

Discovery of five conserved β -defensin gene clusters using a computational search strategy

Brian C. Schutte*[†], Joseph P. Mitros*, Jennifer A. Bartlett[†], Jesse D. Walters[‡], Hong Peng Jia*, Michael J. Welsh[§], Thomas L. Casavant^{†,‡}, and Paul B. McCray, Jr.*^{†¶}

*Department of Pediatrics, [†]Genetics Ph.D. Program, [‡]Electrical and Computer Engineering, [§]Departments of Internal Medicine and Physiology and Biophysics and the Howard Hughes Medical Institute, University of Iowa, Iowa City, IA 52242

Contributed by Michael J. Welsh, December 21, 2001

The innate immune system includes antimicrobial peptides that protect multicellular organisms from a diverse spectrum of microorganisms. β -Defensins comprise one important family of mammalian antimicrobial peptides. The annotation of the human genome fails to reveal the expected diversity, and a recent query of the draft sequence with the BLAST search engine found only one new β -defensin gene (*DEFB3*). To define better the β -defensin gene family, we adopted a genomics approach that uses HMMER, a computational search tool based on hidden Markov models, in combination with BLAST. This strategy identified 28 new human and 43 new mouse β -defensin genes in five syntenic chromosomal regions. Within each syntenic cluster, the gene sequences and organization were similar, suggesting each cluster pair arose from a common ancestor and was retained because of conserved functions. Preliminary analysis indicates that at least 26 of the predicted genes are transcribed. These results demonstrate the value of a genomewide search strategy to identify genes with conserved structural motifs. Discovery of these genes represents a new starting point for exploring the role of β -defensins in innate immunity.

The β -defensins are defined by a six-cysteine motif [usual spacing, C-X₆-C-X₄-C-X₉-C-X₆-C-C (1, 2)] and a large number of basic amino acid residues. Their coding sequences consist of two exons. The first exon includes the 5' untranslated region and encodes the leader domain of the preproprotein; the second exon encodes the mature peptide with the six-cysteine domain. As this project began, five human β -defensin genes (3–9) and six mouse β -defensin genes (10–14) were reported or submitted to GenBank (accession no. AF318068). However, we suspected the existence of additional β -defensin genes because of the high frequency of gene duplication within β -defensin clusters (15). Thus, we hypothesized that analyzing the genomic sequence (16–18) with multiple computational tools (19, 20) would identify novel defensin genes.

Materials and Methods

BLAST-Based Searches. Genomic search strategies for human gene discovery were applied to the GenBank nonredundant, high throughput genomic sequences, and expressed sequence tag (EST) databases by using the BLASTP and TBLASTN programs (19), using National Center for Biotechnology Information (NCBI) web site tools (www.ncbi.nlm.nih.gov/blast/). Similar approaches were used to query the Celera mouse genome assembly (www.celera.com). The initial queries for the search used the amino acid sequences for the known human defensins (*DEFB1*, *DEFB2*, *DEFB3*, and *DEFB4*) (3–7) and two *HE2/EP2* sequences (8, 9) and the known mouse β -defensins (*Defb1*, *Defb2*, *Defb3*, *Defb4*, *Defb5*, and *Defb6*) (10–14) and GenBank (accession no. AF318068). NCBI and Celera default parameters were used in the searches, and any potential hits were curated manually.

For each novel β -defensin gene identified by using the HMMSEARCH program (described below), additional iterative BLAST searches were performed against the human and mouse

databases to identify additional related sequences and search for ESTs to confirm that the sequences were transcribed.

Construction of Hidden Markov Models (HMMs) for the Six-Cysteine β -Defensin Motif. The complementary strategy used to identify β -defensin genes used a quantitative sequence analysis with the HMM (20). For this purpose, we defined core human and mouse β -defensin amino acid sequences containing the six-cysteine motif and sorted them according to their scores in HMMs trained on defensin motifs. Initially, 12 second-exon, 6-cysteine motifs derived from human and mouse β -defensin sequences were defined by manual inspection of full-length β -defensin domain sequences. These motifs were aligned by using the CLUSTALW program (21) and trimmed of extra amino acids extending on both sides of a 38-aa core. We used these 12 aligned sequences as input for the HMMER 2.1.1 suite of software at Washington University (St. Louis) (<http://hmm.wustl.edu>; ref. 20) to build the first of our HMM β -defensin models. We constructed the first model by using the program HMMBUILD and then used HMMCALIBRATE to calibrate E-value scores. Ultimately, three models were built: all 33 human second-exon sequences, all 36 mouse second exons, and a combined model of human and mouse second exons (69 sequences). These models were used to search all of the contigs of the Univ. of California (Santa Cruz) Golden Path assembly (<http://genome.cse.ucsc.edu/goldenPath/01apr2001>) and the July 30, 2001, NCBI assembly (<ftp://ncbi.nlm.nih.gov/genomes/H.sapiens>) by using the HMMSEARCH program. The assemblies, once downloaded to a local file server, were translated into all six reading frames for searching, as HMMER does not currently have the native capability to search nucleotide databases. We followed the same procedure for all contigs from the mouse genomic sequence that contained β -defensin genes as revealed by BLAST analysis. A Linux cluster consisting of 32 dual-Pentium-III (550 MHz) processors, each with 2 giga-bytes of RAM, was used to perform the searching in parallel, taking roughly 90 min. We used the default cutoff values for filtering the results of these queries and performed postprocessing of the results (using custom Perl programs) to remove motifs that did not contain the six-cysteine target motif. Before this filtering, hundreds of false domains were reported. Also, we observed multiple HMMER hits of the same gene, indicating that the present state of the assembled draft sequence contains multiple copies of the same regions.

Assembly of Human and Mouse β -Defensin Genomic Clusters. To generate a continuous DNA sequence for some analyses, the sequences from the human and mouse defensin containing

Abbreviations: HMM, hidden Markov model; EST, expressed sequence tag.

[¶]To whom reprint requests should be addressed. E-mail: paul-mccray@uiowa.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

BAC clones and genomic contigs were aligned by using the SEQUENCHER program (Gene Codes, Ann Arbor, MI).

Analysis of Predicted β -Defensin Peptide Sequences: Alignment and Phylogeny. The multiple sequence alignment and dendrogram construction were performed by using the program PILEUP from the Wisconsin Package software (Accelrys, San Diego). The amino acid sequences were predicted from the known, related, and predicted β -defensin genes in human and/or mouse and included two residues before and after the six-cysteine domain. The comparison matrix was set at Blosum62 with a gap creation penalty of 8 and a gap extension penalty of 2.

Results

To discover new β -defensin genes, we constructed an HMM (20) with the mature peptide sequences of known β -defensin genes. Then we used this model [HMMER software (20)] to screen ~4 Mb of genomic DNA sequence around the β -defensin locus on human chromosome 8p23-p22. This search found 11 genes, including the five known β -defensin genes (*DEFB1-4* and *HE2/EP2*) plus six novel genes (*DEFB5-9* and *DEFBp1*) (Fig. 1). When these novel sequences were used in a BLAST search of the draft human genomic sequence, we discovered another β -defensin gene, *DEFB10*. Surprisingly, *DEFB10* maps to chromosome 6p12, whereas all previously identified human defensin genes localized to chromosome 8p23-p22 (22, 23). We reseeded the HMM with the predicted peptide sequences from the new genes and used it to analyze the genomic DNA sequence derived from the BAC clone that contains *DEFB10* at the 6p12 locus (GenBank accession no. NT.007402). This analysis revealed four more β -defensin genes (*DEFB11-14*) (Fig. 1). Thus, the *DEFB10-14* genes represent a second β -defensin gene cluster. Subsequent iterations of the BLAST/HMMER process identified 15 additional β -defensins (*DEFB15-29*) (Fig. 1). These genes are located on two sequence contigs, one on chromosome 20q11.1 and one on chromosome 20p13, and represent two more β -defensin gene clusters.

Finally, to analyze the entire human genome with HMMER, we used all 31 human β -defensin genes for the HMM. This analysis revealed 40 more sequences that encode a 6-cysteine domain. However, the position of the cysteines and the cationic charge density were consistent with β -defensin genes in only two cases, *DEFB30* and *DEFB31* (Table 1, which is published as supporting information on the PNAS web site, www.pnas.org). In addition, these two genes are located on the same BAC clone that ambiguously maps to multiple regions of the genome, including chromosome 8p23-p22 (Fig. 1). Because the sequence of this clone is not contiguous with the 8p23-p22 contigs containing the known β -defensin locus, it may represent a fifth cluster in the human genome. Significantly, this HMMER search detected only 13 of the 31 previously identified β -defensin genes. These results suggest that, like BLAST searches, genomewide searches with HMMER alone are not sufficient for identifying all β -defensin genes. Further BLAST and HMMER analyses did not detect additional sequences in the human genome. In total, this iterative, computational method discovered 28 novel human β -defensin genes (Fig. 1 and Table 1).

To search for novel β -defensin genes in the mouse genome, we used this hybrid process to screen assembled mouse sequence contigs in the Celera database (18). This analysis discovered 43 new genes (Table 1) in five clusters, on chromosomes 1, 8, and 14, and two on 2. Significantly, these regions are syntenic to the regions in the human genome that contain the β -defensin clusters (www.ncbi.nlm.nih.gov/Homology), indicating that the syntenic clusters have a common origin. Supporting this hypothesis, we found the highest sequence similarity between gene products from syntenic clusters (Fig. 2 and Table 1). Moreover, the order and orientation for many genes in the syntenic clusters

Gene	Chr location	Gene type	Amino acid sequence of six-cysteine domain					
			1	2	3	4	56	
<i>DEFB1</i>	8p23-p22	K	YNCVS--SGGQCLYSACPIFTKIQGTTCYR-GKAKCCK					
<i>DEFB2</i>	8p23-p22	K	VTCLK--SGAICHNFVCPRRYQIGTGL-PGTKCCNK					
<i>DEFBp1</i>	8p23-p22	S	*CVLNVCSSTLSKIQGTGTH-DNIRCCCK					
<i>DEFB3</i>	8p23-p22	K	YYCRV--FGGRCVAVSLCLP*EIQIG*G*ST-RGRCCRRKK					
<i>DEFB4</i>	8p23-p22	K	RICGY--GTARCKK-CRSQERYRGRCPN--TYACLRK					
<i>DEFB5</i>	8p23-p22	P	ESCKL--GRGCKK-E-CLENEKPDGNCRL--NFLCCRQAI					
<i>DEFB6</i>	8p23-p22	R	EKCNK--LKGTCNN-CGNEELIALCQK--SLFCRTTIQPC+5					
<i>DEFB7</i>	8p23-p22	P	TNC (+21) CEAE-CLTFEVRIGGCFAELAPFCNNRKKH					
<i>DEFB8</i>	8p23-p22	P	EICER--PNSGCHDF-CLETEEIVVGRCLN--SRPCCLPLGQ+11					
<i>DEFB9</i>	8p23-p22	P	GHCLN--LSGVCNDVCKVVDQIGACRR--RMCAPAWWIL+21					
<i>EP2C</i>	8p23-p22	R	VDCNR--SEGFCQEQY-CNYMETQVGYCSK-KKDACLH					
<i>EP2D/E</i>	8p23-p22	R	TICRM--QQGICLFFHSGEKKRDIQSD-PWNRCVSNTE					
<i>DEFB10</i>	6p12	P	ERCEK--VRGICMTF-CDDVEYDY-YCIK-WRSQCV					
<i>DEFB11</i>	6p12	P	RECRV--GNGQCINQ-CHENEIRIAYCIR-PGTTCCLQQ					
<i>DEFB12</i>	6p12	P	KSCIA--IGGRCINQ-CDDSEFRISYCAL-PTTCCVTECDP					
<i>DEFB13</i>	6p12	P	REQLL--VRGACNPE-CNSWEVYVYVYCNV---NFCQAVWE					
<i>DEFB14</i>	6p12	P	DRCTK--RYGRCDRD-CLESEKPDGNCRL--PKTCCTEKL					
<i>DEFB15</i>	20q11.1	P	RRCYV--GTGRCKNS-CKEIEKKECKGE--KHICVVPHEED+16					
<i>DEFB16</i>	20q11.1	P	NPCEL--YQGMCA-N-CHEYEIQYLTCPL--DQCCCLKLSVK+28					
<i>DEFB17</i>	20q11.1	P	KSCWI--IKGRCNN-CMPEGVKPKCN--GKCCIPSNVS					
<i>DEFB18</i>	20q11.1	R	KKCNW--RSGCKKQ-CMDGEAVDTCNK--LRACCIPSNED+62					
<i>DEFB19</i>	20q11.1	R	LRCMG--NSGICNAS-CKNEQPYLYRNN--CQSCCLQSYM+22					
<i>DEFB20</i>	20q11.1	R	VECWV--DGHCHLL-CMDGEDSIIRCN--RKRCCVPSY+32					
<i>DEFB21</i>	20q11.1	P	MFCWG--KSGRCNT-CKEVEYILCKT--EAKCCVDPKVV+19					
<i>DEFB22</i>	20q11.1	P	ETCNW--FSGCHDE-CLNNEVYVYVYVCS--GKLCCLPKDQ+9					
<i>DEFB23</i>	20q11.1	R	QRQWN--LYGRCYR-CSKKEVYVYVYIN--NEMCVVPKYQ+8					
<i>DEFB24</i>	20q11.1	P	KRCNK--GQCAQTY-CTRQETMELCPD--ASLCCLSVALK+10					
<i>DEFB25</i>	20p13	R	QKCNW--NVGCKRR-CLDTEYILLCKN--KLSCCISIS+94					
<i>DEFB26</i>	20p13	R	KKCLN--DVGICCKK-CKPEM(+8)CGK--QRDCCVPADR+46					
<i>DEFB27</i>	20p13	R	KKCNWYVQGHCKRI-CRVNEVPEALCEN--GRYCCNLINKEL+39					
<i>DEFB28</i>	20p13	P	KKCFN-KVTGYCKK-CKVGEVEYIGLS--GKLCANDEEE+34					
<i>DEFB29</i>	20p13	R	RRCLM--GLRGCNDH-CNVDEEIQRCM--KFCVGVPKVV+123					
<i>DEFB30</i>	A	P	KQCIA--LKGVCNDLCLSTLDDTIGICNE--GKRCCKRWW					
<i>DEFB31</i>	A	P	DECPV--EYVHCLK-CNADESAIRCAD--FSTCCRLKI					
Consensus			++Cxx--xxG+C++x-C+xxE+xIxxCxx--x+CC					

Fig. 1. Multiple sequence alignment of human β -defensin proteins. The amino acid sequences were predicted from genomic sequence (Table 1). The chromosomal location of each gene is indicated, except where mapping was ambiguous (A). The sequences were aligned as described in *Materials and Methods* followed by minor manual manipulations to maximize sequence alignment and clustering of genes by chromosome. We classified β -defensin genes as known (K), if evidence exists that they are transcribed and that their protein product demonstrates antimicrobial activity; related (R), if evidence exists that they are transcribed but their protein product has not been tested for antimicrobial activity; predicted (P), if no evidence exists that they are transcribed; and pseudogenes (S), if the DNA sequence is highly similar to a β -defensin gene, but the predicted amino acid sequence lacks an ORF across the six-cysteine motif. The consensus sequence shows cysteine residues (yellow highlight), positively charged residues (blue +) and other residues (red) if they are represented in greater than 50% of all predicted β -defensin proteins.

are conserved (Fig. 3). Given the syntenic relationship between the human and mouse clusters and the conservation of genes within them, we conclude that each pair originated from a common ancestral gene cluster (13, 22). In addition to the six-cysteine domain and density of cationic residues that define the β -defensin peptides, comparisons of the predicted amino acid sequences for the new human (Fig. 1) and mouse (Table 1) β -defensin genes reveal two new features, a conserved glycine between cysteines 1 and 2 and a previously unrecognized glutamic acid between cysteines 3 and 4 (Fig. 1).

To test whether these predicted genes are transcribed, we performed BLAST searches against the six-frame translation of the EST database (dbEST). At least one EST was found for 12 human and 14 mouse predicted genes (Table 1), representing all clusters. The absence of some β -defensin genes in dbEST is not surprising. For example, the β -defensin gene *DEFB3* is not found in dbEST; transcripts from this gene are rare until inflammatory stimuli induce expression (5, 6). We note that presence of

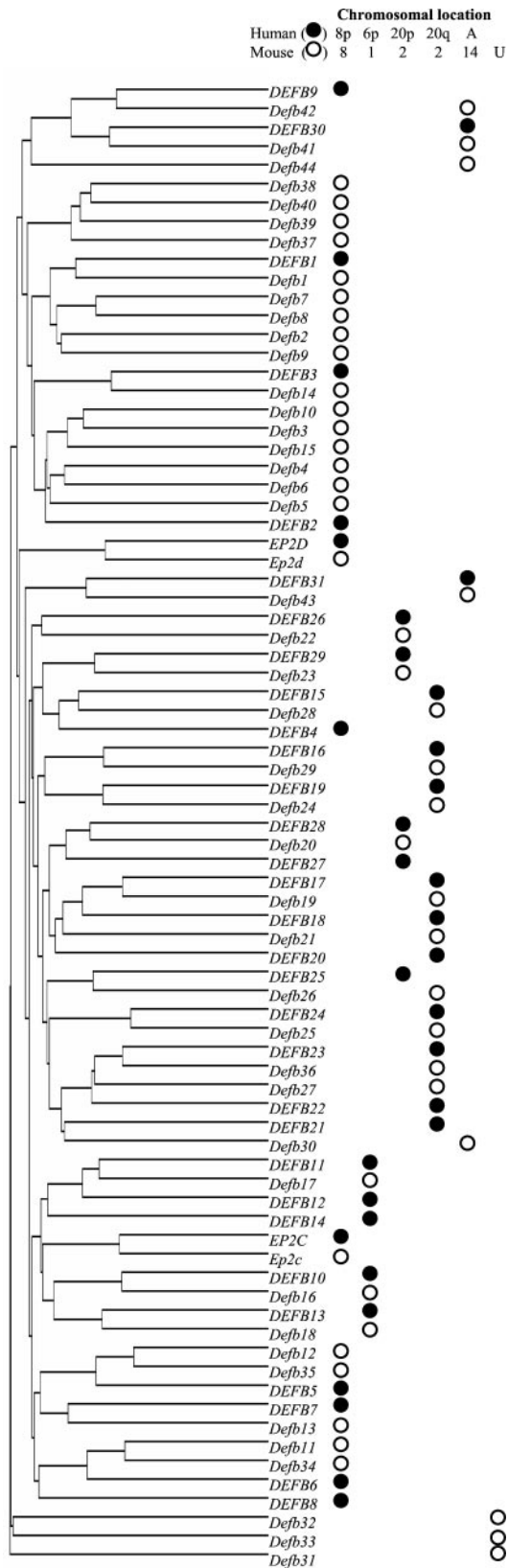


Fig. 2. Dendrogram of predicted β -defensin proteins. The length of each branch is inversely related to their similarity. The tree was constructed with the partial amino acid sequences predicted from the indicated human (●) and mouse (○) known, related, and predicted genes (Table 1). The genomic location for each gene is indicated. In some cases, the location of the genes was ambiguous (A) or unknown (U).

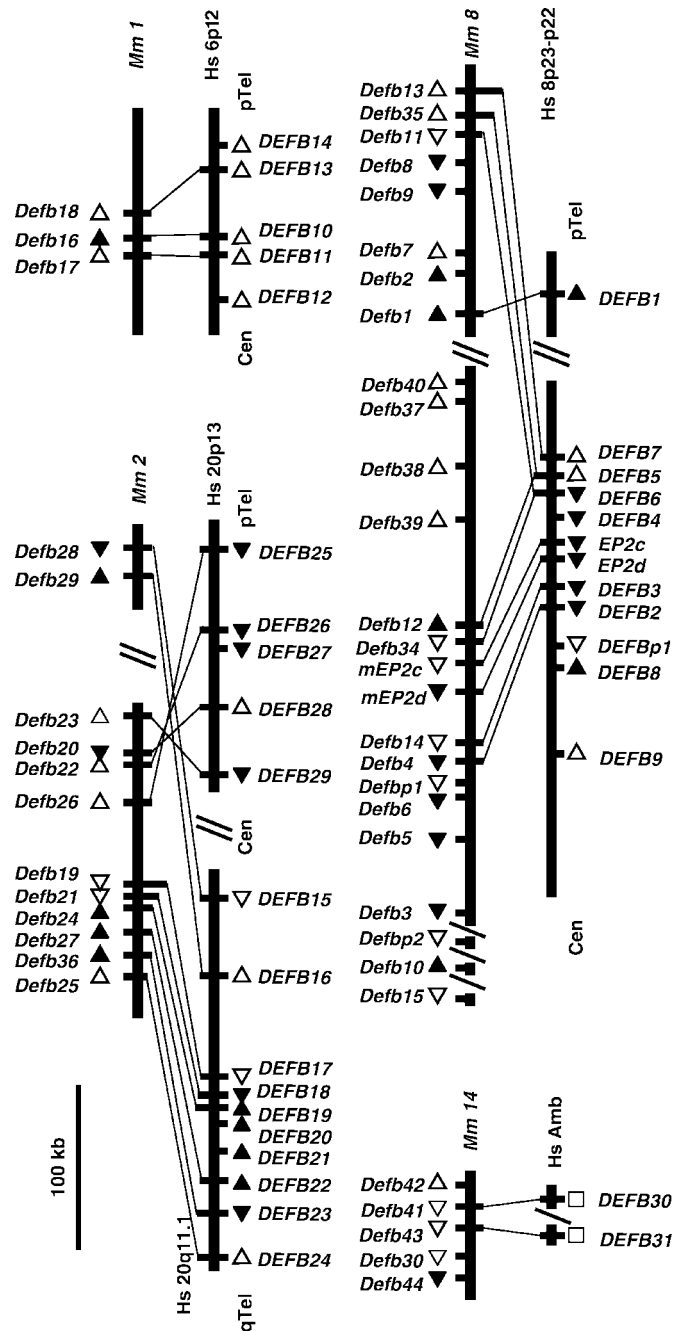


Fig. 3. Genomic structure of five conserved β -defensin gene clusters. Vertical bars represent assembled genomic DNA sequence contigs (see Table 1 for GenBank and Celera accession nos.) from the indicated human (Hs) and mouse (Mm) chromosome. Slanted lines represent gaps in the genomic sequence of known (single) or unknown (double) length. The telomere (Tel) and centromere (Cen) orientation of human DNA sequence contigs was deduced from the position of genetic markers. The orientation of the mouse DNA sequence contigs was deduced from the alignment of human and mouse gene homologs. The direction of transcription is indicated for known and related genes (▲) and predicted genes and pseudogenes (△) or is unknown (◻). Thin lines connect human and mouse genes with highest sequence similarity (Fig. 2).

β -defensin transcripts alone may not accurately predict protein expression and additional studies are required to verify the production of functional β -defensin peptides by these new genes.

We aligned the human and mouse EST sequences with the genomic sequence to verify that the EST sequences encode

full-length β -defensin proteins. Eight human EST sequences split into the familiar two-exon structure of a β -defensin gene. The first exon encoded the leader domain and the second exon encoded the six-cysteine domain. The other four human EST sequences failed to extend into an exon 1 sequence (*DEFB25*) or the first exon did not encode a leader domain (*DEFB8*, *DEFB22*, *DEFB31*). The latter three genes may represent transcribed pseudogenes or *EP2*-like genes. The *EP2* gene (also called HE2) has two promoters and produces multiple transcripts. Each alternative *EP2* transcript encodes a different protein isoform, including predicted proteins with a six-cysteine β -defensin domain (8, 24, 25). Further expression studies are needed to distinguish between these two possibilities. We observed similar results with the mouse EST sequences—12 ESTs split into the expected two-exon structure and two ESTs lacked a leader domain in the first exon. Finally, we identified ESTs with partial identity to the human *DEFB9* and mouse *Defb23*. However, the intron/exon sequences indicated that the transcripts were produced from the opposite strand. Consequently, we do not know whether a β -defensin gene is produced at these two loci. These preliminary expression studies together with the conservation of the five sequence clusters suggest that many of the 28 human and 39 mouse predicted β -defensin genes encode functional products.

Discussion

These findings are a proof of principle for a pioneering genome-wide search strategy that identifies genes with conserved structural motifs. The study highlights the complementary nature of the BLAST and HMMER analysis tools and demonstrates their potential synergy for mining genomic databases and identifying new members of gene families. None of the novel sequences identified were included in previous annotations of the human and mouse genomes, indicating that they were missed. Based on this example, the current estimates of 30,000–40,000 genes in the human and mouse genome (16, 17) may be a significant underestimate. We speculate that novel genes in other families are not counted with conventional strategies but could be detected by using this approach. Excellent candidates include gene families whose products are secreted, such as α -defensins, chemokines, growth factors, and neuropeptides. Like the β -defensin gene products, such peptides use disulfide bridges that are stable in the high oxidation environment outside the cell. This bonding strategy allows the primary structure of a peptide to be short and diverse, because only a few amino acids are required to maintain tertiary and quaternary structure. Consequently, gene families that encode such peptides may be refractory to conventional search strategies.

Some of the predicted protein products of the new β -defensin genes exhibit unique features. Many have subtle variations in the typical spacing between the cysteine residues

(see Table 1). However in *DEFB7*, there are 21 amino acid residues between the first and second cysteines rather than the more typical 6, a finding of unknown functional significance. A unique feature of the sequences within the chromosome 20 cluster and their mouse homologs is the longer C-terminal tails following the six-cysteine domain (Fig. 1). Further studies are needed to learn whether these C-terminal sequences are posttranslationally modified in the mature peptides. This structure is reminiscent of some chemokine gene products that share both functional and structural features with the β -defensins (26, 27). Both have a structural core of three anti-parallel β -sheets secured by disulfide linkages (28). The β -defensins and the chemokine CCL20 (MIP-3 α) engage the CCR6 receptor on selected immune effector cells (27), and some chemokines exhibit antimicrobial activity (29). Chemokines are also encoded by small genes that exist in the genome as gene clusters, with functionally related sequences present within a cluster, suggesting that the gene family expanded by duplication events (30). Thus, it is possible that features of functional significance are conserved within this larger family of β -defensin sequences and the chemokines.

Expansion of the β -defensin gene family has significant implications for future studies of innate host defense. The genomic conservation between human and mouse clusters suggests functional conservation and supports use of the mouse as a model organism to study this gene family. However, the size of the gene family provides the potential for redundant functions and may limit studies of mice that lack the function of a single gene or cluster. The search strategy outlined here focused on finding β -defensin second exons—the second exon encodes the mature peptide. A similar approach could be used to discover all first exon-coding sequences and the associated regulatory elements that confer cell specificity and responsiveness to inflammatory stimuli and pathogens. Addition of a large number of proteins to this family should also provide new opportunities to investigate the relationship between antimicrobial peptide structure, function, and specificity. Thus, this study helps lay the groundwork to understand better the complex interaction between innate host defense and the diversity of microorganisms in our environment.

We thank Margaret Malik, Andrea Penisten, and Autumn Bradley for technical assistance. We thank John Stokes, Beverly Davidson, and Tomas Ganz for critically reviewing the manuscript. These data were generated in part through the use of the Celera Discovery System and Celera Genomics-associated databases. We acknowledge the support of National Institutes of Health Grant NHLBI HL-61234 (to M.J.W., P.B.M., and B.C.S.), National Institutes of Health/National Institute of Child Health and Human Development Grant P30-HD27748 (to B.C.S.), and the Howard Hughes Medical Institute. T.L.C. is supported by National Institutes of Health/National Cancer Institute Grant CA-85188 and by a Cystic Fibrosis Genome Analysis grant. M.J.W. is an Investigator of the Howard Hughes Medical Institute.

- Huttner, K. M. & Bevins, C. L. (1999) *Pediatr. Res.* **45**, 785–794.
- Tang, Y. Q. & Selsted, M. E. (1993) *J. Biol. Chem.* **268**, 6649–6653.
- Bensch, K. W., Raida, M., Magert, H. J., Schulz-Knappe, P. & Forssmann, W. G. (1995) *FEBS Lett.* **368**, 331–335.
- Harder, J., Bartels, J., Christophers, E. & Schroder, J.-M. (1997) *Nature (London)* **387**, 861–862.
- Harder, J., Bartels, J., Christophers, E. & Schroder, J. M. (2001) *J. Biol. Chem.* **276**, 5707–5713.
- Jia, H. P., Schutte, B. C., Schudy, A., Linzmeier, R., Guthmiller, J. M., Johnson, G. K., Tack, B. F., Mitros, J. P., Rosenthal, A., Ganz, T. & McCray, P. B., Jr. (2001) *Gene* **263**, 211–218.
- Garcia, J. R., Krause, A., Schulz, S., Rodriguez-Jimenez, F. J., Kluver, E., Adermann, K., Forssmann, U., Frimpong-Boateng, A., Bals, R. & Forssmann, W. G. (2001) *FASEB J.* **5**, 1819–1821.
- Kirchhoff, C., Osterhoff, C., Habben, I., Ivell, R. & Kirchhoff, C. (1990) *Int. J. Androl.* **13**, 155–167.
- Frohlich, O., Po, C., Murphy, T. & Young, L. G. (2000) *J. Androl.* **21**, 421–430.
- Huttner, K. M., Kozak, C. A. & Bevins, C. L. (1997) *FEBS Lett.* **413**, 45–49.
- Morrison, G. M., Davidson, D. J. & Dorin, J. R. (1999) *FEBS Lett.* **442**, 112–116.
- Bals, R., Wang, X., Meegalla, R. L., Wattler, S., Weiner, D., Nehls, M. C. & Wilson, J. M. (1999) *Infect. Immun.* **67**, 3542–3547.
- Jia, H. P., Wovk, S. A., Schutte, B. C., Lee, S. K., Vivado, A., Tack, B. F., Bevins, C. L. & McCray, P. B., Jr. (2000) *J. Biol. Chem.* **275**, 33314–33320.
- Yamaguchi, Y., Fukuhara, S., Nagase, T., Tomita, T., Hitomi, S., Kimura, S., Kurihara, H. & Ouchi, Y. (2001) *J. Biol. Chem.* **14**, 31510–31514.
- Hughes, A. L. & Yeager, M. (1997) *J. Mol. Evol.* **44**, 675–682.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) *Nature (London)* **409**, 860–921.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001) *Science* **291**, 1304–1351.
- Marshall, E. (2001) *Science* **292**, 822–823.

19. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
20. Eddy, S. R. (1998) *Bioinformatics* **14**, 755–763.
21. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
22. Liu, L., Zhao, C., Heng, H. H. Q. & Ganz, T. (1997) *Genomics* **43**, 316–320.
23. Bevins, C. L., Jones, D. E., Dutra, A., Schaffzin, J. & Muenke, M. (1996) *Genomics* **31**, 95–106.
24. Hamil, K. G., Sivashanmugam, P., Richardson, R. T., Grossman, G., Ruben, S. M., Mohler, J. L., Petrusz, P., O’Rand, M. G., French, F. S. & Hall, S. H. (2000) *Endocrinology* **141**, 1245–1253.
25. Frohlich, O., Po, C. & Young, L. G. (2001) *Biol. Reprod.* **64**, 1072–1079.
26. Zlotnik, A. & Yoshie, O. (2000) *Immunity* **12**, 121–127.
27. Yang, D., Chertov, O., Bykovskaia, S. N., Chen, Q., Buffo, M. J., Shogan, J., Anderson, M., Schroder, J. M., Wang, J. M., Howard, O. M. & Oppenheim, J. J. (1999) *Science* **286**, 525–528.
28. Perez-Canadillas, J. M., Zaballos, A., Gutierrez, J., Varona, R., Roncal, F., Albar, J. P., Marquez, G. & Bruix, M. (2001) *J. Biol. Chem.* **276**, 28372–28379.
29. Cole, A. M., Ganz, T., Liese, A. M., Burdick, M. D., Liu, L. & Strieter, R. M. (2001) *J. Immunol.* **167**, 623–627.
30. Rollins, B. J. (1997) *Blood* **90**, 909–928.