Published in final edited form as:
Biochem Biophys Res Commun. 2005 July 29; 333(2): 383–389.

The evolution of a genetic locus encoding small serine proteinase inhibitors ⋆

Adam Clauss[a,*], **Hans Lilja**[a,b], and **Åke Lundwall**[a]

a *Department of Laboratory Medicine, Clinical Chemistry, Lund University, University Hospital MAS, Malmö, Sweden*

b *Memorial Sloan-Kettering Cancer Center, Departments of Clinical Laboratories, Urology, and Medicine, New York, NY, USA*

Abstract

We previously identified a locus on human chromosome 20 that encompasses 14 genes of postulated WFDC-type proteinase inhibitors with a potential role in innate immunity. In an extended study, homologous loci are here described on mouse chromosome 2, rat chromosome 3, and dog chromosome 24. As in humans, the murine and canine loci are divided into two sub-loci separated by 0.2 Mb. The majority of genes are conserved in all species, but there are also species-specific gains and losses of genes, e.g., several duplications have yielded four SLPI genes in the rat and, most surprisingly, there is no murine elafin gene. Two human pseudogenes were identified due to the discovery of functional rodent genes. The conservation of different WFDC domains varies considerably, and it is hypothesized that this reflects a dual role of WFDC inhibitors in natural immunity, which is directed both against microbes and proinflammatory cells.

Keywords

Protease inhibitor; Serine protease; Kallikrein; WFDC; Innate immunity; Inflammation; Epididymis; Genital tract; Respiratory tract; Evolution

A characteristic structural motif consisting of four conserved disulphide bridges was first discovered in whey acidic protein (WAP), the major protein in milk whey of rodents [1]. The same motif is present in several other secreted proteins containing what is now known as a WAP four-disulphide core (WFDC) domain. Notably, the WFDC domain is present in some small serine proteinase inhibitors, such as secretory leukocyte protease inhibitor (SLPI) and elafin (also known as PI3) [2,3]. The WFDC domains of these proteins are composed of around 40 amino acid residues that form a relatively flat structure, with the peptide chain folded into a central two-stranded β-sheet surrounded by two peptide strands that are joined at one end of the molecule, forming a looped structure that binds to the active site and inhibits serine proteases [4,5]. The target enzymes of SLPI and elafin are endogenous proteases like elastase and cathepsin G, secreted by proinflammatory cells. Presumably, they also inhibit exogenous proteases secreted by microorganisms, as they display anti-bacterial, anti-fungal, and anti-viral properties [6,7]. However, it has been suggested that some of the anti-microbial properties might result from a defensin-like mechanism, rather than resulting directly from the protease inhibition [8]. In any case, the WFDC proteins seem to have a biological role that is closely related to innate immunity and inflammation.

⋆The novel nucleotide sequence data reported in this paper have been deposited with the GenBank sequence data bank; *Wfdc5* AY669505, *Wfdc8* AY669506 and AY757956, *Wfdc16* AY541526, Wfdc13 AY542486 and AY542487, and Wfdc16 AY542488 and AY542489.

* Corresponding author. Fax: +46 40337043. E-mail address: adam.clauss@med.lu.se (A. Clauss).

In an earlier study, we found that genes encoding SLPI and elafin are homologous to genes of the major coagulum proteins in semen, e.g., semenogelin I and semenogelin II in human semen [9,10]. In a more recent analysis of DNA sequences surrounding the semenogelin locus on human chromosome 20, we discovered a 678-kb region encompassing a total of 14 genes that code for proteins with WFDC domains [10,11]. Among them are the previously known genes of SLPI, elafin, HE4, and eppin. The typical gene carries a 5′ exon that codes for a signal peptide, followed by one or more exons that each code for a WFDC domain, and finally a 3′ exon with no or very limited coding information that contains the polyadenylation signal. The majority of the genes code for a single WFDC domain, but some carry more than one. There are also examples of mixed inhibitors, like eppin and WFDC8, which—in addition to the WFDC domain—carry a domain with the sequence motif of a kunitz-type serine protease inhibitor [11,12].

In this report, we have extended our studies on the WFDC locus to non-primate mammals. The structural conservation allows an assessment of the relative importance of certain genes and motifs and it might also provide insights into whether the target enzymes are endogenous or exogenous proteases. Because of this we have mapped in detail the mouse WFDC locus and then identified homologous genes in rat and dog. Finally, we have made a comparison of the locus in humans, mice, rats, and dogs.

Materials and methods

Nomenclature

We have discussed the nomenclature of WFDC-encoding genes with a representative of the Human Gene Nomenclature Committee and revised the nomenclature used in our previous publications [10,11]. The gene symbol *WAP* has been replaced by the approved root symbol *WFDC*. In general, the hierarchical numbering used in the previous paper has been left intact, but there are a few exceptions: the former *WAP1* is now *WFDC5*, the former *WAP2* is now *WFDC12*, and the former *WAP10* and *WAP12* are now *WFDC10A* and *WFDC10B*, due to the close relationship of these two genes. Rodent genes are written in italics with a beginning uppercase letter followed by lower case letters and numbers, as recommended by the Mouse Gene Nomenclature Committee.

Identification and sequencing of genes and transcripts

Genes encoding WFDC domains were identified in genomic DNA following analysis of six reading frames as described [11]. RT-PCR was carried out as published [10], with primers designed from nucleotide sequences located on separate exons to ensure that the size of PCR products from spliced transcripts differed from those of non-spliced transcripts or chromosomal DNA (Table 1). The transcripts for DNA sequencing were generated by rapid amplification of cDNA ends (RACE) with the SMART RACE amplification kit (Clontech, Palo Alto, CA, USA), essentially as described [11]. The primers were the same as used for the RT-PCR. Automated DNA sequencing was done on an ABI Prism 3100 Genetic Analyser using the Big Dye Terminator Cycle Sequencing Ready reaction kit (Applied Biosystems, Foster City, CA, USA). Nucleotide and peptide sequences were analysed using Emboss (http://emboss.sourceforge.net), and the signal peptide cleavage sites were predicted by the SignalP 3.0 server at the Center for Biological Sequence Analysis (http//www.cbs.dtu.dk/servises/SignalP/).

Results

The mouse WFDC locus was assigned to chromosome 2, cytogenic region H3, due to the presence in this region of *Slpi* and the genes encoding mouse seminal vesicle secreted proteins

(SVS). An extended search of the chromosomal region allowed us to identify a locus of 566 kb, encompassing 23 WFDC-coding exons located in 16 genes. Many of the genes were fully confirmed by expressed sequence tags (EST) retrieved by BLAST search of the mouse EST database at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). The remainder, mouse genes, *Wfdc5*, *Wfdc6b*, *Wfdc8*, *Wfdc13*, and *Wfdc16*, were confirmed by DNA sequencing of RACE products. The majority of the mouse genes were identified as murine orthologues of already known human genes (supplement 1). As in humans, the mouse locus is divided into two sub-loci separated by a gap of 206 kb containing unrelated genes. The more centromeric sub-locus is 179 kb and has *Wfdc5* at the 5′ end and *Slpi* at the 3′ end. It also contains *Wfdc12* and the genes encoding SVS II–VI, the latter in a position that is homologous with that of the human semenogelin locus. In the human genome, *PI3*, encoding elafin, is located next to the semenogelin I gene (*SEMG1*). In the mouse genome, we could not identify a *PI3* at the homologous position, but instead we found two new WFDC genes further upstream. These genes are not orthologous with any of the previously characterized human WFDC genes, but their sequence similarity of 76% suggests that they are the result of a relatively late duplication —hence, they were given the symbols *Wfdc15a* and *Wfdc15b* (supplement 2). A search of the human WFDC locus with the novel gene sequences identified a human gene, which we designated *WFDC15*, that presumably is related to a precursor of *Wfdc15a* and *Wfdc15b*, as the sequence similarity is the same to both of them (Fig. 1A). Several mutations were identified that drastically affect the translation product of the human gene: a mutated start codon resulting in a truncation of the signal peptide and a deletion leading to a shift of reading frame that corrupts the WFDC motif and also leads to a premature stop of translation. These features indicate that human *WFDC15* is a pseudogene.

The more telomeric sub-locus is 180 kb and extends from *Wfdc2* on the centromeric side to *Wfdc3* on the telomeric side. The organization is the same as that of the human telomeric sub-locus, and we identified orthologues to all genes present in the human sub-locus, with the exception that there is only one *Wfdc10*—the duplication leading to *WFDC10A* and *WFDC10B* occurred after the split of the lineages leading to murines and primates. Two new genes were also discovered at the murine telomeric sub-locus. One of these is located 12 kb downstream of *Wfdc8* and seems to be the result of a duplication of *WFDC6*. The novel gene *Wfdc6b* is 91% similar in sequence to the gene, *Wfdc6a*, located at the same relative position as the human *WFDC6* (supplement 3). Like their neighbouring genes *Wfdc8* and *Spin-lw1* (encoding eppin), the two mouse *Wfdc6* genes each code for a kunitz domain in addition to the WFDC domain. A homologous sequence was also identified in human *WFDC6*, but mutations affecting the reading frame and one of the conserved Cys codons suggest that the exon is a pseudoexon (Fig. 1B). This was also verified by several failed attempts to amplify a transcript encompassing both the kunitz and the WFDC domain by RT-PCR with RNA from epididymis or testis. The second new mouse gene at the telomeric sub-locus is situated 15 kb upstream of *Wfdc9*. The gene, *Wfdc16*, differs from most other genes at the locus by having an extra exon located between the WFDC-coding exon and the 3′ exon containing the polyadenylation signal. The nucleotide sequence of *Wfdc16* was used for a search of the human WFDC locus, which led to the identification of the corresponding human gene (Fig. 1C). However, the human *WFDC16* contains several mutations—affecting both the splice acceptor site and the coding sequence of the exon of the WFDC domain—that indicate it is a pseudogene.

In addition to the novel genes, we also found that several of the WFDC genes were differently organized in mice and humans. The *WFDC3*, which in humans contains four WFDC-coding exons, has only the last two in mouse. Sequences that are homologous with the first two exons of human *WFDC3* were found in the mouse gene, but they are presumably pseudoexons as they carry several in-frame stop codons (supplement 4). Another example of a truncated mouse gene is *Wfdc13*, which is lacking the two last exons of the human gene. Instead, the same exon that codes for the WFDC domain also carries the stop codon and all 3′ non-translated

nucleotides. We also found mouse genes that carried more exons than their human counterparts. Mouse *Wfdc2*, encoding epididymal protein HE4, has an extra exon located between the two exons that code for WFDC domains. In mouse *Wfdc8*, there is a second exon that is not present in the human gene and that carries the initiation Met and codes for 38 amino acid residues. Analysis of the predicted amino acid sequence suggests that the protein is not preceded by a signal peptide. We also detected two splice forms of mouse *Wfdc8* that differ due to the presence or absence of the third WFDC domain.

The locus on rat chromosome 3 is very similar to that of the mouse and almost all genes are preserved, including the duplicated *Wfdc6* and the novel genes *Wfdc15a*, *Wfdc15b*, and *Wfdc16* (Fig. 2). However, we could not find a *Wfdc13* orthologue. A most interesting finding is that there have been multiple duplications of *SLPI* that seem to be unique to the rat. This has resulted in four genes, designated *Slpi1–4*, with paralogous WFDC domains that are 58–95% similar in sequence (supplement 5). The N-terminal half of the molecules, encompassing the protease binding site, is less conserved than the C-terminal half that carries the residues of the core β-sheet. The second WFDC domain of Slpi4 contains the amino acid sequence CLMLNPPN, which is identical to the conserved elastase-binding loop of human, dog, and mouse SLPI.

The dog genome is less completely sequenced than those of the other species. As a consequence, there are some ambiguities at the dog WFDC locus with respect to the localization of contigs and genes, e.g., *WFDC5* and *PI3* are juxtaposed and not intervened by *WFDC15* which is located on another contig, while the comparison with other species suggests that *WFDC15* should be located between *WFDC5* and *PI3*. However, from the available data it can be concluded that the dog locus on chromosome 24 is more similar to the human than to the rodent locus, e.g., there are *PI3* and *WFDC10B* orthologues, *WFDC3* carries four WFDC domains, and *WFDC6* is not duplicated. Major differences from the human locus are that *WFDC15* presumably is expressed and that there seems to be no dog *WFDC12*. We also discovered two dog *WFDC9* genes that may be the result of a canine-specific duplication, as they show 90% sequence conservation in the exon that codes for the signal peptide. Both are probably pseudogenes, as the exon encoding the WFDC domain either contains deleterious mutations or is missing. The dog locus also contains three canine WFDC genes that were assigned to *WFDC16* because of homology with the human pseudogene. These are probably functional genes, as each carries an exon that codes for a typical WFDC motif, which is surrounded by correct splice donor and acceptor sites. The sequence similarity of the genes ranges between 83% and 90% in nucleotides flanking the WFDC motif, while the WFDC motifs display a conservation of only 37–63%. Another interesting duplication was found on the contig that also contains *WFDC15*. It carries a sequence that is 74% similar to a 2 kb region of *PI3*, encompassing a part of the second intron, the third exon, and the 3′ flanking sequence. Located upstream of this sequence is an exon that codes for a WFDC domain that is only 47% similar to the domain in *PI3*. Further upstream there is no similarity at all with *PI3*—thus, there seems to have been a duplication of *PI3* that was followed by recombination that eliminated the 5′ part of the gene.

A comparison of WFDC sequences from different species provided some interesting results. As might be expected, the conservation of the WFDC domains varies. Some of them, e.g., WFDC12, WFDC9, and the amino-terminal domain of WFDC8, display a rapid evolution; only around half of the amino acid residues are conserved. In contrast, close to 80% of the residues are retained in the preserved C-terminal domains of WFDC3 (Table 2). Another interesting observation is that amino-terminal domains are less conserved than C-terminal domains in most of the proteins that contain multiple WFDC domains.

Discussion

In this study, we have characterized the WFDC loci of mouse, rat, and dog. With the information already available on the human counterpart, the organization of the locus is now known in four mammalian species. The general layout of the WFDC locus, with the division into two sub-loci, is conserved in all species. The rodent homologues of the human centromeric sub-locus also contain the genes of the seminal vesicle secreted proteins SVS II–VI. At the human sub-locus, the equivalent position is occupied by the genes encoding semenogelin I and II (*SEMG1* and *SEMG2*), further supporting the earlier proposal that the seminal vesicle secreted proteins in humans and rodents are related, in spite of a very limited—or even absent—similarity in primary structure [13]. The conserved location of the seminal vesicle-expressed genes at the WFDC locus also lends support to the hypothesis that they have evolved from a progenitor that was a *WFDC* gene [11]. In the dog, we could not identify equivalents to the genes encoding seminal vesicle secreted proteins—which came as no surprise as dogs lack seminal vesicles. However, recent studies have shown expression of semenogelin I and II in several human tissues outside of the male genital tract [14], and the existence of a dog homologue is therefore not fully ruled out; it might have escaped recognition due to gaps in the available nucleotide sequences.

A recent paper described two novel mouse cDNAs that code for WFDC proteins [15]. The two proteins, designated SWAM1 and SWAM2, are identical to the products of mouse *Wfdc15b* and *Wfdc12* in this report. In the same paper, it was suggested that the human orthologue of SWAM1 is the product of the predicted gene *LOC149709*, situated between *SEMG2* and *SLPI*. The predicted transcript (Accession No. XM_086637) does not code for a WFDC motif, but inserting one nucleotide and substituting another would create a sequence that can be translated to yield the human SWAM1 motif described by Hagiwara et al. These mutations—and at least two more that affect the formation of a proper signal peptide—prohibit the synthesis of a functional WFDC inhibitor. Therefore, we assume that *LOC149709* is a pseudogene that is only distantly related to *SWAM1/Wfdc15b* and that the closest relative in humans is the *WFDC15* pseudogene.

Several of the genes at the locus display rapid evolution, as revealed by the poor conservation of certain WFDC domains due to point mutations. However, there are other mechanisms that have also created large differences between proteins originating from orthologous genes. Splicing has, for instance, created a rodent HE4 that contains a unique structure at the centre of the molecule [16,17]. A second example of splice-dependent species variation is seen in *WFDC8*. In the human gene, the signal peptide was predicted to originate from codons located in the first and the second exons. In the mouse gene, the first exon is presumably non-translated as there are in-frame stop codons. Instead, the translation starts in the second exon [11], which is not present in the human gene. Coding then proceeds in the third exon and gives rise to a long stretch of hydrophobic residues that is followed by a site with homology to the predicted signal peptidase cleavage site of human WFDC8. The predicted signal peptide of human WFDC8 is unusually long and in the mouse protein it would be even longer. Furthermore, the signal peptide of WFDC8 is also unique in that it is encoded by two exons, whereas it is a single exon in the other WFDC genes at the locus. All this suggests that Wfdc8 may not have a signal peptide and that the hydrophobic sequence close to the amino terminus could instead serve as membrane anchor of a type II integral membrane protein.

As might have been expected, the WFDC loci of rat and mouse are very similar. It was therefore a great surprise to detect four *Slpi* paralogues in the rat. One of the protein products (Slpi4) is presumably an elastase inhibitor, as it carries the conserved protease binding loop in the second WFDC domain. The remaining three rat *Slpi* probably encode protease inhibitors with less affnity for elastase, but as they are unique to the rat they presumably also regulate the activity

of serine proteases that are specific to the rat. In a recent analysis of the rat glandular kallikrein locus, we demonstrated that several simple serine proteases, closely related to tissue kallikrein, are specific to the rat [18]. We have previously hypothesized about a linkage between the WFDC locus and the kallikrein locus on human chromosome 19. It is therefore reasonable to speculate that the novel Slpi-related molecules in the rat may regulate the activity of rat-specific glandular kallikreins.

As expected from the phylogeny, the dog WFDC locus is more similar to the human than to the rodent locus. However, it was a surprise to find that one of the major differences between the human and canine loci is that the latter is lacking *WFDC12*. We have previously demonstrated that this gene is highly expressed in the human prostate [10]. Because of the similarity in anatomy and physiology of the prostate in humans and dogs—with development of benign hyperplasia and cancer of the prostate as unique consequences [19]—one would have expected conservation of genes expressed in the prostate. Perhaps the neighbouring *WFDC15*, which is silenced in humans, has taken over the role of *WFDC12* in the dog. However, one should also bear in mind that there might have been errors in the assembly of DNA sequences into the contig predicted to hold dog *WFDC12*. In humans, the genes are located in the order *WFDC5*, *WFDC12*, *WFDC15-ps*, and *PI3*. The dog contig that is homologous with this region only carries *WFDC5* and *PI3; WFDC15* is located on another contig. This implies that there might be errors afflicting the DNA sequence in the region where *WFDC12* is expected to reside. Thus, a dog *WFDC12* should not be excluded until the DNA sequence of this genomic region is finalized.

Several studies have demonstrated important roles of SLPI and elafin in natural immunity to microbial infection. In more recent studies, both eppin and SWAM1 have also displayed anti-bacterial properties [15,20]. These results indicate that genes at the WFDC locus are anti-microbial and part of the innate immune system. Many anti-microbial defence molecules display a rapid evolution, e.g., as seen with defensins [21]. We found rapid evolution among some of the WFDC genes as well, and an attractive hypothesis could be that the sequence variation is the result of adaptation to proteolytic challenges from microorganisms that are endemic to different animal species. However, there are also examples of very conserved WFDC domains, and one could speculate that these regulate conserved endogenous proteases secreted at sites of inflammation. Thus, WFDC proteins could have a dual role in natural immunity by simultaneously inhibiting microbial proteases to prevent microbial invasion and inhibiting proteases from proinflammatory cells to regulate the magnitude of the inflammatory response. The known properties of SLPI are in favour of this hypothesis [22], as the C-terminal domain inhibits elastase secreted from proinflammatory cells, while the N-terminal domain displays anti-tryptic activity of unknown specificity that could very well be directed against proteases of microbial origin. Future studies, focused on the inhibitory spectrum of WFDC proteins coupled to their anti-microbial and anti-inflammatory properties, will reveal the full importance of products from this recently discovered genetic locus in innate immunity.

Supplementary materials

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding for this study was provided by Alfred Österlunds foundation, DOD CDMRP (Contract # W81XWH-05-4-0124) PCRP; NCI Spore Pilot Program 7 (Contract # P50-CA92629); the Swedish Cancer Society; (Grant # 3555); and European Union STREP-contract # LSHC-CT-2004-50301 (P-Mark).

References

1. Hennighausen LG, Sippel AE. Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. Nucleic Acids Res 1982;10:2677–2684. [PubMed: 6896234]
2. Thompson RC, Ohlsson K. Isolation, properties, and complete amino acid sequence of human secretory leukocyte protease inhibitor, a potent inhibitor of leukocyte elastase. Proc Natl Acad Sci USA 1986;83:6692–6696. [PubMed: 3462719]
3. Wiedow O, Schroder JM, Gregory H, Young JA, Christophers E. Elafin: an elastase-specific inhibitor of human skin. Purification, characterization, and complete amino acid sequence. J Biol Chem 1990;265:14791–14795. [PubMed: 2394696]
4. Grutter MG, Fendrich G, Huber R, Bode W. The 2.5 Å X-ray crystal structure of the acid-stable proteinase inhibitor from human mucous secretions analysed in its complex with bovine alpha-chymotrypsin. EMBO J 1988;7:345–351. [PubMed: 3366116]
5. Tsunemi M, Matsuura Y, Sakakibara S, Katsube Y. Crystal structure of an elastase-specific inhibitor elafin complexed with porcine pancreatic elastase determined at 1.9 Å resolution. Biochemistry 1996;35:11570–11576. [PubMed: 8794736]
6. Hiemstra PS, Maassen RJ, Stolk J, Heinzel-Wieland R, Steffens GJ, Dijkman JH. Antibacterial activity of antileukoprotease. Infect Immun 1996;64:4520–4524. [PubMed: 8890201]
7. Simpson AJ, Maxwell AI, Govan JR, Haslett C, Sallenave JM. Elafin (elastase-specific inhibitor) has anti-microbial activity against gram-positive and gram-negative respiratory pathogens. FEBS Lett 1999;452:309–313. [PubMed: 10386612]
8. Sallenave JM. Antimicrobial activity of antiproteinases. Biochem Soc Trans 2002;30:111–115. [PubMed: 12023836]
9. Lundwall A, Ulvsback M. The gene of the protease inhibitor SKALP/elafin is a member of the REST gene family. Biochem Biophys Res Commun 1996;221:323–327. [PubMed: 8619854]
10. Lundwall AA, Clauss A. Identification of a novel protease inhibitor gene that is highly expressed in the prostate. Biochem Biophys Res Commun 2002;290:452–456. [PubMed: 11779191]
11. Clauss A, Lilja H, Lundwall A. A locus on human chromosome 20 contains several genes expressing protease inhibitor domains with homology to whey acidic protein. Biochem J 2002;368:233–242. [PubMed: 12206714]
12. Richardson RT, Sivashanmugam P, Hall SH, Hamil KG, Moore PA, Ruben SM, French FS, O'Rand M. Cloning and sequencing of human Eppin: a novel family of protease inhibitors expressed in the epididymis and testis. Gene 2001;270:93–102. [PubMed: 11404006]
13. Lundwall A. The cloning of a rapidly evolving seminal-vesicle-transcribed gene encoding the major clot-forming protein of mouse semen. Eur J Biochem 1996;235:424–430. [PubMed: 8631362]
14. Lundwall A, Bjartell A, Olsson AY, Malm J. Semenogelin I and II, the predominant human seminal plasma proteins, are also expressed in non-genital tissues. Mol Hum Reprod 2002;8:805–810. [PubMed: 12200457]
15. Hagiwara K, Kikuchi T, Endo Huqun Y, Usui K, Takahashi M, Shibata N, Kusakabe T, Xin H, Hoshi S, Miki M, Inooka N, Tokue Y, Nukiwa T. Mouse SWAM1 and SWAM2 are antibacterial proteins composed of a single whey acidic protein motif. J Immunol 2003;170:1973–1979. [PubMed: 12574366]
16. Bingle L, Singleton V, Bingle CD. The putative ovarian tumour marker gene HE4 (WFDC2), is expressed in normal tissues and undergoes complex alternative splicing to yield multiple protein isoforms. Oncogene 2002;21:2768–2773. [PubMed: 11965550]
17. Kappler-Hanno K, Kirchho3 C. Rodent epididymal cDNAs identified by sequence homology to human and canine counterparts. Asian J Androl 2003;5:277–286. [PubMed: 14695977]
18. Olsson AY, Lilja H, Lundwall A. Taxon-specific evolution of glandular kallikrein genes and identification of a progenitor of prostate-specific antigen. Genomics 2004;84:147–156. [PubMed: 15203212]
19. Johnston SD, Kamolpatana K, Root-Kustritz MV, Johnston GR. Prostatic disorders in the dog. Anim Reprod Sci 2000;60–61:405–415.

20. Yenugu S, Richardson RT, Sivashanmugam P, Wang Z, O'Rand MG, French FS, Hall SH. Antimicrobial activity of human EPPIN, an androgen-regulated, sperm-bound protein with a whey acidic protein motif. Biol Reprod 2004;71:1484–1490. [PubMed: 15229136]

21. Semple CA, Rolfe M, Dorin JR. Duplication and selection in the evolution of primate beta-defensin genes. Genome Biol 2003;4:R31. [PubMed: 12734011]

22. Ohlsson, K.; Rosengren, M.; Stetler, G.; Brewer, M.; Hale, KK.; Thompson, RC. Pulmonary Emphysema and Proteolysis: 1986. Academic Press, Inc; 1987. Structure, genomic organization, and tissue distribution of human secretory leukocyte-protease inhibitor (SLPI): a potent inhibitor of neutrophil elastase; p. 307-322.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.bbrc.2005. 05.125.

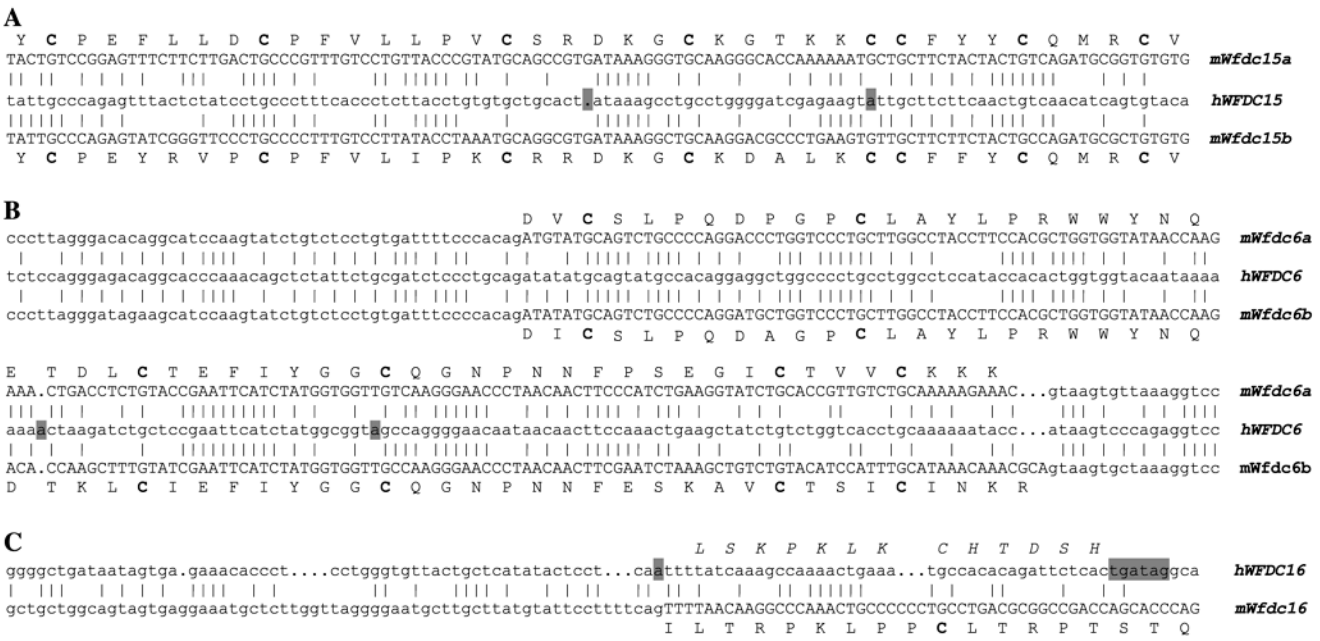

Fig 1.

The nucleotide sequences were aligned using the program bl2seq. Exon sequences are written in capital letters and translations are given with the one-letter code. Gaps are indicated by dots and conserved nucleotides are indicated by vertical bars. The conserved Cys residues are indicated by bold letters. (A) Identification of a human pseudogene with homology to mouse*Wfdc15a* and*Wfdc15b*. The frame-shifting deletion and the mutation of the Cys codon are indicated by shaded boxes. (B) Identification of a pseudoexon in human*WFDC6*, with homology to exons encoding the kunitz domain of mouse Wfdc6a and Wfdc6b. The frame-shifting insertion and the mutated nucleotide of the conserved Cys codon are shaded. (C) Identification of a human pseudogene with homology to mouse*Wfdc16*. The mutated splice acceptor site and the two in-frame stop codons are shaded.

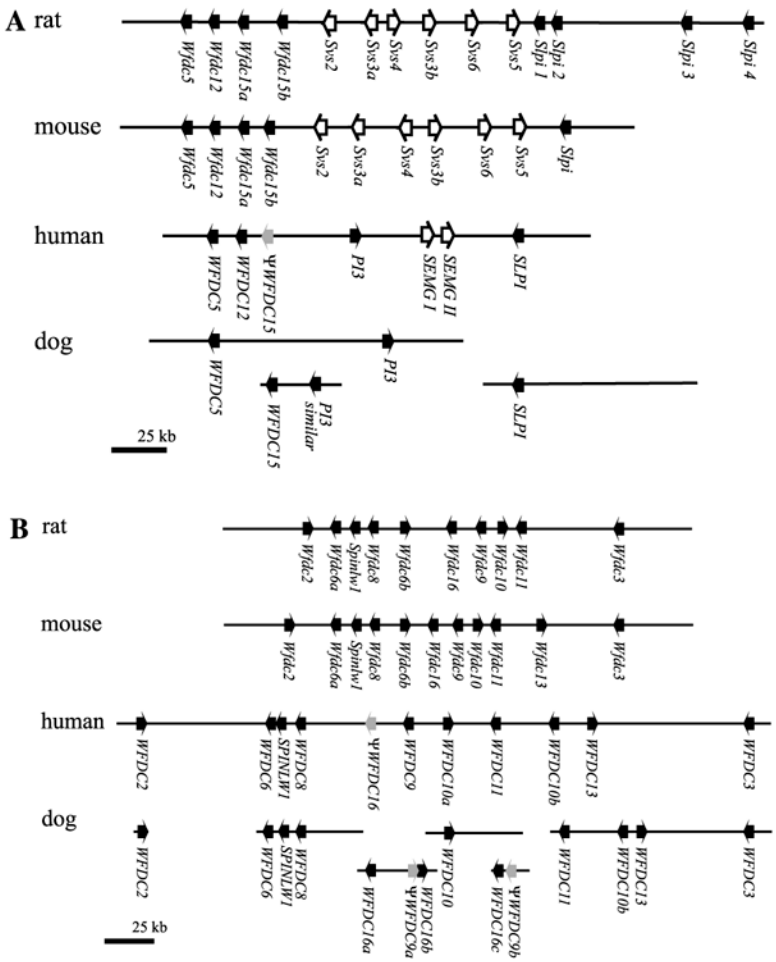

Fig 2.

Schematic illustration of the WFDC loci in man, mouse, rat, and dog. Equally sized arrows indicate the direction of transcription and approximate location of the genes. Black arrows indicate functional genes and arrows shaded in grey indicate pseudogenes. The horizontal bars illustrating the dog locus symbolize DNA contigs of the genomic sequence, which is not yet fully assembled. (A) Loci homologous with the human centromeric sub-locus. Semen coagulum proteins are illustrated with unfilled arrows. (B) Loci homologous with the human telomeric sub-locus. The broken line of the dog locus symbolizes the DNA contigs, which are not fully assembled.

Table 1

Sequences of the oligonucleotides used as primers for RT-PCR and RACE

Gene	Direction	Sequence 5′ → 3′
Wfdc5	Forward	CCTGGCTTTGGAGACTCACTTGCCT
	Reversed	TGGCACAGCACCGCTTCTTGCCTGAGCA
Wfdc8	Forward	CGGACTTCAACTGTAAGGCGCATTTCA
	Reversed	CATGCAAGCGTTTCTACAGTCAGTCTTGG
Wfdc6b	Forward	CTGTGGAGCATCCAGGAACCTGCGCT
	Reversed	GCTGAGGTCCAGGCATTTCCTCCCACA
Wfdc16	Forward	CGTTAGGCCTGCCAGTGTTGGCAAGATGGAA
	Reversed	GCAGCAGTGAAAGTCCTGCTCACAGTCC
Wfdc13	Forward	GACCTGTGTCACCTTTGCAGCTCCTGCT
	Reversed	AGCACTGAAGCGGTTCAGGGCATTCCTC

Table 2

Similarity of WFDC domains, presented as percent identical amino acids in aligned human (H), mouse (M), rat (R), and dog (D) sequences

	WFDC5d1			**WFDC5d2**			**WFDC12**			**WFDC15a**			**WFDC15b**		
	M	**R**	**D**	**M**	**R**	**D**	**M**	**R**	**D**	**M**	**R**	**D**[a]	**M**	**R**	**D**[a]
H	63	68	78	81	81	73	54	65	—	—	—	—	—	—	
M		82	58		86	65		76	—		89	68		76	57
R			65			68			—			70			62
	elafin			SLPId1			SLPId2			HE4d1			HE4d2		
	M	R	D	M	R[b]	D	M	R[b]	D	M	R	D	M	R	D
H	—	—	76	66	68	54	58	63	76	51	46	83	73	73	80
M		—	—		87	56		79	53		80	57		92	65
R			—			54			61			51			68
	WFDC6a			WFDC6b			Eppin			WFDC8d1			WFDC8d2		
	M	R	D	M	R	D	M	R	D	M	R	D	M	R	D
H	57	54	54	47	51	—	70	70	82	54	49	57	70	70	73
M		95	49		74	42		86	66		81	41		100	78
R			49			57			63			38			78
	WFDC8d3			WFDC16			WFDC9			WFDC10a			WFDC10b		
	M	R	D	M	R	D	M	R	D	M[c]	R[c]	D	M[c]	R[c]	D
H	78	75	75	—	—	—	75	72	—	66	74	60	54	51	69
M		92	72		71	34		92	—		80	66		—	56
R			69			37			—			69			56
	WFDC11			WFDC13			WFDC3d3			WFDC3d4					
	M	R	D	M	R	D	M	R	D	M	R	D			
H	61	64	81	55	—	73	85	88	85	82	82	88			
M		94	61		—	70		97	88		97	85			
R			61			—			85			85			

The suffxes d1 to d4 indicate the individual WFDC domains in multidomain proteins. A dash indicates that no comparison was made due to a missing domain in one of the species.

[a] The single dog WFDC15 is used in both comparisons.

[b] Comparison is made using rat Slpi4.

[c] The single mouse and rat Wfdc10 is used in both comparisons.