Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 352 (2005) 20 – 29
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Evolution of the spermadhesin gene family

Bianca Haase[a], Christian Schlötterer[b], Mahnaz Ekhlasi Hundrieser[c], Heidi Kuiper[a], Ottmar Distl[a], Edda Töpfer-Petersen[c], Tosso Leeb[a,*]

[a]Institute of Animal Breeding and Genetics, School of Veterinary Medicine Hannover, Bünteweg 17 p, 30559 Hannover, Germany
[b]Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, A-1210 Vienna, Austria
[c]Institute of Reproductive Medicine, School of Veterinary Medicine Hannover, 30559 Hannover, Germany

Received 20 February 2005; received in revised form 18 March 2005; accepted 7 April 2005

Received by M. Di Giulio

Abstract

Spermadhesins belong to a novel family of secretory proteins of the male genital tract. They are major proteins of the seminal plasma and have been found peripherally associated to the sperm surface. So far, they have only been detected in ungulates, specifically in pig, cattle, and horse, respectively. Spermadhesins form a subgroup of the superfamily of proteins with a CUB-domain that has been found in a variety of developmentally regulated proteins. The structure and function of the spermadhesins have been investigated in the pig. They are multifunctional proteins showing a range of ligand-binding abilities, e.g. to carbohydrates, phospholipids, and protease inhibitors, suggesting that they may be involved in different steps of fertilization. We report here the genomic organization of the porcine spermadhesin gene cluster as well as a detailed comparative analysis with respect to other mammalian species. The porcine spermadhesin genes are located on SSC 14q28–q29 in a region syntenic to HSA 10q26. The pig contains five closely linked spermadhesin genes, whereas only two spermadhesin genes are present in the cattle genome. Inactive copies of spermadhesin genes are still detectable in the human, chimp, and dog genome while the corresponding region was lost from the rodent genomes of mouse and rat. Within the pig, the five spermadhesin genes contain both highly diverged and highly conserved regions. Interestingly, the pattern of divergence does not correlate with the position of the exons. Evolutionary analyses suggest that the pattern of diversity is shaped by ancestral variation, recombination, and new mutations.

Keywords: Spermadhesin; Evolution; Reproduction; Gene amplification; Male-biased genes

1. Introduction

Spermadhesins are a group of 12–16 kDa polypeptides found in the seminal plasma and peripherally associated with the sperm surface of pig, bull, and horse (Reinert et al., 1996). They are multifunctional proteins exhibiting a palette of ligand-binding affinities to oligosaccharides and sulfated polysaccharides such as heparin and also to the serine protease inhibitor and to phospholipids (reviewed in Töpfer-Petersen et al., 1998). The amino acid sequence of these proteins does not show any discernible similarity with known carbohydrate-recognition domains (CRD), indicating that spermadhesins may belong to a novel group of animal lectins. Spermadhesins form a subgroup of a superfamily of proteins with a single CUB domain of 16 functionally diverse proteins that has been found in a

Abbreviations: aSFP, acidic seminal fluid protein; BAC, bacterial artificial chromosome; CFA, canine chromosome; CRD, carbohydrate-recognition domain; CUB, name of protein domain named after the first three members, in which this domain was found (*c*omplement subcomponents C1s/C1r, *U*egf, *B*mp1); DMBT1, deleted in malignant brain tumor 1; FnII, fibronectin type II; HSA, human chromosome; kb, kilobase; Mb, megabase; MMU, murine chromosome; nt, nucleotide; PAC, P1-derived artificial chromosome; PSP-I, porcine seminal protein I; PSP-II, porcine seminal protein II; RNO, rat chromosome; SPADH1, spermadhesin 1; SPADH2, spermadhesin 2; SPADHP1, spermadhesin pseudogene 1; SPADHP2, spermadhesin pseudogene 2; SSC, pig chromosome.

* Corresponding author. Tel.: +49 511 9538874; fax: +49 511 9538582.
E-mail address: Tosso.Leeb@tiho-hannover.de (T. Leeb).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.04.015

variety of developmentally regulated processes (Bork and Beckmann, 1993). The crystal structures of two members of the spermadhesin family were recently elucidated and allowed for the first time insight into the architecture of a CUB domain (Romero et al., 1996, 1997). The overall structure of the subunit consists of a β-sandwich made up of two sheets, each containing four anti-parallel strands and one parallel β-strand.

In the pig five members of the spermadhesin family were identified, termed AQN-1, AQN-3, AWN, PSP-I, and PSP-II (Sanz et al., 1991, 1992a,b; Kwok et al., 1993). Porcine spermadhesins are synthesized by the epididymis and accessory glands of the male genital tract. Interestingly, porcine AWN is not only expressed in the male epididymis and seminal vesicle but also in the Fallopian tube of the female genital tract (Ekhlasi-Hundrieser et al., 2002). In other species than pigs, the diversity and expression level of spermadhesin proteins is greatly reduced. In cattle, only two spermadhesins termed SPADH1 (aSFP) and SPADH2 (Z13) were identified (Wempe et al., 1992; Tedeschi et al., 2000). In the horse one spermadhesin homolog to boar spermadhesin AWN was reported (Reinert et al., 1996). In animals outside the ungulate order, neither spermadhesin proteins nor spermadhesin-specific transcripts could be detected so far. Only after the proteins had been identified, some spermadhesin cDNAs were cloned but for at least two members (cattle *SPADH2* and horse *AWN*) no cDNA sequences were available up to now.

The analysis of the genome sequences from human, mouse and rat revealed that about 80% of the mammalian protein coding genes are well conserved in a 1:1 orthologous relationship between different mammalian species. The remaining 20% of the mammalian genes are those where gene duplications and gene losses are observed between species. Odorant receptor genes as well as genes with functions in immune defense or reproduction are frequently among these fast evolving genes (IHGSC, 2004). So far, gene births and gene losses were mainly investigated between primates and rodents. We investigate here the genomic organization and evolution of the spermadhesin genes that are active only in ungulate species. To our knowledge this is the first study where the evolution of ungulate-specific genes is investigated.

2. Materials and methods

2.1. Cloning of the porcine spermadhesin gene cluster

Two PCR primer pairs were developed from the publicly available porcine *AQN-1* and *AQN-3* genes (sequences are available on request). These primer pairs were used to screen the TAIGP714 PAC library (http://www.rzpd.de/) and one PAC clone for each of the primers pairs was isolated. Additionally, a ^{32}P-labeled fragment of the porcine *AWN* cDNA was used to screen the RPCI-44 BAC library according to the BACPAC protocols (http://bacpac.chori.org/). Seven positive BAC clones were isolated after this screen. End sequences and insert sizes were determined from all nine spermadhesin clones.

2.2. Sequencing of the porcine spermadhesin gene cluster

The two clones RP44-374O13 and TAIGP714I2397 were selected for sequencing. DNA from the clones was isolated using the Qiagen large construct kit (Qiagen, Hilden, Germany). The purified BAC or PAC DNA was sheared to approximately 2–5 kb using a nebulizer and shotgun subclone plasmid libraries were prepared in the vector pGEM-4Z (Promega, Mannheim, Germany). Shotgun plasmid subclones were sequenced approximately to sixfold coverage of the BAC or PAC clone using either the ThermoSequenase kit (Amersham Biosciences, Freiburg, Germany) and a LICOR 4200 automated sequencer (LICOR Biosciences, Bad Homburg, Germany) or the DYEnamic ET Terminator kit (Amersham Biosciences, Freiburg, Germany) and a MegaBACE 500 capillary sequencer (Amersham Biosciences, Freiburg, Germany). After sequencing a random collection of plasmid subclones, the remaining gaps were closed by a primer walking strategy until both strands were completely sequenced. One polyC-stretch was extremely difficult to sequence and in this region only single stranded coverage could be achieved. In the case of RP44-374O13 a high degree of segmental duplication was observed. Therefore, directed restriction fragments were cloned and sequenced in addition to the random shotgun clones to ensure correct long range assembly. Sequence data were assembled with Sequencher 4.2 (GeneCodes, Ann Arbor, MI, USA). Full-length cDNA sequences of the porcine spermadhesin genes were obtained by combining 3′-EST data with experimentally derived 5′-RLM-RACE sequences as described (Giese et al., 2002).

2.3. Bovine spermadhesin cDNAs

To obtain bovine full-length cDNAs, the genomic sequences of the bovine spermadhesin genes were retrieved from the TraceArchive of bovine WGS sequences. The two bovine spermadhesin genes were assembled from the individual WGS sequences. The *SPADH1* exons were derived by aligning the bovine genomic sequence to a known partial *SPADH1* cDNA (Accession No. M84603). 3′-RACE experiments were performed to determine the previously unknown *SPADH2* bovine cDNA sequence and to annotate the *SPADH2* exon boundaries. The missing 5′-ends of the bovine sequences were taken from the bovine genomic sequences based on the extremely high similarity of the bovine first exons and upstream flanking sequences to the porcine spermadhesin genes.