Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 339 (2004) 49–59

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Molecular evolution of epididymal lipocalin genes localized on mouse chromosome 2

Kichiya Suzuki[a,b,*,1], Jean-Jacques Lareyre[b,c,1], Diego Sánchez[d], Gabriel Gutierrez[e], Yoshihiko Araki[b,c], Robert J. Matusik[a,b,f], Marie-Claire Orgebin-Crist[b,c,f]

[a] *Department of Urologic Surgery, Vanderbilt University School of Medicine, Room A-1302 MCN, 1161 21st Avenue South, Nashville, TN 37232, USA*
[b] *Center for Reproductive Biology Research, Vanderbilt University School of Medicine, 1161 21st Avenue South, Nashville, TN 37232, USA*
[c] *Department of Obstetrics and Gynecology, Vanderbilt University School of Medicine, 1161 21st Avenue South, Nashville, TN 37232, USA*
[d] *Instituto de Biología y Genética Molecular, Departamento de Bioquímica y Biología Molecular y Fisiología, Universidad de Valladolid-CSIC, Valladolid 47005, Spain*
[e] *Departamento de Ciencias Ambientales, Universidad Pablo de Olavide, Seville 41013, Spain*
[f] *Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, 1161 21st Avenue South, Nashville, TN 37232, USA*

Received 8 January 2004; received in revised form 15 May 2004; accepted 1 June 2004
Available online 17 August 2004
Received by W. Martin

Abstract

We previously identified two murine secretory proteins, mE-RABP(Lcn5) and mEP17(Lcn8), belonging to the lipocalin family and specifically expressed in the epididymis. The genes are contiguous and localized on mouse chromosome 2. We now show that five other related lipocalin genes, *Lcn9*, *Lcn10*, *Lcn11*, *Lcn12*, and *Lcn13*, that evolved by in situ tandem duplication are present on the same locus. *Lcn9*, *Lcn10*, *Lcn12*, and *Lcn13* genes, like *Lcn5* and *Lcn8* genes, are specifically expressed in the mouse epididymis. However, each gene has a distinct spatial expression within the epididymis and different regulation. Analysis of the human genome sequence shows the presence of genes encoding lipocalins with genomic organization, chromosomal arrangement, and orientation similar to that of the corresponding murine genes, indicating that the epididymal cluster is evolutionary conserved. A phylogenetic analysis of the new epididymal proteins reveals their spread position in the lipocalin protein family tree. This suggests the preservation of the regulatory sequences, while protein sequences have greatly diverged, reflecting functional diversity and possibly multifunctionality. In terms of the cluster ancestry, epididymal expression possibly appeared in a PGDS-like lipocalin in amniotes, and the duplications generating the cluster occurred at least in the common ancestor of rodents and primates. The presence and conservation of a cluster of five genes encoding epididymal lipocalins, differently regulated and regionalized in the epididymis, strongly suggests that these proteins may play an important role for male fertility.

Keywords: Epididymis; Male reproductive tract; Gene cluster; Gene expression; Gene phylogeny

1. Introduction

Lipocalins are an ancient and functionally diverse family of proteins found in bacteria, protoctists, plants, arthropods, and chordates (Sanchez et al., 2003). The lipocalin family is defined by a highly conserved three-dimensional structure despite low overall sequence similarity. The structure consists of an eight-stranded antiparallel β-barrel forming a cup-shaped ligand binding pocket or calyx. Lipocalins generally bind small hydrophobic ligands such as retinoids, fatty acids, steroids, odorants, and pheromones and interact with cell surface receptors. They function as transport proteins delivering these ligands to specific cells. They have been implicated in the modulation of immune and inflammatory responses, cell proliferation and differentiation, and animal behavior (reviewed in Akerstrom et al., 2000).

Abbreviations: mE-RABP, murine epididymal retinoic acid-binding protein; mEP17, murine epididymal lipocalin 17 kDa; PCR, polymerase chain reaction; PBS, phosphate-buffered saline; RACE, rapid amplification of cDNA ends; Mb, megabases(s).

* Corresponding author. Department of Urologic Surgery, Vanderbilt University School of Medicine, Room A-1302 MCN, 1161 21st Avenue South, Nashville, TN 37232, USA. Tel.: +1-615-322-7484; fax: +1-615-343-7797.
E-mail address: kichiya.suzuki@vanderbilt.edu (K. Suzuki).

[1] These authors contributed equally to the work presented in this manuscript.

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.06.027

One of the members of the lipocalin family is the epididymal retinoic acid binding protein (E-RABP), first identified as one of the major, androgen-dependent proteins secreted by the rat epididymis. An orthologue was reported in the mouse. Both the rat and the mouse E-RABP bind 9-*cis* and *all-trans* retinoic acid, suggesting that E-RABP may be involved in retinoid trafficking and be part of the anti-inflammatory response in the epididymis (reviewed in Orgebin-Crist et al., 2002).

E-RABP is present as a single copy gene in both the rat and the mouse genome (Brooks et al., 1986; Orgebin-Crist et al., 2002). The gene structure is similar, but not identical in both species. The murine gene has seven exons characteristic of several lipocalin family members, but the rat gene has an additional exon. However, the exon/intron boundaries of the first seven exons are identical in both the rat and mouse genes. The fluorescent in situ hybridization (FISH) technique revealed that the mE-RABP gene is localized on the [A3-B] locus of mouse chromosome 2 (Lareyre et al., 1998). A paralogue of mE-RABP, termed mEP17, was found 1.7 kb upstream from the mE-RABP gene (Lareyre et al., 2001). Interestingly, the genes of several other members of the lipocalin family are also found in the same region of chromosome 2 (Chan et al., 1994). Since these genes have a genomic organization similar to that of the mE-RABP gene, it is likely that these genes are evolutionarily related and some may have evolved from the duplication of an ancestral gene within the [A3-B] locus of murine chromosome 2.

The genome sequencing projects are now providing evidence that gene duplication events have contributed significantly to gene family expansion and to genome evolution. The Mouse Genome Sequencing Consortium has identified 25 gene clusters containing mouse genes that are descendents of a single ancestral gene and for which at least one gene persists in the human genome (Waterston et al., 2002). Interestingly, of these 25 clusters, 14 contain genes involved in reproduction and five contain genes involved in immune surveillance.

In this study, we have identified an epididymis-specific cluster of at least six lipocalin genes on mouse chromosome 2 and human chromosome 9 with similar gene structure, sequence, chromosomal order, and orientation, suggesting that functional constraints maintained this cluster through 75 million years (Myr) of evolutionary distance. We have found that each member of the murine gene cluster has a different pattern of gene expression within the epididymis. Finally, we have used protein sequence and gene structure to perform a phylogenetic analysis of the epididymal lipocalin cluster and determine their place in the evolutionary history of lipocalins.

The Human Gene Nomenclature Committee (HGNC) has adopted the symbol Lcn for all lipocalins. E-RABP and EP17 have been assigned the symbols Lcn5 and Lcn8, respectively. To maintain continuity, they will be referred to as E-RABP(Lcn5) and EP17(Lcn8) in this paper. The new lipocalins in the cluster will be referred to by the Lcn symbols.

2. Materials and methods

2.1. Animals

Male B6D2F1 or ICR mice were purchased from Harlan Sprague–Dawley, (Indianapolis, IN). All experiments were conducted in accordance with the National Institutes of Health Guidelines for Care and Use of Animals in the Laboratory. When required, bilateral castration, unilateral castration or hormone replacement was performed as described previously (Suzuki et al., 2003).

2.2. Rapid amplification of cDNA ends (RACE)

Epididymis of B6D2F1 mice were isolated, frozen in liquid nitrogen and stored at −80 °C until used. Total RNA was extracted using CsCl cushion method. First strand cDNA synthesis was performed using 1 μg of caput epididymidis total RNA to prepare 5′-RACE-Ready cDNA and 3′ -RACE Ready cDNA (primers and reverse transcriptase were supplied from Smart RACE cDNA Amplification Kit, Clontech Laboratories, Palo Alto, CA). Ten microliters of cDNA were diluted with 100 μl H_20 (11-fold dilution), and then 2.5 μl of the cDNA was used in a primary PCR reaction with the RACE linker primer and gene-specific primers (for mLcn9 5′-RACE 5′-ATGGTACTACTAC-TAGTCCTGGGCC-3′, 3′-RACE 5′-GAACCAAAG-GAGGAACTTTATTTGG-3′; for mLcn10 5′-RACE, 5′-AGGCTCTTGTAGTTGCTGCC-3′, 3′-RACE, 5′-GCTA-GAGACAAGAGGAAGCTGG-3′; for mLcn11 5′-RACE, 5′-ATCCGAGCTCATCTGTGGAC-3′; for mLcn12 5′-RACE 5′-TCCTGGTCAGGCAGATGAA-3′, 3′-RACE, 5′-TTCAGGTGACCAACTCCATG-3′; mLcn13 5′RACE, 5′-GTAGTGCTCATTCACCGACG-3′, 3′ RACE 5′ -GGAA-GAAGGAGACTTGGAGACC-3′) using Advantage 2 PCR kit (Clontech) with the following parameters: 15 s at 94 °C, 30 s at 55 °C and 30 s at 72 °C for 25 cycles; then 5 min at 72 °C. The resulting PCR products were cloned into pGEM-T Easy vector (Promega, Madison, WI) for subsequent sequence determination. Sequencing was performed using an ABI automated sequencer Model 3700 (Applied Biosystems, Foster City, CA). Nucleotide sequences were analyzed using the Vector NTI (InforMax, Bethesda, MD). K-Estimator (Comeron, 1999) was used to estimate the values of nucleotide substitutions per site (synonymous [K_s] and nonsynonymous [K_a] for coding regions) between orthologous or paralogous genes based on the sequence alignments.

2.3. Northern blotting

Total RNA samples from various tissues were extracted from ICR mice using RNeasy Midi Kit (Qiagen, Valencia, CA). A total of 10 μg of total RNA was separated by 1% agarose gel and blotted to Hybond N^+ (Amersham Biosciences, Piscataway, NJ) by capillary method. The cDNA frag-

ments were cloned from epididymal total RNA using gene specific primers (for mLcn9, ATGGTACTACTAC-TAGTCCTGGGCC and 5′-TTGTCAAGCAGTATG-GAGCG-3′; for mLcn10, 5′-ATGCCTCGACAG TTTTCAGG-3′ and 5′-TCAGGCTCTCTGCAGCCAGC-3′; for mLcn11, RACE linker primer and 5′-ATCCGAGCT-CATCTGTGGAC-3′; for mLcn12, RACE linker primer and 5′-TCCTGGTCAGGCAGATGAA-3′; for mLcn13, RACE linker primer and 5′-GTAGTGCTCATTCACCGACG-3′; for mPGDS, 5′-ACCTGCTCTGCTCTGAGCAAATG-3′ and 5′-CTTTTATTTCTGAGTGACAGAGCAAAGGA-3′; for mLcn2, 5′-ATGGCCCTGAGTGTCATGTG-3′ and 5′-GTTGTCAATGCATTGGTCGG-3′). These cDNA fragments were labeled with [^{32}P]dCTP using rediprimeII (Amersham Biosciences). Hybridization and serial washing were performed as previously described (Suzuki et al., 2003).

2.4. In situ hybridization

Tissues were fixed and sectioned at 7 μm to ensure that all epididymal segments were present on the sections. The mLcn9 and mLcn10 cDNA fragments were cloned by RT-PCR as described above. RNA probe preparation, hybridization, and serial washing were performed as described previously (Suzuki et al., 2003).

3. Results and discussion

3.1. Identification of 5 novel genes encoding lipocalin on the mouse chromosome 2

We have previously isolated and described the *mEP17(Lcn8)* and *mE-RABP(Lcn5)* genes encoding two lipocalins specifically expressed in the mouse epididymis. The FISH technique demonstrated the chromosomal localization of these genes on the [A3-B] locus of mouse chromosome 2 (Lareyre et al., 1998). Using the mouse genome draft from the Celera Discovery System™ (Celera Genomics, Rockville, MD) and National Center for Biotechnology Information (NCBI, Bethesda, MD), we confirmed and restricted the chromosomal localization of these genes at 25.9 Mb from centromere on the [A3] locus of mouse chromosome 2 (Fig. 1A). Because this locus was suspected to be a hot spot for lipocalin genes, the flanking regions of the *mE-RABP(Lcn5)* gene were analyzed for new genes. The existence of one new lipocalin gene was predicted 164 kb upstream, and four other genes were found 25, 42, 119, and 165 kb downstream from the *mE-RABP(Lcn5)* gene transcription initiation site (Fig. 1A). These 5 new putative genes, *mLcn9*, *mLcn10*, *mLcn11*, *mLcn12*, and *mLcn13* were referred as mCG18728/9230102I19Rik, mCG18740/LOC332578, mCG18743/LOC227630, mCG18730/9230102M18Rik, and mGC131106/MGC41397, respectively in the Celera and NCBI database. The existence of *mLcn9*, *mLcn10*, and *mLcn12* was also supported by the isolation of expressed sequence tags (EST) from a mouse adult epididymis (AK020305; AK020338; and AK020307) and pregnant mouse uterus BC027556).

3.2. Analysis of the novel lipocalin genes

Based on these data, the molecular cloning of the full-length cDNA of the novel genes was carried out by RACE from the mouse epididymal total RNA to determine and confirm their genomic organization. Sequence data of the 5′-RACE and 3′-RACE products were integrated and computer analyzed. The structure of each gene was determined by comparison of the cDNA and genomic sequences provided by Celera and NCBI. A full-length cDNA corresponding to *mLcn9* was isolated and found identical to the AK020305 EST clone. The *mLcn9* gene length was 2.4 kb and had seven exons (Fig. 1B). A full-length cDNA corresponding to mLcn10 was cloned allowing identification of the exon 1 that was lacking in the AK020338 EST clone. The gene length was 3.4 kb and had six exons (Fig. 1B). A full-length cDNA corresponding to *mLcn13* was isolated. It contained an additional 36 bp upstream of exon 1 when compared to that of the BC027556 EST clone. The *mLcn13* gene length was 3.2 kb and had seven exons (Fig. 1B). A full-length cDNA corresponding to *mLcn12* was isolated and exons 2–6 were identical to EST clone AK20307. The gene length was 3.1 kb and had six exons (Fig. 1B). A 5′-RACE cDNA fragment corresponding to *mLcn11* was isolated and its predicted gene organization was found to contain six exons. All exon/intron boundaries on each gene were consistent with the consensus splice junction sequences and were similar to that of the *mE-RABP(Lcn5)* and *mEP17(Lcn8)* genes (Lareyre et al., 2001; Orgebin-Crist et al., 2002). The intron phase was also conserved between these six genes. The size of exons 2, 4, and 5 that encode conserved lipocalin motifs was similar to that of the corresponding exons found on the *mE-RABP(Lcn5)* and *mEP17(Lcn8)* genes (Fig. 1B). In addition, a potential androgen receptor binding site was located 329 nucleotide upstream from the translation initiation codon of the *mLcn12* gene, and 216 nucleotides upstream from that of the *mLcn10* gene (Fig. 1B).

3.3. Analysis of the predicted novel lipocalin proteins

The open reading frame of the *mLcn12*, *mLcn10*, *mLcn13*, *mLcn11*, and *mLcn9* genes encoded 193 amino acids (aa), 182 aa, 176aa, 170 aa, and 178 aa, respectively (Fig. 1C), and the predicted molecular mass of the precursors was 21,913, 20,651, 19,995, 19,440, and 20,501, respectively. All lipocalins exhibit a highly hydrophobic transmembrane domain at the N-terminal end that is typical of a signal peptide (Fig. 1C). The weight matrix method and neural networks algorithms confirmed the presence of a putative signal peptide by identifying a cleavage site between amino acids 19–20, 19–20, 19–20, 16–17, and 16–17 for the

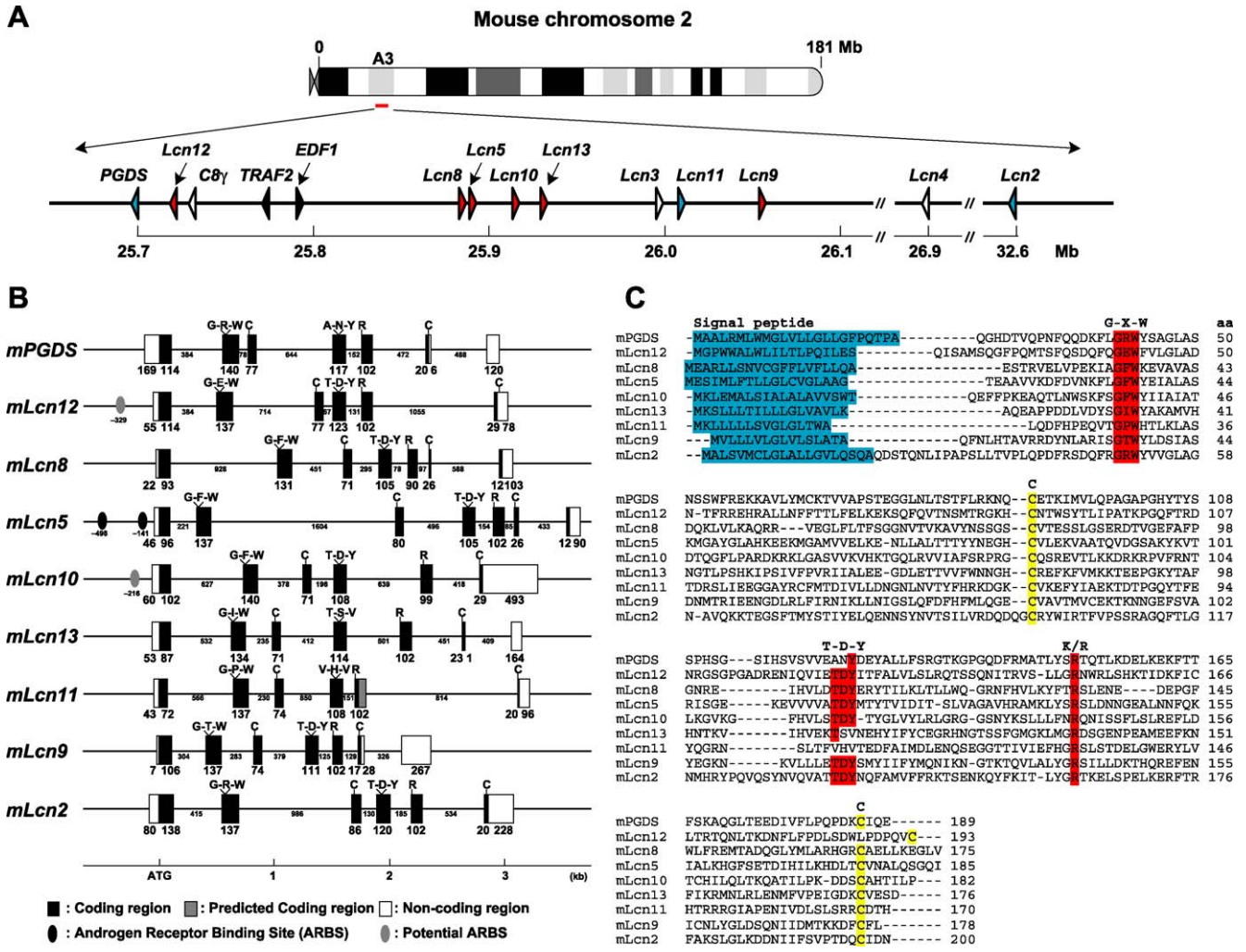

Fig. 1. (A) Location and orientation of the lipocalin genes on mouse chromosome 2. Distances from centromere are shown in Mb based on the NCBI. *Mus musculus* genome released on January 27, 2003. Color codes and gene symbols are as follows: red triangles, epididymal lipocalin genes; blue triangles, lipocalin genes expressed in the epididymis and other tissues; white triangles, lipocalin genes expressed in other tissues; black triangles, other genes. PGDS: prostaglandin D_2 synthase; C8γ: complement 8 γ; TRAF2, TNF receptor associated factor 2; EDF1, endothelial differentiation-related factor 1; Lcn3, vomeronasal secretory protein I precursor (VNSPI); Lcn4, vomeronasal secretory protein II precursor (VNSPII). (B) Gene organization of lipocalin genes. Closed boxes represent coding regions and open boxes show noncoding regions. Gray boxes are predicted coding region based on the NCBI database. Note that structurally important lipocalin motifs located on exons 2–5 [G-X-W, T-D-Y, K/R and cystein residues (C)] are highly conserved. (C) Multiple sequence alignment of epididymal lipocalin genes. The predicted amino acid sequences were aligned. Color codes are as follows: blue indicates putative signal peptide; red indicates lipocalin motifs (G-X-W, T-D-Y, and K/R); yellow indicates cystein residue. Numbers of the amino acid residues are shown on the right.

mLcn12, mLcn10, mLcn13, mLcn11, and mLcn9, respectively (Fig. 1C). Further, this is consistent with lipocalin family members being secreted proteins. The novel genes clearly encoded lipocalins harboring the conserved G-X-W in exon 2, T-D-Y in exon 4 (except mLcn11 and mLcn13) and K/R in exon 5 as well as cysteine residues in exons 3 and 5 that are involved in the formation of a disulfide bond (Fig. 1B and C). When comparing to other known proteins, the highest match for mLcn12 was mouse Lcn2 (20% identity). mLcn13 shows the highest similarity (38% identity) to mLcn3 (VNSPI), which is localized 65 kb downstream from the mLcn13 gene. Similarly, mLcn9 has the highest similarity (38% identity) with mouse major urinary protein IV, which is localized on mouse chromosome 4. Otherwise, all the novel lipocalins share lower similarity (20–30% identities) with each other. This is typical in the lipocalin family when overall sequence similarity is low but shared sequence motifs and three dimensional structure similarity is high (Akerstrom et al., 2000).

3.4. Tissue-specific expression, hormonal regulation, and ontogenic expression of the epididymal lipocalin genes

Northern blot analyses were carried out to determine the tissue-specific expression of the novel genes (Fig. 2A). The *mLcn9*, *mLcn10*, *mLcn12*, and *mLcn13* genes expressed approximately 1 kb transcript only in the epididymis. No gene expression was detectable, even after

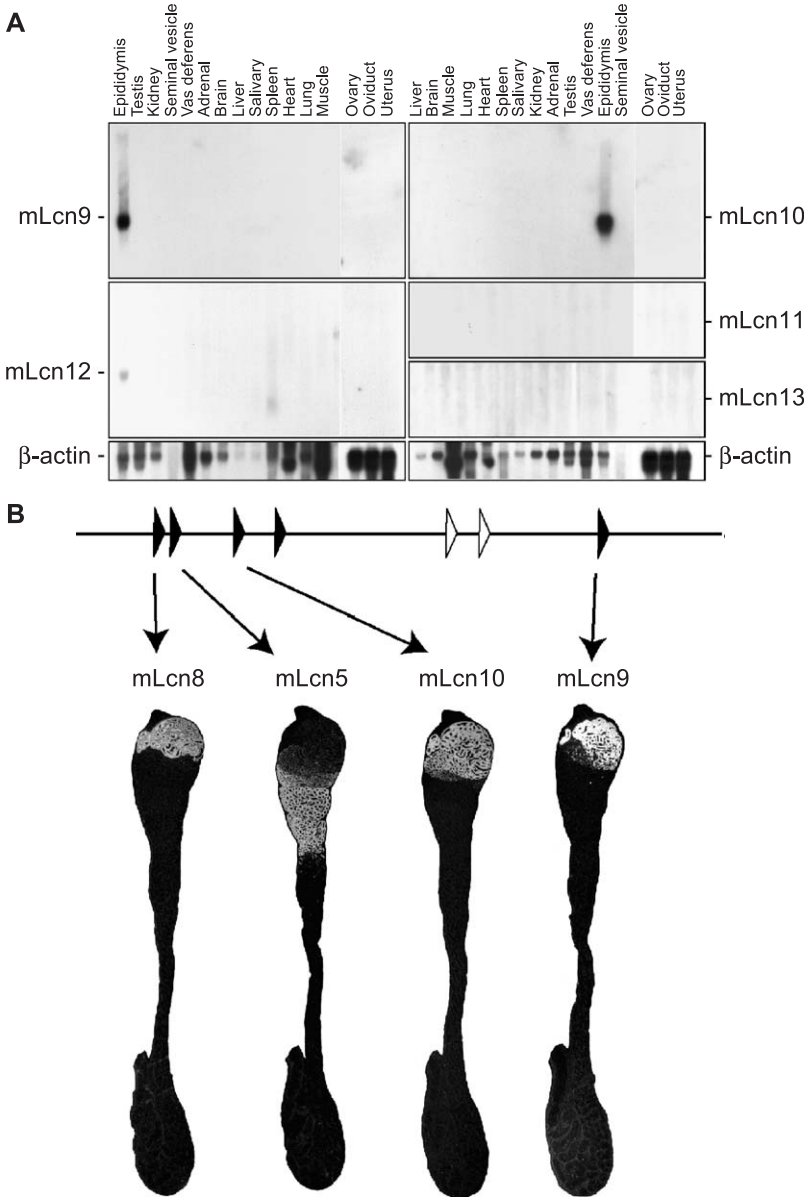

Fig. 2. (A) Epididymis-specific expression of the epididymal lipocalin genes. Ten micrograms of total RNA were extracted from adult male and female tissues (8–10 weeks old of age) and hybridized with the [^{32}P]radiolabeled cDNA probes corresponding to each gene. Mouse β-actin cDNA probes were used as a positive control to check RNA loadings. Membranes were exposed to the X-ray films for over night. (B) Region specificity of the epididymal lipocalin gene expression. Serial sections of the epididymis were hybridized with [^{35}S]labeled antisense RNA probes for the *mLcn9* and *mLcn10* genes. The *mLcn5* and *mLcn8* RNA probes were also used to compare region specific gene expression. Note that all epididymal lipocalin genes demonstrate different expression pattern. Sections were exposed to emulsion for 1 week.

long-term exposure, in the other male or female mouse tissues tested. The mLcn11 gene was not detectable in any tissue by Northern blot even after 10 days exposure, but were detected in the epididymis, testis, vas deferens, and prostate among other tissues by nested RT-PCR, but not in kidney, lung, brain, and small intestine (data not shown). The *mLcn9* and *mLcn13*, like *mEP17(Lcn8)*, gene expression disappeared after castration, but unilateral castration, or testosterone supplementation for 1 week failed to restore gene expression (Fig. 3A). The *mLcn10* gene expression also disappeared after castration, but unilateral castration only partially maintained, and testosterone administration only partially restored gene expression (Fig. 3A). The *mLcn12*, *mLcn2* and *mPGDS* gene expression, like *mE-RABP(Lcn5)*, decreased after castration; however, testosterone supplementation restored its expression completely (Fig. 3A).

During postnatal development, gene expression of *mEP17(Lcn8)*, *mLcn9*, *mLcn10* and *mLcn13* was not detectable before 3 weeks (a period corresponding to the absence of testicular fluid and spermatozoa in the lumen of the epididymal duct and by low levels of circulating androgens)

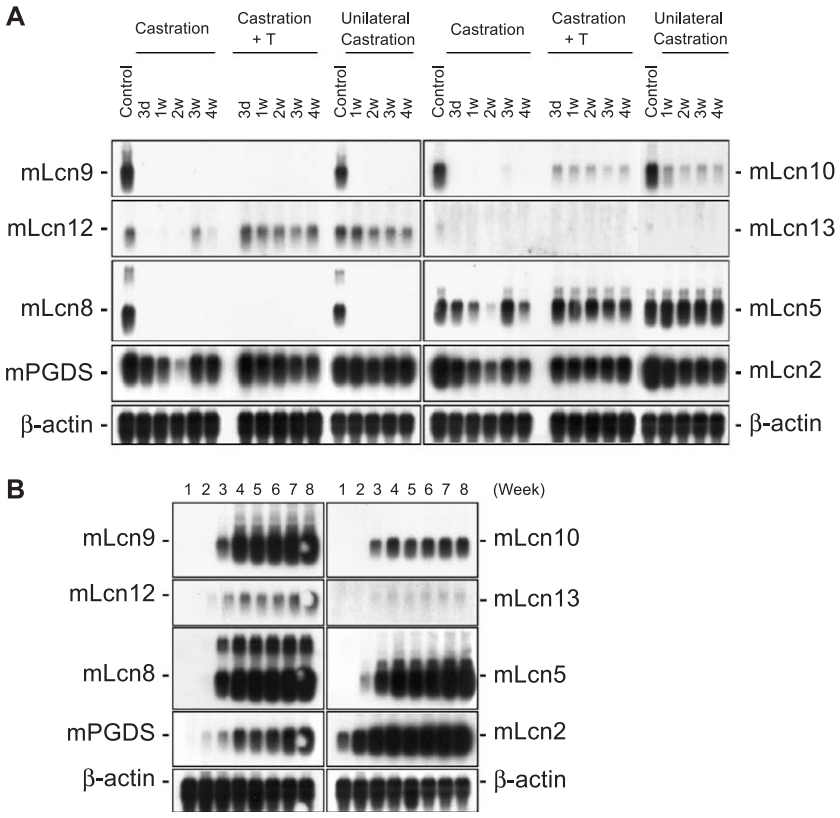

Fig. 3. Hormonal regulation and ontogenic expression of the epididymal lipocalin genes. (A) Hormonal regulation of the epididymal lipocalin genes. Three adult male mice (8–10 weeks old of age) were prepared for each time point of treatments. Treatments are as follows: for bilateral castration treatment (Castration), adult mice were castrated via scrotum route and sacrificed after 3 days (3 days), 1, 2, 3, or 4 weeks (1–4 weeks); for bilateral castration plus testosterone supplementation treatment (Castration + T), bilaterally castrated adult mice (3 days–4 weeks) were supplemented with testosterone propionate for 1 week (2 μg/g body weight/day); for unilateral castration, testis on one side was removed and mice were maintained for 1–4 weeks. Intact side epididymis was used as a control. (B) Ontogenic expression of the epididymal lipocalin genes. Total RNA (10 μg) was isolated from the epididymis of male mice at 1, 2, 3, 4, 5, 6, 7, and 8 weeks of age.

(Jean-Faucher et al., 1985) and increased from 3 to 4 weeks of age to reach a plateau (Fig. 3B). In contrast, *mE-RABP(Lcn5)*, *mLcn12*, *mPGDS* and *mLcn2* expression was detected before 3 weeks and increased thereafter (Fig. 3B).

In summary, our results demonstrate that *mLcn9 and mLcn13*, like the *mEP17(Lcn8)* gene, are strictly dependent on testicular factors circulating within the luminal fluid and that the *mLcn12* gene, like *mE-RABP(Lcn5)* is dependent on androgens circulating in the serum, whereas *mLcn10* gene expression is regulated both by circulating androgens and by testicular factors supplied from the luminal fluid.

Interestingly, this epididymis-specific lipocalin cluster is flanked by other members of the lipocalin family, such as *Lcn11*, *PGDS*, and *Lcn2* (Chan et al., 1994), that are expressed in other tissues besides the epididymis (Sorrentino et al., 1998; Chu et al., 2000), while the genes of the epididymal cluster are expressed only in the epididymis. This suggests that the epididymis-specific lipocalin cluster may contain a locus control region that enhances lipocalin gene expression only in the epididymis, while genes flanking either side of this epididymal gene cluster show expression in other tissues besides the epididymis.

3.5. Lipocalins encoded by the epididymal lipocalin gene cluster are differentially expressed within the epididymis

In situ hybridization technique using serial sections (Fig. 2B) was used to demonstrate the spatial expression pattern of the *mLcn9* and *mLcn10* genes within the epididymis. Expression of the *mLcn9* and *mLcn10* genes was detected in the epithelium lining the initial segment (*mLcn9* and *mLcn10*) and segment 2 (*mLcn10*) of the mouse epididymis. Such a highly restricted expression pattern also has been observed for two other members of the epididymis-specific lipocalin gene cluster, *mEP17(Lcn8)* and *mE-RABP(Lcn5)*. *mEP17(Lcn8)* gene expression is restricted to the initial segment of the caput epididymidis, although a checkerboard expression pattern has been observed at the boundary between the initial segment and segment 2 (Lareyre et al., 2001). The *mE-RABP(Lcn5)* gene is expressed with an increasing gradient from segments 2 to 5 of the mid and distal caput epididymidis (Orgebin-Crist et al., 2002), whereas the rat orthologue gene termed *ESPI* is expressed in the whole caput (Garrett et al., 1991). In contrast, the

expression of the other lipocalin genes present on mouse chromosome 2 is not restricted to the caput epididymidis. *mPGDS* is expressed with an increasing gradient from the distal caput to the cauda epididymidis (Fouchecourt et al., 2002), and *mLcn2* is expressed in the caput and corpus epididymidis (Chu et al., 2000). One could hypothesize that the clustered organization of the *mEP17(Lcn8)*, *mE-RABP(Lcn5)*, *mLcn9* and *mLcn10* genes may be required for tissue and region-specific expression. Indeed, it has been shown that 5 kb of the 5′ flanking region of the *mE-RABP(Lcn5)* gene and 5.3 kb of the 5′ flanking region of the *mEP17(Lcn8)* gene contain all the information to target foreign gene expression to either the distal caput epididymidis (*Lcn5*) or the initial segment (*Lcn8*) in transgenic mice (Lareyre et al., 1999; Suzuki et al., 2003). Further studies will be required to identify DNA regions involved in the regionalization of the expression of the epididymal lipocalin gene cluster. Taken together, our data demonstrate that, within the lipocalin gene cluster on mouse chromosome 2, there is a cluster of 6 lipocalin genes expressed specifically in the epididymis [*mE-RABP(Lcn5)*, *mEP17(Lcn8)*, *mLcn9*, *mLcn10*, *mLcn12*, and *mLcn13*] and within the epididymis their expression is highly regionalized.

3.6. The epididymal lipocalin gene cluster is evolutionarily conserved in mouse and human

A computational analysis of the human genome was carried out using the Celera and NCBI database to identify human orthologous genes corresponding to mouse epididymal lipocalin genes. Five putative genes [*E-RABP(LCN5)*, *EP17(LCN8)*, *LCN9*, *LCN10*, and *LCN12*] encoding homologous lipocalins (Fig. 4) were identified. These genes were localized on human chromosome 9q34 (Fig. 4). The genomic organization, chromosomal arrangement, and orientation of these human genes were found similar to those of the murine genes. The human counterparts of mouse Lcn12, Lcn8, Lcn10, and Lcn9 shared 79%, 75%, 76%, and 68% overall similarity, respectively. Notably, the E-RABP(LCN5) lipocalin was globally less conserved (46% similarity) in human. However, the presence of two potential androgen receptor binding sites, similar to that described previously in the proximal promoter of the *mE-RABP(Lcn5)* gene (Lareyre et al., 2000), was observed in the 5′-flanking region of the human E-RABP(LCN5) gene, suggesting that the hormonal regulatory mechanisms of *E-RABP(Lcn5)* gene expression may be conserved from mouse to human.

Gene evolution can be assessed by using the rate of nonsynonymous nucleotide substitutions per nonsynonymous site (K_a) and the rate of synonymous nucleotide substitutions per synonymous site (K_s) (Li, 1997). A K_a/K_s ratio higher than 1 is indicative of positive selection. A K_a/K_s ratio lower than 1 is indicative of purifying selection. When the sequences of the coding exons of *Lcn2*, *E-RABP(Lcn5)*, *EP17(Lcn8)*, *Lcn9*, *Lcn10*, *Lcn12* and *PGDS* were aligned, the mouse and human orthologous genes have an average K_a/K_s ratio of 0.48 ± 0.04 indicative of purifying selection. Considering paralogues within either the mouse or human gene cluster, the K_a/K_s ratio of 0.97 ± 0.12 and 0.97 ± 0.09, respectively, close to the K_a/K_s ratio of 1, is indicative of neutral selection. However, the murine paralogues *Lcn2–Lcn8*, *Lcn8–Lcn12* and *Lcn9–Lcn12* have K_a/K_s ratios of 1.68, 2.23, and 1.92, respectively, while the human paralogues *LCN5–LCN9* have a K_a/K_s ratio of 2.53. These higher K_a/K_s ratios may reflect some positive selection.

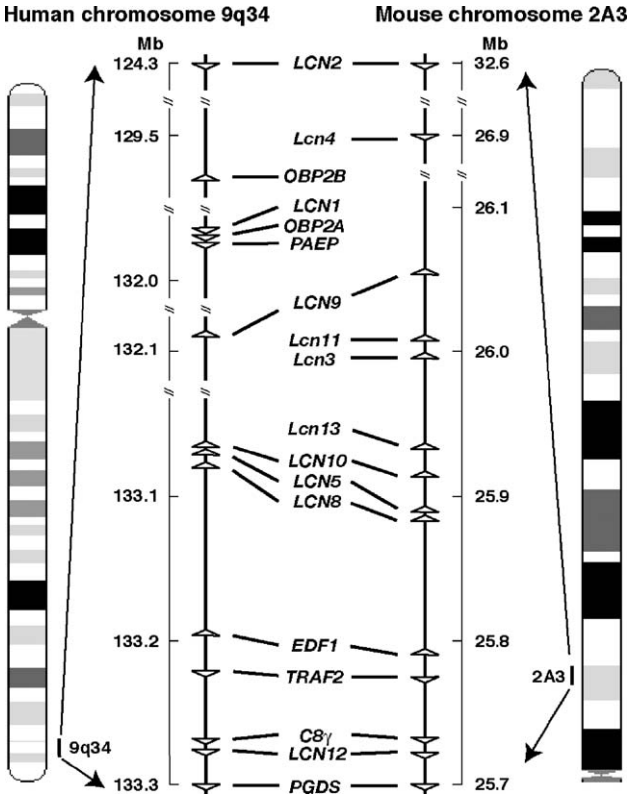

Fig. 4. Syntenic alignment of human chromosome 9 and mouse chromosome 2. Left vertical bar represents assembled human genomic sequence contig based on NCBI *Homo sapiens* genome released on April 10, 2003. Gene locations are shown in Mb. Triangles indicate gene orientation. Symbols are as follows: LCN1, tear lipocalin; OBP2A and B, odorant-binding protein 2A and B; PAEP, progesterone-associated endometrial protein; EDF1, endothelial differentiation-related factor 1, TRAF2, TNF receptor-associated factor 2; C8γ, complement 8 γ; PGDS, prostaglandin D_2 synthase.

Our work has identified orthologous genes of the mouse epididymal gene cluster on human chromosome 9q34, which suggests that lipocalins are expressed in human epididymis. The remarkable conservation observed in the murine and human epididymal lipocalin cluster implies that functional constraints have maintained its integrity during evolution. The value of the K_a/K_s ratio of orthologous genes of the epididymal lipocalin cluster reflects this purifying selection.

3.7. The epididymal lipocalin cluster was generated by gene duplication

The genomic clustering, sequence conservation, gene orientation, splicing junction, and codon phases of the six epididymis-specific lipocalin genes within the 400-kb distance localized between 25.7 and 26.1 Mb on mouse chromosome 2 suggest that these lipocalins have evolved by gene duplication (Fig. 1A). Gene duplication is believed to play an important role in biological evolution (Ohno, 1970), but the mechanisms underlying the passage from a state of complete redundancy to one of functional diversity are still under debate. However, the co-optation of a new function (neofunctionalization) and the subfunctionalization of regulatory or protein coding regions could explain the maintenance of duplicate genes (Lynch and Force, 2000). The lipocalin cluster on mouse chromosome 2 shows the hallmark of subfunctionalization: *mPGDS*, *mLcn2*, and *mLcn11* are expressed in epididymis, testis, and other tissues, but expression of *mE-RABP(Lcn5)*, *mEP17(Lcn8)*, *mLcn9*, *mLcn10*, *mLcn12* and *mLcn13* is restricted only to the epididymis. Furthermore, members of the epididymis-specific cluster display a striking region-specific expression within the epididymis. In addition, some lipocalins function exclusively as lipophylic ligand-binding proteins, but other lipocalins such as PGDS are multifunctional with both ligand binding and enzymatic activities (Urade and Hayaishi, 2000). Therefore, it is likely that from a common ancestral gene expressed in epididymis and other tissues the epididymal lipocalin cluster on mouse chromosome 2 has evolved by subfunctionalization to include new gene functions and expression patterns. It is also likely that there is some redundancy in the epididymal lipocalin cluster and that functional analysis will require contiguous gene deletion. The subfunctionalization process could be inferred from the higher K_a/K_s ratio observed in some paralogous mouse and human epididymal lipocalins (see above). However, the pattern of positive selection expected in genes underlying subfunctionalization might be obscured by 30–50 Myr of natural selection operating over these sequences (Hughes, 1999), which are surpassed by the 75 Myr of divergence experienced by primates and rodents.

The epididymal lipocalin cluster is one of a number of gene clusters found in the murine genome (Waterston et al., 2002). It has been estimated that 20% of the mouse genome comprises gene families that have undergone expansion by duplication and the percentage may be higher since the procedure used by the genome sequence projects may underestimate the extent of gene duplication (Bailey et al., 2002). A high proportion of clusters identified in the mouse and human genomes contain genes involved in reproduction and immune surveillance, respectively. These clusters include cystatins, defensins, and other members of the lipocalin family (i.e., odorant binding proteins). Interestingly, a cluster of five novel β-defensin genes, present on human chromosome 20p13 (Rodriguez-Jimenez et al., 2003), and a cluster of three cystatin-related genes is present on human chromosome 20p11 and mouse chromosome 2 (Hsia and Cornwall, 2003) encode proteins highly expressed in the epididymis. Both the β-defensin and cystatin related proteins display a highly regionalized expression pattern. Therefore, the clustered organization of the genes encoding epididymal lipocalins may not be unique and will be a model of choice to dissect the molecular mechanisms that control the spatial and temporal expression of a large number of related genes.

3.8. Phylogeny and evolution of the epididymal lipocalin gene cluster

The relationships of the epididymal lipocalins studied in this work with other family members were explored by building a phylogenetic tree based on amino acid sequence alignments. In order to decrease the tree-building computational time, we appended the new mouse and human Lcn9-13 to a monophyletic set of chordate lipocalins (Clades IV–XIV; Ganfornina et al., 2000) that are paralogous according to phylogenetic analyses using protein sequence and gene structure, and to their chromosomal location in several organisms (Salier, 2000). We thus aligned 113 chordate lipocalins including the new epididymal lipocalins, and reconstructed a phylogenetic tree using the maximum likelihood-based method as previously described (Ganfornina et al., 2000). This tree is shown in Fig. 5.

The epididymal lipocalins Lcn8 and Lcn10 group with E-RABP(Lcn5) in a strongly supported clade. In spite of the unresolved position of human LCN5 in the E-RABP(Lcn5) clade, the supported monophyletic relationship of each Lcn5, Lcn8, and Lcn10 in rodents and humans suggests a tandem gene duplication of epididymal lipocalins in their common ancestor. This agrees with the contiguous chromosomal location of the three lipocalins in both the human and mouse genomes (Fig. 4).

An intriguing result of the phylogeny is the relationship of Lcn9 and Lcn11 with lipocalins implicated in chemoreception of varied ligands and other lipocalins of unknown function. Similarly, Lcn13 groups with a different clade of chemoreception lipocalins. Because of the genomic clustering of Lcn9, Lcn11, and Lcn13 to other epididymal lipocalins, and the expression in the epididymis, their phylogenetic relationship to chemoreception lipocalins suggests the direct inheritance of the regulatory regions involved in epididymal expression.

Finally, our phylogeny relates Lcn12 to the PGDS-Lcn2 clade. This result, and their shared epididymal expression, suggests that Lcn12 represents a Lcn2 gene duplicated at least in the common ancestor of primates and rodents.

The gene structure (see Fig. 1B) also supports the evolutionary relationship of the epididymal lipocalins, since they all bear six homologous exons in their CDS, while the related E-RABP(Lcn5) and EP17(Lcn8) show an additional C-terminal exon. These results support the hypothesis that

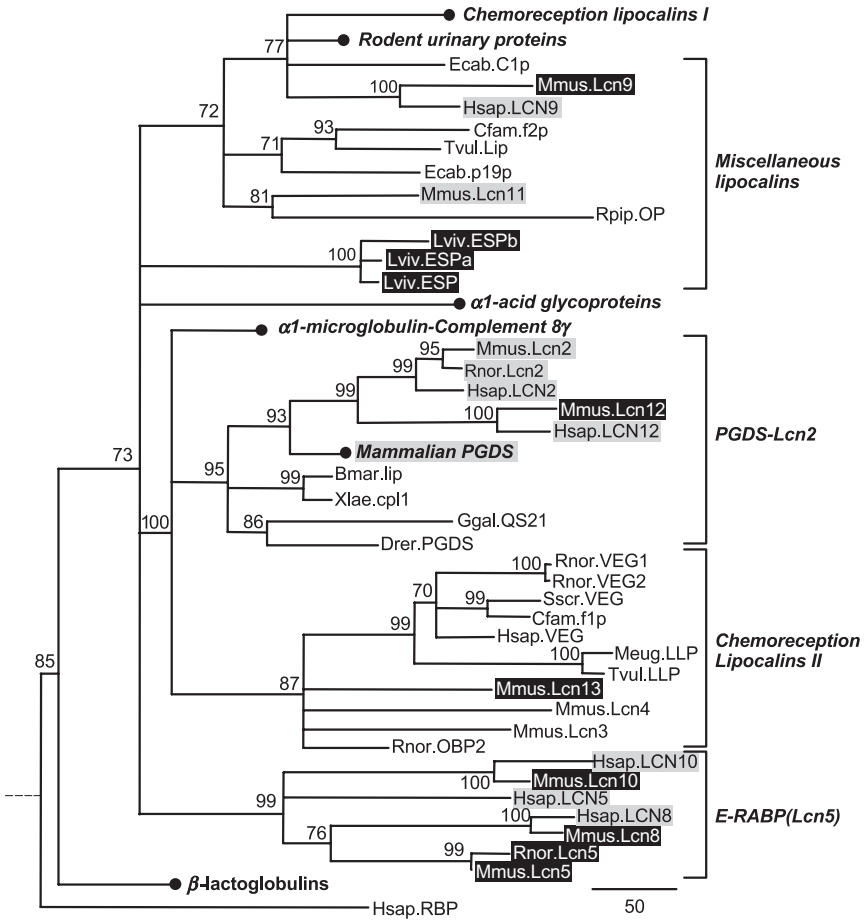

Fig. 5. Phylogenetic tree of chordate lipocalins rooted with the human RBP sequence as an outgroup. Proteins are named using an abbreviated species name followed by a functional label (Ganfornina et al., 2000). LBP values >70 are indicated at each node. The rest of the nodes were represented as polytomies. The lipocalin clades not containing epididymal lipocalin sequences are resumed to their main clade node (shown as a black dot). The lipocalins demonstrated to be epididymis-specific are marked by black boxes. Other lipocalins expressed in epididymis and other tissues or with unknown tissue expression (as the human genes) are marked by gray boxes. The scale bar represents branch length (amino acid substitutions/100 residues).

the presence of the sixth intron in the CDS of some chordate lipocalins reflects a high genomic plasticity in the C-termini (Sanchez et al., 2003). This also agrees with the presence of both 6-exon and 7-exon lipocalins in the well-supported monophyletic clade of functionally related epididymal Lcn5, Lcn8, and Lcn10.

A final question is the character of the ancestral lipocalin that gave rise to the duplicate genes we report here. Our study suggests that this lipocalin was present at least in the common ancestor of primates and rodents, but the polytomies present in our tree prevent an inference of ancestry for a particular lipocalin clade. However, the organismal representation of the epididymal lipocalins can shed light on this issue. We present evidence for the identification of the cluster in humans and mice. Also, lipocalins related to mE-RABP (Lcn5) are present in marsupial (Lamont et al., 1998), rat, rabbit, and boar epididymis (reviewed in Orgebin-Crist et al., 2002). More recently, a cDNA encoding the monkey orthologue of mEP17(Lcn8) has been identified (Liu et al., 2001). Lcn2 has been found in mouse epididymis (Chu et al., 2000), and PGDS is present in zebrafish and Fugu (Lareyre, unpublished data). In lizards, the monophyletic relationship of three epididymal lipocalins (Lviv.ESPs; Morel et al., 2000) (Fig. 5) suggests that the current array of lizard epididymal lipocalins arose by duplications in the reptilian lineage.

The existence of Lviv.ESP proteins suggests the presence of epididymal lipocalins in the common ancestor of amniotes in the late Carboniferous. In relation to marine chordates, teleosts lack an epididymis and cartilaginous fishes have a primitive epididymis (Jones and Jones, 1982). However, the presence of PGDS so far found in the former, as well as in many other chordate classes, suggests that a PGDS-like protein could be the ancestral protein that became expressed in the epididymis. Whether this expression appeared in marine or in terrestrial chordates remains to be elucidated. Whatever the case may be, the PGDS-like ancestral protein experienced rounds of duplications that gave rise to the cluster of epididymal lipocalins present in extant chordate lineages.

4. Conclusion

In conclusion, the present study provides evidence that gene duplication of a PGDS-like lipocalin gave rise to a new epididymal lipocalin family as amniotes appeared during evolution. Internal fertilization is a common feature of amniotes and this process leads to new complex and changing environmental conditions for male gametes. It is well established that the epididymis is required for sperm survival and ability to fertilize an oocyte within the female genital tract. This specialized function has likely required new gene functions. The gene duplication, which led to an epididymal lipocalin gene cluster, may have allowed an early adaptative response to a new environment. Its conservation through 250 Myr strongly suggests that the epididymal lipocalins are required for male fertility. Moreover, the existence of an epididymal lipocalin gene cluster implies partial redundancy between cluster members as is commonly observed in other gene clusters.

Acknowledgements

This work was supported in part by the National Institute of Child Health and Human Development (HD36900) and by funds from the Ernst Schering Research Foundation/ Rockefeller Foundation.

The nucleotide sequences reported in this paper have been deposited to the GenBank with accession numbers AF435738 for mLcn9, AF435737 for mLcn10, AY301267 for mLcn11, AY299493 for mLcn12, AY360148 for mLcn13, AY301269 for LCN5, AY301268 for LCN8, AY301270 for LCN9, AY301271 for LCN10, AY301272 for LCN12.

References

Akerstrom, B., Flower, D.R., Salier, J.P., 2000. Lipocalins: unity in diversity. Biochim. Biophys. Acta 1482, 1–8.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., Eichler, E.E., 2002. Recent segmental duplications in the human genome. Science 297, 1003–1007.

Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P., Tiver, K.K., 1986. Molecular cloning of the cDNA for two major androgen-dependent secretory proteins of 18.5 kilodaltons synthesized by the rat epididymis. J. Biol. Chem. 261, 4956–4961.

Chan, P., Simon-Chazottes, D., Mattei, M.G., Guenet, J.L., Salier, J.P., 1994. Comparative mapping of lipocalin genes in human and mouse: the four genes for complement C8 γ chain, prostaglandin-D-synthase, oncogene-24p3, and progestagen-associated endometrial protein map to HSA9 and MMU2. Genomics 23, 145–150.

Chu, S.T., Lee, Y.C., Nein, K.M., Chen, Y.H., 2000. Expression, immunolocalization and sperm-association of a protein derived from 24p3 gene in mouse epididymis. Mol. Reprod. Dev. 57, 26–36.

Comeron, J.M., 1999. K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics 15, 763–764.

Fouchecourt, S., Chaurand, P., DaGue, B.B., Lareyre, J.J., Matusik, R.J., Caprioli, R.M., Orgebin-Crist, M.-C., 2002. Epididymal lipocalin-type prostaglandin D2 synthase: identification using mass spectrometry, messenger RNA localization, and immunodetection in mouse, rat, hamster, and monkey. Biol. Reprod. 66, 524–533.

Ganfornina, M.D., Gutierrez, G., Bastiani, M., Sanchez, D., 2000. A phylogenetic analysis of the lipocalin protein family. Mol. Biol. Evol. 17, 114–126.

Garrett, S.H., Garrett, J.E., Douglass, J., 1991. In situ histochemical analysis of region-specific gene expression in the adult rat epididymis. Mol. Reprod. Dev. 30, 1–17.

Hsia, N., Cornwall, G.A., 2003. Cres2 and Cres3: new members of the cystatin-related epididymal spermatogenic subgroup of family 2 cystatins. Endocrinology 144, 909–915.

Hughes, A.L., 1999. Adaptive Evolution of Genes and Genomes. Oxford Univ. Press, New York, NY.

Jean-Faucher, C., Berger, M., de Turckheim, M., Veyssiere, G., Jean, C., 1985. Testosterone and dihydrotestosterone levels in the epididymis, vas deferens and preputial gland of mice during sexual maturation. Int. J. Androl. 8, 44–57.

Jones, N., Jones, R.C., 1982. The structure of the male genital system of the Port Jackson shark, *Heterodontus portusjacksoni*, with particular reference to the genital ducts. Aust. J. Zool. 30, 523–541.

Lamont, A.E., Clarke, H., Cooper, N.J., Holland, M.K., Breed, W.G., 1998. Protein synthesis and secretion by the epididymis of the brushtail possum, *Trichosurus vulpecula*. J. Reprod. Fertil. 114, 169–177.

Lareyre, J.J., Mattei, M.G., Kasper, S., Ong, D.E., Matusik, R.J., Orgebin-Crist, M.-C., 1998. Genomic organization and chromosomal localization of the murine epididymal retinoic acid-binding protein (mE-RABP) gene. Mol. Reprod. Dev. 50, 387–395.

Lareyre, J.J., Thomas, T.Z., Zheng, W.L., Kasper, S., Ong, D.E., Orgebin-Crist, M.-C., Matusik, R.J., 1999. A 5-kilobase pair promoter fragment of the murine epididymal retinoic acid-binding protein gene drives the tissue-specific, cell-specific, and androgen-regulated expression of a foreign gene in the epididymis of transgenic mice. J. Biol. Chem. 274, 8282–8290.

Lareyre, J.J., Reid, K., Nelson, C., Kasper, S., Rennie, P.S., Orgebin-Crist, M.-C., Matusik, R.J., 2000. Characterization of an androgen-specific response region within the 5′ flanking region of the murine epididymal retinoic acid binding protein gene. Biol. Reprod. 63, 1881–1892.

Lareyre, J.J., Winfrey, V.P., Kasper, S., Ong, D.E., Matusik, R.J., Olson, G.E., Orgebin-Crist, M.-C., 2001. Gene duplication gives rise to a new 17-kilodalton lipocalin that shows epididymal region-specific expression and testicular factor(s) regulation. Endocrinology 142, 1296–1308.

Li, W.H., 1997. Molecular Evolution. Sinauer, Sunderland, MA.

Liu, Q., Hamil, K.G., Sivashanmugam, P., Grossman, G., Soundararajan, R., Rao, A.J., Richardson, R.T., Zhang, Y.L., O'Rand, M.G., Petrusz, P., French, F.S., Hall, S.H., 2001. Primate epididymis-specific proteins: characterization of ESC42, a novel protein containing a trefoil-like motif in monkey and human. Endocrinology 142, 4529–4539.

Lynch, M., Force, A., 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154, 459–473.

Morel, L., Dufaure, J.P., Depeiges, A., 2000. The lipocalin sperm coating lizard epididymal secretory protein family: mRNA structural analysis and sequential expression during the annual cycle of the lizard, *Lacerta vivipara*. J. Mol. Endocrinol. 24, 127–133.

Ohno, S., 1970. Evolution by Gene Duplication. Springer-Verlag, Berlin.

Orgebin-Crist, M.-C., Lareyre, J.J., Suzuki, K., Araki, Y., Matusik, R.J., Ong, D.E., 2002. Retinoids and epididymal function. In: Robaire, B., Hinton, B.T. (Eds.), The Epididymis—From Molecules to Clinical Practice. Kluwer Academic/Plenum Publishers, New York, pp. 339–352.

Rodriguez-Jimenez, F.J., Krause, A., Schulz, S., Forssmann, W.G., Conejo-Garcia, J.R., Schreeb, R., Motzkus, D., 2003. Distribution of new human β-defensin genes clustered on chromosome 20 in functionally different segments of epididymis. Genomics 81, 175–183.

Salier, J.P., 2000. Chromosomal location, exon/intron organization and evolution of lipocalin genes. Biochim. Biophys. Acta 1482, 25–34.

Sanchez, D., Ganfornina, M.D., Gutierrez, G., MarOn, A., 2003. Exon–intron structure and evolution of the lipocalin gene family. Mol. Biol. Evol. 20, 775–783.

Sorrentino, C., Silvestrini, B., Braghiroli, L., Chung, S.S., Giacomelli, S., Leone, M.G., Xie, Y., Sui, Y., Mo, M., Cheng, C.Y., 1998. Rat prostaglandin D2 synthetase: its tissue distribution, changes during maturation, and regulation in the testis and epididymis. Biol. Reprod. 59, 843–853.

Suzuki, K., Araki, Y., Zhu, M.Y., Lareyre, J.J., Matusik, R.J., Orgebin-Crist, M.-C., 2003. The 5′-flanking region of the murine epididymal protein of 17 kilodaltons gene targets transgene expression in the epididymis. Endocrinology 144, 877–886.

Urade, Y., Hayaishi, O., 2000. Biochemical, structural, genetic, physiological, and pathophysiological features of lipocalin-type prostaglandin D synthase. Biochim. Biophys. Acta 1482, 259–271.

Waterston, R.H., et al., 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.