and two pseudogenes, indicating that no significant gain or loss occurred in class II genes after divergence of Old World monkeys and hominids from their last common ancestor.

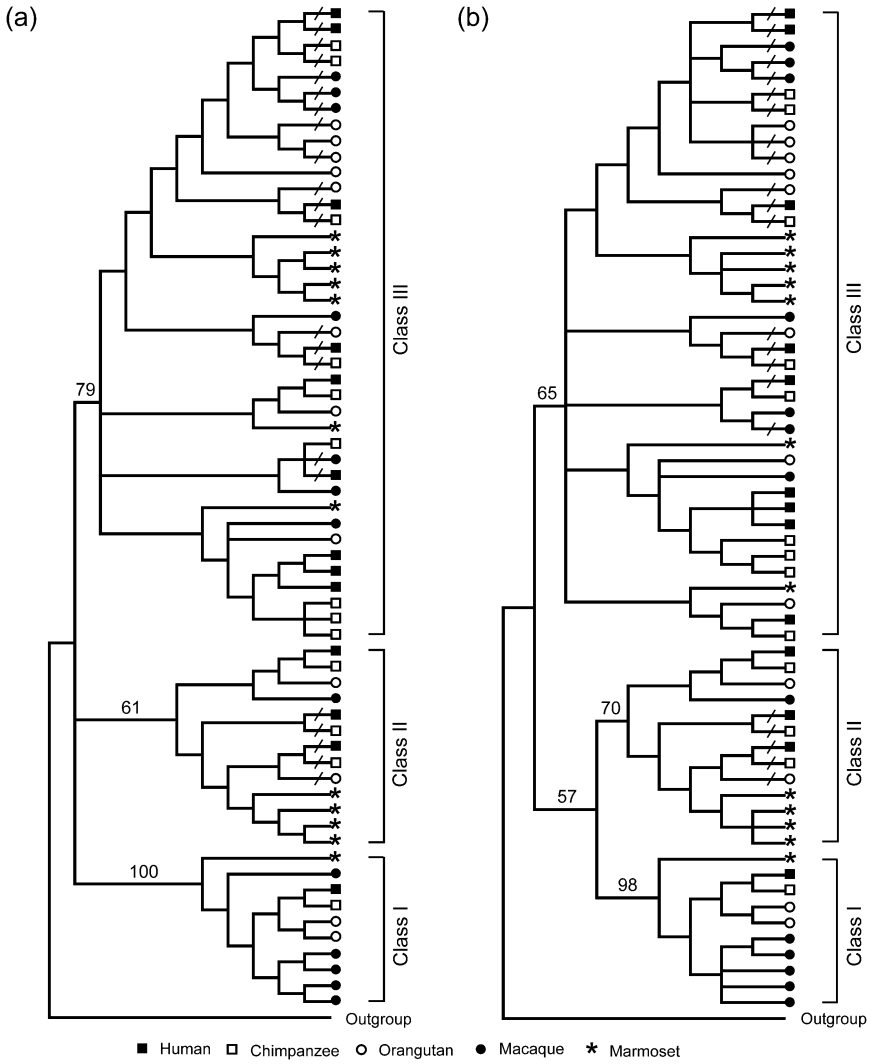

FIG. 1. Phylogenetic tree condensed at the 50% bootstrap value level for α-defensin genes of five primate species. The phylogenetic trees are calculated by the (a) NJ and (b) MP methods, respectively. The pseudogenes are shown by "/" symbol. In both trees, a mouse α-defensin sequence is used as an outgroup .

Identification of Phylogenetic Class-Specific Amino Acid Residues in α-Defensins

To identify whether there are specific amino acid residues or motifs that distinguish the three phylogenetic classes, we analyzed the amino acid sequences of α-defensins. The class-specific consensus sequences were generated by identifying the most commonly used amino acid residues or motifs in each class of α-defensin (fig. 2). From the alignment of the encoded proteins of all functional α-defensin genes, we identified five molecular markers, which can distinguish the α-defensin sequences belonging to classes I, II, and III. Interestingly, all five molecular markers are located in the prosegment regions of α-defensins. Although the identified molecular markers are mostly class specific, in certain cases, the same amino acid residue(s) in a particular position are shared by two phylogenetic classes. For example, class I α-defensin sequences possess Ser residues at position 22, whereas at the same position classes II and III consensus sequences contain Pro residues. At positions 39–40, 49–50, 57–60, and 70–72, the class I sequences have TQ, DL(X), NGLS, and QAR motifs, respectively (fig. 2). Here, "X" represents any amino acid that appeared due to the substitution of consensus residue at particular position. In contrast, the motifs present at the same positions in class II α-defensin sequences are AQ, DF, DASS, and T(R)R(T), whereas relatively less conserved motifs are found at the same positions of class III sequences.

Table 2. Number of α-Defensin Genes for Three Phylogenetic Classes in Each Primate Species.

	Class I		Class II		Class III	
Species	**Func**	**Pseu**	**Func**	**Pseu**	**Func**	**Pseu**
Human	1	0	1	2	4	5
Chimpanzee	1	0	1	2	5	4 (1)
Orangutan	2	0	1	2	4	5 (1)
Macaque	5	0 (1)	1	2	3	6
Marmoset	1	0	4	2	7	0

NOTE.—The phylogenetic classes in this table are defined in fig. 1. The number in the parenthesis is the number of partial gene. Func, functional genes; Pseu, pseudogenes.

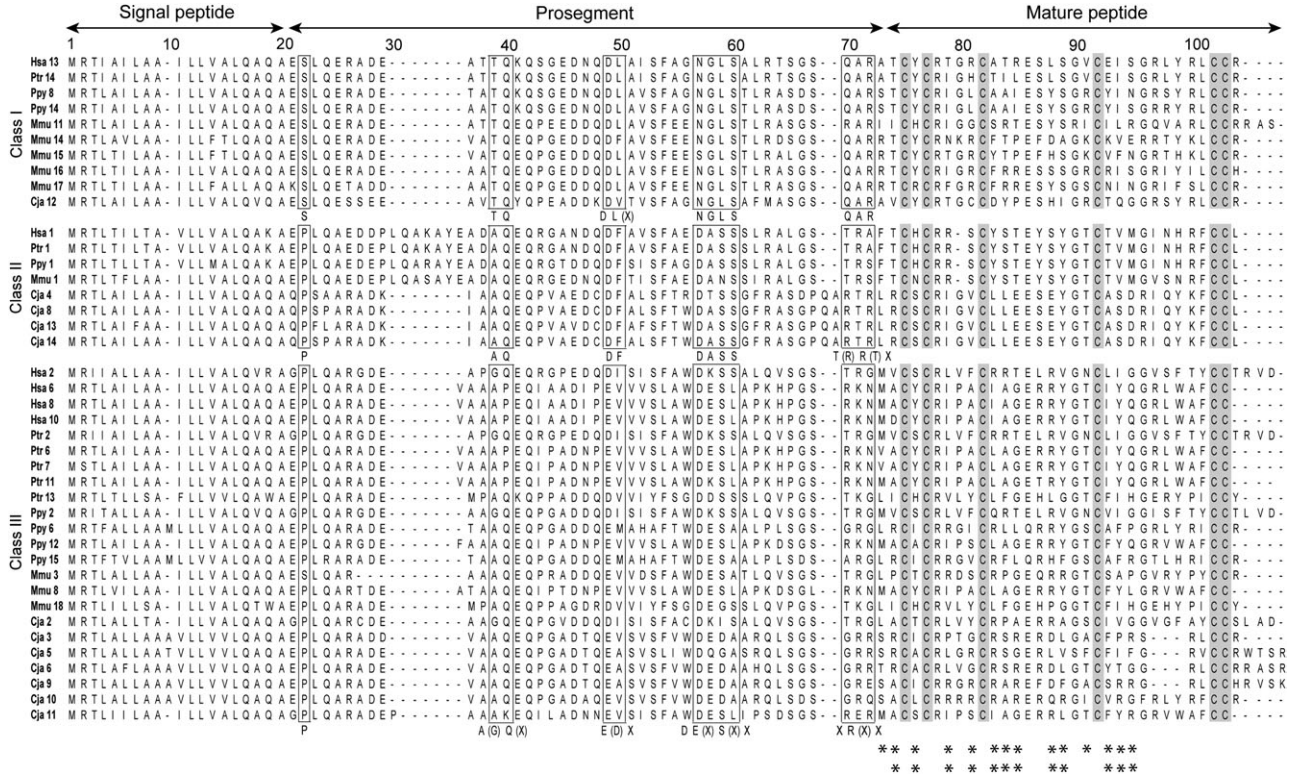

FIG. 2. Multiple sequence alignment of the functional α-defensin amino acid sequences. The conserved Cys residues are highlighted in gray and the amino acid residues or motifs that distinguish the three phylogenetic classes (classes I, II, and III) are marked with boxes. The probable consensus sequences (shown below the boxes) are the most commonly used amino acid residues (≥50%) or motifs in each α-defensin class. If only one specific residue appeared due to a substitution of the consensus residue at a particular position, that residue is shown in parentheses after the consensus sequence. If multiple residues are found in a particular position, an "X" has been used, either within parentheses after the consensus sequence or without parentheses (where no consensus found), which represents any amino acid. The potential positively selected sites are indicated by asterisks below the alignment. These sites are detected by both M2 and M8 models (see text for description). * and ** indicate significance at 95% and 99% levels, respectively.

Structural Characteristics of Primate α-Defensins

We analyzed the amino acid sequences of signal peptide, propeptide, and mature peptide regions to understand the structural features of primate α-defensin peptides. In all primate sequences studied, the signal peptide is the most conserved segment, followed by the prosegment and mature peptide regions (supplementary fig. S3, Supplementary Material online). It has been shown that the prosegment region of defensins contains several amino acids, which are involved in folding and functional inhibition of the mature peptides (Zou et al. 2008; Figueredo et al. 2009). Most of the negatively charged residues are conserved in the prosegment of primate defensins (E20, D27, E28, E34, D39, and E42) (numbering is according to sequence Hsa 6 in fig. 2 excluding gaps). In addition to the negatively charged amino acids, three regions that are predominantly occupied by hydrophobic residues (positions 29–33, 35–38, and 43–49) are also reasonably well conserved. These observations suggest that although the prosegment sequence is not highly conserved, specific structural characteristics are preserved most probably due to functional constraints.

The most well-studied region of α-defensins is the mature peptide. Defensin structures have been solved using both X-ray crystallography and nuclear magnetic resonance. To date, structures for several α-defensins have been reported from human and other mammals (McManus et al. 2000; Szyk et al. 2006). The overall fold of the α-defensin monomer is composed of three β-strands arranged into an antiparallel β-sheet. These architectural elements are restrained in their relative orientations by three disulfide bridges, C66–C94, C68–C83, C73–C93, and one salt bridge formed by the side chains of R69 and E77 (numbering is according to sequence Hsa 6 in fig. 2 excluding gaps). These residues are well conserved in all primate defensins (fig. 2). These observations indicate that despite the high level of divergence at their primary amino acid sequences, primate α-defensins fold very similar to each other. This notion is also supported by theoretical models built by homology (fig. 3). These models suggest that primate defensins display the canonical α-defensin disulfide arrangement and a similar fold, but differ markedly in surface charge distribution and loop sizes/orientations (fig. 3). The electrostatic surface analysis also supports the above notion and shows the amphipathicity of these proteins (supplementary fig. S4, Supplementary Material online). All primate α-defensin structures contain several Arg residues distributed fairly evenly in the primary

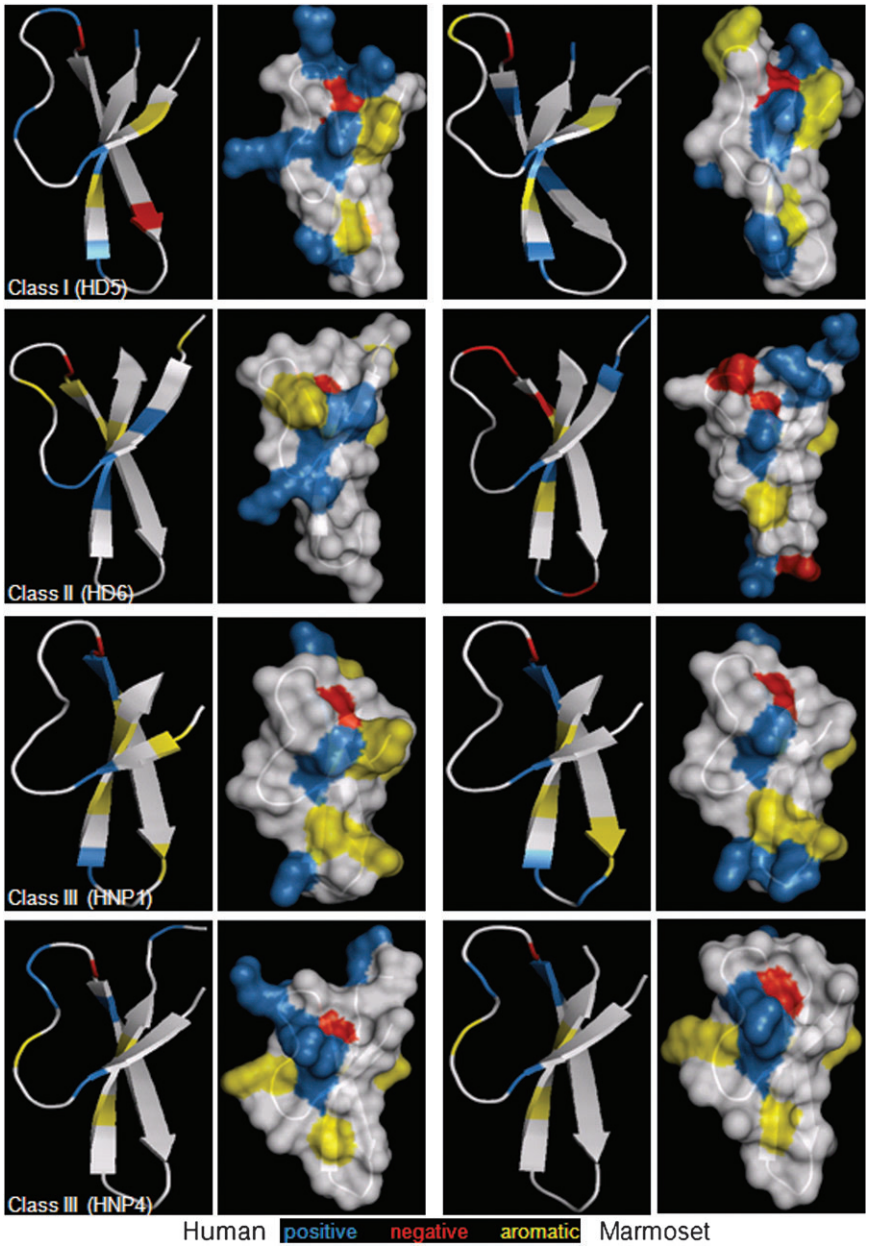

FIG. 3. Primate α-defensins fold similarly but have different composition of charged and surface amino acids. The theoretical 3D models for the marmoset proteins were constructed using as templates the experimentally resolved structures of human α-defensins, which are shown on the left panels for comparison. For convenience, the cartoon representations of the molecules are also shown inside the semitransparent surfaces. Three sets of amino acids are shown with different colors (blue, positive; red, negative; yellow, aromatic). The Protein Data Bank code of the templates used are 1ZMP for class I, 1ZMQ for class II, 2PM4 for class III (HNP1), and 1ZMM for class III (HNP4).

sequence; however, when the protein is folded, most of these basic residues are located on one face. As has been observed for other primate defensins (Vasudevan et al. 2008), the positively charged surface is distinct and separate from the hydrophobic region. The positively charged surface distinguishes both paralogous and orthologous defensin proteins (supplementary fig. S4, Supplementary Material online). Furthermore, the different classes can be differentiated by the differential presence of one or two small negatively charged regions (supplementary fig. S4, Supplementary Material online). These variations in α-defensin electrostatic surface distributions imply that these proteins have distinct mechanisms for their mode of antimicrobial action. In addition to the charge distributions, it has been hypothesized that the large hydrophobic surfaces of defensins may play a role in hydrophobic interactions with the membrane hydrocarbons of the target cell (Vasudevan et al. 2008). Thus, the differences in hydrophobicity (supplementary fig. S5, Supplementary Material online) and the distribution of these surfaces, together with the presence/absence of exposed aromatic amino acids (figs. 2 and 3; supplementary fig. S4, Supplementary Material online),

may be additional determinants of functional specificity and differentiation between the different primate α-defensins.

Genomic Organization of α-Defensin Cluster

For relatively short and highly evolving sequences, the establishment of true orthologous relationships is often difficult using conventional phylogenetic tree–building methods (Das et al. 2008a, 2008b). In the case of α-defensin sequences, which are short (≤100 aa) and evolve rapidly (Lynn et al. 2004; Patil et al. 2004), we used three different methods to identify orthologous relationships: 1) phylogenetic analysis, 2) reciprocal Blast best hits, and 3) comparison of the flanking repetitive elements of α-defensin genes (see Materials and Methods). For the third method, we have identified the repetitive elements in the entire α-defensin cluster and compared the repetitive elements flanking the 5′ and 3′ sides of each α-defensin gene. One example of the comparison of flanking repetitive elements of α-defensin gene is shown diagrammatically in figure 4. In our study, we regarded as orthologs, genes for which orthology was supported by at least two methods. These analyses allowed us to establish orthologous relationships between α-defensin genes in all primates, with the exception of some marmoset sequences. The estimated divergence time between New World and Old World Monkeys is ~3 My (Steiper and Young 2006). It is possible that due to their relatively long divergence time, establishing true orthologous relationships between α-defensin genes of marmoset and Old World Monkeys is difficult. Hence, the orthology between α-defensin genes is established for four primate species (human, chimpanzee, orangutan, and macaque) as they are separated from each other by a relatively short divergence time.

Using the orthologous sequences, we have carried out a comparative genomic analysis of α-defensin clusters of human, chimpanzee, and macaque. Due to the incompleteness of the genome assembly (see supplementary table S1, Supplementary Material online), the complete genomic organization of the orangutan α-defensin cluster could not be established. Therefore, we excluded orangutan from the comparative genomics analysis of the α-defensin locus. Because of this limitation, certain conclusions can be drawn only tentatively. The physical maps of α-defensin genes in human, chimpanzee, and macaque are shown in figure 5. The results of the analysis reveal conservation in the synteny of five genes (i.e., Hsa1, Hsa2, Hsa3, Hsa4, and Hsa5 and their orthologous sequences), located at the 3′ regions of α-defensin clusters in the human, chimpanzee, and macaque (fig. 5). Although no duplication or deletion of genes is observed in this region, one α-defensin gene (Mmu2) became nonfunctional in the macaque lineage. Similarly, Hsa3 and Ptr3 became nonfunctional in the human–chimpanzee lineage, whereas the ortholog in macaque (Mmu3) is functional. In contrast to the conservation observed in the 3′ region, in the 5′ regions of α-defensin clusters, we identified a few events of species-specific gene duplication or deletion.

To gain a better understanding of the evolution of the α-defensin cluster, we analyzed the distribution of α-defensin functional genes and pseudogenes in the three phylogenetic

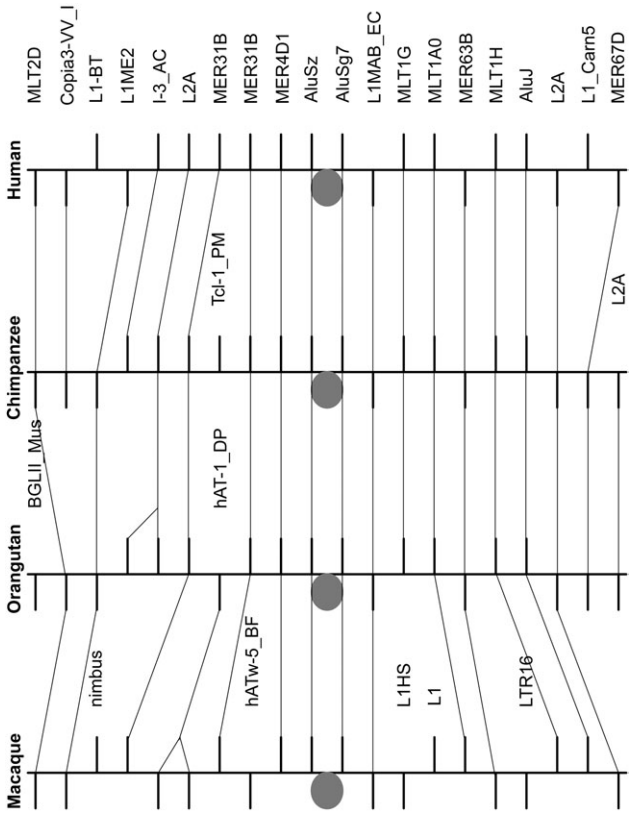

FIG. 4. A representative example of flanking repetitive elements that were used to determine the orthologous relationships between α-defensin genes. Ten flanking repetitive elements from the 5′ and 3′ ends of Hsa1, Ptr1, Ppy1, and Mmu1 genes are shown. The α-defensin gene is shown as circles and the repetitive elements as rods. Rods above and below the lines indicate that the repetitive elements are located on different strands. The lines connecting the rods show orthologous and paralogous relationships of repetitive elements. The figure is not drawn in scale.

classes. This analysis showed that the genes in all three phylogenetic classes are intermingled in the genome (fig. 5). The genomic distribution of classes I and III α-defensin genes in macaque is quite different from that of hominids (human and chimpanzee). In contrast, it is noticeable that the functional and pseudogenes belonging to class II are conserved (i.e., no duplication or deletion) from macaque to human.

Examination of Positive Selection

Intrigued by the high divergence among α-defensin genes, we investigated whether this observation could be explained by positive selection. To examine if positive selection acts on α-defensin genes, we carried out three LRT using six codon-substitution models for heterogeneous selection pressure at amino acid positions (Yang et al. 2000; Yang et al. 2005; Yang 2007): 1) M1a versus M2a, 2) M7 versus M8, and 3) M8a versus M8 (tables 3 and 4). For this analysis, we used all the potentially functional genes from five primate species and compared both orthologous and paralogous defensin genes. This analysis suggests the action of positive selection on some specific sites of α-defensins

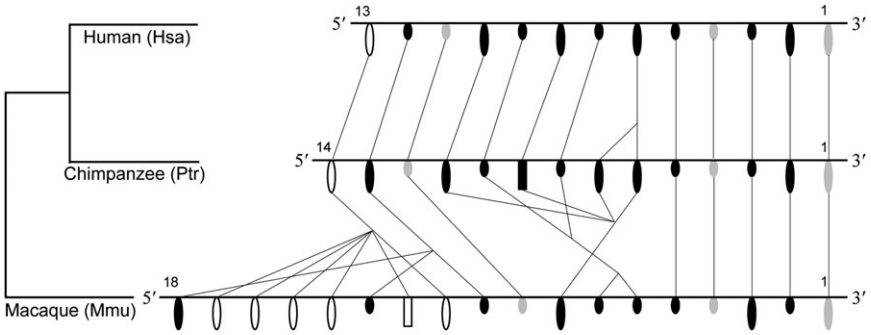

Fig. 5. Chromosomal localization of α-defensin genes and their orthologous and paralogous relationships in human (Hsa), chimpanzee (Ptr), and macaque (Mmu). The large oval and the small oval shapes represent functional genes and pseudogenes, respectively. The rectangular shapes indicate partial genes. The open, gray, and black shapes represent class I, class II, and class III genes (as defined by the phylogenetic analysis), respectively. The lines connecting the shapes show orthologous and paralogous relationships of α-defensin genes. The figure is not drawn in scale.

(table 3 and 4 and fig. 2). Except for one site at the boundary of prosegments and mature peptides, all of the other sites that could be positively selected were detected in the mature peptides. Our results are similar with previous observations by Patil et al. (2004) and Lynn et al. (2004). These results indicate that the presence of multiple potential positively selected sites in the mature peptide might be functionally important for the broad range of antimicrobial activities of α-defensin.

Discussion

In this study, we identified the α-defensin gene repertoires in human, chimpanzee, orangutan, macaque, and marmoset based on currently available genome sequences and analyzed the genomic organization using comparative genomic and evolutionary approaches. We found that the ratio between functional and nonfunctional α-defensin genes is nearly identical for hominids (human, chimpanzee, and orangutan) and Old World monkeys (macaque), whereas that of marmoset (New World monkey) is quite different. The latter species also showed the lowest fraction of pseudogenes compared with all species studied (table 1). In several multigene families, a considerable number of pseudogenes has been described (Piontkivska and Nei 2003; Nei 2007; Das et al. 2010). These genes have accumulated nonsense mutations, frameshift deletions and insertions, or single nucleotide substitutions within functionally important sites, which disrupt the expression of functional proteins (Kawasaki et al. 1997; Das et al. 2008a). It is, therefore, possible that comparatively higher accumulation of mutations in the α-defensin locus of catarrhine primates (human, chimpanzee, orangutan, and macaque) may have led to the higher number of pseudogenes compared with that of platyrrhine (or New World monkey) primate species (marmoset).

Our analysis suggests that the α-defensin genes of primates fall into three major phylogenetic classes (classes I, II, and III). The presence of all three classes in the marmoset indicates that their divergence occurred before the separation of New World and Old World monkeys and these classes have persisted for ~43 Myr in the primate genomes. The comparative analysis of the α-defensin genomic clusters suggests that several genomic rearrangements occurred in these genomic regions. In all classes, with the exception of class II genes, the differences in the number of

Table 3. Parameter Estimates and Log-Likelihood Values Under Models of Variable ω Ratios Among Sites.

Model	*p*	Parameter	ln*L*	d_N/d_S	Positively Selected Sites
M0: one ratio	**1**	**$\omega = 1.141$**	**−3854.68**	**$= \omega$**	
M1a: neutral	**1**	**$p_0 = 0.465$, $\omega_0 = 0.173$;**			
		$p_1 = 0.535$, $\omega_1 = 1.000$	**−3755.88**	**0.6158**	
M2a: selection	**3**	**$p_0 = 0.324$, $\omega_0 = 0.170$;**			
		$p_1 = 0.462$, $\omega_1 = 1.000$;			
		$p_2 = 0.214$, $\omega_2 = 4.078$	**−3672.53**	**1.3913**	
M7: beta	**2**	**$p = 0.500$,**			73, **74, 76, 79, 81, 83, 84,**
		$q = 0.315$	**−3761.11**	**0.6136**	**85, 88, 89,** 91, **93, 94, 95**
M8: beta and ω	**4**	**$p_0 = 0.777$, $p = 0.497$,**			
		$q = 0.320$; $p_1 = 0.223$,			73, **74, 76, 79, 81, 83, 84, 85,**
		$\omega = 3.874$	**−3673.45**	**1.3358**	87, **88, 89,** 91, **93, 94, 95**
M8a: beta and ω = 1	**4**	**$p_0 = 0.528$, $p = 1.154$,**			
		$q = 4.021$; $p_1 = 0.472$,			
		$\omega = 1.000$	**−3752.41**	**0.5856**	

NOTE.—Bold number indicates significance at the 99 % level and the other indicates significance at the 95% level. The d_S and d_N stand for the numbers of synonymous and nonsynonymous substitutions per site, respectively.

Table 4. Likelihood Ratio Statistics ($2\Delta \ln L$).

Models	$2\Delta \ln L$	*P* value
M1a versus M2a	166.709442	$<10^{-10}$
M7 versus M8	175.301622	$<10^{-10}$
M8a versus M8	157.9137	$<10^{-10}$

class I and class III genes between the hominids and macaque lineages have been generated mainly by repeated tandem gene duplication within each genomic cluster (fig. 5). In the phylogenetic trees, all three classes contain α-defensin genes from five primate species and the phylogenetically distantly related genes are more or less intermingled in the genomic cluster. Hence, considering the phylogeny and the genomic organization of α-defensin genes, we can infer that the α-defensin multigene family is mainly subject to the birth-and-death model of evolution rather than to concerted evolution. In the birth-and-death model, new genes are originated by repeated gene duplication and by accumulation of mutation some of them may acquire a new function and remain in the genome for a long time, whereas others become pseudogenes or are deleted from the genome (Nei et al. 2000; Su and Nei 2001; Rooney 2004; Das et al. 2008b). In contrast, the concerted evolution model proposes that the genes in a multigene family of a species are homogenized over some period of time by gene conversion or unequal crossing-over, causing higher sequence similarity of genes within species than between species (Liao 1999; Nikolaidis and Nei 2004).

In catarrhine primates, the α-defensin cluster is located in the subtelomeric region of the chromosome. Due to incompleteness of the genome assembly, the chromosomal location for marmoset α-defensin cluster is not available. The subtelomeres harbor several multigene families, including the olfactory receptor and immunoglobulin heavy chain variable region genes that exhibit a fairly high level of sequence divergence as well as multiple duplication and deletion events (Linardopoulou et al. 2001; Das et al. 2008b). Recent studies suggest that the subtelomeric and the pericentromeric regions might be less constrained than other genomic regions for recombination, duplication, gene conversion, point mutation, and translocation (Linardopoulou et al. 2005; Webber and Ponting 2005). Thus, these regions may facilitate the birth and death of their harboring genes. The multiple duplications and the acquisition of several point mutations observed in the α-defensin locus are consistent with this notion. Due to point mutations, α-defensin genes have been diversified and some of them became pseudogenes in certain lineages. In fact, we found that less than 50% of the α-defensin genes are functional genes in human, chimpanzee, orangutan, and macaque. Moreover, our analyses, together with previous studies (Lynn et al. 2004; Patil et al. 2004), imply that the functional α-defensin genes may have evolved under positive selection. In our analysis, we have detected potential positively selected sites in the α-defensin matured region, most of which are consistent with the sites predicted to be under positive selection by Lynn et al. (2004). The experimental evidence available for primate's α-defensins shows that the variation in the mature peptides leads to the diversity in their potency against different pathogens. For example, although the primary amino acid sequences of Hsa6 (HNP1), Hsa8 (HNP2), and Hsa10 (HNP3) are identical, except for a single amino acid alteration at the N-terminal end of the mature peptide (Ala to Asp), the Hsa6 (HNP1) and Hsa8 (HNP2) are different from Hsa10 (HNP3) in the activities to kill *Candida albicans* (Lehrer et al. 1988). The selective replacements of Arg residues in Hsa13 (HD5) at positions 81 (numbering is according to the alignment of α-defensins presented in fig. 2) significantly decrease the antimicrobial activity, whereas a replacement of Arg to His at position 85 as found in a patient suffering from Crohn's disease (a chronic inflammatory disease) severely reduce the bactericidal activity of Hsa13 (HD5) (de Leeuw et al. 2009). In macaque, the matured regions of functional class I α-defensins are diverged from each other except for canonical conservation of six Cys residues and one Arg residue (fig. 2). Tanabe et al. (2004) showed that these α-defensins are expressed in the intestine of macaque and that their matured regions are highly variable in bactericidal activities. It is, therefore, possible that positive selection in the specific regions of α-defensin may facilitate diverse profiles of antimicrobial activity and maximize host ability to defend against infections.

The α-defensin gene classes are defined by phylogenetic relationships only in this study and therefore the classification may not be related to gene function or expression pattern. However, considering the 100% sequence similarity with the deposited human α-defensin sequences in NCBI, we found that human class I and class II genes are primarily expressed in the paneth cells of intestine, whereas class III genes are expressed mainly in neutrophils (see supplementary table S2, Supplementary Material online). To test whether the phylogenetic classification reflects any structural and functional clustering, we undertook a systematic analysis of the translation products of all functional α-defensin genes. Our analysis reveals that certain peptide residues differentiate the α-defensin sequences belonging to classes I, II, and III (fig. 2). Hence, corresponding conserved nucleotide sequences define phylogenetic class-specific amino acid residues or motifs whose structures have been maintained across species and evolutionary barriers. Interestingly, all of the class-specific signatures are located in the prosegments and not in the matured peptides. It has been reported that the acidic prosegment may be important for neutralization, processing, and/or folding of the cationic C-terminal matured peptide (Valore and Ganz 1992; Liu and Ganz 1995). The arrangement of negatively charged residues (Glu, Asp) in the prosegment is functionally important for the interaction of the prosegment and the mature peptides (Liu and Ganz 1995; Zou et al. 2008). It is possible that the conservation in phylogenetic class-specific amino acid residues or motifs in the propeptides might have some functional importance. Moreover, our results suggest that despite the high level

of divergence at the primary amino acid sequences of the mature peptide, primate α-defensins fold in a very similar fashion to each other (fig. 3). The differences observed in the surface charge distribution and the loop sizes/orientations (fig. 3; supplementary fig. S4, Supplementary Material online) together with the divergent hydrophobicity patterns and the presence or absence of exposed aromatic amino acids (figs. 2 and 3; supplementary fig. S5, Supplementary Material online) suggests functional differentiation between the different α-defensin classes. The latter supposition is in accordance with the birth-and-death model of evolution, which postulates that new genes may acquire a new function. Therefore, it seems that the evolution of the α-defensins in primates is driven by two opposing forces, one diversifying and the other stabilizing the specific amino acid residues or motifs due to functional constraints.

Supplementary Material

Supplementary tables S1 and S2 and figures S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Jan Klein, Masatoshi Nei, and Masafumi Nozawa for valuable comments and suggestions. This work was supported by the National Institutes of Health (grant AI72435 to M.D.C.), Georgia Research Alliance and the California State University, Fullerton (CSUF) (start-up money to N.N.). C.M. was supported by the Faculty Development Center Grant from CSUF.

References

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201.

Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA. 2008. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* 70: 611–625.

Boniotto M, Tossi A, DelPero M, Sgubin S, Antcheva N, Santon D, Masters J, Crovella S. 2003. Evolution of the beta defensin 2 gene in primates. *Genes Immun.* 4:251–257.

Das S. 2009. Evolutionary origin and genomic organization of microRNA genes in immunoglobulin lambda variable region gene family. *Mol Biol Evol.* 26:1179–1189.

Das S, Mohamedy U, Hirano M, Nei M, Nikolaidis N. 2010. Analysis of the immunoglobulin light chain genes in zebra finch: evolutionary implications. *Mol Biol Evol.* 27:113–120.

Das S, Nikolaidis N, Klein J, Nei M. 2008a. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proc Natl Acad Sci U S A.* 105:16647–16652.

Das S, Nozawa M, Klein J, Nei M. 2008b. Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. *Immunogenetics* 60:47–55.

de Leeuw E, Rajabi M, Zou G, Pazgier M, Lu W. 2009. Selective arginines are important for the antibacterial activity and host cell interaction of human alpha-defensin 5. *FEBS Lett.* 583:2507–2512.

Eck RV, Dayhoff MO. 1966. Atlas of Protein Sequence and Structure. Silver Springs (MD): National Biomedical Research Foundation.

Figueredo SM, Weeks CS, Young SK, Ouellette AJ. 2009. Anionic amino acids near the pro-alpha-defensin N terminus mediate inhibition of bactericidal activity in mouse pro-cryptdin-4. *J Biol Chem.* 284:6826–6831.

Ganz T. 2003. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol.* 3:710–720.

Ganz T, Lehrer RI. 1994. Defensins. *Curr Opin Immunol.* 6:584–589.

Garcia AE, Osapay G, Tran PA, Yuan J, Selsted ME. 2008. Isolation, synthesis, and antimicrobial activities of naturally occurring theta-defensin isoforms from baboon leukocytes. *Infect Immun.* 76:5883–5891.

Hollox EJ, Armour JA. 2008. Directional and balancing selection in human beta-defensins. *BMC Evol Biol.* 8:113.

Holm L, Park J. 2000. DaliLite workbench for protein structure comparison. *Bioinformatics* 16:566–567.

Jo S, Vargyas M, Vasko-Szedlar J, Roux B, Im W. 2008. PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.* 36:W270–W275.

Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Schmeits JL, Wang J, Shimizu N. 1997. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res.* 7:250–261.

Kim C, Gajendran N, Mittrucker HW, Weiwad M, Song YH, Hurwitz R, Wilmanns M, Fischer G, Kaufmann SH. 2005. Human alpha-defensins neutralize anthrax lethal toxin and protect against its fatal consequences. *Proc Natl Acad Sci U S A.* 102:4830–4835.

Klotman ME, Chang TL. 2006. Defensins in innate antiviral immunity. *Nat Rev Immunol.* 6:447–456.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics.* 7:474.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 157:105–132.

Lehrer RI, Ganz T, Szklarek D, Selsted ME. 1988. Modulation of the in vitro candidacidal activity of human neutrophil defensins by target cell metabolism and divalent cations. *J Clin Invest.* 81:1829–1835.

Liao D. 1999. Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet.* 64:24–30.

Linardopoulou E, Mefford HC, Nguyen O, Friedman C, van den Engh G, Farwell DG, Coltrera M, Trask BJ. 2001. Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum Mol Genet.* 10:2373–2383.

Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* 437: 94–100.

Liu L, Ganz T. 1995. The pro region of human neutrophil defensin contains a motif that is essential for normal subcellular sorting. *Blood* 85:1095–1103.

Lynn DJ, Bradley DG. 2007. Discovery of alpha-defensins in basal mammals. *Dev Comp Immunol.* 31:963–967.

Lynn DJ, Lloyd AT, Fares MA, O'Farrelly C. 2004. Evidence of positively selected sites in mammalian alpha-defensins. *Mol Biol Evol.* 21:819–827.

McManus AM, Dawson NF, Wade JD, Carrington LE, Winzor DJ, Craik DJ. 2000. Three-dimensional structure of RK-1: a novel alpha-defensin peptide. *Biochemistry* 39:15757–15764.

Menendez A, Brett Finlay B. 2007. Defensins in the immunology of bacterial infections. *Curr Opin Immunol.* 19:385–391.

Michaelson D, Rayner J, Couto M, Ganz T. 1992. Cationic defensins arise from charge-neutralized propeptides: a mechanism for avoiding leukocyte autocytotoxicity? *J Leukoc Biol.* 51:634–639.

Morrison GM, Semple CA, Kilanowski FM, Hill RE, Dorin JR. 2003. Signal sequence conservation and mature peptide divergence

within subgroups of the murine beta-defensin gene family. *Mol Biol Evol.* 20:460–470.
Nei M. 2007. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci U S A.* 104:12235–12242.
Nei M, Gu X, Sitnikova T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A.* 94:7799–7806.
Nei M, Rogozin IB, Piontkivska H. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci U S A.* 97:10866–10871.
Niimura Y, Nei M. 2005. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci U S A.* 102:6039–6044.
Nikolaidis N, Nei M. 2004. Concerted and nonconcerted evolution of the Hsp70 gene superfamily in two sibling species of nematodes. *Mol Biol Evol.* 21:498–505.
Nozawa M, Kawahara Y, Nei M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. *Proc Natl Acad Sci U S A.* 104:20421–20426.
Ota T, Nei M. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol.* 11:469–482.
Patil A, Hughes AL, Zhang G. 2004. Rapid evolution and diversification of mammalian alpha-defensins as revealed by comparative analysis of rodent and primate genes. *Physiol Genomics.* 20:1–11.
Piontkivska H, Nei M. 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Mol Biol Evol.* 20:601–609.
Presicce P, Giannelli S, Taddeo A, Villa ML, Della Bella S. 2009. Human defensins activate monocyte-derived dendritic cells, promote the production of proinflammatory cytokines, and up-regulate the surface expression of CD91. *J Leukoc Biol.* 86:941–948.
Rooney AP. 2004. Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in Apicomplexans. *Mol Biol Evol.* 21:1704–1711.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4:406–425.
Selsted ME. 2004. Theta-defensins: cyclic antimicrobial peptides produced by binary ligation of truncated alpha-defensins. *Curr Protein Pept Sci.* 5:365–371.
Semple CA, Maxwell A, Gautier P, Kilanowski FM, Eastwood H, Barran PE, Dorin JR. 2005. The complexity of selection at the major primate beta-defensin locus. *BMC Evol Biol.* 5:32.
Semple CA, Rolfe M, Dorin JR. 2003. Duplication and selection in the evolution of primate beta-defensin genes. *Genome Biol.* 4:R31.
Steiper ME, Young NM. 2006. Primate molecular divergence dates. *Mol Phylogenet Evol.* 41:384–394.
Su C, Nei M. 2001. Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. *Mol Biol Evol.* 18:503–513.
Szyk A, Wu Z, Tucker K, Yang D, Lu W, Lubkowski J. 2006. Crystal structures of human alpha-defensins HNP4, HD5, and HD6. *Protein Sci.* 15:2749–2760.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A.* 101:11030–11035.
Tanabe H, Yuan J, Zaragoza MM, Dandekar S, Henschen-Edman A, Selsted ME, Ouellette AJ. 2004. Paneth cell alpha-defensins from rhesus macaque small intestine. *Infect Immun.* 72:1470–1478.
Tang YQ, Yuan J, Osapay G, Osapay K, Tran D, Miller CJ, Ouellette AJ, Selsted ME. 1999. A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science* 286:498–502.
Tran D, Tran P, Roberts K, Osapay G, Schaal J, Ouellette A, Selsted ME. 2008. Microbicidal properties and cytocidal selectivity of rhesus macaque theta defensins. *Antimicrob Agents Chemother.* 52:944–953.
Valore EV, Ganz T. 1992. Posttranslational processing of defensins in immature human myeloid cells. *Blood* 79:1538–1544.
Vasudevan S, Yuan J, Osapay G, Tran P, Tai K, Liang W, Kumar V, Selsted ME, Cocco MJ. 2008. Synthesis, structure, and activities of an oral mucosal alpha-defensin from rhesus macaque. *J Biol Chem.* 283:35869–35877.
Webber C, Ponting CP. 2005. Hotspots of mutation and breakage in dog and human chromosomes. *Genome Res.* 15:1787–1797.
Wehkamp J, Salzman NH, Porter E, et al. (17 co-authors). 2005. Reduced paneth cell alpha-defensins in ileal Crohn's disease. *Proc Natl Acad Sci U S A.* 102:18129–18134.
Xiao Y, Hughes AL, Ando J, Matsuda Y, Cheng JF, Skinner-Noble D, Zhang G. 2004. A genome-wide screen identifies a single beta-defensin gene cluster in the chicken: implications for the origin and evolution of mammalian defensins. *BMC Genomics.* 5:56.
Yang D, Biragyn A, Kwak LW, Oppenheim JJ. 2002. Mammalian defensins in immunity: more than just microbicidal. *Trends Immunol.* 23:291–296.
Yang D, Chertov O, Bykovskaia SN, et al. (11 co-authors). 1999. Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* 286:525–528.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 22:1107–1118.
Yudin AI, Generao SE, Tollner TL, Treece CA, Overstreet JW, Cherr GN. 2005. Beta-defensin 126 on the cell surface protects sperm from immunorecognition and binding of anti-sperm antibodies. *Biol Reprod.* 73:1243–1252.
Zhang L, Yu W, He T, et al. (13 co-authors). 2002. Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science* 298:995–1000.
Zou G, de Leeuw E, Lubkowski J, Lu W. 2008. Molecular determinants for the interaction of human neutrophil alpha defensin 1 with its propeptide. *J Mol Biol.* 381:1281–1291.