CALL FOR PAPERS | Comparative Genomics

Cross-species analysis of the mammalian β-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract

Amar A. Patil,1 Yibin Cai,1 Yongming Sang,2 Frank Blecha,2 and Guolong Zhang1

1Department of Animal Science, Oklahoma State University, Stillwater, Oklahoma; and 2Department of Anatomy and Physiology, Kansas State University, Manhattan, Kansas

Submitted 3 May 2005; accepted in final form 14 July 2005

Patil, Amar A., Yibin Cai, Yongming Sang, Frank Blecha, and Guolong Zhang. Cross-species analysis of the mammalian β-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. Physiol Genomics 23: 5–17, 2005. First published July 20, 2005; 10.1152/physiolgenomics.00104.2005.—Mammalian β-defensins are an important family of innate host defense peptides with pleiotropic activities. As a first step to study the evolutionary relationship and biological role of the β-defensin family, we identified their complete repertoires in the human, chimpanzee, mouse, rat, and dog following systemic, genome-wide computational searches. Although most β-defensin genes are composed of two exons separated by an intron of variable length, some contain an additional one or two exons encoding an internal pro-sequence, a segment of carboxy-terminal mature sequences or untranslated regions. Alternatively, spliced isoforms have also been found with several β-defensins. Furthermore, all β-defensin genes are densely clustered in four to five syntenic chromosomal regions, with each cluster spanning <1.2 Mb across the five species. Phylogenetic analysis indicated that, although the majority of β-defensins are evolutionarily conserved across species, subgroups of gene lineages exist that are specific in certain species, implying that some β-defensins originated after divergence of these mammals from each other, while most others arose before the last common ancestor of mammals. Surprisingly, RT-PCR revealed that all but one rat β-defensin transcript are preferentially expressed in the male reproductive tract, particularly in epididymis and testis, except that Defb-4, a human β-defensin-2 ortholog, is more restricted to the respiratory and upper gastrointestinal tracts. Moreover, most β-defensins expressed in the reproductive tract are developmentally regulated, with enhanced expression during sexual maturation. Existence of such a vast array of β-defensins in the male reproductive tract suggests that these genes may play a dual role in both fertility and host defense.

defensin; antimicrobial peptide; innate immunity; fertility; comparative genomics

DEFENSINS AND DEFENSIN-LIKE molecules comprise a diverse group of cationic antimicrobial peptides characterized by the presence of multiple cysteine residues and a highly similar tertiary structure known as the defensin motif (14, 28, 35, 46, 60). Defensin-like genes have been discovered in many species including plants, fungi, arthropods, mollusks, reptiles, birds, and mammals (13, 14, 28, 36, 47, 52, 55, 57). Defensins in vertebrate species are further classified into three families, namely α-, β-, and θ-defensins, based on the spacing pattern of six cysteine residues (10, 14, 28). Although the evolutionary relationship between vertebrate defensins and defensins in other species remains unclear, phylogenetic analysis revealed that a primordial β-defensin gene is the common ancestor for all vertebrate defensins (57). β-Defensins are believed to have evolved before the divergence of mammals from birds (57) and gave rise to α-defensins in glires and primates after they diverged from other mammalian species (36). θ-Defensins further originated from α-defensins after separation of primates from other mammals (33, 36).

In contrast to α- and θ-defensins, which are produced primarily in the granules of either leukocytes or intestinal Paneth cells, β-defensins are unique in that they are primarily produced by nongranular mucosal epithelial cells lining the respiratory, gastrointestinal, and genitourinary tracts (14, 28, 35, 46, 60). In addition to their broad-spectrum antimicrobial activity against bacteria, fungi, and certain enveloped viruses, β-defensins have been recognized recently to resemble chemokines both structurally and functionally (14, 28, 46, 60). Recent studies also revealed that certain β-defensins are actively involved in sperm maturation and capacitation (41, 65, 68, 70), physiological properties that are not related to host defense.

All β-defensin genes encode a precursor peptide that consists of a hydrophobic, leucine-rich signal sequence, a pro-sequence, and a mature six-cysteine defensin motif at the carboxy terminus (14, 28). In most cases, β-defensin precursors are encoded in two separate exons separated by an intron of variable length, with one exon encoding the signal and pro-sequence and the other encoding the mature mature peptide (14, 28). It is generally believed that mature β-defensins need to be cleaved from the pro-sequence by proteolytic enzymes to become biologically active, although such proteases are yet to be identified.

A number of putative β-defensins were reported recently in humans and mice (47); however, later identification of a novel human sequence implied the possible existence of additional β-defensins in those two species (42). In addition, the evolutionary relationships among β-defensin genes across mammalian species are poorly understood, although limited data are available on β-defensin evolution within species or among the primates (9, 22, 32, 48).

Recent availability of complete genome sequences of a number of phylogenetically distinct mammalian species provides an ex-
cellent opportunity to identify the entire β-defensin gene family in these species. Here, we report the discovery of complete repertoires of β-defensins in the human, chimpanzee, mice, rat, and dog by using a systemic computational search strategy that we developed recently (36, 57). We showed that β-defensin genes form four to five syntenic clusters in these mammals. Although the majority of β-defensin genes are evolutionarily conserved across rodents, canines and primates, a few gene lineages exist only in certain species. We also examined the tissue expression patterns of the entire β-defensin gene family in the rat by RT-PCR. To our surprise, all but one of the β-defensins were found to be preferentially expressed in the male reproductive system, particularly in testis and different segments of epididymis. The only exception, Defb4, is more restricted to the respiratory and upper gastrointestinal tracts, with virtually no expression in the male reproductive system. Furthermore, we provide evidence that many β-defensins are developmentally regulated, and the expression levels are elevated with age. These findings argue that the major biological functions of β-defensins may be related to reproduction and fertility, in addition to innate host defense.

MATERIALS AND METHODS

Computational search for novel β-defensins. To identify potential novel β-defensins in the human, mouse, rat, dog, and chimpanzee, systemic computational searches were performed essentially as we described (36, 57). In brief, all known β-defensin peptide sequences were individually queried against expressed sequence tags (EST), nonredundant sequences (NR), unfinished high-throughput genomic sequences (HTGS), and whole genome shotgun sequences (WGS) in GenBank by using the TBLASTN program (1) with the default settings on the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST). We then examined all potential hits for the presence of the characteristic β-defensin motif or conserved signal peptide and pro-sequence. Additional iterative BLAST searches were performed for every novel β-defensin sequence identified as above described until no more novel sequences were revealed. Because mammalian defensins tend to form clusters (47, 48), all genomic sequences containing β-defensins were also retrieved from GenBank to discover potential novel sequences with distant homology. The nucleotide sequences between two neighboring defensin genes were translated into six open reading frames and individually compared with the two defensin peptide sequences for the presence of β-defensin motif or signal peptide/pro-sequence by use of the BLASTP program (1) on the NCBI web site (http://www.ncbi.nlm.nih.gov/blast/bl2seq/) and/or ClustalW program (version 1.82) (53) on the European Bioinformatics Institute web site (http://www.ebi.ac.uk/clustalw).

Prediction of full-length coding sequences and genomic structures of β-defensins. In most cases, signal and pro-sequence of β-defensins are encoded in an exon separated from the mature peptide-encoding exon. If either the signal/pro-sequence or β-defensin motif of a novel gene was missing in a genomic sequence, a 5- to 15-kb flanking sequence was retrieved to identify the full-length coding sequence and to derive the structural organization of that novel β-defensin gene by using a combination of GenomeScan (61), GENSCAN (5), and/or GeneWise2 (4). All cDNA sequences of novel β-defensins described in this study have been submitted to GenBank, and the accession numbers are listed in Supplemental Table S1 (available at the Physiological Genomics web site)1.

Chromosomal mapping of β-defensin gene clusters. The BLAT program (25) was used to determine the relative position and orientation of each defensin in the genome through the University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu). Individual defensins were searched against latest versions of the assembled genomes of human (NCBI build 35, version 1), chimpanzee (NCBI build 1, version 1), mouse (NCBI build 32), rat [Baylor College of Medicine (BCM) version 3.1], and dog (NCBI build 1, version 1), released in June 2004, November 2003, September 2004, June 2003, and August 2004, respectively.

Sequence alignment and phylogenetic analysis of β-defensins. Multiple sequence alignments were carried out by using the ClustalW program (version 1.82) (53). The neighbor-joining method (43) was used to construct the phylogenetic tree by calculating the proportion of amino acid differences (p-distance), and the reliability of each branch was tested by 1,000 bootstrap replications.

Semi-quantitative RT-PCR analysis of tissue and developmental expression patterns of β-defensins. A total of 29 different tissue samples (see Fig. 6B) were harvested from healthy, 2-mo-old Sprague-Dawley rats. Total RNA was extracted using TRIzol (Invitrogen). For each RNA sample, 4 μg were reverse transcribed with random hexamers and Superscript II RT by using a first-strand cDNA synthesis kit (Invitrogen) according to the instructions. Subsequent PCR was performed with DNA Engine (model PTC-200, MJ Research) essentially as described (36, 57, 69). Briefly, 1/40 of the first-strand cDNA from each tissue was used to amplify β-defensins and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) with genespecific, exon-spanning primers (Supplemental Table S2). The PCR program used was 94°C denaturation for 2 min, followed by different cycles of 94°C denaturation for 20 s, 55°C annealing for 20 s, and 72°C extension for 40 s, followed by a final extension at 72°C for 5 min. The number of PCR cycles was optimized for each gene to ensure linear amplification (Supplemental Table S2). Specificity of each PCR product was confirmed by cloning the PCR product into the pGEM-T Easy vector (Promega), followed by sequencing of the recombinant plasmid.

To examine developmental expression of the reproductive tract-specific β-defensins, testes were harvested from Sprague-Dawley rats on 1, 15, 30, 60, and 90 days after birth, and epididymis were collected on 7, 15, 30, 60, and 90 days after birth, with two to four animals per time point. RNA isolation and semi-quantitative RT-PCR were performed as described above. Signal intensity of each gene-specific PCR product was quantified by using Sciion Imaging software (http://www.scioncorp.com), followed by normalization against the GAPDH signal amplified from the same cDNA sample, and the results were expressed as relative level of expression. Animal protocols were approved by the Institutional Animal Care and Use Committee of Oklahoma State University.

RESULTS

Discovery of mammalian β-defensin gene repertoires. On the basis of the conservation of the characteristic six-cysteine-containing β-defensin motif, an iterative, genome-wide homology search strategy was developed to identify complete repertoires of the β-defensin gene family in the human, chimpanzee, mouse, rat, and dog, as we previously described (36, 57). Because such computational searches are biased toward the defensin motif, the nucleotide sequences identified were primarily composed of the exons encoding the cysteine-spanning region. To identify the first exon sequences, 5–15 kb of finished or unfinished genomic sequence containing each putative defensin gene were retrieved from GenBank and used for computational prediction by using GenomeScan (61), GENSCAN (5), and/or GeneWise2 (4). As a result, a number of novel, full-length defensin precursor sequences with the conserved six cysteines and signal peptides have been identified in...
each species and submitted to GenBank. All new genes were named by following the recommendations of the Human Genome Organization (HUGO) and Mouse Gene Nomenclature Committees and kept consistent with the published sequences.

In the case of the rat, a total of 42 β-defensin (Defb) genes and a pseudogene (Defb16-ps) were discovered, including four known ones (24, 29, 41, 68) (Fig. 1). Among all rat genes, 12 β-defensins, namely Defb1, -4, -14, -22, -24, -27, -29, -33, -36, -39, -51, and -52, were found in the expressed sequence tag (EST) database (data not shown), from which full-length peptide precursor sequences can be deduced (Fig. 1). All the remaining genes were identified from WGS and HTGS sequences with first exons being predicted, except for Defb19. Consistent with earlier observations, the predicted signal sequences are highly homologous with known β-defensin genes in other mammalian species, rich in leucine and hydrophobic residues (Fig. 1). The accuracy of the exon-intron boundary prediction for each gene has also been confirmed by direct sequencing of RT-PCR product with exon-spanning primers (Supplemental Table S2). Inability to identify the signal sequence for Defb19, Defb55, and Defb56-ps was eventually confirmed by sequencing of RT-PCR product amplified from appropriate tissues. The failure to amplify these three genes most likely resulted from no or extremely low expression in healthy animals.

Although a number of β-defensins have been described in humans and mice with the use of hidden Markov models (HMM) (47), later discovery of another novel sequence in the human (42) prompted us to resequence the human and mouse genome sequences for the possible presence of additional new sequences, by using the strategy that we described above. Albeit of low throughput, our method yielded a total of 39 human and 52 mouse β-defensin genes and pseudogenes, including four novel human (DEFB133, -134, -135, and -136) and six new mouse β-defensins (Defb49, -50, -51, -52, -53, and -54-ps). All new β-defensin sequences consist of the characteristic signal sequence and a conserved defensin motif (Fig. 2A). Moreover, most of their orthologs are present in dogs and/or rodents except for mouse Defb53 and Defb54-ps, which appear to be mouse specific (Fig. 2B). To further support their

![Fig. 1. Amino acid sequence alignment of rat β-defensins. Dashes are inserted to optimize the alignment, and conserved residues are shaded.](http://physiolgenomics.physiology.org.org by 10.220.33.4 on August 14, 2017)
authenticity, all new genes were found to locate inside the same genomic clusters as others (Fig. 3).

We also predicted first exons for most human (Supplemental Fig. S2) and mouse genes (Supplemental Fig. S3), which were lacking in the earlier publication (47). The predicted sequences and their intron-exon boundaries are consistent with the rat orthologous genes, which have been confirmed by direct sequencing. Such sequences are also in agreement with a few full-length human and mouse β-defensins as described in the follow-up studies (3, 32, 39, 42, 48, 58, 66).

We also searched the chimpanzee genome and found a repertoire of 37 β-defensin genes that share 91–100% identity to their human orthologs in each case (data not shown), except that DEFB110 and DEFB128 orthologs were not found in the chimpanzee genomic sequences currently available in GenBank. Because of their conservation in dogs and rodents, there is little reason to suspect that DEFB110 and DEFB128 are missing in the chimpanzee, but instead it is most likely due to low coverage (4-fold) of the current genome sequences.

Alignment of all β-defensins revealed a high degree of diversity in the primary amino acid sequence in the mature region, despite the conservation of six cysteine residues, indicative of their possible nonredundant biological functions. Noticeably, the DEFB126/CBD126/Defb122 lineage (Fig. 1 and Supplemental Figs. S1–S3) and dog-specific CBD141 (Supplemental Fig. S1) contain an additional four amino acids between the third and fourth cysteine compared with most other β-defensins. Moreover, three canine- and rodent-specific β-defensin lineages, namely CBD138/Defb33, CBD139/Defb52, and CBD140/Defb51, consist of three to four extra amino acids between the second and third cysteine but two to three fewer amino acids between the third and fourth cysteine (Fig. 1 and Supplemental Figs. S1 and S3). Conceivably, such variations are expected to alter their overall tertiary structures and possibly functional properties from conventional β-defensins.

Strikingly, in contrast to most characterized β-defensins that contain less than five amino acids following the last cysteine, the majority of the family members in fact have much longer and mostly linear carboxy-terminal tails. The role of such long tails remains largely a mystery. However, several β-defensins were recently found to be heavily glycosylated in vivo in the tail region (30, 41, 65, 68). Furthermore, rat Defb22/E-3/2D6 and rhesus DEFB118/ESC42 contain lectin and trefoil-like motifs, respectively, which potentially facilitate oligomerization and interactions of proteins with target cell membrane (30, 41).

Chromosomal clustering of β-defensins. To identify genomic organization of β-defensins, individual β-defensin genes were searched against the most current versions of the assembled human, mouse, rat, chimpanzee, and dog genomes by using the BLAT program (25) through the UCSC Genome Browser. It is apparent that β-defensin genes in each species are clustered densely in four to five different chromosomal regions, with each cluster spanning between 55 and 1,118 kb (Fig. 3). All clusters are syntenic across species, because orthologous genes are extensively conserved with the same order and orientation for most genes in each cluster (Fig. 3).

In the rat genome (BCM version 3.1), all 43 β-defensin genes have been mapped and form four distinct clusters expanding 812 kb, 78 kb, 55 kb, and 193 kb continuously on chromosomes 16q12.5, 15p12, 9q12-q13, and 3q41, respectively (Fig. 3). Such syntenic clusters are also conserved in mice (NCBI build 32) with four clusters spanning 1,118 kb, 71 kb, 56 kb, and 164 kb on chromosomes 8q11.13-A2, 14q31, 1q3, and 2q11, respectively (Fig. 3). However, the genomic contig containing mouse Defb33, Defb51, and Defb52 has not been assigned to a specific region on mouse chromosome 8. The location of mouse Defb48-ps currently remains unmapped. Mouse Defb32, reported earlier (47), was found on chromosome 6. However, we suspect that
Defb32 is unlikely to be an authentic β-defensin gene, because 1) only five instead of six cysteines are present in the putative mature region, 2) such a cysteine-containing region is negatively instead of positively charged, 3) no characteristic first exon sequence could be identified within a 15-kb upstream sequence, and 4) no orthologs can be found in other mammals. It is also noteworthy that we could not find mouse Defb31 and Defb47-ps (47) in either EST.
WGS, HTGS, or assembled genome sequences that are publicly available in GenBank. The canine genome (NCBI build 1, version 1) also encodes four β-defensin clusters expanding 356 kb, 76 kb, 99 kb, and 373 kb on the syntenic regions on chromosomes 16, 25, 12, and 24, respectively (Fig. 3). However, human β-defensin genes in the most current genome assembly (NCBI build 35, version 1) were annotated in five separate clusters on three different chromosomes, with two on chromosome 8p23.1 (Fig. 3, A and B), one on 6p12.3 (Fig. 3C), and two on 20p13 and 20q11.21 (Fig. 3D). Apparently, two human β-defensin clusters on the p and q arms of chromosome 20 share a common ancestor with a continuous syntenic cluster in rodents and dogs (Fig. 3D). In contrast, another two human gene clusters on 8p23.1 are only separated by <4 Mb, but their counterparts in rodents and dogs are located on two different chromosomes (Fig. 3, A and B). The current chimpanzee genome (NCBI build 1, version 1) also encodes five major β-defensin clusters in almost perfect synteny with human genes (Supplemental Fig. S4). It is noteworthy that the α-defensin cluster is located within the largest β-defensin cluster in rodents and humans but is apparently missing in the canine genome (Fig. 3A), consistent with the fact that α-defensins have only been discovered in glires and primates (36).

Earlier studies indicated that a varied copy number of α- and β-defensin genes are present on chromosome 8p23.1 of human individuals (19, 51). Because of such complexity, current annotation of several human β-defensins in this region, such as DEFBl07, -108, -109, -130, and -131, shows an obvious discrepancy in the location and order of their orthologous genes with other species (Fig. 3, A and B). Thus the accuracy of the current human genome assembly in this region remains to be refined and corrected in the future (51).

Nevertheless, our current version of human and mouse β-defensin genomic clusters manifests a significant improvement over an earlier attempt (47). For example, several genes in human 8p23.1 and their orthologous loci in mice have been reordered, which agrees with their synteny in other species (Fig. 3A) and is also consistent with more recent studies (48, 51). Moreover, an ambiguous cluster containing DEFB130 and DEFB131 (47) has been mapped to human chromosome 8p23.1, with three additional novel genes (DEFB134, -135, and -136) being added (Fig. 3B). Another novel human β-defensin gene (DEFB133) has also been mapped to 6p12.3, and its orthologs have been found and localized in rodent and canine genomes (Fig. 3C).

Structural organizations and transcriptional flexibility of β-defensin genes. A comparison of all available ESTs with genomic sequences of β-defensins across mammalian species revealed that, similar to enteric α-defensins (35), most β-defensin genes are composed of two separate exons, with the first exon encoding the signal and pro-sequence and the second exon encoding primarily the mature sequence containing the defensin motif. The intron sizes of β-defensins vary greatly, spanning 1–10 kb in most cases, in contrast with α- and θ-defensins, consisting of introns of usually <1 kb (36). However, we also found an additional five distinct patterns of the β-defensin gene structure (Fig. 4). For example, mouse Defb21 (EST accession no. AK076875), mouse Defb44 (AK079042), rat Defb29 (CK839317), rat Defb51 (BM390654), and canine CBD102 (DN346173), CBD139 (CO687407), and CBD140 (CO692331) have EST sequences showing the presence of an additional one or two exons encoding the 5′-untranslated region (UTR), reminiscent of myeloid α-defensins (35).

On the other hand, mouse Defb30 (AK078987) contains an additional one exon encoding the 3′-UTR. Mouse Defb28 (AV044615) and canine CBD122 (BM537999) are unique in that they are composed of three exons, with the last exon encoding a segment of the mature sequence downstream of the six cysteines and the first two exons encoding the signal, pro-sequence, and majority of mature sequences. A few genes, such as SPAG11c/CSPAG11c/Spag11c (11, 12) and DEFB105/CBD105/Defb12 (data not shown), exhibit a three-exon structure, containing an extra intron presumably encoding an additional segment of pro-sequence. Interestingly, rat Defb52 (CR466423) is composed of four exons, with three exons encoding the entire open reading frame and the fourth exon encoding a part of 5′-UTR. Additional gene structures are expected to be revealed with availability of more mRNA and EST sequences for β-defensins.

It is well known that the SPAG11 gene is capable of differentially utilizing different exons to produce multiple isoforms of β-defensin-like sequences (11, 12). Such alternative splicing now does not appear to be unique to the SPAG11 gene. We have found two EST sequences (AK020304 and BB787829) that consist of the first exon of mouse Defb17 and the second exon of Defb16, suggesting that these two genes are actually alternatively spliced isoforms. Such splicing is most likely evolutionarily conserved, because of the absence of signal sequence for its orthologous gene, DEFB110/CBD110/Defb16, in the human, chimpanzee, dog, and rat. Therefore, primates DEFB110 and DEFB111, canine CBD110 and CBD111, and rat Defb16-ps and Defb17 are likely to be two isoforms of the same gene. Similarly, human DEB119 and DEFB120 have been shown recently to share the same first exon encoding the signal sequence (39). CBD122, on the other hand, has three isoforms sharing identical signal, pro-sequence, and the majority of mature sequences, with the only difference being in the carboxy-terminal tail after the last cysteine (44).
Collectively, acquisition of additional exons and frequent alternative splicing add additional layers of complexity in β-defensin gene transcription and regulation that presumably allow the host to better cope with invading infections.

**Phylogenetic analysis of β-defensins.** To reveal the evolutionary relationships of all known β-defensins including newly identified ones in rodents, dogs, and primates, we sought to construct a phylogenetic tree by using the neighbor-joining method (43). Interestingly, genes within a genomic cluster tend to form a separate clade from the genes in other clusters (data not shown), implying that many of them most likely evolved by gene duplication events. Indeed, a number of highly similar paralogous genes are present in each species and physically located adjacent to each other. Consistent with the genetic mapping data (Fig. 3), many β-defensins have obvious orthologs with minimum sequence variation across four mammalian species and are clustered together (Fig. 5), meaning that these lineages arose before the last common ancestor of rodents, canines, and primates. Some even form well-supported clusters with different chicken β-defensins, suggesting their existence before the divergence of mammals from birds (57). The conservation of these genes during evolution may be indicative of their functional significance.

Conversely, subgroups of β-defensins also exist that are specific in mice, rats, and dogs, implying that these genes originated after these species separated and have undergone different evolutionary patterns. For example, two well-supported clades of rodent-specific genes (clade I: Defb2, -9, -10, and -11; and clade II: Defb37, -38, -39, -40, and -50) exist (Fig. 5). High sequence similarity within each clade supports the notion that these rodent-specific β-defensin genes arose from gene duplication and positive diversifying selection as previously described (32, 48). In contrast to the clade II genes with the orthologs present in both rodent species, many mouse β-defensins in clade I have no obvious orthologs in rats, implying that many clade I genes were likely to have emerged only after the mouse-rat split about 41 million years ago (27), whereas all clade II genes apparently appeared before the split. Another large cluster of rodent genes, namely Defb3, -4, -5, -6, -7, and -8, form a common clade with human DEFB4/hBD-2 gene (Fig. 5). Therefore, it is likely that such a gene lineage duplicated and expanded significantly in rodents but remained unchanged in primates and was lost in the canine lineage. Alternatively, multiple such ancestral gene lineages were lost in primates and canines but retained in rodents. However, significant homology of murine genes in this cluster, as well as their physical vicinity (Fig. 3A), supports the former conclusion.

Multiple gene lineages specific to dogs and primates are also present, including DEFB104/CBD104, DEFB108/CBD108,
DEFB114/CBD114, DEFB120/CBD120, DEFB121/CBD121, DEFB133/CBD133, and DEFB134/CBD134 (Figs. 3 and 5). Conversely, CBD138/Defb33, CBD139/Defb52, and CBD140/Defb51 (Fig. 3A) only exist in rodents and dogs but are absent in primates (Fig. 5). Dog-specific genes also exist, such as CBD141 and CBD142 (Fig. 3D). However, no primate-specific β-defensins were found, suggesting that gene duplication did not occur in the primates after their divergence. It is noteworthy that CBD102 is not the canine ortholog of human DEFB4 or murine Defb2 or Defb4, but instead it is specific to the dog, paralogous to CBD103 (Fig. 5 and Supplemental Fig. S1). Therefore, the CBD102 gene most likely evolved from duplication of the CBD103 gene after separation of canines from other mammals.

Tissue expression pattern of rat β-defensins. We next examined the tissue expression patterns of the entire β-defensin gene family in the rat as a first step toward understanding their in vivo biological functions. Semiquantitative RT-PCR was performed with all identified rat β-defensin genes by using a panel of 29 different tissues from healthy, 2-mo-old, Sprague-Dawley rats. Primers were designed from two different exons to differentiate between the amplicons from cDNA and those from genomic DNA and verify predicted exon-intron junctions (Supplemental Table S2). Different numbers of PCR cycles were performed for each gene to ensure the linearity of amplification. Specificity of each amplicon was confirmed by direct sequencing of the RT-PCR product or recombinant plasmid containing the RT-PCR product.

To our surprise, nearly all β-defensins were found to be expressed preferentially in the male reproductive system, particularly in epididymis and testis (Fig. 6A), with no or minimum expression in most other tissues (data not shown). More...
strikingly, the reproductive tract-specific genes are expressed differentially in three different regions of epididymis, namely caput (head), corpus (body), and cauda (tail) (Fig. 6A). For example, Defb12/35, -15/34, -17, -18, -21, -25, and -41, Spag11c, and Spag11e are more abundantly expressed in the caput of epididymis, whereas Defb9, -10, -11, -13, and -40 are restricted to the cauda (Fig. 6A). On the other hand, Defb24 and Defb33 are highly specific to testis (Fig. 6A). Our results are consistent with earlier studies on the expressions of rat Defb1/ RBD-1 (8, 24) and Spag11e/Bin-1b (29) and also largely agree with the expressions of their human and mouse orthologs (39, 42, 58, 59, 66).

In addition to a canonical Spag11c/EP2c transcript (3, 12, 56), an alternatively spliced longer transcript was detected in the rat (Fig. 6A) and has been deposited in GenBank under accession number DQ012093. Such a transcript is unique in that it contains an extra 60 bp at the 3′-end of the first exon, resulting in a longer pro-sequence, which is not found in either human or rhesus monkey (11, 12). To further illustrate the difference in the SPAG11/HE2/EP2 locus between primates and rodents, the d isoform, which is composed of the first two exons of the c form and the last exon of the e form, was not detected in the rat (data not shown) but is readily detectable in the primates (11, 12).

Although the majority of genes showed the most abundant expression in testis or epididymis of healthy adult rats, Defb1 and Defb42 are also expressed at considerable levels in kidney (Fig. 6A) and Defb24 in spleen and ovary (data not shown). Defb36 was found to a lesser extent on mucosal surfaces lining the respiratory, gastrointestinal, and reproductive tracts as well as on skin (Fig. 6B). The only obvious exception is Defb4/ RBD-2, which shows virtually no expression in either testis or epididymis but instead is more restricted to the respiratory tract, particularly in lung (Fig. 6B). Defb4 transcript was also found at low levels in the upper gastrointestinal tract as well as in the vagina (Fig. 6B), which is consistent with the previously published report (24). The biological significance of such concentrated, but apparently differential, production of so many β-defensins in testis and different regions of epididymis remains largely unknown.

It is noted that we failed to detect the expression of Defb3, Defb5, and Defb16-ps following multiple attempts with different combinations of primer pairs. Although we cannot rule out the possibility of a wrong computational prediction of first exons, it is most likely that Defb3 and Defb5 are expressed at extremely low levels in healthy adult rats but may be inducible in response to infection and inflammation. Defb16-ps could be a nonexpressing pseudogene. Alternatively, these three genes could be expressed in tissues other than the ones that we examined. Because of the inability to identify the first exon of Defb19, we could not detect its gene expression by RT-PCR either.

Developmental regulation of rat β-defensins. Semiquantitative RT-PCR was performed to study developmental expression patterns of a few selected rat genes specific to epididymis and/or testis. Rat testes were collected on 1, 15, 30, 60, and 90 days after birth, and epididymides were collected on 7, 15, 30, 60, and 90 days after birth, with 2–4 animals used per time point. As shown in Fig. 7A, expressions of all selected β-defensins (Defb1, -15, -29, -30, -42, and -49) in epididymis are significantly elevated during development, with virtually no expression on day 7 but a rapid peak in expression at 1 mo. Defb42 is slightly different in that it shows little expression up to 1 mo, but the expression level is dramatically enhanced at 2 mo of age and increased up to 3 mo after birth. Such a pattern is reminiscent of rat Spag11e/Bin-1b, whose transcript does not peak in epididymis until 4 mo and remains elevated in rats of 2 yr of age (29).

However, β-defensin genes in testis showed a different pattern. Although most epididymal β-defensin genes were not expressed on day 7, many genes in testis display considerable expression soon after birth, with some (Defb24 and Defb29) expressed constitutively in rats of all ages studied and some (Defb27 and Defb36) upregulated and peaked at 1 mo of age (Fig. 7B). Defb33, on the other hand, behaves similar to epididymal Defb42 (Fig. 7A) in that it is not detected in testis until 1 mo but continues to increase for at least 3 mo after birth (Fig. 7B). It is interesting to note that Defb29 is developmentally regulated in the epididymis (Fig. 7A) but exhibits a constitutive pattern in the testis (Fig. 7B), implying the presence of a certain epididymis-specific enhancer(s) or testis-specific inhibitory factor(s).

The constitutive expression of several β-defensins in testis, but not epididymis, of newborn animals is presumably required...
to protect testis and particularly germ cells from invading infections before the time of sexual maturation. Simultaneous production of such a large array of defensins in various segments of the male reproductive tract with different developmental expression patterns implies that these molecules may play a nonredundant role in the reproductive process.

**DISCUSSION**

Sequence diversity of mammalian β-defensins. Through systemic genome-wide screening, we have discovered a total of 39, 37, 43, 52, and 43 β-defensin genes and pseudogenes in the human, chimpanzee, dog, mouse, and rat, respectively. Alignment of all β-defensins revealed a high degree of conservation in the spacing pattern of six cysteines with a consensus pattern of C-X5–8-C-X5–7-C-X5–13-C-X4–7-CC. The invariantly spaced cysteines form a rigid, triple antiparallel β-sheet structure and are believed to be critically important in maintaining functional activities (14, 28). However, we observed the absence of a canonical cysteine or presence of an extra cysteine in a few putative functional defensins. For example, human DEFB107 and mouse Defb8/Defr1 have a nonconservative mutation in the first cysteine, whereas mouse and rat Defb50 lack the second conserved cysteine (Supplemental Figs. S2 and S3). DEFB133/Defb49 lineage, on the other hand, contains a mutation in either the fifth or sixth cysteine in the mouse and human (Supplemental Figs. S2 and S3) but is fully conserved with six canonical cysteines in the rat (Fig. 1).

Extra cysteines have also been found within the six-cysteine motif of three peptides, namely human DEFB119 and DEFB132 and mouse Defb5 (Supplemental Figs. S2 and S3). In contrast, the instance of cysteines occurring outside the defensin motif appears to be much higher, with multiple members being found in each species. Among 39 human β-defensins, seven such peptides are found (DEFB105, -106, -112, -117, -126, -127, and -133) (Supplemental Fig. S2). Interestingly, five are conserved across species, including DEFB105/CBD105/Defb12/Defb35, DEFB106/CBD106/Defb15/Defb34, DEFB112/CBD112, DEFB117/CBD117/Defb19, and DEFB126/CBD126/Defb22.

It is largely unknown how a missing or an extra cysteine would affect the tertiary structures and functional properties of β-defensins. Obviously, the existence of an odd number of cysteines will potentially oligomerize through intermolecular disulfide bridging. Indeed, such β-defensins as Defb8/Defr1 (6) and DEFB126/Defb22/ESPI3.2/E-3/2D6 (41, 65, 68) have been shown to form covalently bonded homodimers in vitro and/or in vivo. Such dimerization in fact leads to enhanced antibacterial activities in the case of Defb8/Defr1 (6). Not surprisingly, a few β-defensins with six cysteines also have a tendency to form dimers or oligomers, which are believed to facilitate membrane interaction and permeabilization (20, 45).

It is noteworthy that our strategy yielded a total of 10 additional novel β-defensins in humans and mice compared with the strategy based on the HMM (47). Although we believe that the current list most likely represents the complete repertoire of β-defensins in these species, we could not rule out the possibility that additional β-defensin-related genes with remote similarity might be uncovered in these species. In fact, a group of human SPAG11/HE2/EP2 isoforms were recently found to share identical signal and pro-sequences with similar antimicrobial properties with classic β-defensins but differ significantly in the number and cysteine spacing pattern in the mature sequence (3, 56, 62, 63). Such α-defensin-related sequences have also been discovered in mice (21, 23) and rats (36). Nevertheless, the current study represents the most comprehensive attempt to identify mammalian canonical β-defensin gene family members with the characteristic six-cysteine motif.

**Evolutionary relationships of mammalian β-defensins.** β-Defensins are the most ancient family of vertebrate defensins compared with α- and θ-defensins (57). β-Defensin-like sequences have been found in the rattlesnake (34), orange-spotted grouper (GenBank accession no. AY129305), and zebrafish (Zhang G, unpublished results), suggesting that the primordial gene for β-defensins appeared before the fish-primate split about 450 million years ago (27). In contrast, α-defensins were derived from β-defensins after divergence of glires and rodents from other mammals about 91 million years ago, whereas θ-defensins did not appear until primates separated from other mammals around 23 million years ago (27, 33, 36).

Our earlier study discovered a single cluster consisting of 13 β-defensin genes encoded in the entire chicken genome (57). Comparative analysis of the chicken and mammalian β-defensin gene clusters revealed that two clusters on human 8p23.1 and their orthologous loci in other mammalian species (Fig. 3, A and B) are syntenic with the chicken gene cluster, implying that the ancestral genes in these two mammalian clusters evolved before the bird-mammal separation (57). Many genes in these two ancient clusters were demonstrated to have undergone significant repeated duplication and positive diversification in humans and mice after their divergence (32, 48), which further gave rise to other β-defensin clusters in mammals during evolution, presumably as a result of chromosomal translocation and expansion of certain gene lineages.

Consistent with the evolutionarily active nature of the two ancient clusters, subgroups of genes exist in these two regions that are specific in rodents, dogs, and/or primates (Fig. 5). This notion is further reinforced by the presence of multiple highly homologous α- and β-defensins and emergence of α- and θ-defensins in rodents and primates in these regions (Fig. 3A). In contrast, the remaining genomic clusters are rather static throughout evolution with the presence of orthologs across mammalian species, except that CBD141 and CBD142 on chromosome 24 are specific to dogs and were likely duplicated and diverged from adjacent genes (Fig. 3D) after the dog diverged from other mammals.

Collectively, these results suggest that the evolution of mammalian β-defensins is an extremely dynamic and active process. Individual gene lineages were derived at different evolutionary times. Although some β-defensins evolved before divergence of mammals from fish, many appear to be mammal specific, with a few emerging as recently as the rat-mouse split, which occurred only 41 million years ago (27). However, no gene duplication occurred before the separation of chimpanzee from humans about 5.5 million years ago (27) because of the complete orthology of β-defensin genes in these two species (Supplemental Fig. S4).

What is the driving force for the presence of such a large number of divergent β-defensins in mammals? One plausible explanation is that sequence diversification may confer func-
tional novelties on different β-defensins for each species to better cope with a variety of microbial threats from the environment. Consistent with this hypothesis, different β-defensins (15, 16, 18) and even orthologous genes in different primates (2) have been found to differ quite dramatically in their antimicrobial potency and spectrum. Furthermore, expression of a large array of β-defensins in the male reproductive tract is presumably to safeguard the reproductive process, enhance fertility, and sustain species survival, as detailed in the next section.

Dual role of region-specific expression of β-defensins in the male reproductive tract. Obviously, optimal reproductive function ensures survival of the species, which needs full protection of germ cell precursors in testis and sperm in different stages of maturation in epididymis and after deposition in the female reproductive tract. Infection of testis and epididymis may result in temporary or permanent loss of fertility by disrupting the specialized environment of these organs conducive for sperm storage and maturation (37). Because adaptive immunity is largely absent in the male reproductive system, the host must evolve alternative and effective mechanisms of protection of sperm from infectious agents.

Region-specific expression of all but one of the rat β-defensins in the male reproductive system (Fig. 6) suggests that these molecules may constitute an essential component in maintaining the normal reproductive process. In line with this, many testis- and epididymis-specific β-defensins were shown to be antimicrobial and capable of protecting sperm from infections (3, 8, 29, 56, 62, 64). In addition to being microbiocides, β-defensins expressed in different regions of epididymis appear to be actively involved in reproduction. These defensins are mainly produced by epithelial cells lining the epididymal duct, regulated by androgens, and secreted in luminal plasma, and they bind preferentially to the surface of maturing but not immature sperm (29, 30, 41, 65, 68, 70). Rat Spag11e/Bin-1b has been shown to enhance sperm maturation by inducing Ca2+ uptake and subsequent motility and progressive movement of immature spermatozoa (70). Moreover, another epididymis-specific β-defensin, macaque DEFB126/ESP13.2, coasts the entire ejaculated sperm and masks zona pellucida ligands on the sperm surface but becomes dissociated when sperm are fully capacitated, suggesting that DEFB126 may be an important decapacitation factor on the sperm surface that needs to be removed before sperm-zona interaction and fertilization (54, 65).

Aside from β-defensins, α-defensins have also been reported in the male reproductive tract (8, 17). Human cathelincidin LL-37/hCAP-18, which belongs to another important family of vertebrate antimicrobial peptides (40, 67), is expressed by epididymal epithelia and is abundantly present in the seminal fluid with ability to bind to sperm (31). Furthermore, the female reproductive tract also produces α- and β-defensins albeit at low levels (26, 38, 50). Coupled with the fact that many β-defensins in epididymis and testis are developmentally regulated, with elevated expression at the time of puberty and sexual maturation (Fig. 7), these results strongly favor the argument that β-defensins, perhaps together with many other antimicrobial peptides, have a dual function in both innate host defense and fertility.

Defb4/RBD-2 represents the only β-defensin that is expressed most abundantly in lung with virtually no expression in the reproductive tract (Fig. 6B), and thus it may play an important role in airway defense. Its human ortholog, DEFB4/hBD-2, was shown to be the predominant β-defensin in human neonatal lung and developmentally regulated (49), and its concentration is inversely correlated with the severity of lung disease in cystic fibrosis patients (7). Further studies are needed to identify the regulatory mechanisms that specify the expression of defensins in the male reproductive tract vs. respiratory tract. Functional divergence of β-defensins produced in the same and different regions of the reproductive tract also remains to be investigated, and these molecules may have potential as fertility and contraceptive drugs.

ACKNOWLEDGMENTS
We thank Lin Liu and Kathy Swenson for assistance with rat tissue harvesting and Yanjing Xiao for initial help with the developmental regulation of rat β-defensins.

GRANTS
This work was supported by the Oklahoma Center for the Advancement of Science and Technology Grant HR03-146 and Oklahoma Agricultural Experiment Station Project H-2507.

REFERENCES


