INFECTION AND IMMUNITY, Jan. 2011, p. 459–473 Vol. 79, No. 1
0019-9567/11/$12.00 doi:10.1128/IAI.00996-10
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

Strain-Specific Polymorphisms in Paneth Cell α-Defensins of C57BL/6 Mice and Evidence of Vestigial Myeloid α-Defensin Pseudogenes▽†

Michael T. Shanahan,‡ Hiroki Tanabe,§ and André J. Ouellette*

Department of Pathology and Laboratory Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, California

Received 13 September 2010/Returned for modification 12 October 2010/Accepted 17 October 2010

Paneth cells at the base of small intestinal crypts secrete microbicidal α-defensins, termed cryptdins (Crps) in mice, as mediators of innate immunity. Proteomic studies show that five abundant Paneth cell α-defensins in C57BL/6 mice are strain specific in that they have not been identified in other inbred strains of mice. Two C57BL/6-specific peptides are coded for by the *Defcr20* and *-21* genes evident in the NIH C57BL/6 genome but absent from the Celera mixed-strain assembly, which excludes C57BL/6 data and differs from the NIH build with respect to the organization of the α-defensin gene locus. Conversely, C57BL/6 mice lack the Crp1, -2, -4, and -6 peptides and their corresponding *Defcr1*, *-2*, *-4*, and *-6* genes, which are common to several mouse strains, including those of the Celera assembly. In C57BL/6 mice, α-defensin gene diversification appears to have occurred by tandem duplication of a multigene cassette that was not found in the mixed-strain assembly. Both mouse genome assemblies contain conserved α-defensin pseudogenes that are closely related to functional myeloid α-defensin genes in the rat, suggesting that the neutrophil α-defensin defect in mice resulted from progressive gene loss. Given the role of α-defensins in shaping the composition of the enteric microflora, such polymorphisms may influence outcomes in mouse models of disease or infection.

Mammalian α-defensins are cationic antimicrobial peptides (AMPs) with microbicidal activities against many species of bacteria and fungi, antiviral functions, and diverse immunomodulatory effects (16, 26, 56). Although α-defensin genes are absent in the cattle and dog genomes (13, 42), the peptides are expressed in distant mammalian species, including humans and primates as well as additional Euarchontoglires, such as horses (5, 6), elephants, opossum, tenrecs (3, 30), and the platypus (65, 66).

Bone marrow promyelocytes and Paneth cells in the crypts of Lieberkühn of the small intestine are the two major sites of α-defensin gene expression and peptide biosynthesis. In neutrophils, α-defensins in azurophil granules contribute to nonoxidative killing of ingested microbes following phagolysosomal fusion (55). Paneth cells secrete α-defensins into the crypt lumen, and they have a key role both in mediating enteric innate immunity and in determining the composition of the small intestinal microbiome (46, 47). Myeloid and Paneth cell α-defensin genes differ: genes expressed by Paneth cells consist of two exons separated by a single intron of ~500 bp (20, 21, 56), in contrast to 3-exon myeloid α-defensin genes which have an additional intron that interrupts the 5′-untranslated region near the translation initiation site (17, 28). Curiously, the mouse is the only species known to express Paneth cell α-defensins but to lack them in neutrophils (12).

Aside from three canonical structural features, including six invariantly placed Cys residue positions, a canonical Arg-Glu salt bridge, and a highly conserved Gly residue position, the primary structures of α-defensins are remarkably diverse (26, 50, 56). In most species, several distinct isoforms occur at high levels in neutrophils and also in Paneth cells (22, 38). The extensive divergence of α-defensin primary structures, even between closely related species, often complicates identification of peptide orthologs. In the mouse, for example, 26 Crp isoform cDNAs have been described (19, 41), but only 6 Crps have been detected at the peptide level in outbred Swiss mice (10, 36, 38, 54) and/or in a variety of inbred strains, including 129/SvJ (20), C3H/HeN (18, 20), C3H/HeJ (8, 38), NMRI/KI, FVB, and BALB/c (23), and FVBJ (58). In addition to Crp-coding *Defcr* genes, Paneth cells of inbred and outbred mice also express two-exon *Defcr-rsN* genes (provisional gene designation, where $N = 1$ to 12), which code for Cys-rich sequence 1C (CRS1C) and 4C (CRS4C), peptides that are unique to the mouse as a species (18, 19, 39). Although the CRS1C and CRS4C peptides differ markedly from α-defensins, the *DefcrN* (provisional designation for the gene coding for mouse Paneth cell α-defensins, where $N = 1$ to 27) and *Defcr-rsN* genes have nearly identical first exons. Although the literature implies that inbred strains of mice have the same complement of α-defensin and CRS peptides, new evidence shows otherwise.

Previously (58), two distinctive α-defensins were reported in C57BL/6 small bowel, consisting of two Crp4-related variants that had not been found earlier in cloning and peptide studies of 129/SvJ, C3H/HeJ, BALB/cJ, or outbred Swiss mice (8, 10, 18, 20, 32, 33, 38, 54). These observations and

* Corresponding author. Mailing address: Department of Pathology and Laboratory Medicine, Keck School of Medicine of the University of Southern California, USC Norris Cancer Center, NRT 7514, 1450 Biggy Street, Los Angeles, CA 90089-9601. Phone: (323) 442-7959. Fax: (323) 442-7962. E-mail: aouellet@usc.edu.

‡ Present address: University of North Carolina, Chapel Hill School of Medicine, Center for Gastrointestinal Biology and Disease, Room 7317 BRB, Chapel Hill, NC 27599.

§ Present address: Research and Development Division, Health Policy Bureau, Ministry of Health, Labour and Welfare, 1-2-2, Kasumigaseki, Chiyoda-Ku, Tokyo 100-8916, Japan.

† Supplemental material for this article may be found at http://iai.asm.org/.

▽ Published ahead of print on 1 November 2010.

widespread use of C57BL/6 mice as a reference strain for genetic manipulation provided rationale for characterizing the α-defensin proteome of the C57BL/6 strain and for comparing the α-defensin loci in the NIH C57BL/6 and Celera mixed-strain assemblies.

We report that five abundant Crps and pro-Crps in C57BL/6 mouse small bowel differ from those characterized in other mouse strains and that C57BL/6 polymorphisms are evident at the genetic level. Comparative genomic analyses suggest that C57BL/6 α-defensin gene expansion and diversification occurred by tandem duplication of a multigene cassette. Also, vestigial myeloid α-defensin pseudogenes that are related to rat myeloid α-defensin genes occur in both mouse genome assemblies, providing evidence for the progressive loss of previously functional neutrophil α-defensin genes in mice.

MATERIALS AND METHODS

Reagents and animals. Crp4 and pro-Crp4 peptides were prepared in the *Escherichia coli* BL21 expression system (2, 48, 58). All procedures on mice were performed in compliance with the policies of the Institutional Animal Care Committees of the University of California, Irvine (UCI). Six-week-old adult male C57BL/6 mice were purchased from Charles River Breeding Laboratories, Inc. (North Wilmington, MA). Mice were housed in specific pathogen-free facilities under 12-h cycles of light and dark and had free access to standard rodent chow and water. Intestinal RNA was isolated from individual flushed, full-length, full-thickness adult small bowel from the pylorus to the ileocecal valve by extraction with guanidine isothiocyanate and with phenol, chloroform, and isoamyl alcohol (37).

Protein purification and analysis. Whole mouse small intestine was extracted with 30% (vol/vol) acetic acid and clarified by ultracentrifugation (27, 52, 54, 61). Samples dissolved in 5% (vol/vol) acetic acid were separated by P-30 Bio-Gel permeation chromatography (Bio-Rad, Hercules, CA) (31, 54) developed with 5% (vol/vol) acetic acid at a flow rate of 0.5 ml/min. Samples of column fractions were analyzed by acid urea-polyacrylamide gel electrophoresis (AU-PAGE) and by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (Voyager-DE STR; PerSeptive Biosystems, Foster City, CA) (31). Fractions containing nominal 4- to 6-kDa and also 8- to 10-kDa peptides were selected for further purification and analysis of Crps and pro-Crps, respectively. Because size separations by molecular weight using P-30 gel permeation chromatography are of low resolution, it should be noted that the size ranges are nominal and do not represent definitive size fractions. Lyophilized samples were applied to analytical reversed-phase high-performance liquid chromatography (RP-HPLC) using Vydac 218TP54 columns in 0.1% (vol/vol) trifluoroacetic acid. Putative Crps were eluted using a 23% to 33% (vol/vol) acetonitrile gradient for 100 min, and putative pro-Crps were resolved using a 25% to 40% (vol/vol) acetonitrile gradient for 150 min, monitoring both chromatograms by A_{230} (Fig. 1). Preliminary evidence of HPLC fractions containing putative α-defensins was obtained by assaying individual peptide peaks to identify fractions with peptides containing six Cys residues. Briefly, fraction samples were reduced with 5 mM dithiothreitol at 95°C for 15 min and then alkylated using a 3-fold molar excess of iodoacetamide for 1 h. The method increases the molecular mass of each Cys residue position by + 58 atomic mass units (AMU) and thus by 347 AMU for tentative identification of putative α-defensins or pro-α-defensins (31, 61).

Purified α-defensins and pro-α-defensins were identified by comparing experimentally determined atomic masses with known peptides or Crps deduced from cDNA or genomic sequences. Lyophilized peptide samples were dissolved in 10 μl of 5% (vol/vol) acetic acid containing 3 M urea and electrophoresed in 12.5% (wt/vol) acid-urea (AU)-PAGE (51) for 1 h at 100 V and 3 h at 250 V. Peptides were visualized by staining gels with 0.05% (wt/vol) Coomassie blue in 50% (vol/vol) methanol and 10% (vol/vol) acetic acid and destained in 50% (vol/vol) methanol and 10% (vol/vol) acetic acid. Peptide bands were excised, macerated, and sonicated in 50 μl of extraction solution consisting of 50% (vol/vol) formic acid, 25% (vol/vol) acetonitrile, and 15% (vol/vol) 2-isopropanol for 30 min at ambient temperature. After centrifugation, supernatants were dried by centrifugation *in vacuo*, dissolved in 5 μl of 50% (vol/vol) acetonitrile–5% (vol/vol) trifluoroacetic acid, and 0.5-μl samples were mixed with equal volumes of 10 mg/ml α-cyano-4-hydroxy-cinnamic acid, and analyzed by MALDI-TOF MS.

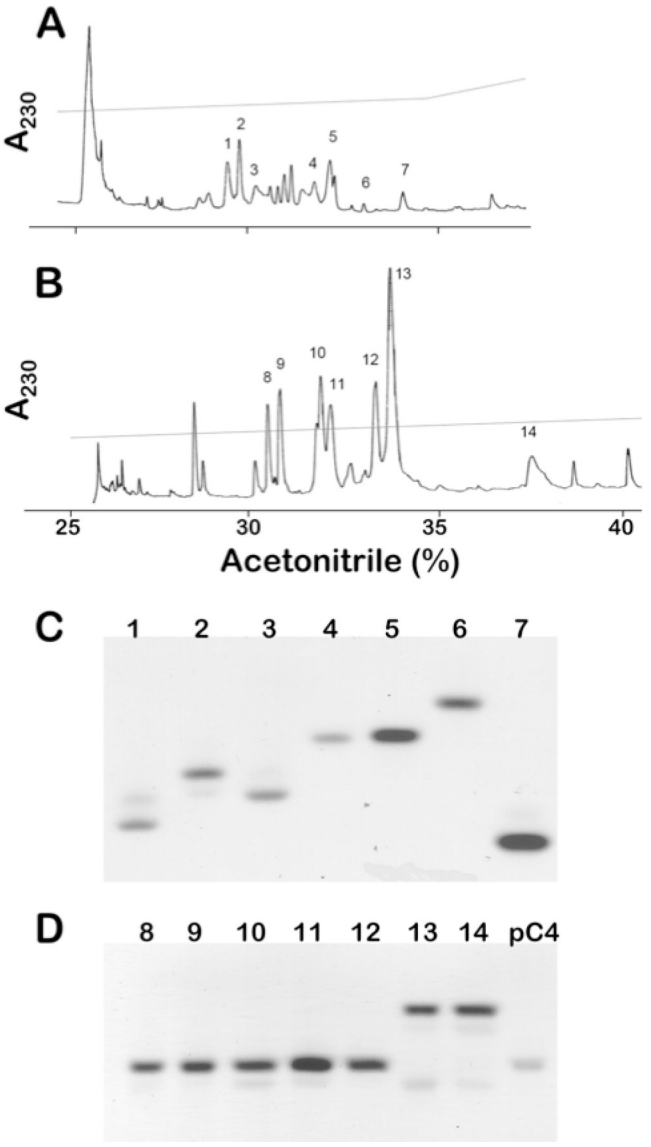

FIG. 1. Purification of C57BL/6 mouse Paneth cell α-defensins and their precursors. Chromatograms depict C_{18} RP-HPLC separations of C57BL/6 α-defensins (A) and pro-α-defensin (B) peptides. Peptides with apparent molecular masses of 4 to 5 kDa from P-30 gel permeation chromatography were applied to a Vydac 218TP54 C_{18} RP-HPLC column and developed with a 23% to 33% (vol/vol) acetonitrile gradient for 100 min. Seven peaks, numbered 1 to 7, were identified tentatively by MALDI-TOF MS as containing candidate α-defensin peptides (Materials and Methods). In panel B, peptide fractions from P-30 chromatography containing peptides with apparent molecular masses of 8 to 9 kDa were applied to the C_{18} column in 0.1% (vol/vol) trifluoroacetic acid, and proteins were resolved with a 25% to 40% (vol/vol) acetonitrile gradient for 150 min. As in panel A, MALDI-TOF MS analyses identified seven peaks, numbered 8 to 14, as containing putative pro-Crps. Following final purification of the peptides shown in panels A and B, 2-μg samples of recombinant peptides were analyzed by AU-PAGE and stained with Coomassie blue. In panel C, the lanes are as follows: 1, Crp3; 2, Crp20; 3, Crp23; 4, Crp27; 5, Crp24; 6, Crp5; and 7, Crp4. In panel D, the lanes are as follows: 8 and 9, pro-Crp20; 10, pro-Crp3; 11, pro-Crp23; 12, pro-Crp27; 13, pro-Crp24; 14, pro-Crp5 (Fig. 2); and pC4, recombinant pro-cryptdin-4.

Analysis of C57BL/6 pro-α-defensins. To identify putative pro-Crps purified from C57BL/6 small intestine, samples consisting of 5 μg of each purified peptide were incubated with or without 0.5 molar equivalent of MMP7 in a mixture of 1 mM HEPES, 15 mM NaCl, and 0.5 mM $CaCl_2$ (pH 7.4) for 18 h at 37°C (2, 58, 67). Samples of digests were analyzed by AU-PAGE and MALDI-TOF MS as described in the previous section. For analyses of MMP7 digests of pro-Crp5, approximately 200-ng quantities of complete digests were subjected to 5 or more cycles of N-terminal peptide sequencing at the former UCI Biomedical Protein and Mass Spectrometry Resource Facility.

Bactericidal peptide assays. *Escherichia coli* ML35, *Staphylococcus aureus* 710a, *Listeria monocytogenes* 10403S, the *Salmonella enterica* serovar Typhimurium Δ*PhoP* mutant (provided by Samuel I. Miller, University of Washington), and *Vibrio cholerae* 0395 were tested for peptide sensitivity in bactericidal peptide assays. These are laboratory strains rather than fresh clinical isolates, and they were selected because their *in vitro* responses to α-defensins and additional antimicrobial peptides are well characterized (29, 60, 61). Individual peptides were incubated at the concentrations shown with 1×10^6 CFU/ml of log-phase bacterial cells in 50 μl buffer consisting of 10 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid), pH 7.4] supplemented with 1% (vol/vol) of Trypticase soy broth (TSB) at 37°C for 1 h. After the incubation mixtures had been diluted 100-fold, they were plated on semisolid media using a Spiral Biotech Autoplater 4000 (Spiral Biotech, Bethesda, MD). Bacterial cell survival was determined by counting bacterial CFU after overnight growth.

Bioinformatics analyses. To identify genes coding for C57BL/6 mouse Crps, nonredundant nucleotide, protein, and mouse expressed sequence tag (est_mouse) GenBank databases using the BLASTP and BLASTN programs at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) as well as the RIKEN mouse cDNA database (http://fantom.gsc.riken.jp/4/). The query peptide sequences used for searching were Crp1, Crp4, Crp5, and Crp20, the product of the C57BL/6 *Defcr-20* gene and previously termed Crp4(B6a) (58).

α-Defensin gene locus analysis in mouse and rat. To identify and determine the relative positions of α-defensin genes in the mouse and rat genomes, the NCBI and Celera mouse (NCBI build 37.1) and rat genome assemblies (NCBI build 3.4) were interrogated using TBLASTN at the Mouse Genome Resources page at NCBI (http://www.ncbi.nlm.nih.gov/genome/guide/mouse). cDNA sequences corresponding to full-length Crp1 (*Defcr1*; NM_010031) and Crp4 (*Defcr4*; NM_010039) mRNAs were used to search for rat genomic sequences with similarities to mouse enteric α-defensin gene sequences. Similarly, CRS1C and CRS4C genes and sequences similar to these genes were located by querying the same databases using full-length CRS1C-1 (*Defcr-rs1*; NM_007844) and CRS4C-1 (*Defcr-rs2*; NM_007847) sequences in genomic BLAST searches. Queries with full-length cDNA sequences for rat myeloid α-defensin RatNP-1/2 (*Defa1/2*; NM_173329), rat enteric α-defensin RD-5 (*Defa5*; AF115768), and *Defa1/2* exon 1, were used to further characterize murine α-defensin gene organization. The chromosome locations of the α-defensin (*Defcr*), CRS1C, and CRS4C (*Defcr-rsN*) genes, as well as genes with apparent similarities to these α-defensin related families, were recorded using the Map Viewer program (http://www.ncbi.nlm.nih.gov/mapview). The coding functions of genomic sequences located by BLAST searches were assigned by translating sequences in six reading frames and locating the canonical Cys spacings that define the α-defensin, CRS1C, and CRS4C peptides. Unassigned or unannotated genes were designated according to the extent of similarity of the deduced genomic sequence translation products to previously described mouse and rat α-defensin and CRS peptides using BLASTP at NCBI.

In the C57BL/6J assembly, *Defcr* genes were found in two nonoverlapping contigs (NT_039455.7 and NT_039457.7) that are separated by an unsequenced gap of approximately 2.0 Mb (Table 1). An additional contig, termed "ungenomic chromosome 8 specific contig" (NT_166309.1), has not been aligned on the assembly to our knowledge (Table 1; see Table S3 in the supplemental material). In the Celera mixed-strain assembly, α-defensin-coding sequences are located on eight nonoverlapping chromosome 8 contigs (NW_001030882.1, NW_001030885.1, NW_001030886.1, NW_001030887.1, NW_001030888.1, NW_001030889.1, NW_00103891.2, and NW_00103892.1) separated by unsequenced gaps of an average 268 kb (Table 2; see Table S4 in the supplemental material). Additional "ungenomic contigs" (NW_001031579.1, NW_001034677.1, NW_001034895.1, NW_001037941.1, NW_001071933.1, NW_001072513.1, NW_001072962.1, NW_001073383.1, and NW_001073917.1) also include α-defensin genes in the Celera assembly (Table 2; see Table S4 in the supplemental material).

Phylogenetic analysis of mouse and rat α-defensin, CRS1C, and CRS4C genes. To determine the evolutionary relationships of murine α-defensin genes, the nucleotide sequences of intron 1 of mouse and rat enteric α-defensin, CRS1C, and CRS4C genes, intron 2 sequences of rat myeloid α-defensin genes, and intron 2 sequences of apparent mouse myeloid α-defensin pseudogenes were acquired from GenBank. Multiple sequence alignments were performed using ClustalW from the MEGA software version 4.0 (25, 59). Phylogenetic trees were constructed using the neighbor-joining method (45) by calculating nucleotide (*p*-distance) differences. One thousand bootstrap replications were used to test the reliability of each branch.

(Please note that all gene symbols [*DefcrN*, *DefcrN-ps*, *Defcr-rsN*, and *DefmapsN*] used in this report should be considered provisional. The nomenclature is currently under revision by the Mouse Genome Informatics [MGI], Rat Genome Database [RGD], and HUGO Gene Nomenclature Committee [HGNC] nomenclature committees and researchers in the field [1].)

RESULTS

In a previous study, two novel Paneth cell α-defensins, then termed Crp4(B6a) and Crp4(B6b), were purified from C57BL/6 mouse small intestine (58). Those peptides are coded for by the *Defcr20* and *Defcr21* genes, respectively. Alignments of their deduced precursor sequences showed that the *Defcr20*- and *Defcr21*-coded proregions were most similar to pro-Crp4, which is coded by the *Defcr4* gene cloned from the 129/SvJ strain (58). Previous analyses of intestinal cDNAs and genes from outbred Swiss mice and inbred strains, including 129/SvJ, C3H/HeN, C3H/HeJ, NMRI/KI, FVB, BALB/c, and FVBJ had not detected these Crp4-related coding sequences, prompting a proteomics and genomics study of C57BL/6 α-defensins.

Polymorphisms in C57BL/6 mouse Paneth cell α-defensin peptides. To assess C57BL/6 α-defensin diversity, protein extracts of C57BL/6 small intestine were separated by gel permeation chromatography, and samples of column fractions were analyzed by AU-PAGE and by MALDI-TOF MS (see Materials and Methods). Fractions containing nominal 4- to 6-kDa peptides of high mobility in AU-PAGE (data not shown) were combined for further RP-HPLC purification and MS analysis of putative α-defensins. Nominal 8- to 10-kDa peptides were subjected to separate pro-α-defensin isolation and analysis. Seven putative Crps (Fig. 1A, peaks 1 to 7) and pro-Crps (Fig. 1B, peaks 8 to 14) were identified by MS, and each was purified by using analytical C_{18} RP-HPLC (data not shown) to apparent homogeneity as judged by chromatograms and analytical AU-PAGE analyses (Fig. 1C and D). Individual, isolated Crp peptides migrated in AU-PAGE with mobilities characteristic of mouse α-defensins (38, 51, 54), and putative pro-Crps comigrated with recombinant pro-Crp4 (Fig. 1D). Individual peptides were reduced and reacted with iodoacetamide, and the atomic masses of each candidate peptide increased by ~348 AMU (Fig. 2B), corresponding to carboxyamidomethyl group modification of molecules containing six Cys residues, a characteristic feature of α-defensins and their precursors (see Materials and Methods). Peptide fractions with peaks evident in Fig. 1A and B but which did not conform to these α-defensin or pro-α-defensin criteria were not analyzed further, but it remains possible that additional C57BL/6 α-defensins exist.

Paneth cell α-defensins specific to C57BL/6 mice. Analyses of isolated C57BL/6 Paneth cell α-defensins (Fig. 1A, lanes 1 to 7) showed that most abundant C57BL/6 Crps differ from those identified previously in other mouse strains. Comparisons with known Crps or Crps deduced from cDNAs showed that masses acquired for C57BL/6-derived Crps corresponded to distinct Crp isoforms (Fig. 2A), of which Crps 20, 21, 23, 24,

TABLE 1. NIH C57BL/6 defensin locus from telomere to centromere

Gene identity or provisional symbol	Orientation[a]	Chromosome 8 position	Intergenic distance (kb)	MGI[b] designation
NT_039455.7				
Defb5	Cen	19247592–19250828		
Defma1-ps	Tel	19279000–19280000	29.2	
Defb3	Cen	19293361–19295339	13.3	
Defb54-ps	Cen	19300677–19304794	9.5	
Defma2-ps	Tel	19318972–19319352	14.2	*Defa-ps13*
Defma3-ps	Cen	19334863–19335554	15.5	
Defma4-ps	Cen	19358436–19359276	2.9	
Defma5-ps	Cen	19399300–19399900	40	
Defma6-ps	Cen	19422519–19423270	22.6	*Defa-ps14*
Defb8	Tel	19447606–19445769	24.3	
Defma7-ps	Tel	19473822–19473027	25.4	*Defa-ps15*
sim RpL19-ps	Cen	19492902–19493633	19.1	
Defb7	Cen	19495097–19497775	1.5	
NT_039457.7				
sim Defb52	Tel	22041664–22040297	15	
Defb51	Tel	22057370–22056690	26.2	
CRS4C-6	Tel	22084389–22083536	80.8	AY761185
Defcr21	Cen	22165224–22166196	19.2	*Defa21*
Defma8-ps	Cen	22185400–22185900	8.8	
Defcr23a	Cen	22194745–22195694	9	*Defa23*
Defcr5a	Cen	22204698–22205682	12.6	
Defma9-ps	Cen	22217800–22218000	5.5	*Defa-ps3*
Defcr25	Cen	22224119–22224965	6	*Defa25*
ribo L21	Cen	22231012–22231564	2.9	
CRS1C-2	Tel	22234497–22235775	4	
Zn MYM	Cen	22239793–22241841	16.7	
Hyb-Defma1-ps	Cen	22258800–22259050	15.8	*Defa-ps4*
vDefcr24	Cen	22274295–22274509	27.4	
Defcr22	Cen	22301954–22302925	19	*Defa22*
Defma10-ps	Cen	22321900–22322400	8.9	
Defcr23b	Cen	22331291–22332556	8.7	*Defa-rs7*
vDefcr5	Cen	22341236–22342220	16.7	
Defma11-ps	Cen	22358600–22358850	5.6	*Defa-ps5*
vDefcr18-ps (*Crpi*)	Cen	22365083–22365927	5.6	*Defa-ps6*
CRS1C-ps	Tel	22371500–22372500	4.9	*Defa-ps7*
Zn-MYM	Cen	22376351–22377865	11.6	
Defcr26	Cen	22389424–22390231	20.8	
Defma12-ps	Cen	22410300–22410600	15.1	
Defcr3	Cen	22427120–22427963	9.1	*Defa3*
Defcr5b	Cen	22437071–22438055	10.9	*Defa5*
Defma13-ps	Cen	22448800–22449050	5.2	
vDefcr2-ps	Cen	22455223–22456123	7.9	
ribo L21	Cen	22464000–22474000	1.5	
CRS1C(-4a)	Tel	22465565–22466700	3.3	*Defa-rs1*
Zn-MYM	Cen	22470000–22471000	11.3	
Hyb-Defma2-ps	Cen	22482263–22485701	10.1	*Defa-ps16*
Gap		22495820–22545830	50	
Defma14-ps	Cen	22549100–22549350	3.6	
vDefcr2,7,18-ps	Cen	22555539–22556383	5.5	*Defa-ps18*
ribo L21	Cen	22564864–22565148	8.5	
CRS1C-1	Tel	22565897–22567035	0.8	
Zn-MYM	Cen	22571500–22572000	4.5	
Hyb-Defma3-ps	Cen	22583721–22587169	11.5	*Defa-ps17*
Defcr20A	Cen	22619726–22620709	32.6	*Defa20*
Defcr20B	Cen	22639500–22640500	18.8	
vDefcr2,16-ps	Cen	22666760–22667760	26.3	
Defcr5c	Cen	22675839–22676823	9.1	
Defma15-ps	Cen	22687050–22687300	10.5	
vDefcr2,18	Cen	22693497–22694340	5.5	
ribo L21	Cen	22702812–22703096	8.5	
CRS1C-1	Tel	22703845–22704983	0.8	
Zn-MYM	Cen	22705159–22709922	0.2	
Defcr26	Cen	22728653–22729355	18.7	*Defa26*
Defma16-ps	Cen	22749500–22749750	20.9	

Continued on following page

TABLE 1—*Continued*

Gene identity or provisional symbol	Orientation[a]	Chromosome 8 position	Intergenic distance (kb)	MGI[b] designation
Defcr17	Cen	22766205–22767209	15.3	*Defa17*
Defcr5d	Cen	22776222–22777206	9	
Defma17-ps	Cen	22799100–22799900	22.1	*Defa-ps9*
vDefcr18-ps	Cen	22805444–22806420	5.4	*Defa-ps1*
CRS1C-3,-6	Tel	22812994–22814120	6.6	AY761184
Zn-MYM	Cen	22818500–22819000	4.4	
Hyb-Defma4-ps	Cen	22832541–22835992	13.5	*Defa-ps10*
Defcr24	Cen	22845007–22845850	9	*Defa24*
Defma18-ps	Cen	22859651–22860443	13.8	*Defa-ps11*
Defb1	Cen	22887069–22905658	26.6	
NT_166309.1				
Defma19-ps	Unk	110050–111004		
CRS4C-6	Unk	7684–8700		

[a] Cen, centromere; Tel, telomere; Unk, unknown.
[b] MGI, Mouse Genome Informatics.

and 27 have not been reported in other mice. Potential C57BL/6-specific Crp cDNA sequences also were retrieved from the RIKEN C57BL/6 mouse small intestinal cDNA database (http://fantom.gsc.riken.jp/4/) by BLASTP searches using Crps 1 to 6 as query sequences (see Table S1 in the supplemental material). Except for Crps 3 and 5 (Fig. 1A, peaks 1 and 7, respectively), the *m/z* values of five C57BL/6 Crps differed from those of previously described mouse α-defensins (Fig. 2A). For example, the mass of the peak 2 peptide is greater than that of any Crp characterized from other strains (54), corresponding to Crp20 (Fig. 2A and B). The unusual mass of Crp20 results from a mutation of the MMP7 cleavage site at the proregion-defensin junction, thus extending the peptide N terminus (58). C57BL/6 mouse Paneth cell α-defensins identified also included Crp21, Crp23, and Crp24 (Fig. 2A and B; see Table S1 in the supplemental material). The last of the purified C57BL/6 α-defensins was inferred to be Crp27, because it matched the product deduced from a C57BL/6 small bowel cDNA sequence (Fig. 2B). The corresponding Crp27 gene has not been identified, and no gene symbol has been assigned. However, the purification of both Crp27 and pro-Crp27 peptides shows that a Crp27 gene exists, perhaps within one of the unsequenced gaps of the α-defensin locus in the NIH C57BL/6 assembly (see Fig. 4A). The known or deduced Crps and the strains in which they are known to occur are summarized in Tables 1 and 2 and in Tables S1 and S2 in the supplemental material, showing that five of seven abundant α-defensins detected in C57BL/6 mouse small bowel have not been described previously in other mouse strains.

Pro-Crps specific to C57BL/6 mouse small intestine. Individual peptides purified from the nominal 8- to 10-kDa P-30 fractions (Fig. 1B) were identified as pro-α-defensins by MALDI-TOF MS analyses after *in vitro* proteolysis with MMP7, the activating convertase for mouse Paneth cell pro-α-defensins (58, 67). For example, MMP7 digestion of peaks 8 and 9 generated major products of 4,945.9 AMU and 4,947.6 AMU, respectively, the mass of Crp20, identifying both peaks as pro-Crp20 (Fig. 2B). By similar MS analyses of major MMP7 cleavage products, peaks 10 to 13 were identified as pro-Crp3, pro-Crp23, pro-Crp27, and pro-Crp24, respectively (Fig. 2A and B). The major product of MMP7 proteolysis for peak 14 had a mass of 4,314.8 AMU, corresponding to Crp5 and showing that peptide 14 is pro-Crp5.

Because the pro-Crp5 primary structure diverges substantially from pro-Crps whose processing by MMP7 has been described (2, 58, 67), we determined the MMP7 processing sites within pro-Crp5 by N-terminal sequencing of MMP7-digested pro-Crp5. Four N-terminal sequences were obtained: (i) DPIHKT, the N terminus of pro-Crp5(20–93), (ii) ISFGGQ, resulting from the Ser-43 ↓ Ile-44 cleavage event, (iii) LHEELS, resulting from Ala-53 ↓ Leu-54 cleavage; and (iv) LSKKLI, resulting from cleavage at Glu-57 ↓ Leu-58, the Crp5 N terminus. These cleavage positions are consistent with the general preference of MMP7 for Leu or Ile at the P′ position of the cleavage site. Thus, MMP7 processes pro-Crp5 at proregion residue positions that correspond to pro-Crp cleavage sites characterized previously, even though the amino acids at the Crp5 peptide N terminus are distinct from those of other known pro-Crps (2, 58).

Bactericidal peptide activities of C57BL/6 Crps. Because the abundant C57BL/6 peptides are variants of known sequences, we compared their activities to Crp4 in *in vitro* bactericidal peptide assays. Although Crp4 is the most potent of the known mouse Paneth cell α-defensins (38, 54), the bactericidal peptide activities of Crps 20, 24, and 27 (Fig. 2A) were equivalent to that of Crp4 at 5 μg/ml or less peptide against the Gram-positive species assayed (Fig. 3C and D). On the other hand, Crps 20, 24, and 27 were less bactericidal against certain Gram-negative bacteria, including the attenuated *S. enterica* serovar Typhimurium Δ*PhoP* strain and *E. coli* ML35 (Fig. 3A and B). Full-length or N-terminally truncated forms of Crps 3, 5, 20, 21, 24, and 27 exist in C57BL/6 colonic lumen (J. R. Mastroianni et al., unpublished observations), showing that the prevalent Crps in C57BL/6 mice are both bactericidal and persist after secretion (31). The data shown in Fig. 3 are representative of three replicate dose-sensitivity assays for each target organism and were highly consistent. The results support the view that C57BL/6 Crps and Crps 3 and 5, which are common to C57BL/6 and varied inbred strains, also mediate innate immunity (47). Also, because the activities of Crps 20, 24, and

TABLE 2. Celera defensin locus from telomere to centromere

Gene identity or provisional symbol	Chromosome 8 position	Orientation[a]	MGI[b] designation
Genomic contigs			
NW_001030882.1			
Defb5	19371528–19374763	Cen	
Defma1-ps	19403200–19404000	Tel	
Defb3	19417528–19419519	Cen	
Defb3or8	19424935–19429044	Cen	
Defma2-ps	19443200–19443700	Tel	
Defma3-ps	19459000–19459800	Cen	
Defma4-ps	19467900–19468300	Tel	
Defma5-ps	19475800–19476600	Tel	
Defma6-ps	19491600–19492400	Tel	
sim Defb3	19506310–19508061	Cen	
Defma7-ps	19526100–19526900	Cen	
sim Np 37	19530367–19531629	Tel	
Defma8-ps	19540700–19541500	Cen	
Defma9-ps	19572668–19573419	Cen	
Defb8	19612444–19617523	Tel	
Defma10-ps	19642100–19642900	Tel	
sim Rp L19	19657583–19658313	Cen	
Defb7	19659779–19662429	Cen	
NW_001030885.1			
Defma11-ps	20693500–20694000	Cen	
Defcr23		Cen	
NW_001030886.1			
Defcr4	20982620–20983555	Tel	*Defa4*
Defcr1	20990361–20991325	Tel	*Defa1*
Defma12-ps	21003600–21004300	Tel	
Defcr26	21017880–21018691	Tel	*Defa26*
CRS1C(−4)	21034312–21035438	Tel	
sim Rp L21	21038461–21039032	Tel	
vDefcr18	21047000–21048500	Tel	
Defma13-ps	21054200–21054700	Tel	
Defcr5	21063322–21064303	Tel	*Defa5*
Defcr7	21076594–21077585	Tel	*Defa7*
sim scf25	21097318–21112127	Cen	
Defcr2	21119631–21120603	Cen	*Defa2*
CRS4C-ps	21148015–21148572	Cen	
vCRS4C-5	21179862–21180687	Tel	
CRS4C-4,-4a	21194825–21195846	Cen	
NW_001030887.1			
CRS1C-2	21511964–21513307	Tel	
NW_001030888.1			
CRS1C-1	21795363–21796498	Tel	
sim Rp L21	21797250–21797534	Tel	
NW_001030889.1			
Defma14-ps	22071113–22072066	Tel	
Ag recA-ps	22090000–22102000	Tel	
Defb48ps	22097825–22110441	Cen	
sim EMO2-ps	22102246–22104904	Tel	
hypo sim CG6004-PB	22105707–22109529	Tel	
Defb33	22116428–22118939	Cen	
sim Defb52	22121922–22123289	Tel	
Defb51	22141263–22141943	Tel	
CRS4C-6	22167834–22168841	Tel	
CRS4C-ps	~22255000	Tel	
hypo sim veg-ps	22276526–22278591	Cen	
CRS4C-1	22281990–22282819	Cen	
NW_001030891.1			
Defcr22	22918740–22919712	Cen	
NW_001030892.1			
Zn MYM	23200400–23201330	Cen	
Hyb-Defma1-ps	23214334–23217785	Cen	
Defcr24	23226803–23227643	Cen	
Defma15-ps	23241453–23242243	Cen	
Defb1	23268870–23284384	Cen	
Nongenomic contigs			
NW_001072513.1			
Defma16-ps	9765–10168		
NW_001034677.1			
Defma17-ps	6328–6550		
NW_001034895.1			
Defma18-ps	7639–7862		
NW_001073383.1			
Defma19-ps	10814–12249		
NW_001037941.1			
Defma20-ps	9127–10036		
NW_001073917.1			
Defma21-ps	53227–54142		
NW_001072962.1			
Defma22-ps	10436–11547		
NW_001031579.1			
Defma23-ps	1993–2928		

[a] Cen, centromere; Tel, telomere.
[b] MGI, Mouse Genome Informatics.

27 against the *S. enterica* serovar Typhimurium Δ*PhoP* strain, which is highly susceptible to defensin-mediated bactericidal effects, and *E. coli* ML35 are reduced relative to Crp4 (Fig. 3A and B), we speculate that they may be more permissive toward Gram-negative species of the microflora. Because most C57BL/6 Crps are distinct from Crps of other mouse strains (Fig. 2), we compared Crp-coding *Defcr* genes and their organization in the NIH (C57BL/6) and Celera (mixed-strain) genomic assemblies.

α-Defensin and CRS4C gene polymorphisms in C57BL/6 mice. The region of mouse chromosome 8 containing α-defensin (*Defcr*) and defensin-related CRS1C and CRS4C (*Defcr-rsN*) genes (1, 19, 40) was searched using full-length Crp, CRS1C, and CRS4C cDNA query sequences (see Materials and Methods and Tables S1 and S2 in the supplemental material).

Two mouse genome reference assemblies were studied. The first was the C57BL/6J reference assembly produced by the NIH consortium, and the second was the "alternative" assembly of Celera Genomics from whole-genome shotgun sequencing of the A/J, DBA/2J, 129×1/SvJ, and 129S1/SvImJ inbred strains and from which C57BL/6 sequences are excluded. The characterized or deduced α-defensin and α-defensin-related (i.e., CRS1C and CRS4C) precursors in both assemblies are shown in Fig. S1A to C in the supplemental material, with documentation of their chromosomal positions shown in Tables 1 and 2 and in the more complete data in Tables S3 and S4 in the supplemental material. The genes coding for C57BL/6-specific α-defensins were located and analyzed for potential polymorphisms and gene copy numbers (see Materials and Methods). Because one mouse α-defensin pseudogene, *vDefcr2-ps*, is currently annotated as "similar to the

A

Peak	C57BL/6 Peptides	Peptide Primary Structure
1	Crp3	LRDLVCYCRKRGCKRRERMNGTCRKGHLMYTLCCR
2	Crp20	LHEKSSRDLICYCRKGGCNRGEQVYGTCSGRLLFCCRRRHRH
3	Crp23	LRDLVCYCRTRGCKRRERMNGTCRKGHLIYTLCCR
4	Crp27	LRDLVCYCRARGCKGRERMNGTCSKGYLLYMLCCR
5	Crp24	LRDLVCYCRARGCKGRERMNGTCSKGHLLYMLCCR
6	Crp21	LSRDLICLCRNRRCNRGELFYGTCAGPFLRCCRRRR
7	Crp5	LSKKLICYCRIRGCKRRERVFGTCRNLFLTFVFCCS

B

Peak	Peptide	Mass (A.M.U.)		Accession #	Gene
		Actual	Predicted		
1	Crp3	4288.2	4274.0	NM_007850	*Defcr3*
2	Crp20	4949.9	4949.5	AK008107, AK008459	*Defcr20*
3	Crp23	4227.9	4228.9	XP_125338	*Defcr23*
4	Crp27	4077.0	4086.7	AV070188	*Defcr27*
5	Crp24	4061.7	4060.7	XM_125364	*Defcr24*
6	Crp21	4329.6	4328.9	AK008266	*Defcr21*
7	Crp5	4314.1	4315.1	NM_007851	*Defcr5a-e*
8	proCrp20	8548.1	8554.3	AK008107	*Defcr20*
9	proCrp21	8531.1	8528.2	AK008459	*Defcr21*
10	proCrp3	8504.3	8481.4	NM_007850	*Defcr3*
11	proCrp23	8456.7	8449.4	XP_125338	*Defcr23*
12	proCrp27	8303.4	N/A	N/A	N/A
13	proCrp24	8269.4	8265.2	XM_125364	*Defcr24*
14	proCrp5	8434.2	8438.4	NM_007851	*Defcr5a-e*

FIG. 2. Distinct α-defensin peptides identified in C57BL/6 mouse small intestine. Following purification by C_{18} RP-HPLC (Fig. 1), peptides were identified as putative α-defensins by cysteine content after iodoacetamide modification (see Materials and Methods). Comparisons of those masses with C57BL/6-derived Crp-encoding cDNA sequences enabled the primary structures of the isolated peptides to be deduced. In panel A, the identities of seven abundant C57BL/6-specific α-defensins (Fig. 1A and C) are identified and shown with their primary structures. In panel B, the experimentally determined molecular masses of the purified peptides (Fig. 1) are listed beside the molecular masses of the deduced products of the Crp-encoding cDNAs that are most comparable to the molecular mass of the purified peptide.

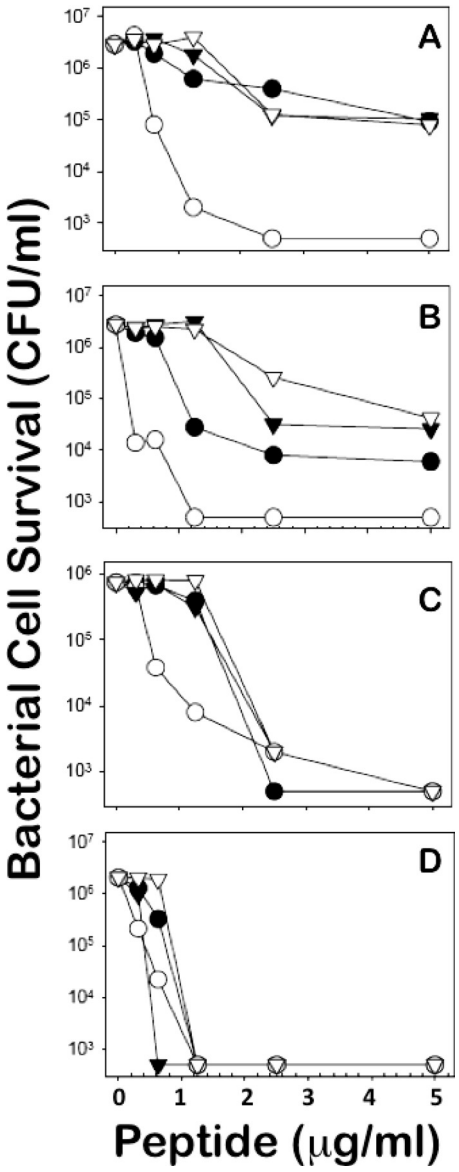

FIG. 3. *In vitro* bactericidal peptide activity of C57BL/6-derived α-defensins. Exponentially growing cells of the *Salmonella enterica* serovar Typhimurium *phoP*Δ strain (A), *Escherichia coli* ML35 (B), *Staphylococcus aureus* (C), and *Listeria monocytogenes* (D) were exposed to the peptide concentrations shown at 37°C in 50 ml 10 mM PIPES buffer supplemented with 1% TSB for 1 h. Symbols: ●, Crp20; ○ Crp4; ▽, Crp27; and ▼, Crp24. Following peptide exposure, the bacteria were plated on Trypticase soy agar and incubated overnight at 37°C, and bacterial cell survival was determined by counting the CFU/ml at each peptide concentration. The panels are representative of three independent experiments in each case. Values at or below 1×10^3 CFU/ml signify that no colonies were detected.

rat NP-1 myeloid α-defensin gene" (*Defa1*; not shown), BLAST searches also were performed using rat α-defensin cDNA sequences.

As recently reported (1), the C57BL/6 mouse α-defensin genes are organized as two sets that span approximately 4.5 Mb at chromosomal position 8pA2 in proximity to β-defensin gene clusters (Fig. 4). Comparisons of α-defensin, α-defensin-like, and CRS genes in the NIH and Celera assemblies disclosed several differences between the C57BL/6 and mixed-strain α-defensin loci. For example, in the NIH C57BL/6 assembly, 16 α-defensin (*Defcr*), 5 CRS1C (*Defcr-rs1*), and 2 CRS4C (*Defcr-rsN*) genes have been located and annotated (Table 1), and in the Celera mixed-strain assembly, 7 *DefcrN*, 1 *Defcr-rs1*, and 2 *Defcr-rs2* to *-12* genes have been annotated (Table 2). The determined or predicted primary structures of α-defensin (*Defcr*) and α-defensin-related (*Defcr-rs*) coding sequences of both assemblies are summarized in Fig. S1A to C in the supplemental material. The C57BL/6J genome includes the *Defcr20a*, *Defcr20b*, and *Defcr21* genes that were not detected in the current Celera assembly build (Fig. 4A and Tables 1 and 2; see Tables S1 and S2 in the supplemental material). Conversely, C57BL/6 mice lack the *Defcr1* (Crp1), *Defcr2* (Crp2), *Defcr4* (Crp4), *Defcr6* (Crp6), and *Defcr7* (Crp7) genes found in the Celera assembly strains as well as in 129/SvJ, C3H/HeJ, FVB, BALB/cJ, and outbred Swiss mice (38).

Crp27 has been characterized in C57BL/6 ileum at the cDNA and peptide levels (Fig. 2; accession no. AV070188), and the peptide has been isolated from C57BL/6 colonic lumen (J. R. Mastroianni et al., unpublished data). The Crp27 coding gene has yet to be located in either assembly but is likely to be in an unsequenced gap in the locus (1). Thus, C57BL/6 mice express a distinct panel of Paneth cell α-defensins (Fig. 2 and Tables 1 and 2; see Fig. S1A to C in the supplemental material), and these comparisons of the genomic assemblies provide the genetic basis for the differences (Fig. 4A and B and Tables 1 and 2).

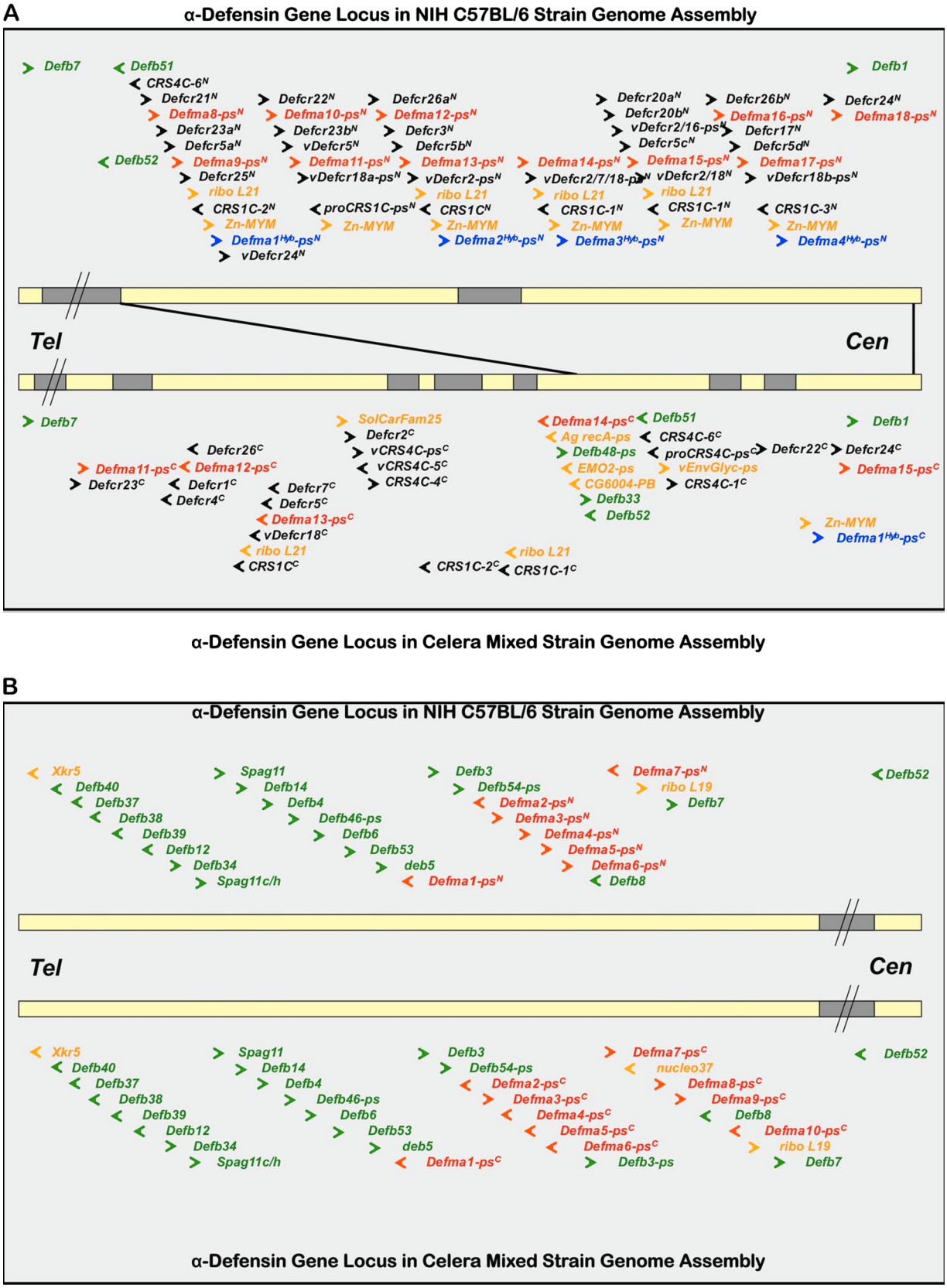

FIG. 4. Mouse α-defensin gene loci in NIH and Celera assemblies. The ~ 4.7-Mb region of mouse chromosome 8 containing all mouse α-defensin, CRS1C, and CRS4C genes is depicted as two horizontal lines corresponding to the C57BL/6 (upper) and Celera Genomics, Inc., mixed-strain (lower line) assemblies. In panel A, that part of the α-defensin gene locus located between the two β-defensin gene clusters of the

A repeated multigene cassette in the C57BL/6 mouse α-defensin gene locus. The mouse α-defensin gene locus exhibits variability in the arrangement of *Defcr* and nondefensin coding genes (1). This is most evident for the genes clustered between *Defb51*/*52* and *Defb1* in the NIH C57BL/6J assembly (Fig. 4A). The approximately 800 kb of DNA, interrupted by an unsequenced gap of approximately 50 kb, provided the longest sequenced chromosomal region for examination of α-defensin gene organization (Fig. 4A and Table 1; see Table S3 in the supplemental material). In this region of the chromosome, a unit consisting of 8 to 10 genes or pseudogenes is repeated approximately six times. For example, beginning with the most telomeric cassette of the cluster, the order of 10 genes identified was as follows: *Defcr21*→*Defma8-ps*→*Defcr23a*→*Defcr5a*→*Defma9-ps*→*Defcr25*→*ribo L21*→*CRS1C-2*→*Zn-MYM*→*Defma1*^{Hyb}*-ps* (Fig. 4A and Tables 1 and 2). (Please note that *DefmaN-ps* genes are vestigial mouse myeloid α-defensin genes as described in the following section.) Although *ribo L21* or *Defma1*^{Hyb}*-ps* genes are not found in every repeat, the relative position of *DefcrN*^{N}, *DefmaN*^{N}*-ps*, *Defcr5*^{N}, *CRS1C*^{N}, and *Zy-MYM*^{N} genes is conserved in the six repeated cassettes (Fig. 4A). In certain *Defcr5* gene copies, the deduced signal peptides and proregions contain substitutions compared to the 129/SvJ *Defcr5* gene (1, 19) (see Fig S1A to C in the supplemental material). However, every *Defcr5* gene repeat codes for the same Crp5 peptide previously reported (54), a primary structure that is distinct from all known α-defensins. The fourth repeated Crp5 gene sequence is incomplete, and we speculate that the 5′ region of that multigene cassette is in the unsequenced gap telomeric of *Defma14*^{N}*-ps*.

The α-defensin loci of the NIH and Celera assemblies differ markedly. For example, the current mixed-strain assembly build contains only a single *Defcr5* gene copy, in contrast to the five in the NIH C57BL/6 genome. Consistent with peptide studies (Fig. 1 and 2), the Celera assembly lacks genes corresponding to C57BL/6-specific α-defensins Crps 20, 21, and 27, but it does contain *Defcr1*, *-2*, *-4*, *-6*, and *-7* and additional CRS4C-coding genes not found in the C57BL/6 genome (Fig. 4A and Table 2; see Fig S1A to C in the supplemental material). Also, nondefensin genes interspersed with *Defcr* genes in the Celera assembly are more numerous and differ from those in the C57BL/6 mouse α-defensin locus (Fig. 4A and Table 2; see Table S4 in the supplemental material). Because gaps in the Celera assembly are more numerous and extensive than in the NIH build, genes and cDNAs cloned from the constituent strains have not been ordered on the locus. In contrast to these disparities, the arrangement of *DefbN* and *DefmaN-ps* genes within the telomeric β-defensin cluster between *Xkr5* and *Defb7* are nearly the same in both assemblies (Fig. 4B and Tables 1 and 2). Thus, the gene cluster more distal from the telomere seems to be less stable. Perhaps, *cis*-acting elements that flank the repeated cassette promote recombination in that region of the C57BL/6 locus to facilitate gene duplications that expanded the α-defensin gene family. Thus, the arrangement and diversity of α-defensin-coding *Defcr* genes in C57BL/6 mice diverge markedly at this location from those of the other inbred strains investigated.

Vestigial mouse myeloid α-defensin genes. Most mammals, including the rat, express both myeloid and enteric α-defensins (9, 11), but mouse neutrophils lack α-defensins (12). To investigate this unusual feature of the mouse, the genome assemblies were queried with cDNAs for rat myeloid α-defensins RatNP-1/2 (*Defa1*/*2*; NM_173329), enteric α-defensin RD-5 (*Defa5*; AF115768), and *Defa1*/*2* exon 1 as a specific marker for myeloid α-defensin genes. The rat α-defensin genes at chromosome 16q12.5 were found within a single contig of ~230 kb, where they are arrayed between two β-defensin gene clusters and flanked by *Defb51* and by *Defb1* (1) (see Fig. S2 and Table S5 in the supplemental material). The expression sites of the 18 rat α-defensin genes and pseudogenes were inferred to be enteric or myeloid on the basis of their respective two- or three-exon gene structures, although because rat *Defa3* and *Defa12-ps* are expressed both in bone marrow and in small bowel (41), gene structure may not be a fully reliable criterion for predicting exclusive sites of expression in rats.

The mouse genome contains vestigial myeloid α-defensin genes that are related to their functionally orthologous genes in the rat (Fig. 4 to 7), but none is predicted to produce functional peptide. Mouse α-defensin pseudogenes (19 in the NIH assembly and 23 in the Celera assembly) have extensive nucleotide sequence similarity to myeloid α-defensin genes from the rat (Fig. 4 and 5 and Tables 1 and 2). These pseudogenes are provisionally termed *DefmaN-ps* for "*def*ensin *m*yeloid *a*lpha-*ps*eudogene," where $N = 1$ to 23, numbering from the telomere toward the centromere. Each *DefmaN-ps* gene identified consists of three exons. The deduced *DefmaN-ps* first exons are highly similar to first exons of the rat myeloid α-defensin genes, as shown by both a TCoffee alignment (Fig. 6) and by individual pairwise BLAST comparisons of all first exons (data not shown). Seven of these three-exon vestigial myeloid C57BL/6 genes and 10 genes in the mixed-strain assembly occur within the conserved telomeric β-defensin gene cluster between *Defb53* and *Defb7* and are separated from the α-defensin, CRS1C, and CRS4C genes that are flanked by *Defb51*/*52* and *Defb1* (Fig. 4A and B and Tables 1 and 2). The

proximal arm of mouse chromosome 8 is delineated. The genomes are oriented with the telomere (Tel) to the left. White line segments represent regions of the chromosome that are currently mapped and annotated. Chromosomal regions that have yet to be described are represented by black line segments. The two depicted genome selections, which include sequences between chromosomal positions 18.9 Mb and 23.6 Mb, are proportionally scaled, with the exception of those regions indicated by a line break (//). Individual genes and their transcriptional orientations are indicated by colored arrows as follows: β-defensin genes, green; mouse Paneth cell α-defensin (*DefcrN*), CRS1C, and CRS4C (*Defcr-rsN*) genes, black; vestigial myeloid α-defensin (*Defma-psN*) pseudogenes related to functional myeloid α-defensins of the rat, red; mouse Paneth cell α-defensin (exon 1) and apparent α-defensin hybrid (*DefmaN-ps*$^{Hyb N/C}$) pseudogenes, blue; and nondefensin genes, orange. The black lines that connect the two assemblies show that the α-defensin locus between the two β-defensin clusters (i.e., bordered by connecting lines in the NIH assembly) corresponds to only a portion of the α-defensin gene locus sequenced in the Celera assembly. In panel B, the chromosome region containing a conserved cluster of rat myeloid α-defensin-like pseudogenes within the telomeric β-defensin gene cluster is depicted.

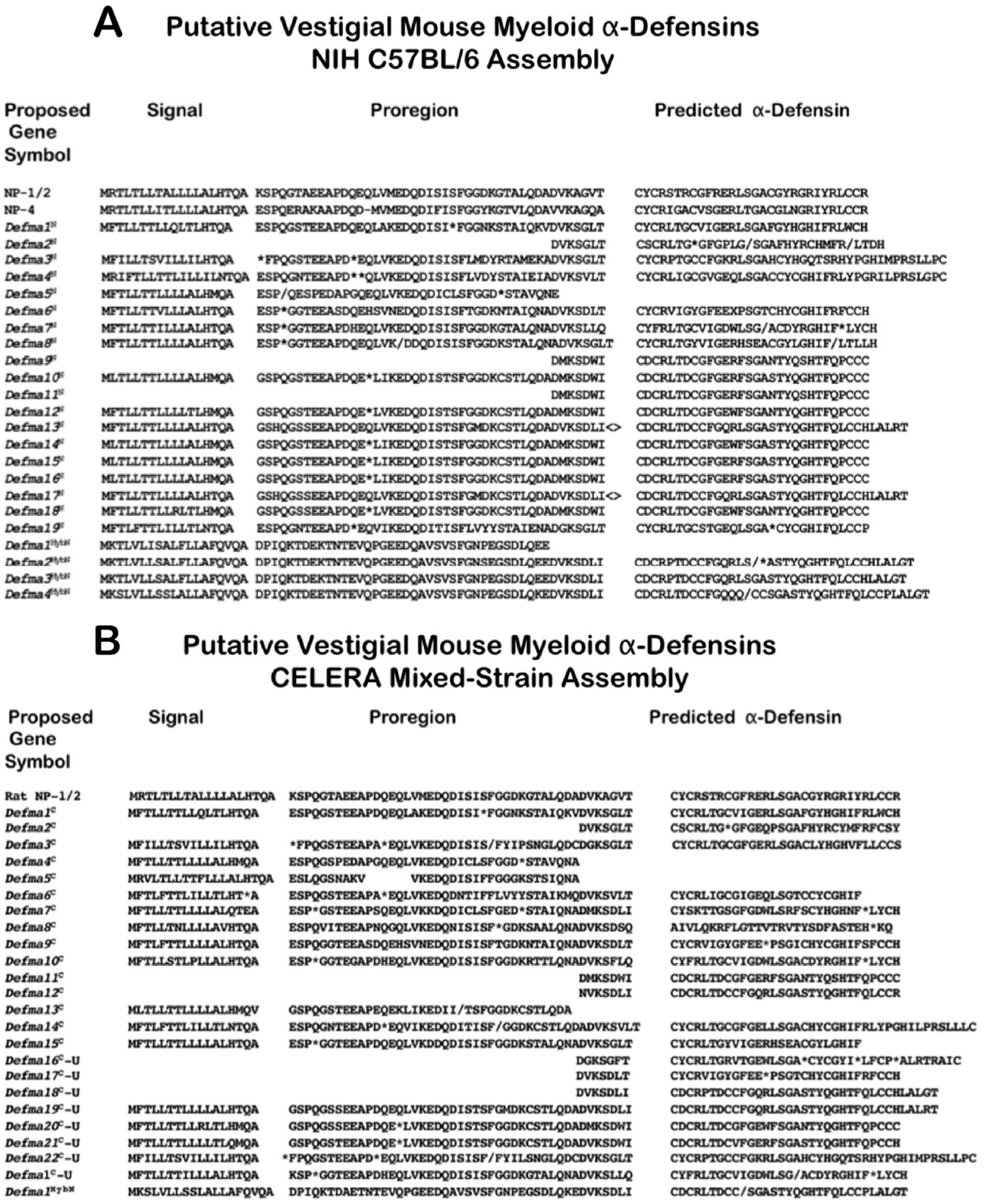

FIG. 5. Primary structures of deduced products of vestigial myeloid α-defensin-like pseudogenes in the mouse genome. The deduced products of mouse vestigial myeloid α-defensin pseudogenes identified in the NIH (A) and Celera (B) mouse genome assemblies are aligned. The corresponding provisional gene symbols (Fig. 4A and B and Tables 1 and 2) used (*DefmaN-ps*^{N} or *-ps*^{C}) are based on gene positions, with the most telomeric gene assigned as *Defma1-ps*$^{N/C}$. *DefmaN-ps*$^{N/C}$ sequences are shown aligned with RatNP-1/2 (i.e., the *Defa1/2* gene product) as a reference. Deduced gene products with extensive similarity to *Defcr*-coded prepro regions were designated *DefmaN-ps*$^{HybN/C}$ to denote that they are chimeras of mouse Crp and other α-defensin pseudogene sequences. The products of *DefmaN-ps* gene second exons and the first exons of *DefmaN-ps*^{Hyb} genes are deduced prepro regions. The deduced products of *DefmaN-ps* gene third exons and the second exons of *DefmaN-ps*^{Hyb} genes are α-defensins. Asterisks denote stop codons, hyphens denote gap positions, angle brackets denote mutated splice junctions, and the slashes denote frameshifted sequences.

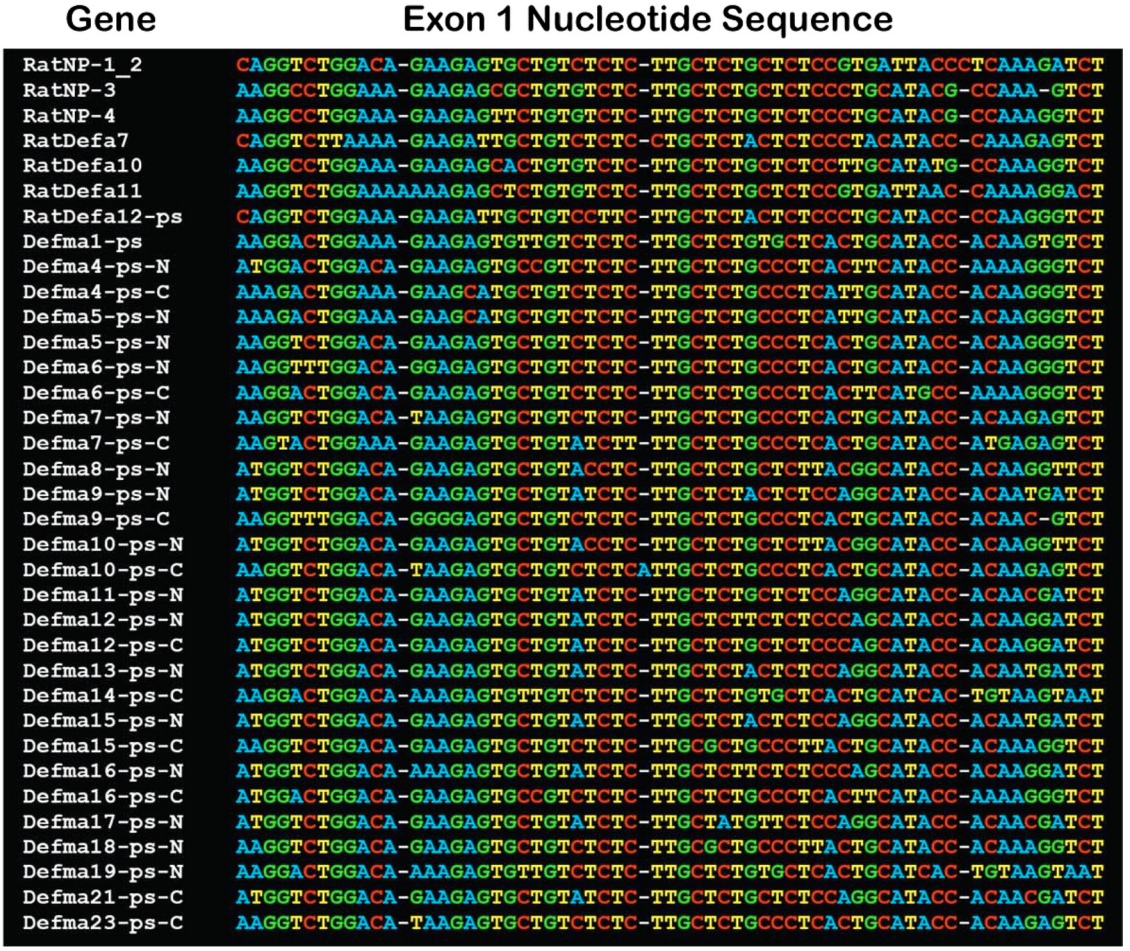

FIG. 6. Alignment of rat myeloid and mouse vestigial myeloid α-defensin gene first exons. The first exons of rat myeloid α-defensin genes (see Fig. S2 and Table S5 in the supplemental material) were aligned with the deduced first exons of *DefmaN-ps* genes (Tables 1 and 2; see Tables S4 and S5 in the supplemental material) by using TCoffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi), and each nucleotide was assigned a specific color (34, 35). Provisional gene symbol designations are as shown in Fig. 5. The *DefmaN-ps* gene first exons show extensive (82 to 90%) nucleotide sequence identity with rat myeloid α-defensin gene first exons when subjected to pairwise BLAST alignment separately (data not shown).

predicted translation products of protein-coding exons 2 and 3 of *DefmaN-ps* genes are deduced prepro-α-defensins that are terminated by premature stop codons or disrupted by frameshift mutations (Fig. 5; see Tables S3 and S4 in the supplemental material). In *Defma13-ps* and *Defma17-ps*, genes with apparent uninterrupted reading frames, the 5′- and 3′-splice sites deviate from canonical sequences. Therefore, we infer that all predicted *DefmaN-ps* gene transcripts are defective (Fig. 5), and many *DefmaN-ps* pseudogenes appear to have become inactivated by independent events.

In addition to *DefmaN-ps* genes, chimeric *DefcrN^{Hyb}-ps* genes that contain both rat and mouse genetic elements were evident (Fig. 4A and 5 and Tables 1 and 2). In these two-exon, deduced enteric pseudogenes (four in the NIH assembly and one in the Celera assembly), exon 1 has extensive nucleotide sequence identity with mouse *Defcr* gene first exons, but the deduced exon 2 products lack primary structural similarity to known rat or mouse α-defensins (Fig. 5).

Phylogeny of mouse α-defensin genes. Evolutionary relationships between *DefcrN* and *Defcr-rsN* (CRS4C and CRS1C coding genes) were investigated by analysis of their single introns along with introns of rat α-defensin genes and pseudogenes (Fig. 7) (see Materials and Methods). To analyze *DefmaN-ps* and rat myeloid α-defensin gene (*DefaN* and *RatNPN*; see Fig. S2 and Table S5 in the supplemental material) relationships, second introns that separate protein-coding exons 2 and 3 (56) were compared (see Materials and Methods). Mouse *Defcr* genes from both assemblies sort into related subgroups that cluster with *Defcr1* and *-3*, *Defcr4* and *-5*, or *Defcr2*, based on intron similarities (41). The *Defcr-rsN* genes, which code for CRS1C and CRS4C peptides, are more closely related to rat enteric *Defa6*, *Defa8*, and *Defa9* genes than to any α-defensin gene in the mouse (Fig. 7). Because the first exons of *Defcr* and *Defcr-rs* genes share 95% nucleotide sequence identity (18, 19), these phylogenetic relationships were unexpected yet consistent with mouse α-defensin and CRS1C/4C-coding genes being evolutionarily distinct (Fig. 7) (41). Thus, even though CRS1C and CRS4C cDNAs and peptides are only found in mice, they are phylogenetically closer to Paneth cell α-defensin genes of the rat. Consistent with their

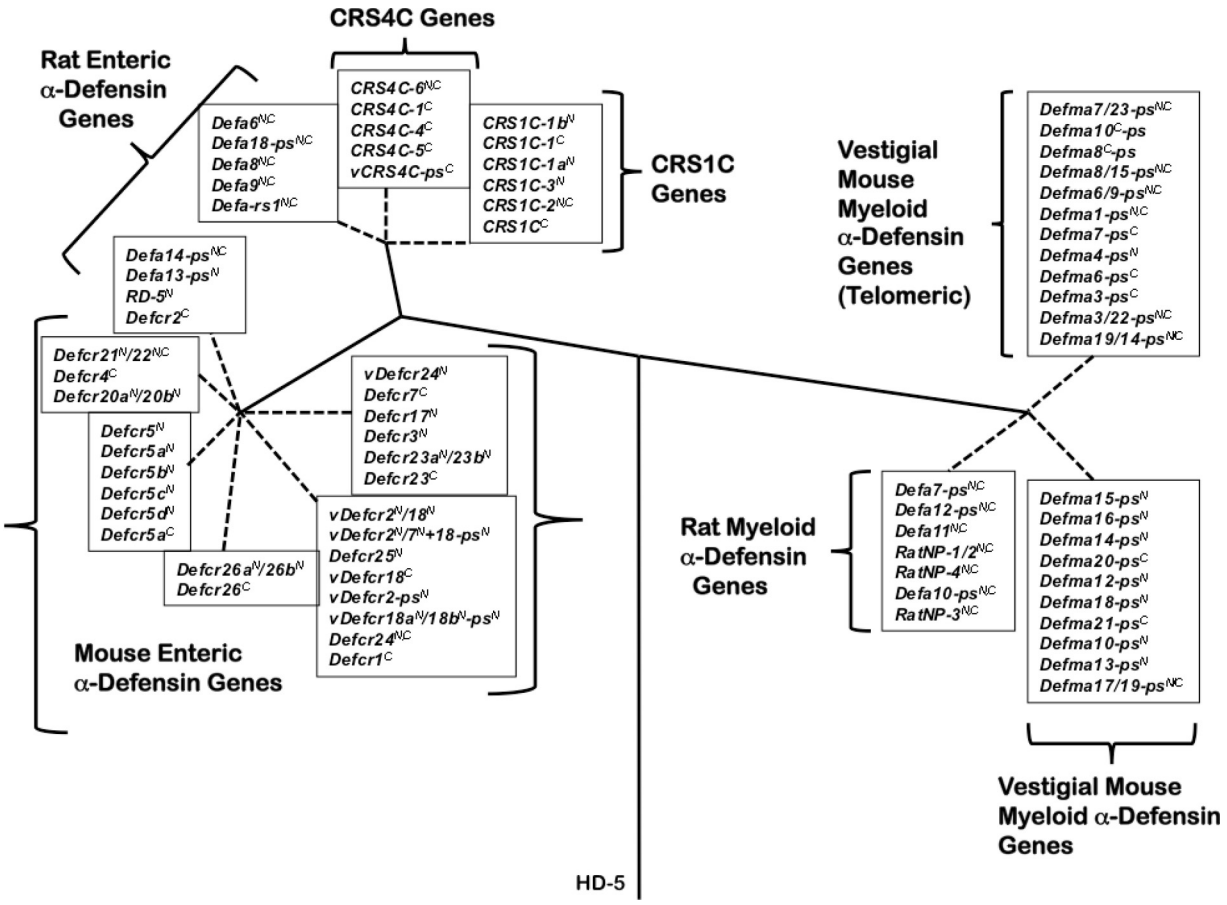

FIG. 7. Phylogenetic relationships between rat and mouse α-defensin, CRS1C, and CRS4C genes. The introns of mouse Paneth cell α-defensin, CRS1C, and CRS4C genes from both assemblies, introns of rat enteric α-defensin genes, second introns of rat myeloid α-defensin genes, and *DefmaN-ps* genes were used to construct the phylogenetic tree (see Materials and Methods). The tree was rooted with the intron of human α-defensin-5 (HD-5), and construction of the tree involved the calculation of the proportion difference (*p*-distance) of aligned nucleotide sites of the entire intron sequences according to the neighbor-joining method. One thousand bootstrap replications were used to test the reliability of each branch. Solid lines maintain phylogenetic distances, but dashed lines do not in order to maintain the legibility of the sequences of the tree.

three-exon structures, vestigial mouse myeloid α-defensin genes (*DefmaN-ps*) are more closely related to rat myeloid α-defensin genes than to Paneth cell α-defensin genes of the mouse or rat (Fig. 7). These data provide further evidence that *DefmaN-ps* pseudogenes represent vestigial myeloid α-defensin-coding genes that may have originated in a common ancestor of mice and rats.

DISCUSSION

The primary structures of the most abundant Paneth cell α-defensins in C57BL/6 mice differ from those of other mouse strains. Evidence for this resulted from characterizing α-defensin and pro-α-defensin proteins isolated from C57BL/6 mouse small intestinal protein extracts and by matching their experimental molecular masses to α-defensins deduced from cDNA and gene sequences (Fig. 1 and 2). The *in vitro* bactericidal peptide activities of C57BL/6-specific Crps and those of other strains are comparable (Fig. 3), suggesting that they also mediate innate immunity in the C57BL/6 mouse small bowel lumen, although their specific effects on particular species of the resident microflora may differ. The genetic basis for C57BL/6 α-defensin proteome differences from other mouse strains was apparent by comparisons of the NIH C57BL/6 and Celera mixed-strain genomic assemblies (Fig. 4 and Tables 1 and 2; see Fig. S1A to C in the supplemental material). The mixed-strain and NIH assemblies were strikingly different, especially with regard to copy numbers for certain genes and a repeated multigene cassette found only in the NIH assembly (Fig. 4 and Tables 1 and 2). The different orientations of several marker genes, including *Defcr5*, *ribo L21*, and several of the *DefmaN-ps* genes, further illustrate the divergence of the locus in these inbred strains. In contrast to the expansion and diversification of the α-defensin proteomes, the organization of β-defensin gene clusters is much more highly conserved and the primary structures of certain β-defensin peptides exhibit similarities across phylogenetic lines (41, 42, 57, 65). The molecular basis of this disparity remains unknown, but perhaps it relates to receptor-mediated roles that have been identified for members of the β-defensin peptide family (4, 7, 15, 49, 62–64). In contrast to the more centromeric α-defensin gene locus (Fig. 4A), the telomeric defensin cluster containing several

DefmaN-ps genes shows extensive conservation between the two assemblies (Fig. 4B). Finally, both mouse genome assemblies contain approximately 20 α-defensin pseudogenes that are related to functional rat myeloid α-defensin genes but are unproductive vestiges of apparent ancestral myeloid α-defensins (Fig. 5 to 7). Given the diversity of inactivating mutations recorded in the *DefmaN-ps* genes, we speculate that inactivation of individual genes may have occurred subsequent to loss of a general factor needed for α-defensin gene transcription in promyelocytes.

The peptide comparisons performed here in mice show that strains of the same species, exemplified by C57BL/6, may express Paneth cell α-defensins of very different primary structures, yet those remarkably diverse primary structures occur in the context of nine residue positions that are highly conserved in all α-defensins (56). Alignments of α-defensins reveal canonical biochemical features, including (i) conserved spacing and disulfide connectivities of the six Cys residues (53, 56), (ii) a Gly at the position corresponding to residue 19 in Crp4 (68), and (iii) a salt bridge formed by Arg-7 and Glu-15 residues in Crp4 (43, 44). Exclusive of these conserved residue positions, however, α-defensins may differ markedly in amino acid sequence, often precluding identification of gene homologues. The structural constraints introduced by the tridisulfide array impose a triple-stranded β-sheet topology on all α-defensin peptides, and the Cys(I)-Cys(VI) disulfide bond brings the N and C termini into proximity. The varied primary structures of Crp20 and Crp21, particularly their extended electropositive C termini and reduction of the distance between the Cys(IV) and Cys(V) residue positions by three amino acids, may localize cationic side chains near one pole of the peptide surface and influence bactericidal peptide activity, a hypothesis that is testable by site-directed mutagenesis approaches.

As shown here, for example, Crps 20, 21, and 27 have not been found in strains other than C57BL/6, nor have their cDNAs or genes been cloned from Celera assembly strains or from other strains of mice. Peptides isolated from numerous mouse strains other than C57BL/6 (e.g., Crps 1, 2, 4, and 6) (54) are not detected in C57BL/6 small bowel, and anti-Crp21 immune sera react strongly with C57BL/6 mouse Paneth cells but not with Paneth cells from the BALBc/J strain (data not shown). Similarly, small intestinal cDNAs and corresponding genes cloned from Celera assembly strains (e.g., Crps 1 to 16) (38) are absent from the NIH build, and most are not annotated in the mixed-strain assembly (Table 2). Because the *Defcr1* to *-6* genes have been cloned from 129/SvJ mice in lambda phage (20), sequencing through the gaps in the Celera assembly will identify the position and orientation of the corresponding genes at the locus and provide a more comprehensive annotation than was possible for this study. The genomic and proteomic differences between α-defensins of C57BL/6 and other reference strains may be important to consider in studies of innate immune responses to enteric infection (14, 24). For example, evidence of Paneth cell α-defensin induction or repression based on microarray data or using immune reagents based on mixed-strain assembly peptides and genes may be difficult to interpret for experiments performed on the C57BL/6 genetic background.

The genetic, epigenetic, translational, or posttranslational determinants that regulate α-defensin gene activity and peptide abundance remain unknown. The rhesus macaque, olive baboon, and horse express many (i.e., 20 or more) diverse enteric α-defensins based on cDNA cloning studies (5, 61), although annotation of the defensin-coding regions of their genomes is incomplete. In contrast, dogs and cattle lack α-defensin genes altogether, human Paneth cells have a limited α-defensin repertoire consisting only of HD5 and HD6, and rats express only five Paneth cell α-defensin genes. Also, the presence of an unusual number of α-defensin pseudogenes in the mouse, rat, and rhesus genomes suggests that loss of functional α-defensin genes may be a frequent occurrence. Perhaps, inactivation of an individual α-defensin gene or addition of a new gene through duplication and subsequent diversification may not influence enteric immunity appreciably in a species with numerous functional genes. Also, despite the 20 or more α-defensin genes found in the mouse genome(s), only 6 or 7 peptides occur at measureable levels (Fig. 1 and 2) (54), and similar findings have been made in rhesus macaques (R. A. Llenado et al., unpublished observations) and horses (5).

The differential sensitivities of particular bacterial species to specific α-defensins support the view that peptide diversity may confer selective advantage. For example, transgenic mice that express human HD5 (*DEFA5*-transgenic [+/+] mice) are immune to oral infection by virulent *Salmonella enterica* serovar Typhimurium, in contrast to the sensitivity of wild-type mice to this virulent mouse pathogen (46). Furthermore, α-defensins secreted by Paneth cells shape the composition of the mouse small intestinal microbiome, apparently by selecting for peptide-tolerant microbial species as residents in that microbial ecosystem (47). In the distal small bowel of individual *DEFA5*-transgenic (+/+) mice, *Firmicutes* represented 25.5% of bacterial species, compared to 59% in FVB controls, and the percentage of *Bacteroidetes* was found to be 69.3% in the $DEFA5^{+/+}$ transgenic mice but only 35% in FVB controls (47). Thus, the small bowel microflora of mice complemented by production of only a single exogenous α-defensin were genotype dependent and significantly different. The possibility that the distinct Paneth cell α-defensin repertoire of the C57BL/6 mouse strain has a selective impact on the composition of the microflora is suggested by the fact that C57BL/6 and FVB mice differed significantly in the percent distribution of *Firmicutes* species and in their relative content of *Tenericutes* (47). This published comparison supports the idea that the C57BL/6 small bowel luminal environment differs from that of the FVB strain, and we speculate that it may result from differences in the Paneth cell α-defensins secreted by these two strains of mice.

ACKNOWLEDGMENTS

This work was supported by NIH grants AI059346 and DK044632.

We thank Clara Amid, Linda M. Rehaume, Kelly L. Brown, James G. R. Gilbert, Gordon Dougan, Robert E. W. Hancock, and Jennifer L. Harrow for sharing their extensive manual annotation and analysis of the defensin gene cluster in the C57BL/6J mouse reference genome prior to publication (1).

REFERENCES

1. **Amid, C., L. M. Rehaume, K. L. Brown, J. G. Gilbert, G. Dougan, R. E. Hancock, and J. L. Harrow.** 2009. Manual annotation and analysis of the defensin gene cluster in the C57BL/6J mouse reference genome. BMC Genomics **10:**606.
2. **Ayabe, T., D. P. Satchell, P. Pesendorfer, H. Tanabe, C. L. Wilson, S. J.**

Hagen, and A. J. Ouellette. 2002. Activation of Paneth cell alpha-defensins in mouse small intestine. J. Biol. Chem. **277:**5219–5228.
3. **Belov, K., C. E. Sanderson, J. E. Deakin, E. S. Wong, D. Assange, K. A. McColl, A. Gout, B. de Bono, A. D. Barrow, T. P. Speed, J. Trowsdale, and A. T. Papenfuss.** 2007. Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. Genome Res. **17:**982–991.
4. **Biragyn, A., M. Coscia, K. Nagashima, M. Sanford, H. A. Young, and P. Olkhanud.** 2008. Murine beta-defensin 2 promotes TLR-4/MyD88-mediated and NF-kappaB-dependent atypical death of APCs via activation of TNFR2. J. Leukoc. Biol. **83:**998–1008.
5. **Bruhn, O., S. Paul, J. Tetens, and G. Thaller.** 2009. The repertoire of equine intestinal alpha-defensins. BMC Genomics **10:**631.
6. **Bruhn, O., P. Regenhard, M. Michalek, S. Paul, C. Gelhaus, S. Jung, G. Thaller, R. Podschun, M. Leippe, J. Grotzinger, and E. Kalm.** 2007. A novel horse alpha-defensin: gene transcription, recombinant expression and characterization of the structure and function. Biochem. J. **407:**267–276.
7. **Candille, S. I., C. B. Kaelin, B. M. Cattanach, B. Yu, D. A. Thompson, M. A. Nix, J. A. Kerns, S. M. Schmutz, G. L. Millhauser, and G. S. Barsh.** 2007. A beta-defensin mutation causes black coat color in domestic dogs. Science **318:**1418–1423.
8. **Cano-Gauci, D. F., J. C. Lualdi, A. J. Ouellette, G. Brady, N. N. Iscove, and R. N. Buick.** 1993. In vitro cDNA amplification from individual intestinal crypts: a novel approach to the study of differential gene expression along the crypt-villus axis. Exp. Cell Res. **208:**344–349.
9. **Condon, M. R., A. Viera, M. D'Alessio, and G. Diamond.** 1999. Induction of a rat enteric defensin gene by hemorrhagic shock. Infect. Immun. **67:**4787–4793.
10. **Eisenhauer, P. B., S. S. Harwig, and R. I. Lehrer.** 1992. Cryptdins: antimicrobial defensins of the murine small intestine. Infect. Immun. **60:**3556–3565.
11. **Eisenhauer, P. B., S. S. Harwig, D. Szklarek, T. Ganz, M. E. Selsted, and R. I. Lehrer.** 1989. Purification and antimicrobial properties of three defensins from rat neutrophils. Infect. Immun. **57:**2021–2027.
12. **Eisenhauer, P. B., and R. I. Lehrer.** 1992. Mouse neutrophils lack defensins. Infect. Immun. **60:**3446–3447.
13. **Fjell, C. D., H. Jenssen, P. Fries, P. Aich, P. Griebel, K. Hilpert, R. E. Hancock, and A. Cherkasov.** 2008. Identification of novel host defense peptides and the absence of alpha-defensins in the bovine genome. Proteins **73:**420–430.
14. **Foureau, D. M., D. W. Mielcarz, L. C. Menard, J. Schulthess, C. Werts, V. Vasseur, B. Ryffel, L. H. Kasper, and D. Buzoni-Gatel.** 2010. TLR9-dependent induction of intestinal alpha-defensins by Toxoplasma gondii. J. Immunol. **184:**7022–7029.
15. **Funderburg, N., M. M. Lederman, Z. Feng, M. G. Drage, J. Jadlowsky, C. V. Harding, A. Weinberg, and S. F. Sieg.** 2007. Human beta-defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. Proc. Natl. Acad. Sci. U. S. A. **104:**18631–18635.
16. **Ganz, T.** 2003. Defensins: antimicrobial peptides of innate immunity. Nat. Rev. Immunol. **3:**710–720.
17. **Ganz, T., J. R. Rayner, E. V. Valore, A. Tumolo, K. Talmadge, and F. Fuller.** 1989. The structure of the rabbit macrophage defensin genes and their organ-specific expression. J. Immunol. **143:**1358–1365.
18. **Hornef, M. W., K. Putsep, J. Karlsson, E. Refai, and M. Andersson.** 2004. Increased diversity of intestinal antimicrobial peptides by covalent dimer formation. Nat. Immunol. **5:**836–843.
19. **Huttner, K. M., and A. J. Ouellette.** 1994. A family of defensin-like genes codes for diverse cysteine-rich peptides in mouse Paneth cells. Genomics **24:**99–109.
20. **Huttner, K. M., M. E. Selsted, and A. J. Ouellette.** 1994. Structure and diversity of the murine cryptdin gene family. Genomics **19:**448–453.
21. **Jones, D. E., and C. L. Bevins.** 1992. Paneth cells of the human small intestine express an antimicrobial peptide gene. J. Biol. Chem. **267:**23216–23225.
22. **Kaiser, V., and G. Diamond.** 2000. Expression of mammalian defensin genes. J. Leukoc. Biol. **68:**779–784.
23. **Karlsson, J., K. Putsep, H. Chu, R. J. Kays, C. L. Bevins, and M. Andersson.** 2008. Regional variations in Paneth cell antimicrobial peptide expression along the mouse intestinal tract. BMC Immunol. **9:**37.
24. **Kobayashi, K. S., M. Chamaillard, Y. Ogura, O. Henegariu, N. Inohara, G. Nunez, and R. A. Flavell.** 2005. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science **307:**731–734.
25. **Kumar, S., M. Nei, J. Dudley, and K. Tamura.** 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. **9:**299–306.
26. **Lehrer, R. I.** 2007. Multispecific myeloid defensins. Curr. Opin. Hematol. **14:**16–21.
27. **Lehrer, R. I., M. E. Selsted, D. Szklarek, and J. Fleischmann.** 1983. Antibacterial activity of microbicidal cationic proteins 1 and 2, natural peptide antibiotics of rabbit lung macrophages. Infect. Immun. **42:**10–14.
28. **Linzmeier, R., D. Michaelson, L. Liu, and T. Ganz.** 1993. The structure of neutrophil defensin genes. FEBS Lett. **321:**267–273.
29. **Llenado, R. A., C. S. Weeks, M. J. Cocco, and A. J. Ouellette.** 2009. Electropositive charge in alpha-defensin bactericidal activity: functional effects of Lys-for-Arg substitutions vary with the peptide primary structure. Infect. Immun. **77:**5035–5043.
30. **Lynn, D. J., and D. G. Bradley.** 2007. Discovery of alpha-defensins in basal mammals. Dev. Comp. Immunol. **31:**963–967.
31. **Mastroianni, J. R., and A. J. Ouellette.** 2009. Alpha-defensins in enteric innate immunity: functional Paneth cell alpha-defensins in mouse colonic lumen. J. Biol. Chem. **284:**27848–27856.
32. **Menard, S., V. Forster, M. Lotz, D. Gutle, C. U. Duerr, R. L. Gallo, B. Henriques-Normark, K. Putsep, M. Andersson, E. O. Glocker, and M. W. Hornef.** 2008. Developmental switch of intestinal antimicrobial peptide expression. J. Exp. Med. **205:**183–193.
33. **Meyer-Hoffert, U., M. W. Hornef, B. Henriques-Normark, L. G. Axelsson, T. Midtvedt, K. Putsep, and M. Andersson.** 2008. Secreted enteric antimicrobial activity localises to the mucus surface layer. Gut **57:**764–771.
34. **Notredame, C.** 2010. Computing multiple sequence/structure alignments with the T-Coffee package. Curr. Protoc. Bioinformatics **29:**3.8.1–3.8.25.
35. **Notredame, C., D. G. Higgins, and J. Heringa.** 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302:**205–217.
36. **Ouellette, A. J.** 1999. Mucosal immunity and inflamation. IV. Paneth cell antimicrobial peptides and the biology of the mucosal barrier. Am. J. Physiol. **277:**G257–G261.
37. **Ouellette, A. J., and B. Cordell.** 1988. Accumulation of abundant messenger ribonucleic acids during postnatal development of mouse small intestine. Gastroenterology **94:**114–121.
38. **Ouellette, A. J., M. M. Hsieh, M. T. Nosek, D. F. Cano-Gauci, K. M. Huttner, R. N. Buick, and M. E. Selsted.** 1994. Mouse Paneth cell defensins: primary structures and antibacterial activities of numerous cryptdin isoforms. Infect. Immun. **62:**5040–5047.
39. **Ouellette, A. J., and J. C. Lualdi.** 1990. A novel mouse gene family coding for cationic, cysteine-rich peptides. Regulation in small intestine and cells of myeloid origin. J. Biol. Chem. **265:**9831–9837.
40. **Ouellette, A. J., D. Pravtcheva, F. H. Ruddle, and M. James.** 1989. Localization of the cryptdin locus on mouse chromosome 8. Genomics **5:**233–239.
41. **Patil, A., A. L. Hughes, and G. Zhang.** 2004. Rapid evolution and diversification of mammalian α-defensins as revealed by comparative analysis of rodent and primate genes. Physiol. Genomics **20:**1–11.
42. **Patil, A. A., Y. Cai, Y. Sang, F. Blecha, and G. Zhang.** 2005. Cross-species analysis of the mammalian beta-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. Physiol. Genomics **23:**5–17.
43. **Rajabi, M., E. de Leeuw, M. Pazgier, J. Li, J. Lubkowski, and W. Lu.** 2008. The conserved salt bridge in human alpha-defensin 5 is required for its precursor processing and proteolytic stability. J. Biol. Chem. **283:**21509–21518.
44. **Rosengren, K. J., N. L. Daly, L. M. Fornander, L. M. Jonsson, Y. Shirafuji, X. Qu, H. J. Vogel, A. J. Ouellette, and D. J. Craik.** 2006. Structural and functional characterization of the conserved salt bridge in mammalian Paneth cell alpha-defensins: solution structures of mouse cryptdin-4 and (E15D)-cryptdin-4. J. Biol. Chem. **281:**28068–28078.
45. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
46. **Salzman, N. H., D. Ghosh, K. M. Huttner, Y. Paterson, and C. L. Bevins.** 2003. Protection against enteric salmonellosis in transgenic mice expressing a human intestinal defensin. Nature **422:**522–526.
47. **Salzman, N. H., K. Hung, D. Haribhai, H. Chu, J. Karlsson-Sjoberg, E. Amir, P. Teggatz, M. Barman, M. Hayward, D. Eastwood, M. Stoel, Y. Zhou, E. Sodergren, G. M. Weinstock, C. L. Bevins, C. B. Williams, and N. A. Bos.** 2010. Enteric defensins are essential regulators of intestinal microbial ecology. Nat. Immunol. **11:**76–83.
48. **Satchell, D. P., T. Sheynis, Y. Shirafuji, S. Kolusheva, A. J. Ouellette, and R. Jelinek.** 2003. Interactions of mouse Paneth cell alpha-defensins and alpha-defensin precursors with membranes: prosegment inhibition of peptide association with biomimetic membranes. J. Biol. Chem. **278:**13838–13846.
49. **Schmutz, S. M., and T. G. Berryere.** 2007. Genes affecting coat colour and pattern in domestic dogs: a review. Anim. Genet. **38:**539–549.
50. **Selsted, M. E.** 2007. A pocket guide to explorations of the defensin field. Curr. Pharm. Des. **13:**3061–3064.
51. **Selsted, M. E.** 1993. Investigational approaches for studying the structures and biological functions of myeloid antimicrobial peptides. Genet. Eng. (New York) **15:**131–147.
52. **Selsted, M. E., D. M. Brown, R. J. DeLange, S. S. Harwig, and R. I. Lehrer.** 1985. Primary structures of six antimicrobial peptides of rabbit peritoneal neutrophils. J. Biol. Chem. **260:**4579–4584.
53. **Selsted, M. E., and S. S. Harwig.** 1989. Determination of the disulfide array in the human defensin HNP-2. A covalently cyclized peptide. J. Biol. Chem. **264:**4003–4007.
54. **Selsted, M. E., S. I. Miller, A. H. Henschen, and A. J. Ouellette.** 1992. Enteric defensins: antibiotic peptide components of intestinal host defense. J. Cell Biol. **118:**929–936.

55. **Selsted, M. E., and A. J. Ouellette.** 1995. Defensins in granules of phagocytic and non-phagocytic cells. Trends Cell Biol. **5:**114–119.
56. **Selsted, M. E., and A. J. Ouellette.** 2005. Mammalian defensins in the antimicrobial immune response. Nat. Immunol. **6:**551–557.
57. **Semple, C. A., K. Taylor, H. Eastwood, P. E. Barran, and J. R. Dorin.** 2006. Beta-defensin evolution: selection complexity and clues for residues of functional importance. Biochem. Soc. Trans. **34:**257–262.
58. **Shirafuji, Y., H. Tanabe, D. P. Satchell, A. Henschen-Edman, C. L. Wilson, and A. J. Ouellette.** 2003. Structural determinants of procryptdin recognition and cleavage by matrix metalloproteinase-7. J. Biol. Chem. **278:**7910–7919.
59. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
60. **Tanabe, H., X. Qu, C. S. Weeks, J. E. Cummings, S. Kolusheva, K. B. Walsh, R. Jelinek, T. K. Vanderlick, M. E. Selsted, and A. J. Ouellette.** 2004. Structure-activity determinants in Paneth cell alpha-defensins: loss-of-function in mouse cryptdin-4 by charge-reversal at arginine residue positions. J. Biol. Chem. **279:**11976–11983.
61. **Tanabe, H., J. Yuan, M. M. Zaragoza, S. Dandekar, A. Henschen-Edman, M. E. Selsted, and A. J. Ouellette.** 2004. Paneth cell alpha-defensins from rhesus macaque small intestine. Infect. Immun. **72:**1470–1478.
62. **Taylor, K., D. J. Clarke, B. McCullough, W. Chin, E. Seo, D. Yang, J. Oppenheim, D. Uhrin, J. R. Govan, D. J. Campopiano, D. MacMillan, P. Barran, and J. R. Dorin.** 2008. Analysis and separation of residues important for the chemoattractant and antimicrobial activities of beta-defensin 3. J. Biol. Chem. **283:**6631–6639.
63. **Tollner, T. L., A. I. Yudin, A. F. Tarantal, C. A. Treece, J. W. Overstreet, and G. N. Cherr.** 2008. Beta-defensin 126 on the surface of macaque sperm mediates attachment of sperm to oviductal epithelia. Biol. Reprod. **78:**400–412.
64. **Tollner, T. L., A. I. Yudin, C. A. Treece, J. W. Overstreet, and G. N. Cherr.** 2008. Macaque sperm coating protein DEFB126 facilitates sperm penetration of cervical mucus. Hum. Reprod. **23:**2523–2534.
65. **Whittington, C. M., A. T. Papenfuss, P. Bansal, A. M. Torres, E. S. Wong, J. E. Deakin, T. Graves, A. Alsop, K. Schatzkamer, C. Kremitzki, C. P. Ponting, P. Temple-Smith, W. C. Warren, P. W. Kuchel, and K. Belov.** 2008. Defensins and the convergent evolution of platypus and reptile venom genes. Genome Res. **18:**986–994.
66. **Whittington, C. M., A. T. Papenfuss, P. W. Kuchel, and K. Belov.** 2008. Expression patterns of platypus defensin and related venom genes across a range of tissue types reveal the possibility of broader functions for OvDLPs than previously suspected. Toxicon **52:**559–565.
67. **Wilson, C. L., A. J. Ouellette, D. P. Satchell, T. Ayabe, Y. S. Lopez-Boado, J. L. Stratman, S. J. Hultgren, L. M. Matrisian, and W. C. Parks.** 1999. Regulation of intestinal alpha-defensin activation by the metalloproteinase matrilysin in innate host defense. Science **286:**113–117.
68. **Xie, C., A. Prahl, B. Ericksen, Z. Wu, P. Zeng, X. Li, W. Y. Lu, J. Lubkowski, and W. Lu.** 2005. Reconstruction of the conserved beta-bulge in mammalian defensins using D-amino acids. J. Biol. Chem. **280:**32921–32929.

Editor: A. J. Bäumler