Original Article

Cytogenet Genome Res 115:62–69 (2006)
DOI: 10.1159/000094802

Cytogenetic and Genome Research

Characterization of two whey protein genes in the Australian dasyurid marsupial, the stripe-faced dunnart *(Sminthopsis macroura)*

A.A. De Leo[a] C. Lefevre[a, b] D. Topcic[a] E. Pharo[a] J.F. Cheng[c] P. Frappell[d] M. Westerman[e] J.A.M. Graves[a, f] K.R. Nicholas[a]

[a]CRC for Innovative Dairy Products, Department of Zoology, The University of Melbourne, Melbourne
[b]Victorian Bioinformatics Consortium, Monash University, Clayton (Australia)
[c]U.S. Department of Energy, Joint Genome Institute, Walnut Creek, CA (USA)
[d]Department of Zoology; [e]Department of Genetics, LaTrobe University, Bundoora
[f]Comparative Genomics Research Group, Research School of Biological Sciences, Australian National University, Canberra (Australia)

Manuscript received 18 October 2005; accepted in revised form for publication by M. Schmid, 25 January 2006.

Abstract. We report the first isolation and sequencing of genomic BAC clones containing the marsupial milk protein genes Whey Acidic Protein *(WAP)* and Early Lactation Protein *(ELP)*. The stripe-faced dunnart *WAP* gene sequence contained five exons, the middle three of which code for the *WAP* motifs and four disulphide core domains which characterize WAP. The dunnart *ELP* gene sequence contained three exons encoding a protein with a Kunitz motif common to serine protease inhibitors. Fluorescence in situ hybridization located the *WAP* gene to chromosome 1p in the stripe-faced dunnart, and the *ELP* gene to 2q. Northern blot analysis of lactating mammary tissue of the closely related fat-tailed dunnart has shown asynchronous expression of these milk protein genes. *ELP* was expressed at only the earlier phase of lactation and *WAP* only at the later phase of lactation, in contrast to β-lactoglobulin *(BLG)* and α-lactalbumin *(ALA)* genes, which were expressed in both phases of lactation. This asynchronous expression during the lactation cycle in the fat-tailed dunnart is similar to other marsupials and it probably represents a pattern that is ancestral to Australian marsupials.

Marsupials have evolved a unique reproductive strategy of short gestation, birth of immature young and a relatively long lactation period, which contrasts with the relatively long gestation period of eutherian mammals (Tyndale-Biscoe and Janssens, 1988). Whereas eutherians usually produce milk of a constant composition after the expression of the initial colostrum (Jenness, 1986), marsupials progressively change the composition of milk during lactation (Green, 1984; Nicholas, 1988). The components of marsupial milk provide important nutrients and putative growth factors for the development of the young (Ballard et al., 1995; Trott et al., 2003), and may regulate the function of the mammary gland by autocrine/paracrine mechanisms (Nicholas et al., 1997; Hendry et al., 1998). Macropod marsupials, such as the tammar wallaby *(Macropus eugenii)* and red kangaroo *(Macropus rufus)*, are able to feed a pouch young and an older animal at heel with milk of different constituents from adjacent mammary glands (Bailey and Lemon, 1966; Griffiths et al., 1972; Nicholas, 1988; Nicholas et al., 1997). Therefore, milk secretion is likely to be controlled by endocrine stimuli, and other factors intrinsic to the mammary gland (Nicholas et al., 1997, 2001). The molecular mechanisms which regulate marsupial milk composition are not well understood.

The cDNA for the Early Lactation Protein *(ELP)* gene has been isolated from the lactating mammary gland of two Australian marsupials from the order Diprotodontia, the tammar wallaby *(Macropus eugenii)* (Simpson et al., 1998b)

Funding for this project was provided by the CRC for Innovative Dairy Products, Australia.

Request reprints from A.A. De Leo
CRC for Innovative Dairy Products, Department of Zoology
The University of Melbourne, Melbourne, Victoria, 3010 (Australia)
telephone: +61 3 9344 6256; fax: +61 3 8344 7909
e-mail: a.deleo@pgrad.unimelb.edu.au

1424–8581/06/1151–0062$23.50/0

Accessible online at:
www.karger.com/cgr

and brush-tailed possum *(Trichosurus vulpecula)* (Piotte and Grigor, 1996). Marsupial *ELP* is expressed only in early lactation (Piotte and Grigor, 1996; Simpson et al., 1998b), suggesting that it has a specific temporal role in the development of either the pouch young or the mammary gland. The marsupial *ELP* gene belongs to the Kunitz family of serine protease inhibitors and has significant homology to the sequence of bovine colostrum inhibitor (Simpson et al., 1998b). A potential role of the *ELP* protein as a protease inhibitor to protect immunoglobulins in the milk from proteolysis has been suggested (Piotte and Grigor, 1996; Simpson et al., 1998b). However, the function and potential physiological substrate of *ELP* has yet to be established.

Whey Acidic Protein *(WAP)* is a major whey protein gene, expressed throughout lactation in many eutherian species, including rat, mouse (Henninghausen et al., 1982), rabbit (Devinoy et al., 1988), camel (Beg et al., 1986) and pig (Simpson et al., 1998a). However, it is not found in ruminants or primates (including human) milk (Clauss et al., 2002; Rival-Gervier et al., 2003; NCBI Accession AY267466) suggesting that the gene either became inactive due to DNA mutations or was lost after the divergence of these lineages. A recent study using mouse *Wap* gene knockouts has shown that the expression of WAP is necessary for adequate nourishment of the offspring but not functional differentiation of mammary epithelial cells (Triplett et al., 2005). WAP is asynchronously secreted in the milk of diprotodontid marsupials including the tammar (Simpson et al., 2000), red kangaroo (Nicholas et al., 2001) and possum (Demmer et al., 2001). Partial WAP protein sequences have also been determined from the milk of the two monotremes, platypus and echidna (Simpson et al., 2000) but the pattern of secretion in these species is yet to be established. WAP protein secretion has not been characterized in dasyurid marsupials.

Alignment of the amino acid sequence of WAP proteins from many marsupial and eutherian species shows limited sequence identity of amino acids (Simpson et al., 2000). However, these genes are all recognized by the presence of the WAP motif (KXGXCP) and a two-domain structure known as the four disulphide core (4-DSC) domain (Ranganathan et al., 1999), which consists of eight cysteine residues (Hennighausen et al., 1982). The 4-DSC may be a required conformation for the action and stable structure of a specific class of protease inhibitors (Ranganathan et al., 1999).

The aim of the present study was to isolate and sequence large insert Bacterial Artificial Chromosome (BAC) clones containing the *ELP* and *WAP* genes from the dasyurid marsupial species *Sminthopsis macroura* (the stripe-faced dunnart). The presence of the *ELP* and *WAP* genes in dasyurid marsupials (which are only distantly related to the previously studied tammar wallaby and possum) would imply that both genes are ubiquitous at least in Australian marsupials. Furthermore, the long and homologous DNA segments of the BAC clones are ideal for fluorescence in situ hybridization (FISH) mapping to provide the first reliable map position of marsupial milk protein genes.

Materials and methods

Library screening

The BAC library RZPD 688 stripe-faced dunnart *(S. macroura)* (Chapman et al., 2003) was screened for the *ELP* and *WAP* genes. The tammar *WAP* cDNA (Accession AJ005356) and *ELP* cDNA (Accession AJ000490) were labeled with [α-^{32}P]-dCTP and used for probing the libraries as described in Sambrook et al. (1989). Positive signals were identified and DNA extracted from the individual clones as described in Sambrook et al. (1989).

Identifying and sequencing the ELP and WAP genes in the BACs

DNA from the stripe-faced dunnart BACs with either the *ELP* or *WAP* gene were subcloned using the TOPO Shotgun kit (Invitrogen, USA) according to the manufacturer's instructions. Subclones were screened for *ELP* and *WAP* exons, and positives sequenced as described in Sambrook et al. (1989). The entire stripe-faced dunnart BAC containing the *ELP* gene was sequenced by a fluorescent shotgun method. Briefly, plasmid DNA was sequenced from both directions, and sequence reads were then assembled using the Phrap assembler (Ewing and Green, 1998). The 14 kb of DNA surrounding the stripe-faced dunnart *WAP* gene (including 7 kb of DNA 5′ of the first exon) was sequenced by the Australian Genome Research Facility (Queensland, Australia), using primers designed from the tammar *WAP* cDNA sequence.

Bioinformatics and BAC sequence analysis

Gene exon/intron structure was determined by comparing genomic DNA sequence with other marsupial *WAP* and *ELP* cDNAs in the GenBank database. In addition GENESCAN (http://genes.mit.edu/GENSCAN.html) was used to predict the putative cDNA and protein. Predicted protein alignments were carried out using ClustalW (http://clustalw.genome.jp/ or http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html).

A seven-fold coverage of short genomic sequences of approximately 1 kb were available for the didelphid marsupial, the Brazilian short-tailed gray opossum *(Monodelphis domestica)* (at www.ncbi.nlm.nih.gov/BLAST/tracemb.shtml). The gray short-tailed opossum *WAP* and *ELP* sequences were assembled from short genomic sequences retrieved using the cDNA of *WAP* and *ELP* from the stripe-faced dunnart, tammar (Accession AJ000490, AJ005356) and brush-tailed possum (Accession AF275314) to Blast the gray short-tailed opossum genome to obtain matching sequence reads and their mate pairs. The sequences were assembled using phrap (http://www.phrap.org/). Assembled fragments were used to retrieve further the gray short-tailed opossum overlapping sequences together with mate pairs in order to extend local assemblies around *WAP* and *ELP* genes.

RT PCR of cDNA for the milk protein genes

Tissue was excised from the lactating mammary gland of the fat-tailed dunnart, homogenized and total RNA extracted using the Qiagen RNeasy Lipid Tissue mini kit (Melbourne). The total RNA (2 μg) was reverse transcribed (RT) using a SuperScript^{TM} II kit (Invitrogen, Melbourne), following the manufacturer's recommended protocol.

Milk protein gene DNA primers were designed from both eutherian and marsupial β-lactoglobulin *(BLG)* and α-lactalbumin *(ALA)* by aligning sequences from GenBank. However, only marsupial *WAP* genes and marsupial *ELP* genes were aligned for primer design, because eutherian and marsupial *WAP* gene sequences differ greatly, and *ELP* has not been identified in eutherians. Highly conserved regions were selected for the priming region. Lactating mammary gland template cDNA (50 ng), 5 μl 10× buffer, 1 μl 10 mM each forward and reverse primer (Table 1), 1 μl 10 mM dNTPs and 2.5 units Taq Polymerase in a total volume of 50 μl were PCR amplified by initial denaturization at 94°C for 2 min, then 35 cycles at 94°C for 1 min, at 47°C for 30 s, at 72°C for 1 min, and a final extension of 72°C for 10 min. All PCR products showed a single band when electrophoresed on a 0.8% agarose gel and the bands were subsequently excised and DNA extracted using the Qiagen gel extraction kit. The DNA was ligated into pGEM T-easy and clones sequenced using primer sites T7 and SP6 in the vector.

Table 1. Primers used in PCR of cDNA from lactating mammary gland

Gene	Forward primer 5′-3′	Reverse primer 5′-3′
WAP	exon2F AGAAGGCTGGGCGCTGCCC	exon3R CATCGCAGTCCAGGTCAG
ELP	exon1F ATGAAATTCACCATCATTGCCCTCT	exon3R GCAGAAGAGGGAATCATAGAGGG
ALA	exon2F CTGCTCCTCTTTGCTCCTGCT	exon5R ACTGGTCCAGGTCCTCAATG
BLG	exon2F AATTCCCTTGTTTGACATGGAC	exon6R GCTTTCCCTAAAGARGGCAC

Expression of the WAP and ELP genes analyzed by Northern blotting

Total RNA (10 μg, as determined by spectrophotometer) was electrophoresed together with markers of known size and concentration in a 1% agarose gel with 1× MOPS and 1.1% formaldehyde. The RNA was transferred to Zeta probe membrane and the Northern blot was probed (Sambrook et al., 1989) with the full length tammar cDNA for *ELP* (Accession AJ000490), *WAP* (Accession AJ005356), *ALA* (Accession X15211) and *BLG* (Accession L14954-L14960) genes.

Fluorescence in situ hybridization (FISH)

Metaphase spreads were prepared from the stripe-faced dunnart and FISH was carried out as previously described (Toder et al., 1996). The stripe-faced dunnart BACs (237 L6-ELP, 002 F17-WAP) were used as probes. Slides were viewed under a Zeiss Axioplan microscope, images collected with software package Spot Advance and pictures processed using Image J, Confocal Assistant, Adobe Illustrator 7.0, and Adobe Photoshop. At least ten metaphase spreads were captured that had three or four signals of a possible maximum of four.

Results

Isolation and sequencing of the stripe-faced dunnart ELP and WAP genes

One *ELP* positive BAC (237 L6) was recovered from the stripe-faced dunnart BAC library and was completely sequenced by the shotgun method. The BAC was found to be 80 kb in length and to contain the entire *ELP* gene. One *WAP* positive BAC (002 F17) was detected by screening the stripe-faced dunnart library with the *WAP* probe. Approximately 14 kb was sequenced and found to include 7 kb of promoter and the first four exons of the stripe-faced dunnart *WAP* gene.

Comparison of the marsupial ELP sequences

Comparison of a large set of *ELP* EST (cDNA) sequences from the tammar mammary gland pooled from all phases of lactation (the database is held by the Collaborative Research Centre for Innovative Dairy Products, Melbourne, Australia) provided 510 bp of contiguous sequence for *ELP*.

Alignment of the stripe-faced dunnart and gray short-tailed opossum *ELP* genomic sequences with the tammar cDNA from the EST library was carried out to determine the exon/intron boundaries in the genomic regions (Fig. 1a). Stripe-faced dunnart and gray short-tailed opossum genomic *ELP* aligned with tammar *ELP* cDNA predicted that the complete coding sequence of the dunnart and opossum *ELP* is contained in three exons (Fig. 1a). In the stripe-faced dunnart the first exon contained a 5′ untranslated region plus a region coding for the first 24 amino acids of ELP, including a signal peptide sequence. The second exon comprised most of the coding region. Exons 2 and 3 encode the stop codon (TAG). The last nucleotide in exon 2 is T, and the first two nucleotides in exon 3 are AG. Exon 3 contains 154 bp up to the canonical transcription termination signal (AATAAA) at the 3′ end of the gene.

An NCBI BlastP search of the stripe-faced dunnart *ELP* predicted protein revealed the best alignment to be with ELP proteins of other marsupials including the tammar and the brush-tailed possum. The best alignment with the human genome was tissue factor pathway inhibitor 2 *(TFPI2)* located on human chromosome 7q22. The stripe-faced dunnart *ELP* sequence also has homology with eutherian genes coding for a Kunitz motif, including human *LOC 391253* and *WAP8*, bovine trypsin inhibitor, bovine colostrum trypsin inhibitor, and both sheep and bovine trophoblast Kunitz domain proteins.

Identification of other sequences in the stripe-faced dunnart ELP BAC

A second gene, phosphatidyl inositol glycan class T *(PIGT)*, was found in the stripe-faced dunnart *ELP* BAC by BlastP analysis of the GENESCAN predicted protein; stripe-faced dunnart PIGT shared nearly 72% protein identity with mouse, rat and human PIGT. The GENESCAN program searches for conserved start codons, 3′ and 5′ splice sites, canonical TATA boxes, and stop codons. Therefore, the prediction by GENESCAN that the *PIGT* sequence represents a functional gene, together with the sequence similarity between the stripe-faced dunnart and eutherians, strongly suggests that the stripe-faced dunnart *PIGT* is a functional gene rather than a pseudogene. The BAC also contained two identifiable repetitive elements: one with 42% protein identity with a long interspersed sequence element (LINE 1) in human and other eutherian genomes, and the second which showed no significant matches when subjected to BlastP, but a match of 100% with 24 bases of a human MER repeat element revealed by BlastN (Fig. 2).

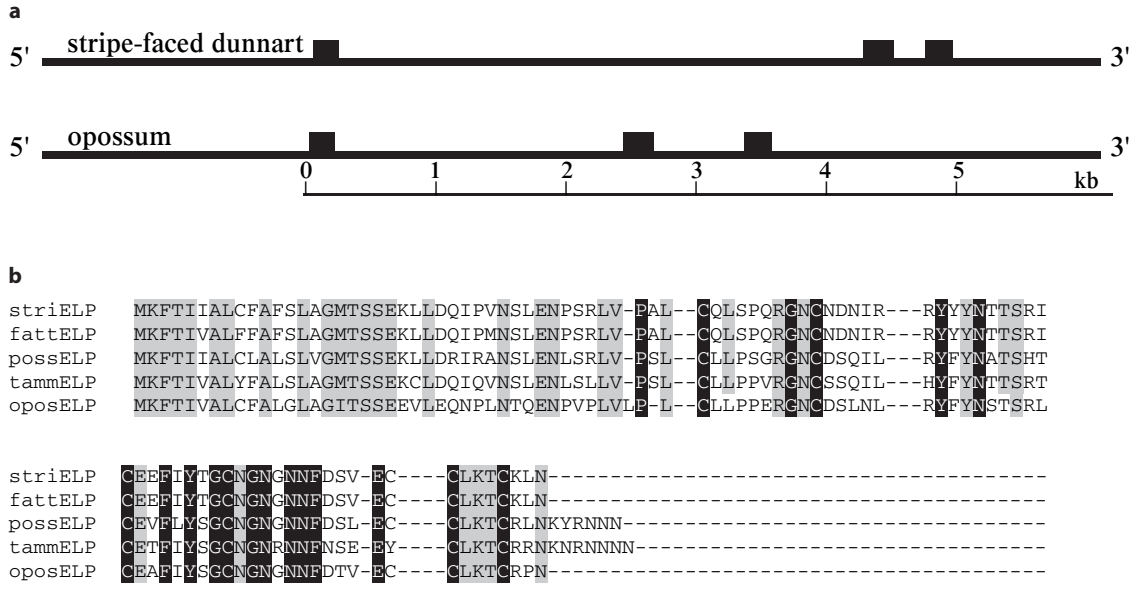


Fig. 1. Structure of the *ELP* gene. (**a**) Comparison of marsupial *ELP* gene structure. The black squares represent exons in the stripe-faced dunnart and opossum. Both species have three exons of almost identical size, but the introns vary in size. (**b**) Alignment of the stripe-faced dunnart ELP protein with other marsupial ELP proteins. The conserved Kunitz motif sequences are shown in white type with a black background. Other matches between marsupial ELP proteins are shown in black type with a grey background. striELP = Stripe-faced dunnart ELP conceptually translated from the BAC sequence; fattELP = fat-tailed dunnart ELP conceptually translated from the cDNA sequence; possELP = brush-tailed possum ELP gi1002806; tammELP = tammar wallaby ELP gi3046758; and oposELP = gray short-tailed opossum ELP, assembled from the short sequences available at www.ncbi.nlm.nih.gov/BLAST/tracemb.shtml.

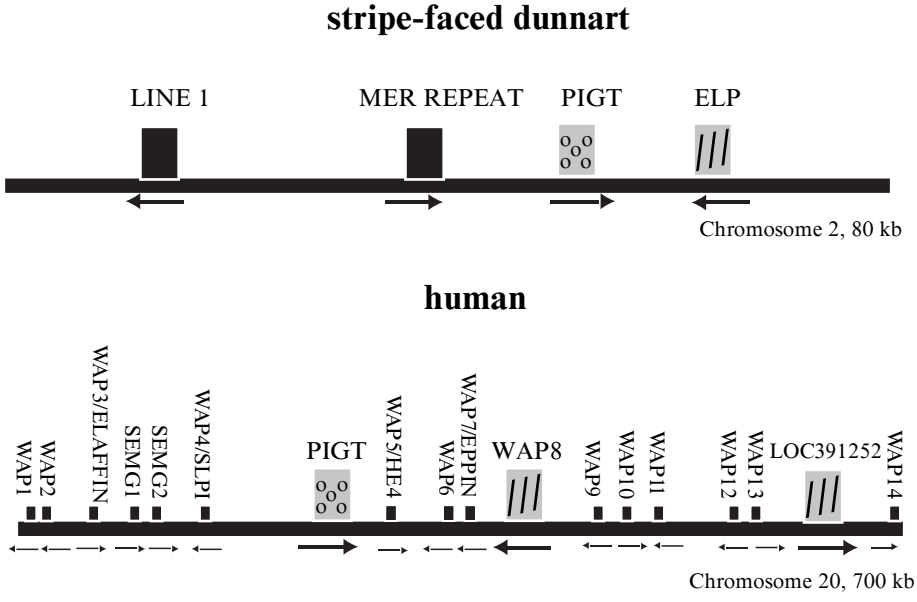
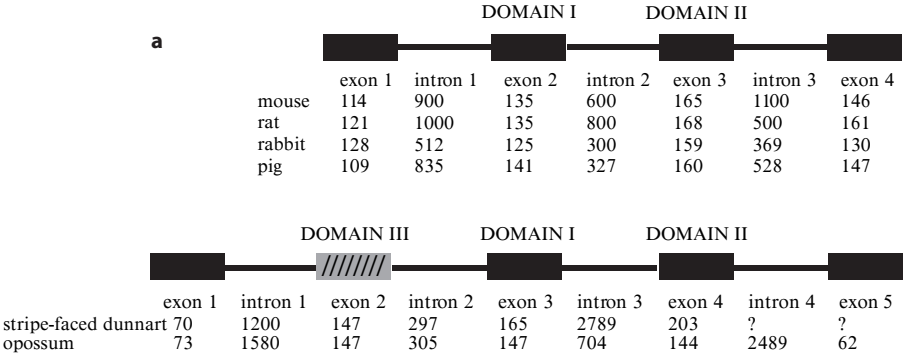


Fig. 2. Comparison of human and stripe-faced dunnart genomic regions surrounding the *ELP*/*PIGT* genes. The arrows represent the direction of transcription. *PIGT* is represented by a grey background with black spots. *ELP* in the stripe-faced dunnart is represented by a grey background with black stripes as are the human genes that have sequence homology to stripe-faced dunnart *ELP*. The black boxes represent other genes. The human map is adapted from Clauss et al. (2002). The stripe-faced dunnart chromosome position has been assigned to chromosome 2 by FISH mapping (see Fig. 4b).

Characterization of marsupial WAP genes

The structures of the stripe-faced dunnart and gray short-tailed opossum *WAP* genomic sequences were determined by comparing sequence with the tammar and possum *WAP* cDNAs in order to identify the five exons. The gray short-tailed opossum *WAP* gene genomic sequence was sourced from the database, as described in Materials and methods, and shown by alignment with the tammar and possum cDNAs to have five exons (Fig. 3a). The limited sequence available from the partial sequencing of the stripe-faced dunnart genomic *WAP* identified the first four of these five exons.



b

```
LEADER PEPTIDE
striWAP  M-QSVQILTLVLL-ALG-AWA-A-QNSVE------
possWAP  MMQSVQILTLVLL-ALG-AWA-A-QNSIE------
tammWAP  M-QPVQILTLVLL-ALG-AWA-A-QESTE------
oposWAP  MMQSVQILTLALL-ALG-SWA-A-QNLEE------

DOMAIN III
striWAP  KACRCPNFRQVLSDRTECKQMCHGDSSCPQDLKCCPQGCNWLCMNTTQE
fattWAP  ----CPNFRQVLSDRTECKQMCHGDSSCPQDLKCCPQGCNWLCMNTTQE
possWAP  KACYCPDFRHVLSDIRDCKQMCNNDASCPQNLRCCQRGCSWLCLNTTQE
tammWAP  KACYCPDFRQVLLDRRDCKQLCNDDASCPQNMRCCQRGCSWLCMNTTQE
oposWAP  KSGYCPVLHPTLSDLIDCKRTCQNDTSCPGEMKCCRQGCSWMCKNPMKE
CONSENSUSK G CP          C  C    C    CC   C   C

DOMAIN I
striWAP  KDGSCPLATFSSSSSEEHHLNPLCG--KACNTDLDC-DGEAKCCV-SS-CGQ-TCTMPIRV-------
fattWAP  KDGSCPLATFSFSSSEEHHLNPLCG--KACHTDLDC-DGEAKCCV-SS-CGQ-TCTMPIRV-------
possWAP  KDGMCPMGTSSSSSSEEQWRNQLCN--STCKTDLDC-AGEAKCCA-SS-CGR-ICSMPIKA-------
tammWAP  KDGLCPVATSHSSSSEEQQRKQLCD--KTCKTDLGC-EGKAKCCA-SS-CGQ-TCFMPVKA-------
oposWAP  KACLCPLAT---SSSEERLP---CGIP--CETDKDCD-GEAKCCVRS--CGRT-CSMPIKV-------
CONSENSUSK G CP           C    C    C     CC    C   C

DOMAIN II
striWAP  KSCQCPAVTKV------CPRN-NWTHTCWLDDHGLGEKKCCF--SGYGRR-CME--PFRDKKELANLKCVSSFGRC
possWAP  KPGRCPAIAEV------CPKNMSWIDNCQSDDHCNRSKKCCS--SGCGQR-CMK--PLPEERGAVTRSFH------
tammWAP  KPGRCPAVTGI------CPEKKSWFHTCQRDDQCKENKKCCS--SACGRR-CTN--PFPEEYEASQDESTLLAL--
oposWAP  KPGRCPVVSEW------CPRNERFIG-CRGDDNCLGMKKCCF--SRCAMR-CMDAIPX------------------
CONSENSUSK G CP           C     C    C    CC    C   C
```

The five exons of the marsupial *WAP* genes contrast with the eutherian *WAP* genes, which have only four exons (Fig. 3a). Domains I and II of the WAP protein are encoded by exons 2 and 3 in eutherians, and exons 3 and 4 in marsupials. Exon 2 in marsupials codes for domain III (Fig. 3a), an extra domain described by studies using 3-dimensional modeling (Simpson et al., 2000). Domains I, II and III contain eight cysteine residues (represented by C in Fig. 3b) which constitute the 4-DSC domain, with the exception of domain II in the stripe-faced dunnart which has only six cysteines. The WAP motif (KXGXCP) is present at the 5′ end of marsupial domains I, II and III. However, the motif in domain II of the stripe-faced dunnart WAP is replaced by KXCXCP. In contrast, eutherians have the WAP motif only in domain II (Fig. 3b).

A BlastP search of the NCBI database with the stripe-faced dunnart predicted WAP protein revealed the best match to be with the eutherian milk WAPs from mouse, rat, camel, rabbit and pig. There is no functional human milk WAP protein as it has become a pseudogene in the human lineage (Clauss et al., 2002). A BlastP search of the human protein database (at NCBI) revealed the two best matches with the stripe-faced dunnart predicted WAP protein to be *WFDC3* and *WFDC5* which are both functional WAP genes located on human 20q13. The stripe-faced dunnart predicted WAP protein is compared to other marsupial milk WAPs in Fig. 3b. Homology is observed in the region of the WAP motif, the 4-DSC consensus sequences and the leader peptide sequences.

Chromosome mapping of marsupial ELP and WAP genes in marsupials

FISH onto metaphase spreads localized the *ELP* gene to chromosome 2q in the stripe-faced dunnart (Fig. 4b), and the *WAP* gene to chromosome 1p (Fig. 4a).

Milk protein gene expression

Studies were carried out to examine if there was asynchronous expression of the *ELP* and *WAP* genes during the lactation cycle in the stripe-faced dunnart. No lactating animals were available from a stripe-faced dunnart *(S. macroura)* colony. However, mammary glands were collected from lactating females of the closely related fat-tailed dunnart *(Sminthopsis crassicaudata)*.

Milk protein gene expression was analyzed by Northern blot analysis of RNA from mammary glands on day 37 of fat-tailed dunnart development, which is the early to middle stage of lactation where the young are permanently attached

Fig. 3. Structure of the *WAP* gene. (**a**) Comparative genomic organization of the *WAP* genes in eutherians and marsupials. The exons common to both marsupials and eutherians are shown in black, the extra exon in marsupials is shown by a grey background with black stripes. In eutherians exons 2 and 3 encode the two 4-DSC domains. In marsupials exons 2, 3 and 4 encode the three 4-DSC domains. Hence, domain III has been found in marsupials, but not in eutherians. The fifth marsupial exon has been identified in opossum, but the stripe-faced dunnart sequence does not extend far enough to reach the fifth exon. (**b**) Alignment of the stripe-faced dunnart WAP protein with other marsupial WAP proteins expressed in the mammary gland. In all marsupials there is homology between the protein domains as well as the leader sequence. The consensus sequence which includes the WAP motif (KXGXCP) and the four disulphide core domain represented by eight cysteines (C) is shown below each domain in bold. In the protein alignment the consensus sequence is shown in white type with a black background and other conserved proteins are shown in black type with a grey background. striWAP = stripe-faced dunnart WAP, conceptually translated from the BAC sequence; fattWAP = fat-tailed dunnart WAP conceptually translated from the partial cDNA sequence from domains III and I; possWAP = possum WAP gi 14582214; tammWAP = tammar WAP gi gi7711019, oposWAP = opossum WAP, assembled from the short sequences available at ww.ncbi.nlm.nih.gov/BLAST/tracemb.shtml.

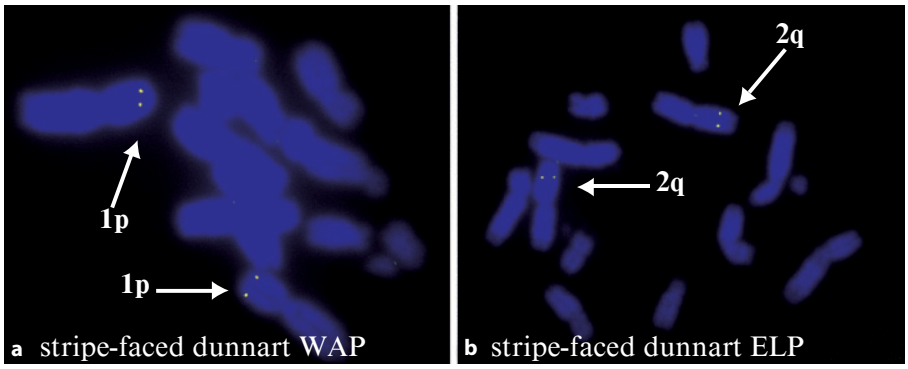


Fig. 4. Mapping of marsupial *WAP* and *ELP* genes on metaphase spreads. (**a**) *WAP* is located on chromosome 1p of the stripe-faced dunnart. (**b**) *ELP* is located on chromosome 2q of the stripe-faced dunnart.

to the teat, and on day 58 when the young are no longer permanently attached to the teat but are still suckling. Expression of the *ELP* gene was detected on day 37, and the *WAP* gene on day 58. In contrast, the *ALA* and *BLG* genes were expressed in lactating mammary gland on both day 37 and 58 (Fig. 5). The PCR products for *ALA, BLG, ELP* and *WAP* genes were sequenced for comparisons with other species.

Homology between ELP and WAP from the two dunnart species was greater than 97%, suggesting that the *ELP* and *WAP* genes in both dunnart species are homologous genes. Comparison of the *ELP* and *WAP* partial cDNA sequence from the fat-tailed dunnart with the *ELP* and *WAP* exon sequence predicted from the stripe-faced dunnart genomic sequence showed overall 1–3 base differences every 100 kb of sequence; seven differences in the 242 nucleotides sequenced from *WAP* (and two differences in the 80 amino acids of the predicted protein translation) and six differences in the 432 nucleotides sequenced from *ELP* (and two differences in the 96 amino acids of the predicted protein translation).

Fig. 5. Milk protein gene expression in the lactating mammary gland of the fat-tailed dunnart. The Northern blot shows that *ALA* and *BLG* are expressed on days 37 and 58. *WAP* is expressed on day 58 only and *ELP* is expressed on day 37 only. The *WAP, ALA* and *BLG* transcripts were detected as single bands at approximately 700 bp. The *ELP* transcript was detected as a single band at about 500 bp.

Discussion

Expression pattern of ELP, WAP, BLG and ALA genes during marsupial lactation cycles

The timing of gestation and lactation in the two dunnart species (the fat-tailed dunnart and the stripe-faced dunnart) is very similar. The stripe-faced dunnart has the shortest gestation (10.5–11 days) known for any mammal (Selwood and Woolley, 1991; Gemmell and Selwood, 1994), and a lactation length of 65–70 days (Frigo and Woolley, 1997). The newborn remains permanently attached to the teat for the first 30–40 days of lactation, and then on day 55 commences eating solid food while still intermittently suckling milk (Frigo and Woolley, 1997). The gestation time for the fat-tailed dunnart (13–16 days) is slightly longer and the duration of lactation is approximately 70 days. Permanent attachment to the teat ceases after 40 days, and after 60 days solid food is consumed by the young (Ewer, 1968; Godfrey and Crowcroft, 1971).

This study has shown that the fat-tailed dunnart *ELP* and *WAP* genes are expressed at different stages of lactation. *ELP* is expressed on day 37, when the young remain permanently attached to the teat and have a diet exclusively of milk, and *WAP* is expressed on day 58 when the young are no longer permanently attached to the teat and have a diet of both solids and milk. This expression pattern is similar to that in diprotodont marsupials (which include the tammar and possum) that diverged about 45 MYA (Woodburne et al., 2003). The tammar and brush-tailed possum *ELP* gene is expressed for 100 days after parturition, when the young remain permanently attached to the teat (Piotte and Grigor, 1996; Simpson et al., 1998b) (Fig. 6). In contrast, *WAP* is expressed in the tammar from day 100 to 200 when the young are in the pouch and intermittently suckling but are no longer permanently attached to the teat (Simpson et al., 2000) (Fig. 6). In possum, *WAP* is expressed during day 80–180, and then declines at permanent pouch exit (Demmer et al., 2001). In contrast to the tammar and fat-tailed dunnart, there is a switch period in the brush-tailed possum, during which *ELP* and *WAP* are both expressed in the mammary gland, then *WAP* continues to be expressed, for about an additional 60 days after solid food is consumed by

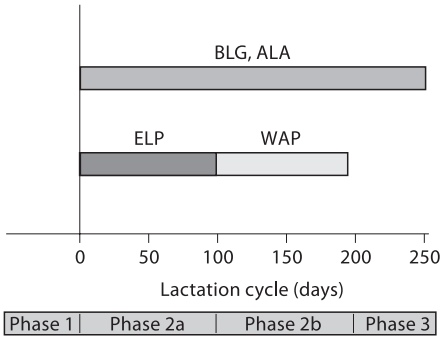


Fig. 6. The timing of expression of the ELP, WAP, BLG and ALA milk proteins during the lactation cycle of the tammar. BLG and ALA are expressed throughout lactation, but ELP and WAP are expressed at specific phases of lactation.

the young (Demmer et al., 2001). Due to the limited availability of tissue samples in the fat-tailed dunnart, the current study could not determine whether there is a similar switch period in this species.

It is interesting that *WAP* and *BLG* are concurrently expressed in the marsupials and monotremes studied to date. When *WAP* was first identified in the milk of mouse, rat (Hennighausen et al., 1982), rabbit (Beg et al., 1986) and camel (Devinoy et al., 1988), it was thought that animals secreted either *WAP* or *BLG* as their major whey protein, but not both. However, a more recent report showed the pig lactating mammary gland secretes *WAP* and *BLG* concurrently (Simpson et al., 1998a), suggesting that not all eutherians express only one of *WAP* or *BLG*. The current study suggests that *WAP* and *BLG* were expressed concurrently in the common mammalian ancestor, and the one or the other of *WAP* or *BLG* was lost subsequently in some eutherian lineages. In the human lineage both the 'milk *WAP*' gene and the *BLG* gene function have been lost and both are now pseudogenes (Clauss et al., 2002).

The evolution of the marsupial WAP and ELP genes

The sequence of the stripe-faced dunnart, tammar, and brush-tailed possum *ELP* and *WAP* genes showed more than 80% homology with the corresponding partial gene sequence from the fat-tailed dunnart. Sequence comparisons suggest that the gene structure, asynchronous expression and function of these genes is common at least to Australian marsupials. This study shows the genomic structure of the *ELP* and *WAP* genes in the gray short-tailed opossum is similar to that of Australian marsupials, therefore, it is likely that both genes are also present in the genome of other American marsupials. However, the expression pattern of these genes during lactation in the gray short-tailed opossum is yet to be studied.

The stripe-faced dunnart ELP protein contains a Kunitz motif with homology at the C terminal end (3′ end) of proteins to other Kunitz proteins. This suggests that the marsupial evolved additional sequences on the 5′ end of the *ELP* gene required for specific functions in marsupial lactation. However, a similar *ELP* gene in monotremes would suggest that the additional 5′ sequences were present in the common mammalian ancestor and lost in the eutherian lineage. Information on a potential platypus *ELP* should be available soon with the platypus genome nearly completely sequenced (http://www.ncbi.nlm.nih.gov/BLAST/tracemb.shtml). The Kunitz motif is common to one of the numerous families of serine protease inhibitors (NCBI protein database), including human TFPI2, trypstatin a mast cell inhibitor of trypsin in rats, isoinhibitor from the garden snail, a number of venom basic protease inhibitors, and basic protease inhibitor from sea turtle. Thus Kunitz motif genes are not exclusive to the mammary gland but are present in many genes expressed in different tissues with various functions.

The *WAP* genes expressed in the mammary gland of eutherians have two 4-DSC domains (Hennighausen et al., 1982; Simpson et al., 1998b), whereas WAP in all marsupials studied to date, as well as platypus and echidna, have only three (Simpson et al., 2000; Demmer et al., 2001). This extra domain is probably ancestral for all mammals, and was lost in the eutherian lineage. In eutherians, the identical distributions of cysteine residues within the two domains suggest that the *WAP* gene evolved by intergenic duplication (Hennighausen et al., 1982). The three domains of marsupials also have similar distributions of cysteine residues, suggesting they arose from two intergenic duplications.

Location of the stripe-faced dunnart WAP and ELP genes

The mouse milk *Wap* gene is located on chromosome 11, flanked by the genes *Ramp3* and *Tbrg4*. This region is equivalent to human chromosome 7p13→p12, where a non-functional human *WAP* relic is also flanked by *RAMP3* and *TBRG4* (Clauss et al., 2002; Rival-Gervier et al., 2003). Functional *WAP*-like genes in the human genome are clustered on chromosome 20q13, and are ubiquitously expressed, including non-lactating breast tissue (Clauss et al., 2002). However, *WAP* has not yet been identified in human milk, and the mouse milk *WAP* gene does not share orthology with the human *WAP* genes on chromosome 20 (Clauss et al., 2002). Therefore, the marsupial milk *WAP*s along with the mouse and other eutherian 'milk *WAP*s' are likely to share orthology with the human 'milk *WAP*' pseudogene. Hence, the finding of stripe-faced dunnart *WAP* on chromosome 1p suggests that stripe-faced dunnart 1p shares a common ancestry with human chromosome 7, where the human *WAP* pseudogene is located.

The finding that the stripe-faced dunnart *PIGT* gene is located next to the stripe-faced dunnart *ELP* gene, and that the sole human *PIGT* gene (and also the mouse *Pigt* gene) are near to two other genes (*WAP8* and *LOC391252*) that also share the Kunitz motif suggests that the putative therian common ancestor may have had a *PIGT* gene and a Kunitz protein motif gene neighboring each other on an ancestral chromosome. This shows that the chromosomal locations of this gene on stripe-faced dunnart chromosome 2q and the human chromosome 20q probably share common ancestry.

Cross-species painting between the stripe-faced dunnart, and gray short-tailed opossum (Rens et al., 2001) reveals which chromosome regions are shared between these species. The gene mapping from this study can therefore predict that in the gray short-tailed opossum genome the *ELP* gene should be found on chromosome 1q, and the *WAP* gene on chromosome 6q near the centromere. As the gray short-tailed opossum genome currently has only scaffold regions assembled (http://www.ensembl.org), which are not yet identified as belonging to specific chromosomes, the gene locations predicted in this study provide valuable reference points to anchor sequence to particular chromosomes in the opossum. The human/marsupial comparative map positions discovered in this milk protein gene study can now be used to advance the construction of a comparative map between human and marsupial genomes, which will be vital for exploring the evolution of the mammalian genome.

GenBank submissions

S. crassicaudata ALA partial cDNA AC DQ518240, *S. crassicaudata BLG* partial cDNA AC DQ518241, *S. crassicaudata ELP* partial cDNA AC DQ518242, *S. crassicaudata WAP* partial cDNA AC DQ5182403, *S. macroura* ELP BAC (237 L6) AC 186006, *S. macroura* WAP BAC (002F17) 14 kb AC DQ518244.

Acknowledgements

We thank Yi Peng and Shibani Mukherjee for performing the ELP BAC sequencing under the NHLBI's Program for Genomic Applications grant to J.F.C.

References

Bailey L, Lemon M: A specific protein difference in the milk from two mammary glands of a red kangaroo. Aust J Exp Biol Med Sci 44:705 (1966).

Ballard FJ, Grbovac S, Nicholas KR, et al: Differential changes in the milk concentrations of epidermal growth factor and insulin like growth factor-I during lactation in the tammar wallaby, *Macropus eugenii*. Gen Comp Endocrinol 98:262–268 (1995).

Beg OU, Von Bahr-Lindstrom H, Zaidi ZH, et al: A camel milk whey protein rich in half-cystine. Primary structure, assessment of variations, internal repeat patterns, and relationships with neurophysin and other active polypeptides. Eur J Biochem 159:195–201 (1986).

Chapman MA, Charchar FJ, Kinston S, et al: Comparative and functional analyses of *LYL1* loci establish marsupial sequences as a model for phylogenetic footprinting. Genomics 81:249–259 (2003).

Clauss A, Lija H, Lundwall A: A locus on human chromosome 20 contains several genes expressing protease inhibitor domains with homology to whey acidic protein. Biochem J 368:233–242 (2002).

De Leo AA, Guedelha N, Toder R, et al: Comparative chromosome painting between marsupial orders: relationships with a 2n = 14 ancestral marsupial karyotype. Chromosome Res 7:509–517 (1999).

Demmer J, Stasiuk SJ, Grigor MR, et al: Differential expression of the whey acidic protein gene during lactation in the brushtail possum *(Trichosurus vulpecula)*. Biochim Biophys Acta 1522: 187–194 (2001).

Devinoy E, Hubert C, Jolivet G, et al: Recent data on the structure of rabbit milk protein genes and on the mechanism of the hormonal control of their expression. Reprod Nutr Dev 28:1145–1164 (1988).

Ewer RF: A preliminary survey of the behaviour in captivity of the dasyurid marsupial, *Sminthopsis crassicaudata* (Gould). Z Tierpsychol 25: 319–365 (1968).

Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194 (1998).

Frigo L, Woolley PA: Growth and development of pouch young of the stripefaced Dunnart, *Sminthopisis macroura* (Marsupialia:Dasyuridae), in captivity. Austral J Zool 45:157 (1997).

Gemmell RT, Selwood L: Structural development in the newborn marsupial, the stripe-faced dunnart, *Sminthopsis macroura*. Acta Anat 149:1–12 (1994).

Godfrey GK, Crowcroft P: Breeding the fat-tailed marsupial mouse, *Sminthopsis crassicaudata*, in captivity. Inter Zoo Yearbook 11:34 (1971).

Green B: Composition of milk and energetics of growth in marsupials. Zool Soc Lond Symp 51: 369 (1984).

Griffiths M, McIntosh DL, Leckie RMC: The mammary gland of the red kangaroo with observations on the fatty acid components of the milk triglycerides. J Zool Lond 166:265 (1972).

Hendry KA, Simpson KJ, Nicholas KR, et al: Autocrine inhibition of milk secretion in the lactating tammar wallaby *(Macropus eugenii)*. J Mol Endocrinol 21:169–177 (1998).

Hennighausen LG, Sippel AE, Hobbs AA, et al: Comparative sequence analysis of the mRNAs coding for mouse and rat whey protein. Nucleic Acids Res 10:3733–3744 (1982).

Jenness R: Lactational performance of various mammalian species. J Dairy Sci 69:869–885 (1986).

Nicholas KR: Asynchronous dual lactation in a marsupial, the tammar wallaby *(Macropus eugenii)*. Biochem Biophys Res Commun 154: 529–536 (1988).

Nicholas K, Simpson K, Wilson M, et al: The tammar wallaby: a model to study putative autocrine-induced changes in milk composition. J Mammary Gland Biol Neoplasia 2:299–310 (1997).

Nicholas KR, Fisher JA, Muths E, et al: Secretion of whey acidic protein and cystatin is down regulated at mid-lactation in the red kangaroo. Comp Biochem Physiol 129:851–858 (2001).

Piotte CP, Grigor MR: A novel marsupial protein expressed by the mammary gland only during the early lactation and related to the Kunitz proteinase inhibitors. Arch Biochem Biophys 330:59–64 (1996).

Ranganathan S, Simpson KJ, Shaw DC, et al: The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. J Mol Graph Model 17:106–113 (1999).

Ren J, Stuart DI, Acharya KR: Alpha-lactalbumin possesses a distinct zinc binding site. J Biol Chem 268:19292–19298 (1993).

Rens W, O'Brien PC, Yang F, et al: Karyotype relationships between distantly related marsupials from South America and Australia. Chromosome Res 9:301–308 (2001).

Rival-Gervier S, Thepot D, Jolivet G, et al: Pig whey acidic protein gene is surrounded by two ubiquitously expressed genes. Biochim Biophys Acta 1627:7–14 (2003).

Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning, a Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

Selwood L, Woolley PA: A timetable of embryonic development, and ovarian and uterine changes during pregnancy, in the stripe-faced dunnart, *Sminthopsis macroura* (Marsupialia: Dasyuridae). J Reprod Fertil 91:213–227 (1991).

Simpson KJ, Bird P, Shaw D, et al: Molecular characterisation and hormone dependent expression of the porcine whey acidic protein gene. J Mol Endocrinol 20:27–35 (1998a).

Simpson K, Shaw D, Nicholas KR: Developmentally regulated expression of a putative protease inhibitor gene in the lactating mammary gland of the tammar wallaby, *Macropus eugenii*. Comp Biochem Physiol 120:535–541 (1998b).

Simpson KJ, Ranganathan S, Fisher JA, et al: The gene for a novel member of the whey acidic protein family encodes three four-disulfide core domains and is asynchronously expressed during lactation. J Biol Chem 275:23074–23081 (2000).

Toder R, Wilcox SA, Smithwick M, et al: The human/mouse imprinted genes *IGF2, H19, SNRPN* and *ZNF127* map to two conserved autosomal clusters in a marsupial. Chromosome Res 4:295–300 (1996).

Triplett AA, Sakamoto K, Matulka LA, et al: Expression of the whey acidic protein (Wap) is necessary for adequate nourishment of the offspring but not functional differentiation of mammary epithelial cells. Genesis 43:1–11 (2005).

Trott JF, Simpson KJ, Moyle RL, et al: Maternal regulation of milk composition, milk production, and pouch young development during lactation in the tammar wallaby *(Macropus eugenii)*. Biol Reprod 68:929–936 (2003).

Tyndale-Biscoe CH, Janssens PA: Introduction, in Tyndale-Biscoe CH, Janssens PA (eds): The Developing Marsupial: Models for Biomedical Research, p 1 (Springer-Verlag, Heidelberg 1988).

Woodburne MO, Rich TH, Springer MS: The evolution of tribospheny and the antiquity of mammalian clades. Mol Phylogenet Evol 28:360–385 (2003).